ISCA: AN ITERATIVE SIMPLICIAL-COMPLEX APPROXIMATION METHOD TO SOLVE k-HYPERPLANE CLUSTERING TO GLOBAL OPTIMALITY

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose the *Iterative Simplicial-Complex Approximation* (ISCA) method to solve to global optimality the k-HC—the k-Hyperplane Clustering problem that asks to find k hyperplanes that minimize the sum of squared 2-norm (Euclidean) distances between each point and its closest hyperplane. ISCA is based on Spatial Branch and Bound (SBB) and approximates the infeasible unit ball of the hyperplane normals featured in k-HC from the inside with a growing simplicial complex, which is adaptively refined, within SBB, by gluing to it new simplices whose apexes are the radial projections onto the unit sphere of infeasible relaxation solutions. We prove that ISCA never discards a feasible solution, thereby certifying global optimality when its gap closes, and that the relaxation given by its initial simplex yields lower bounds within a tight multiplicative factor of $\frac{1}{n^2}$. Experimentally, we show that ISCA improves upon the state of the art as the ambient dimension grows: on the instances with $n \geq 4$, it reduces the mean optimality gap by 25% and it proves a nonzero lower bound on nine instances on which that of the state of the art is still zero, while the converse never occurs; with $n \in \{5, 6, 7\}$, it also solves more instances and is $1.7\times$ faster in geometric mean.

1 INTRODUCTION

Given m points $\{a_1, \ldots, a_m\}$ in $\mathbb{R}^n$, the *k-Hyperplane Clustering* problem, or k-HC, asks for identifying k hyperplanes $H_j := \{x \in \mathbb{R}^n : w_j^\top x = \gamma_j\}$, $j \in [k]$,[1] which minimize the sum of the squares of the Euclidean (2-norm) distances $\frac{|w_j^\top a_i - \gamma_j|}{\|w_j\|_2}$ between each point and the hyperplane closest to it. Differently from center-based clustering paradigms such as k-means (MacQueen, 1967), k-medians (Hakimi, 1964), and k-medoids (Kaufman & Rousseeuw, 1987) (see Ahmed et al. (2020); Ikotun et al. (2023); Miraftabzadeh et al. (2023) for surveys), which measure similarity as the distance between the points of a cluster and a center, k-HC models data lying close to a union of $(n-1)$-dimensional affine subspaces. It is thus a special case of projective (or subspace) clustering, which fits the data with affine subspaces of arbitrary dimension (Aggarwal et al., 1999; Aggarwal & Yu, 2000; Tseng, 2000; Zhang et al., 2009; Wang et al., 2009; Guo & Bai, 2023) and has applications in computer vision (Ma et al., 2008; Elhamifar & Vidal, 2013) and anomaly detection (Bacher et al., 2017; Ben-Gal et al., 2023). k-HC finds applications in sparse component analysis (He et al., 2009; Xu et al., 2017), hyperspectral image classification (Mirzaei, 2023), medical image segmentation (Kumar et al., 2021; 2023), dictionary learning (Zhang et al., 2013), and piecewise-affine regression (Manwani & Sastry, 2015; Yang et al., 2019); variants with other error measures are studied in Amaldi et al. (2013; 2016); Liu et al. (2016).

k-HC was introduced by Bradley & Mangasarian (2000), who also proposed an adaptation of the k-means heuristic to it, which alternates between assigning each point to its closest hyperplane and recomputing the hyperplanes in closed form. He & Cichocki (2007) extended it with an automatic detection of the number of hyperplanes, Amaldi & Coniglio (2013) proposed an exact mixed-integer formulation together with a heuristic that iteratively reassigns borderline points to neighboring clusters, and Tsakiris & Vidal (2017) proposed ℓ_1-based dual algorithms with global optimality

[1] Throughout the paper, we adopt the notation $[\xi] := \{1, \ldots, \xi\}$ for every $\xi \in \mathbb{N}$.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

guarantees for the recovery of a single hyperplane under suitable conditions. k-HC is $\mathcal{NP}$-hard, as fitting m points in $\mathbb{R}^n$ with k hyperplanes with zero error is $\mathcal{NP}$-complete even for $n = 2$ (Megiddo & Tamir, 1982). Since the normals can be scaled to unit norm and the resulting equality can be relaxed w.l.o.g. to an inequality (Bradley & Mangasarian, 2000; Amaldi & Coniglio, 2013), k-HC can be cast as the following mixed-integer nonlinear programming problem:

$$(k\text{-HC}) \qquad \min_{(w,\gamma)} \Big\{ \sum_{i\in[m]} \min_{j\in[k]} \big\{(a_i^\top w_j - \gamma_j)^2\big\} : \|w_j\|_2 \geq 1,\ j \in [k] \Big\}. \tag{1}$$

The constraints $\|w_j\|_2 \geq 1$ are reverse convex: the feasible region of each w_j is $\mathbb{R}^n \setminus \{w : \|w\|_2 < 1\}$. Spatial Branch and Bound (SBB) solvers typically handle them by introducing an auxiliary variable z_{jh} for each term w_{jh}^2, relaxing $z_{jh} = w_{jh}^2$ via its polyhedral envelope over the current bounds of w_{jh}, and imposing $\sum_{h\in[n]} z_{jh} \geq 1$ (Belotti et al., 2009). Over $w_{jh} \in [-1, 1]$, the envelope contains $(w_{jh}, z_{jh}) = (0, 1)$, so that the relaxation admits $w_j = 0$ which, with $\gamma_j = 0$, has objective value 0. Such a lower bound is trivial, and Coniglio (2026) shows that, with midpoint branching, a nonzero one is obtained only after generating at least $2^{k(n-1)}$ branching nodes. Coniglio (2026) mitigates this issue by strengthening the mixed-integer quadratically-constrained quadratic programming (MIQCQP) formulation of Amaldi & Coniglio (2013) with the constraints $\|w_j\|_1 \geq 1$ and $\|w_j\|_\infty \geq \frac{1}{\sqrt{n}}$, which are valid for $\|w_j\|_2 \geq 1$ and exclude $w_j = 0$.

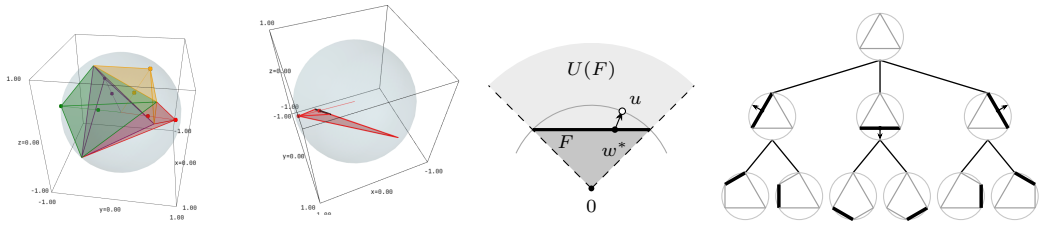
Contributions. We propose a method to solve k-HC to global optimality via SBB which exploits the geometry of its nonconvexity. The method, which we call *Iterative Simplicial-Complex Approximation* (ISCA), approximates the infeasible unit ball from the inside with a simplicial complex, starting from a regular simplex centered at the origin, and refines it within SBB by gluing to it, at each iteration, a simplex whose apex is the radial projection onto the unit sphere of the current relaxation solution. The approximation is thus refined only in the directions visited by the relaxation. We prove that ISCA is correct: each augmentation, i.e., each gluing of a new simplex, yields a valid simplicial complex, the degenerate children that arise are exactly characterized and redundant, and the branching never discards a feasible solution, so that ISCA certifies global optimality when its gap closes. We also show that the relaxation given by the initial simplex yields lower bounds within a tight factor of $\frac{1}{n^2}$. We implement ISCA in SCIP (Hojny et al., 2025) and compare it with the multi-norm method of Coniglio (2026): while the latter is faster in low dimension, ISCA overtakes it as the dimension grows while the number of hyperplanes remains small. On the instances with $n \geq 4$, ISCA reduces the mean gap by 25% (by 54% with $k = 2$), and it proves a nonzero lower bound on nine instances on which that of the multi-norm method is still zero, whereas the converse never occurs.

2 THE ISCA METHOD

Each constraint $\|w_j\|_2 \geq 1$ of (1) removes from the domain of w_j the open unit ball $\{w \in \mathbb{R}^n : \|w\|_2 < 1\}$, i.e., the interior of $B := \{w \in \mathbb{R}^n : \|w\|_2 \leq 1\}$. For each hyperplane $j \in [k]$, ISCA approximates this infeasible ball from the inside with its own (realization of a) finite simplicial complex $\mathcal{S}_j$, i.e., a finite collection of simplices closed under taking faces whose pairwise intersections are empty or common faces; for simplicity, we identify $\mathcal{S}_j$ with the union of its simplices. With k hyperplanes, ISCA thus works with the Cartesian product $\mathcal{S}_1 \times \cdots \times \mathcal{S}_k$ of k (realizations of) simplicial complexes, one per normal. As each of them only constrains its own normal, we first describe the method for $k = 1$, dropping the index j and denoting by $\mathcal{S}$ the complex of the normal w. For $Q \subseteq \mathbb{R}^n$, we call $\overline{Q} := \mathrm{cl}(\mathbb{R}^n \setminus Q)$ the *weak complement* of Q; in particular, $\overline{B} = \{w : \|w\|_2 \geq 1\}$ is the feasible region of w. If $\mathcal{S} \subseteq B$, then $\overline{\mathcal{S}} \supseteq \overline{B}$, and optimizing w over $\overline{\mathcal{S}}$ gives a relaxation of k-HC, which we call its *simplicial relaxation* w.r.t. $\mathcal{S}$. ISCA refines $\mathcal{S}$ iteratively: whenever the simplicial relaxation returns an infeasible w^*, it glues to $\mathcal{S}$ a new simplex that cuts w^* off, and it carries out these refinements within SBB. All proofs are reported in Appendix D.

2.1 GEOMETRIC DESCRIPTION OF THE ISCA METHOD

In ISCA, we choose as the initial approximation of the infeasible ball the regular n-simplex $\Delta = \mathrm{conv}\{v_1, \ldots, v_{n+1}\}$ inscribed in the unit sphere and centered at the origin. We choose a simplex as, among the polytopes containing the origin in their interior, the n-simplex is the one with the fewest vertices and facets, and, as its facets are $(n-1)$-simplices, the same construction applies at every



(a) (b) (c) (d)

Figure 1: (a) A nonconvex simplicial complex obtained by augmentation for $n = 3$. (b) The degenerate case for $n = 3$: w^* lies on the relative boundary of F^*, so that it also belongs to one of the new facets, whose affine hull contains the origin. (c) For $n = 2$, a facet F (thick), the region $\mathrm{conv}(F \cup \{0\})$ (dark), and the up-extension $U(F)$ of Section 2.2 (light), which contains the radial projection u of a relaxation solution $w^* \in F$. (d) An SBB tree for $n = 2$ and $k = 1$: the root has one child per facet of the initial simplex; at each child, the relaxation solution w^* lies on the facet of the node (thick), and the geometric branching splits that facet at u into two new facets, one per grandchild.

refinement. Its vertices are $v_i = \sqrt{(n+1)/n}\, U^\top (e_i - \frac{1}{n+1}\mathbf{1})$, $i \in [n+1]$, where e_i is the i-th unit vector, $\mathbf{1}$ is the all-ones vector, and $U \in \mathbb{R}^{(n+1)\times n}$ is the Helmert matrix, whose j-th column is $(1, \ldots, 1, -j, 0, \ldots, 0)^\top / \sqrt{j(j+1)}$ (with j ones), $j \in [n]$ (see Appendix C). We have $\|v_i\|_2 = 1$, $v_i^\top v_j = -1/n$ for $i \neq j$, and $\sum_i v_i = 0$. Hence, the facet of Δ opposite v_i lies on the hyperplane $v_i^\top w = -1/n$, $\overline{\Delta} = \bigcup_{i\in[n+1]} \{w : v_i^\top w \leq -1/n\}$, and, since $0 \in \operatorname{int} \Delta$, the solution $w = 0$ is infeasible from the start.

Given the current complex $\mathcal{S}$, ISCA solves the simplicial relaxation w.r.t. $\mathcal{S}$, i.e., k-HC with $w \in \overline{\mathcal{S}}$ in place of $\|w\|_2 \geq 1$. By positive homogeneity, it admits an optimal solution on the complex's boundary:

Lemma 1. *If $0 \in \mathcal{S} \subseteq B$, the simplicial relaxation of k-HC w.r.t. $\mathcal{S}$ admits an optimal solution (w^*, γ^*) with $w^* \in \partial\mathcal{S}$.*

If $\|w^*\|_2 = 1$, w^* is feasible for k-HC and, hence, optimal. Otherwise, ISCA identifies a facet $F^* = \mathrm{conv}\{p_1, \ldots, p_n\}$ of $\partial\mathcal{S}$, i.e., an $(n-1)$-simplex on its boundary, that contains w^* (which exists by Lemma 1), computes the radial projection $u := \pi(w^*) := w^*/\|w^*\|_2$ of w^* onto the unit sphere, and glues to $\mathcal{S}$ the simplex $S' := \mathrm{conv}(F^* \cup \{u\})$, whose base is F^* and whose apex is u; we call this operation an *augmentation* of $\mathcal{S}$ at w^*. In the new complex, F^* is no longer on the boundary: it is replaced by the other n facets of S', namely, $F'_i := \mathrm{conv}(\{u\} \cup \{p_\ell : \ell \neq i\})$, $i \in [n]$, each obtained from F^* by replacing its vertex p_i with u; see Figure 1(a). The new complex is, in general, nonconvex, and, along the ray through w^*, it coincides with B: as it contains the unit vector u, the relaxation now only admits normals of norm at least 1 in the direction of w^*, as k-HC does.

If w^* lies in the relative interior of F^*, it lies in the interior of the new complex and is cut off. If it lies on the relative boundary of F^*, which can only happen for $n > 2$ (for $n = 2$, the relative boundary of F^* consists of unit vectors), some of the new facets contain both w^* and u, so that their affine hull contains the origin (Figure 1(b)). We call such facets *degenerate*. Theorem 2 below shows that they can be dropped without loss of generality.

For $n = 2$, the picture is particularly simple: the facets are chords of the unit circle, and an augmentation at w^* inserts the point u of the arc subtended by the chord of w^*, replacing that chord by two shorter ones (Figure 1(d)); the approximation is thus refined, arc by arc, only where the relaxation needs it.

2.2 The SBB aspect of ISCA

Rather than refining a single complex, ISCA lets the SBB tree carry the refinements: each node corresponds to a facet F of the boundary of the complex built along the path from the root, the root has one child per facet of Δ, and a *geometric branching*, i.e., a branching on w, performs the augmentation on the facet of the node and creates one child per nondegenerate new facet, which

makes it a *multiway* branching, with up to n children rather than the usual two; see Figure 1(d). The tree also contains *assignment branchings*, on the binary variables x_{ij} that assign point i to hyperplane j in the formulation that ISCA solves (Section 4). Notably, while constraining the w of a node to its facet F would certainly be correct, it would be inefficient, as the regions of the children would not be contained in that of their parent, which would prevent the solver from warm-starting their relaxations. We constrain it, instead, to the *up-extension* $U(F) := \mathrm{cl}(\mathrm{cone}(F) \setminus \mathrm{conv}(F \cup \{0\}))$ of F (Figure 1(c)), which contains the up-extensions of all its descendants (Theorem 2). As the region of a child is obtained from that of its parent by adding $n+1$ inequalities, the optimal basis of the parent remains dual feasible for the child, and its relaxation can be reoptimized with the dual simplex method.

To impose $w \in U(F)$ at a node, we rely on an H-representation of $U(F)$ constructed from its V-representation. Let $E := [p_1 \cdots p_n]$ have the vertices of F as columns. As the vertices of F are affinely independent, F is degenerate if and only if E is singular. For a nondegenerate F, as $\mathrm{cone}(F) = \{E\lambda : \lambda \geq 0\}$ and $\mathrm{conv}(F \cup \{0\}) = \{E\lambda : \lambda \geq 0,\ \mathbf{1}^\top \lambda \leq 1\}$, we have $U(F) = \{E\lambda : \lambda \geq 0,\ \mathbf{1}^\top \lambda \geq 1\}$, whose vertices and extreme rays are the vertices of F.

In general, converting a V- into an H-representation requires eliminating λ, e.g., by Fourier–Motzkin elimination. Here, it only requires a substitution, as $\lambda = E^{-1}w$ is uniquely determined by w; we call it the *conic coordinates* of w w.r.t. F. Substituting it, we obtain $U(F) = \{w \in \mathbb{R}^n : E^{-1}w \geq 0,\ \mathbf{1}^\top E^{-1} w \geq 1\}$. The first n inequalities describe $\mathrm{cone}(F)$: the rows of E^{-1}, i.e., the dual basis of $p_1, \ldots, p_n$, are the normals of its facets, the i-th of which contains the origin and the vertices of F other than p_i. The last inequality, which we write as $\alpha^\top w \geq 1$ with $\alpha^\top := \mathbf{1}^\top E^{-1}$, is supported by the hyperplane through F.

The following theorem shows that a geometric branching is a valid SBB branching. Let the facet F^* of a node be refined at $u = w^*/\|w^*\|_2$, and let $\bar{\lambda} := E^{-1}u$ be the conic coordinates of u w.r.t. F^*. Then $\bar{\lambda}_i = 0$ if and only if the ray through w^* meets F^* on its face opposite p_i. Hence, as in Section 2.1, if the ray meets F^* in its relative interior, all the n new facets F'_i become children, whereas, if it meets its relative boundary, only the new facets F'_i with $\bar{\lambda}_i = 0$ are degenerate, i.e., have an affine hull that contains the origin (Figure 1(b)); they are discarded, and the other new facets become the children. Their up-extensions are contained in that of the parent, so that each child only adds constraints to it; their cones do not overlap, so that no direction is explored twice; and every feasible w in the region of the parent, i.e., with $\|w\|_2 \geq 1$, lies in the region of some child, so that no feasible solution is lost.

Theorem 2. *Let $F^* = \mathrm{conv}\{p_1, \ldots, p_n\}$ be a nondegenerate facet with unit vertices, let $u \in \mathrm{cone}(F^*)$ be a unit vector other than these vertices, and let $\bar{\lambda} := E^{-1}u$ be its conic coordinates w.r.t. F^*. For $i \in [n]$, let $F'_i := \mathrm{conv}(\{u\} \cup \{p_\ell : \ell \neq i\})$, and let $J := \{i \in [n] : \bar{\lambda}_i > 0\}$. Then:*

(i) valid augmentation: *$S' := \mathrm{conv}(F^* \cup \{u\})$ is an n-simplex contained in B, and $\mathrm{conv}(F^* \cup \{0\}) \cup S' = \bigcup_{i \in J} \mathrm{conv}(F'_i \cup \{0\})$;*

(ii) degeneracy: *F'_i is degenerate if and only if $\bar{\lambda}_i = 0$, i.e., if u lies on the facet of $\mathrm{cone}(F^*)$ that does not contain p_i; moreover, $|J| \geq 2$;*

(iii) partition and nesting: *the cones $\mathrm{cone}(F'_i)$, $i \in J$, cover $\mathrm{cone}(F^*)$ and have pairwise disjoint interiors, and $U(F'_i) \subseteq U(F^*)$ for all $i \in J$;*

(iv) no feasible point is lost: *every $w \in U(F^*)$ with $\|w\|_2 \geq 1$ belongs to $U(F'_i)$ for some $i \in J$.*

The hypothesis on u holds in ISCA: as $w^* \in U(F^*)$ and $\|w^*\|_2 < 1$, w^* is not a multiple tp_i of a vertex, since these belong to $U(F^*)$ only for $t \geq 1$.

In Section 2.1, ISCA solves the simplicial relaxation w.r.t. a single complex $\mathcal{S}$, whereas each node of the tree only sees one facet. The following theorem shows that the two views coincide: with a single hyperplane and no assignment branching, the facets of the active nodes are the boundary facets of a single complex $\mathcal{S}$ (Theorem 2(iii)), and their regions partition $\overline{\mathcal{S}}$, so that the tree solves the simplicial relaxation split into one piece per facet and explores no region twice.

Theorem 3. *Let $\mathcal{S}$ be a complex obtained from Δ by finitely many augmentations, and let $\mathcal{F}$ be the set of the nondegenerate facets of its boundary. Then the up-extensions $U(F)$, $F \in \mathcal{F}$, have pairwise disjoint interiors, and their union is the weak complement $\overline{\mathcal{S}}$ of $\mathcal{S}$.*

With k hyperplanes, the geometric region of a node is a product $U(F_1) \times \cdots \times U(F_k)$, where F_j is a facet of the complex $\mathcal{S}_j$ built along the path from the root; the *root disjunction*, whose regions

are the products of up-extensions of facets of Δ, has $(n+1)^k$ regions, a geometric branching on w_j replaces only the j-th factor, and an assignment branching is an ordinary binary disjunction. Under these rules, the following holds:

Remark 4. ISCA is a valid SBB method: the regions of the root disjunction cover the feasible set of (1) (Theorem 3), a branching on x_{ij} partitions the feasible solutions of its node, and a geometric branching keeps all of them (Theorem 2(iv)). Hence, every feasible solution belongs to the region of an active node or of a node pruned by bound, so that $\min\{\mathrm{LB}, \mathrm{UB}\} \leq \mathrm{OPT} \leq \mathrm{UB}$, where OPT is the optimal value of (1), UB the incumbent value, and LB the smallest lower bound of the active nodes, and ISCA certifies global optimality when $\mathrm{UB} = \mathrm{LB}$.

3 An approximation guarantee

We now bound the quality of the relaxation given by the initial simplex, i.e., in the regions of the root disjunction. The result concerns k-HC with integral assignments, which are implicit in the $\min_j$ below; with fractional x, the relaxation of Section 4 can still have value 0. For a compact convex set $K \subset \mathbb{R}^n$ with $0 \in \operatorname{int} K$, let $p_K(x) := \min\{t \geq 0 : x \in tK\}$ be its gauge, $h_K(y) := \max_{x \in K} x^\top y$ its support function, and $K^\circ := \{y : h_K(y) \leq 1\}$ its polar; then $p_K = h_{K^\circ}$ (Rockafellar, 2015). In the regions of the root disjunction, i.e., in the simplicial relaxation w.r.t. Δ, each $\|w_j\|_2 \geq 1$ is replaced by $w_j \in \overline{\Delta}$, i.e., by $p_\Delta(w_j) \geq 1$, and the following lemma gives p_Δ in closed form:

Lemma 5. $\Delta^\circ = \operatorname{conv}\{-nv_1, \ldots, -nv_{n+1}\}$. *Hence, the gauge of* Δ *is* $p_\Delta(x) = \max_{i \in [n+1]}\{-n\, v_i^\top x\}$.

p_Δ is positively homogeneous and subadditive but not symmetric—hence, it is not a norm (while being a gauge), as norms require symmetry. The next lemma shows that a normal w with $p_\Delta(w) \geq 1$ has $\|w\|_2 \geq \frac{1}{n}$:

Lemma 6. *For all* $x \in \mathbb{R}^n$, $\|x\|_2 \leq p_\Delta(x) \leq n\|x\|_2$. *Moreover,* $\min\{\|x\|_2 : p_\Delta(x) = 1\} = 1/n$.

The next theorem shows that nw is feasible for k-HC at the cost of a factor n^2 in the objective, and that the factor n is tight:

Theorem 7. *For* $K \in \{B, \Delta\}$, *let* $\mathrm{OPT}_K := \min\{\sum_{i \in [m]} \min_{j \in [k]} (a_i^\top w_j - \gamma_j)^2 : p_K(w_j) \geq 1 \; \forall j \in [k]\}$, *where* $p_B = \|\cdot\|_2$. *Then* $\frac{1}{n^2}\mathrm{OPT}_B \leq \mathrm{OPT}_\Delta \leq \mathrm{OPT}_B$, *and the factor* $\frac{1}{n^2}$ *is tight.*

In particular, with the assignment fixed, the root disjunction gives a nonzero lower bound whenever $\mathrm{OPT}_B > 0$, and, as every complex generated by ISCA contains Δ, the guarantee holds at every level. For comparison, with midpoint branching, standard SBB needs at least $2^{k(n-1)}$ branching nodes to exclude $w_j = 0$ (Coniglio, 2026), whereas ISCA excludes it with the $(n+1)^k$ regions of its root disjunction, whose number is polynomial in n for a fixed k.

Differently from a fixed approximation, though, an augmentation at w^* makes the complex exact in the direction of w^*, as it contains $u = w^*/\|w^*\|_2$: ISCA improves the approximation precisely in the directions exploited by the relaxation. This improvement can be quantified locally in terms of the diameter $h(F) := \max\{\|p - p'\|_2 : p, p' \text{ vertices of } F\}$ of the facet F of a node, which shrinks as the node is refined, by the following bound, which is void in the root disjunction, where $h(F) = \sqrt{2 + 2/n} > \sqrt{2}$:

Proposition 8. *Let* $F = \operatorname{conv}\{p_1, \ldots, p_n\}$ *be a facet with unit vertices. Then every* $w \in U(F)$ *satisfies* $\|w\|_2^2 \geq 1 - \frac{1}{2}h(F)^2$. *Consequently, if* $h(F_j) \leq h < \sqrt{2}$ *for the facets* F_j, $j \in [k]$, *of a node, then, for every fixed assignment, the relaxation of the node is within a factor* $1 - \frac{1}{2}h^2$ *of the optimal value of* k*-HC restricted to the region of the node.*

In particular, if the diameters of the facets vanish along a branch of the tree, its node relaxations become exact: with the assignment fixed, the part of the gap of a node that is due to the geometry closes as ISCA refines its facets, and the remaining gap is due to the assignment.

4 Implementation of ISCA

We implement ISCA in C++ on top of SCIP 10.0.3 (Hojny et al., 2025) with SoPlex 8.0.3 as LP solver. SCIP manages the search tree, with presolving and primal heuristics turned off.

At each node, ISCA solves the relaxation of the following formulation, derived from the MI-QCQP formulation of Amaldi & Coniglio (2013) by replacing $\|w_j\|_2 \geq 1$ with $w_j \in U(F_j)$, where F_j is the facet of the node for hyperplane j:

$$\min_{(w,\gamma),x,d} \left\{ \sum_i d_i^2 : \sum_j x_{ij} = 1,\ d_i \geq |w_j^\top a_i - \gamma_j| - M(1-x_{ij}),\ w_j \in U(F_j),\ x_{ij} \in \{0,1\} \right\}, \tag{2}$$

with $i \in [m]$ and $j \in [k]$, complemented with the bounds $w_j \in [-1,1]^n$, $|\gamma_j| \leq \max_{i\in[m]} \|a_i\|_2$, and $0 \leq d_i \leq M$, and with $x_{ij} = 0$ for $j > i$ to break symmetry. The absolute values are formulated with two linear inequalities, and the objective via an epigraph variable $\theta \geq \sum_{i\in[m]} d_i^2$, which SCIP outer-approximates with linear cuts. We set $M := D := \max_{i,\ell\in[m]} \|a_i - a_\ell\|_2$, the diameter of the dataset; this and the bounds on w, γ, and d preserve an optimal solution:

Lemma 9. *Let $M \geq D$. For every assignment of the points to the hyperplanes, k-HC restricted to that assignment has an optimal solution with $\|w_j\|_2 = 1$, $|\gamma_j| \leq \max_{i\in[m]} \|a_i\|_2$, and $|w_j^\top a_i - \gamma_j| \leq M$ for all $i \in [m]$, $j \in [k]$. Hence, Formulation* (2) *with $\|w_j\|_2 \geq 1$ in place of $w_j \in U(F_j)$ has the same optimal value as k-HC.*

As long as the assignment is fractional, however, the relaxation of (2) is weak regardless of w: if, e.g., the points lie in a ball of radius $\frac{M}{2}$ centered at the origin and symmetry is not broken, $x_{ij} = \frac{1}{k}$, $\gamma_j = 0$, and $d_i = 0$ are feasible for every $w_j \in [-1,1]^n$ with $\|w_j\|_2 \leq 1$, since $|w_j^\top a_i| \leq \frac{M}{2} \leq M(1-\frac{1}{k})$ for $k \geq 2$. In our experiments, the lower bound typically stays at zero for thousands of nodes.

ISCA thus interleaves assignment and geometric branching. ISCA first rotates Δ, as a heuristic that adapts the orientation of its root disjunction to the data rather than to the arbitrary coordinates in which Δ is constructed: it uses the vertices $R^\top v_i$, $i \in [n+1]$, where the columns of the orthogonal matrix R are the eigenvectors of $\sum_{i\in[m]} a_i a_i^\top$ in order of decreasing eigenvalue (the rotated simplex is still regular and inscribed in the unit sphere, so that the rotation affects the search tree, not the correctness). ISCA then branches, one hyperplane after the other, on the $n+1$ facets of the rotated simplex, so that the $(n+1)^k$ regions of the root disjunction are the leaves of the first k levels. Afterwards, at a node where both a fractional x_{ij} and a w_j with $\|w_j\|_2 < 1$ exist, it branches on x as long as $\mathrm{LB} \leq 10^{-4}\,\mathrm{UB}$ and, otherwise, it branches geometrically with probability $\min\{\mathrm{Gap}, 1-\mathrm{Gap}\}$, where $\mathrm{Gap} := \frac{\mathrm{UB}-\mathrm{LB}}{\mathrm{UB}}$, and on x otherwise; if only one of the two exists, ISCA branches on it.

A geometric branching refines the facet of the w_j of smallest norm, i.e., of the hyperplane whose normal is the farthest from being feasible. With F this facet, E its vertex matrix, and w_j^* the relaxation solution, ISCA computes the radial projection $u := w_j^*/\|w_j^*\|_2$, its conic coordinates $\bar{\lambda} := E^{-1}u$, and $J := \{i \in [n] : \bar{\lambda}_i > 0\}$, and it creates one child per $i \in J$. The facet F_i' of the i-th child has vertex matrix E_i', i.e., E with its i-th column replaced by u, and the child receives the $n+1$ inequalities $(E_i')^{-1}w_j \geq 0$ and $\mathbf{1}^\top(E_i')^{-1}w_j \geq 1$ as local constraints. As u is a vertex of every F_i', the conic coordinates of $w_j^* = \|w_j^*\|_2\, u$ w.r.t. each F_i' sum to $\|w_j^*\|_2 < 1$, so that w_j^* is cut off at every child. The choice of the x_{ij} is left to SCIP, as is the node selection (SCIP's default best-estimate rule).

The interleaving ends at the nodes where all the assignment variables are fixed. There, the subproblem decomposes into k problems $\min\{\|A_j w_j - \gamma_j \mathbf{1}\|_2^2 : \|w_j\|_2 = 1\}$, where the rows of A_j are the m_j points of cluster j. By the closed form of Bradley & Mangasarian (2000), an optimal w_j is a unit eigenvector of the smallest eigenvalue of $C_j := A_j^\top(I - \frac{1}{m_j}\mathbf{1}\mathbf{1}^\top)A_j$, with $\gamma_j = \frac{1}{m_j}\mathbf{1}^\top A_j w_j$, and the node has value $z := \sum_{j\in[k]} \lambda_{\min}(C_j)$. At a node with a fixed assignment, it uses this solution as an incumbent and closes the node if the incumbent exceeds a lower bound $\sum_{j\in[k]} \eta_j \leq z$ by at most $10^{-7}\max\{1, z\}$. Its first choice of η_j is the eigenvalue computed by LAPACK minus the residual of the eigendecomposition and a margin for rounding errors.

If this is not tight enough, ISCA certifies the η_j in exact rational arithmetic: as the input coordinates are binary64 numbers, i.e., dyadic rationals, for every binary64 $\eta \geq 0$ the matrix $2^h m_j(C_j - \eta I)$ is integer for some $h \in \mathbb{N}$, and $\eta < \lambda_{\min}(C_j)$ holds if and only if its n leading principal minors are positive (Sylvester's criterion), which ISCA checks exactly with the fraction-free elimination of Bareiss (1968) on arbitrary-precision integers. It tries the LAPACK eigenvalue first, then bisects on $[0, \eta]$, and rounds the exact sum of the η_j downward. In our experiments, the exact certificate is invoked in more than half of the runs of ISCA, and it closes every node on which it is invoked.

The inequalities of $U(F)$ are computed with an LU (or, for ill-conditioned E, an SVD) factorization and are accepted only if they satisfy their defining equations up to $5 \cdot 10^{-8}$; otherwise, ISCA branches on x instead, so that no node is discarded because of numerical difficulties. The initial incumbent is the best solution found by the heuristic of Bradley & Mangasarian (2000) from 100 random starts.

Two further ingredients complete ISCA. First, ISCA uses *node-local Big-M* constants: it replaces M by the per-point constants $M_i := \max_{\ell \in [m]} \|a_i - a_\ell\|_2$ and, at each node, it adds the constraint $d_i \geq |w_j^\top a_i - \gamma_j| - M_{ij}(1 - x_{ij})$ as a local cut whenever $M_{ij} := \max_{\ell \in I_j} \|a_i - a_\ell\|_2 < \rho D$, for a given $\rho \in (0, 1]$, where I_j contains the points whose assignment to hyperplane j is not fixed to zero. By the argument in the proof of Lemma 9 (choosing $\gamma_j := a_p^\top w_j$ for some $p \in I_j$ if cluster j is empty), every assignment compatible with the node has an optimal solution satisfying these cuts.

Second, ISCA handles the facets that become *nearly degenerate* deep in the tree, i.e., whose affine hull $\{w : \alpha^\top w = 1\}$ has distance $\frac{1}{\|\alpha\|_2} < \tau$ from the origin for a given threshold $\tau > 0$, and on which further refinement makes little progress. Differently from degenerate facets, nearly degenerate facets are not covered by their siblings, so a node with a nearly degenerate facet can only be discarded if the reported lower bound is corrected to $\min\{\mathrm{LB}, \min_{c \in P} L_c\}$, where P is the set of discarded nodes and L_c is the lower bound of $c \in P$ when it was discarded:

Proposition 10. *With the notation above,* $\min\{\mathrm{LB}, \mathrm{UB}, \min_{c \in P} L_c\} \leq \mathrm{OPT}$.

To keep the correction small, ISCA discards such a node only once $L_c \geq (1 - \varepsilon)\,\mathrm{UB}$, for a given gap budget $\varepsilon \in (0, 1]$, so that the correction costs at most a relative gap of ε. Moreover, ISCA defers this decision, which does not affect the validity of the bound: a node whose facet is found nearly degenerate for the first time is replaced by a copy of itself that is selected only when no other active node is left, so that it is decided against a better incumbent.

5 Computational results

We compare ISCA with `MultiNorm`, the formulation of Coniglio (2026) that strengthens the MI-QCQP formulation with both $\|w_j\|_1 \geq 1$ and $\|w_j\|_\infty \geq \frac{1}{\sqrt{n}}$, $j \in [k]$. Both methods run on SCIP with SoPlex. Every run uses a single core and a time limit of 7,200 seconds. `MultiNorm` runs with SCIP's default settings, keeping the nonconvex constraints $\|w_j\|_2^2 \geq 1$ (handled via spatial branching) and formulating the polyhedral norm constraints with binary variables as done in (Coniglio, 2026). `MultiNorm` is deterministic and is run once per instance on a computing cluster whose machines have Intel Xeon Platinum 8260 CPUs (2.4 GHz), with 32 GB of memory per run. ISCA is randomized (through its choice between geometric and assignment branching) and is run with several seeds per setting of its hyperparameters, either on the same cluster or on a faster one with Intel Xeon Platinum 8358 CPUs (2.6 GHz) and 180 GB of memory per run. For fairness, we multiply every computing time measured on the second cluster by 1.4, as it is between 1.36 and 1.52 times faster on identical `MultiNorm` search trees; a run on the faster cluster that solves its instance to optimality at time t thus counts as such only if $1.4\,t \leq 7{,}200$, and it is otherwise scored with the bound it had reached at time $7{,}200/1.4$.

A run counts as solved only if SCIP reports optimality and the relative gap $|\mathrm{UB} - \mathrm{LB}|/|\mathrm{UB}|$ between the incumbent and the valid lower bound of Proposition 10 is at most 10^{-6}; unsolved runs are charged 7,200 seconds. Following Coniglio (2026), we use synthetic low-noise instances generated with the procedure of Amaldi & Coniglio (2013) (Appendix A): `LowDim` (16 instances) has $n, k \in \{2, 3\}$ and `HighDim` (24 instances) extends n and k up to 5; we also introduce `SuperHighDim` (24 instances), which keeps $k \in \{2, 3\}$ and increases the dimension to $n \in \{5, 6, 7\}$.

We run ISCA with and without its two further ingredients, and we select them per instance among three settings: A (neither the node-local Big-M nor the treatment of nearly degenerate facets), B (node-local Big-M with $\rho = 0.7$), and C (both, with $\rho = 0.9$, $\tau = 10^{-4}$, $\varepsilon = 10^{-6}$, and deferral).

Tables 1 and 2 (and Table 4 in Appendix B, for `LowDim`) report, for each instance, the results of `MultiNorm` and of the ISCA setting with the smallest mean gap over its seeds, with ties broken by the smallest mean time and then by the fewest mean nodes. Every entry is an average over the seeds of the selected setting. As the setting is selected per instance, the ISCA columns are a per-instance best, and a single fixed setting does worse: with setting A alone, ISCA solves to optimality, with all of its seeds, 12, 10, and 3 instances of `LowDim`, `HighDim`, and `SuperHighDim`, respectively.

Table 1: `HighDim` (24 instances): `MultiNorm` vs. `ISCA`. For each instance, the `ISCA` columns report its best setting (column *Set.*). Times in seconds, with the runs on the faster cluster scaled by 1.4. Time, nodes, and gap are means over the seeds, charging 7,200 s to every run that does not solve its instance to optimality; an instance counts as solved to optimality only if every seed solves it. Gap: $|\mathrm{UB} - \mathrm{LB}|/|\mathrm{UB}|$, capped at 100%. Footer: geometric means of time and nodes, and mean gap.

				MultiNorm			ISCA			
Inst.	m	n	k	**Time**	**Nodes**	**Gap**	**Set.**	**Time**	**Nodes**	**Gap**
008	11	4	2	24	24,586	0.0%	A	4	3,832	0.0%
009	12	4	2	30	25,843	0.0%	A	3	2,753	0.0%
000	13	2	4	20	16,503	0.0%	A	272	297,706	0.0%
010	13	4	2	100	84,768	0.0%	A	181	202,669	0.0%
012	13	5	2	308	277,420	0.0%	A	14	12,271	0.0%
001	14	2	4	25	20,710	0.0%	B	106	147,252	0.0%
011	14	4	2	37	28,275	0.0%	C	16	12,730	0.0%
013	14	5	2	286	279,985	0.0%	B	73	72,485	0.0%
002	15	2	4	83	80,444	0.0%	B	366	502,076	0.0%
014	15	5	2	366	346,375	0.0%	A	37	29,945	0.0%
003	16	2	4	40	27,832	0.0%	A	2154	1,094,131	0.0%
004	16	2	5	544	711,689	0.0%	A	7200	6,585,811	92.7%
020	16	4	3	7200	5,340,780	100.0%	B	2935	4,214,610	0.0%
015	16	5	2	503	418,182	0.0%	C	2761	2,118,871	17.0%
005	17	2	5	7200	6,439,692	73.0%	A	7200	6,885,117	100.0%
016	17	3	4	7200	6,143,158	100.0%	B	7200	10,034,432	100.0%
021	17	4	3	7200	6,511,510	100.0%	C	5217	8,420,482	17.7%
006	18	2	5	1232	1,459,526	0.0%	A	7200	6,034,975	79.8%
017	18	3	4	7200	5,152,428	100.0%	B	7200	9,038,473	100.0%
022	18	4	3	7200	6,191,725	100.0%	A	6116	8,663,155	33.0%
007	19	2	5	1374	1,412,870	0.0%	B	7200	6,427,442	100.0%
018	19	3	4	7200	5,067,984	100.0%	B	7200	7,423,209	100.0%
023	19	4	3	7200	5,454,610	100.0%	B	7200	7,771,733	92.5%
019	20	3	4	7200	5,089,020	100.0%	B	7200	6,537,280	100.0%
Mean				592.3	514,578	36.4%		720.0	738,131	38.9%

Table 2: `SuperHighDim` (24 instances): `MultiNorm` vs. `ISCA`; same layout as Table 1.

				MultiNorm			ISCA			
Inst.	m	n	k	**Time**	**Nodes**	**Gap**	**Set.**	**Time**	**Nodes**	**Gap**
000	13	5	2	262	287,287	0.0%	A	21	17,570	0.0%
001	14	5	2	440	395,125	0.0%	B	177	270,139	0.0%
002	15	5	2	417	382,623	0.0%	B	3879	3,536,547	0.0%
008	15	6	2	3883	3,912,155	0.0%	C	60	54,023	0.0%
003	16	5	2	688	595,484	0.0%	C	681	568,390	0.0%
009	16	6	2	4740	3,686,161	0.0%	B	134	94,024	0.0%
010	17	6	2	5120	3,622,022	0.0%	C	4894	6,894,315	32.2%
016	17	7	2	7200	4,669,381	100.0%	C	1497	821,820	0.0%
011	18	6	2	7200	5,586,078	100.0%	B	7200	5,999,279	53.6%
017	18	7	2	7200	4,492,424	100.0%	B	589	387,360	0.0%
004	19	5	3	7200	4,111,395	100.0%	B	7200	6,702,381	100.0%
018	19	7	2	7200	4,250,587	100.0%	B	7200	5,420,312	63.9%
005	20	5	3	7200	3,802,012	100.0%	B	7200	7,176,478	100.0%
019	20	7	2	7200	4,071,212	100.0%	A	7200	6,994,086	65.1%
006	21	5	3	7200	2,717,753	100.0%	B	7200	6,186,791	100.0%
007	22	5	3	7200	3,035,444	100.0%	B	7200	5,753,480	100.0%
012	22	6	3	7200	2,572,368	100.0%	B	7200	5,875,039	100.0%
013	23	6	3	7200	2,644,332	100.0%	A	7200	5,552,211	100.0%
014	24	6	3	7200	2,402,131	100.0%	A	7200	4,748,604	100.0%
015	25	6	3	7200	2,347,285	100.0%	B	7200	4,488,838	100.0%
020	25	7	3	7200	1,708,053	100.0%	B	7200	5,003,649	100.0%
021	26	7	3	7200	2,178,010	100.0%	B	7200	4,594,536	100.0%
022	27	7	3	7200	2,909,772	100.0%	B	7200	4,147,695	100.0%
023	28	7	3	7200	2,046,654	100.0%	B	7200	3,446,627	100.0%
Mean				4244.7	2,255,400	70.8%		2461.6	1,952,857	59.0%

Both ingredients contribute: selecting between settings A and B raises these numbers to 13, 12, and 6, and selecting among all three settings to 13, 12, and 8. The best setting depends on the family: among the instances solved to optimality by ISCA, it is A or B on eleven of the twelve of `HighDim` and, on `SuperHighDim`, C on three of the eight, including instance 016, which no other setting that was run on it solves.

On `LowDim` and `HighDim`, `MultiNorm` is the stronger method. On `LowDim` (Table 4), it solves 15 of the 16 instances to optimality against 13 for ISCA, it is about $2.8\times$ faster and explores about $5\times$ fewer nodes in geometric mean, and its mean gap is 4.8% against 11.9%. On `HighDim`, it solves 15

of the 24 instances to optimality against 12, and it is about $1.2\times$ faster. `HighDim`, however, already shows a split by dimension. On the eight instances with $n \in \{4, 5\}$ and $k = 2$, ISCA is faster on six, by up to an order of magnitude (14 against 308 seconds on instance 012, and 37 against 366 on instance 014), and it is the only method that solves instance 020 ($n = 4$, $k = 3$) to optimality. Conversely, on the eight instances with $n = 2$ and $k \in \{4, 5\}$, `MultiNorm` is faster on seven (both methods time out on the eighth), and ISCA does not solve four of them. This is consistent with the size of the root disjunction of ISCA, $(n + 1)^k$, which, for a fixed k, is polynomial in n but grows exponentially in k.

On `SuperHighDim`, the ranking reverses: ISCA solves 8 of the 24 instances to optimality against 7 for `MultiNorm`, and it is $1.72\times$ faster in geometric mean while exploring only $1.15\times$ fewer nodes. The difference comes from the twelve instances with $k = 2$, where ISCA solves instances 016 and 017 ($n = 7$) to optimality, which `MultiNorm` does not, while `MultiNorm` solves instance 010 to optimality, which ISCA does not.

On the six instances solved by both, ISCA is faster on five (on instance 003, only marginally), by up to more than an order of magnitude (60 against 3,883 seconds on instance 008, and 134 against 4,740 on instance 009), and it is about $9\times$ slower on instance 002. With $k = 3$, neither method solves any instance, and both end with a lower bound of zero. As both relaxations are weak under fractional assignments (Section 4), the assignment rather than the geometry is the bottleneck here. The speed-up of ISCA thus grows with the dimension: $0.36\times$, $0.82\times$, and $1.72\times$ on `LowDim`, `HighDim`, and `SuperHighDim`.

The largest gains of ISCA are in the lower bound. On the 36 instances with $n \geq 4$, it reduces the mean gap from 58.3% to 43.8%, i.e., by 25%, and, on the 20 of them with $k = 2$, from 25.0% to 11.6%, i.e., by 54%, with a nonzero lower bound on each of them. On nine instances with $n \geq 4$, ISCA proves a nonzero lower bound while that of `MultiNorm` is still zero: it solves three of them to optimality (020 of `HighDim`, 016 and 017 of `SuperHighDim`) and ends with gaps between 18% and 93% on the other six (021–023 of `HighDim`, with $n = 4$ and $k = 3$, and 011, 018, and 019 of `SuperHighDim`, with $k = 2$). The converse never occurs with $n \geq 4$, and it occurs only on two instances with $n = 2$ and $k = 5$ (005 and 007 of `HighDim`). ISCA also finds the better incumbent on 19 of the 24 instances that neither method solves.

6 Concluding remarks

We have proposed ISCA, a method to solve k-HC to global optimality via SBB which approximates the infeasible unit ball from the inside with a simplicial complex that is refined, within SBB, only in the directions visited by the relaxation. We have proved that ISCA certifies global optimality when its gap closes, and that the relaxation given by its initial simplex yields lower bounds within a tight factor of $\frac{1}{n^2}$; in particular, ISCA excludes $w_j = 0$ with the $(n + 1)^k$ regions of its root disjunction, rather than the at least $2^{k(n-1)}$ nodes of standard SBB. In our implementation on top of SCIP, the leaves of the tree are closed with lower bounds certified in exact rational arithmetic. Our experiments show that the multi-norm method of Coniglio (2026) is faster in low dimension, whereas ISCA overtakes it as the dimension grows while the number of hyperplanes remains small: with $n \geq 4$, it reduces the mean gap by 25%, and it proves nonzero lower bounds where those of the multi-norm method are still zero.

To face the natural limitations of this research, as directions for future work we identify the extension of ISCA to other problems featuring reverse-convex constraints. The construction of Section 2 and the proof of Theorem 2 carry over to constraints of the form $p_K(w) \geq 1$, where K is a compact convex set with $0 \in \operatorname{int} K$ and p_K is its gauge, once the radial projection $w \mapsto w/\|w\|_2$ is replaced by $w \mapsto w/p_K(w)$; moreover, as the up-extension inequalities are linear, w may be any affine function of the decision variables. This covers, e.g., the Euclidean no-good cuts $\|x - \bar{x}\|_2 \geq \varepsilon$ that exclude a ball around a previously found solution $\bar{x}$, e.g., to enumerate distinct near-optimal solutions, as well as the minimum-distance constraints $\|x_i - x_\ell\|_2 \geq d$ that arise in packing problems and in facility location with separation requirements, with one complex per pair of objects, as ISCA keeps one per hyperplane. Another candidate is the 2-norm classification problem of Mangasarian (1999), whose objective, homogeneous of degree one, turns the factor of Theorem 7 into $\frac{1}{n}$.

AI USE STATEMENT

We used generative AI tools (large language models, through chat interfaces and as coding agents) in the following ways. *Writing:* to edit and polish the text and to draft passages of several sections from the authors' notes, results, and theorem statements; every sentence was revised by the authors. *Mathematical claims and proofs:* to review the draft statements and proofs, to suggest proof steps, corrections, and counterexamples, and to polish their presentation; every statement and every step of every proof in Appendix D was verified by the authors, who are responsible for their correctness. *Implementation:* to port our earlier Python implementation of ISCA to C++ on top of SCIP and to extend it, and to write and debug the scripts that run and analyze the experiments, under the authors' direction; the design of ISCA and of the experiments and the interpretation of the results are the authors'. We have not used generative AI tools to generate the instances, and translation, data cleaning, and qualitative data analysis are not applicable to this work. All reported numbers were produced by the authors' code and checked against the raw logs. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

All the datasets employed in this work are synthetic and contain no personally identifiable information or harmful content. The methods introduced in this paper have a societal impact comparable to that of any other clustering algorithm.

REPRODUCIBILITY STATEMENT

Section 4 describes the implementation of ISCA, Section 5 and Appendix A describe the testbed and the experimental setting, and Appendix D contains complete proofs of all our theoretical results. The source code, the instances, and the raw results of every run will be made available online and provided as an anonymized repository to the reviewers upon request and a link to a repository will be included upon acceptance.

REFERENCES

C. C. Aggarwal and P. S. Yu. Finding generalized projected clusters in high dimensional spaces. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, SIGMOD/PODS00, pp. 70–81. ACM, May 2000. doi: 10.1145/342009.335383.

C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, and J. S. Park. Fast algorithms for projected clustering. *ACM SIGMOD Record*, 28(2):61–72, June 1999. ISSN 0163-5808. doi: 10.1145/304181.304188.

M. Ahmed, R. Seraj, and S. M. S. Islam. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8):1295, August 2020. ISSN 2079-9292. doi: 10.3390/electronics9081295.

E. Amaldi and S. Coniglio. A distance-based point-reassignment heuristic for the k-hyperplane clustering problem. *European Journal of Operational Research*, 227(1):22–29, May 2013. ISSN 0377-2217. doi: 10.1016/j.ejor.2012.09.026.

E. Amaldi, K. Dhyani, and L. Liberti. A two-phase heuristic for the bottleneck k-hyperplane clustering problem. *Computational Optimization and Applications*, 56(3):619–633, May 2013. ISSN 1573-2894. doi: 10.1007/s10589-013-9567-2.

E. Amaldi, S. Coniglio, and L. Taccari. Discrete optimization methods to fit piecewise affine models to data points. *Computers & Operations Research*, 75:214–230, November 2016. ISSN 0305-0548. doi: 10.1016/j.cor.2016.05.001.

M. Bacher, I. Ben-Gal, and E. Shmueli. An information theory subspace analysis approach with application to anomaly detection ensembles. In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 27–39. SCITEPRESS - Science and Technology Publications, 2017. doi: 10.5220/0006479000270039.

E. H. Bareiss. Sylvester's identity and multistep integer-preserving Gaussian elimination. *Mathematics of Computation*, 22(103):565–578, 1968. doi: 10.1090/S0025-5718-1968-0226829-0.

P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. Branching and bounds tighteningtechniques for non-convex minlp. *Optimization Methods and Software*, 24(4–5):597–634, October 2009. ISSN 1029-4937. doi: 10.1080/10556780903087124.

I. Ben-Gal, M. Bacher, M. Amara, and E. Shmueli. A nonparametric subspace analysis approach with application to anomaly detection ensembles. *INFORMS Journal on Data Science*, 2(2):99–115, October 2023. ISSN 2694-4030. doi: 10.1287/ijds.2023.0027.

P. S. Bradley and O. L. Mangasarian. k-plane clustering. *Journal of Global Optimization*, 16(1): 23–32, 2000. ISSN 0925-5001. doi: 10.1023/a:1008324625522.

S. Coniglio. Solving the 2-norm k-hyperplane clustering problem via multi-norm formulations. In *The Fourteenth International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=VJAqqtVXfD`.

E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, November 2013. ISSN 2160-9292. doi: 10.1109/tpami.2013.57.

Y. R. Guo and Y. Q. Bai. Two-dimensional k-subspace clustering and its applications on image recognition. *International Journal of Machine Learning and Cybernetics*, 14(8):2671–2683, February 2023. ISSN 1868-808X. doi: 10.1007/s13042-023-01790-0.

S. L. Hakimi. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459, June 1964. ISSN 1526-5463. doi: 10.1287/opre.12.3.450.

Z. He and A. Cichocki. *An Efficient K-Hyperplane Clustering Algorithm and Its Application to Sparse Component Analysis*, pp. 1032–1041. Springer Berlin Heidelberg, 2007. ISBN 9783540723936. doi: 10.1007/978-3-540-72393-6_122.

Z. He, A. Cichocki, Y. Li, S. Xie, and S. Sanei. K-hyperline clustering learning for sparse component analysis. *Signal Processing*, 89(6):1011–1022, June 2009. ISSN 0165-1684. doi: 10.1016/j.sigpro.2008.12.005.

C. Hojny, M. Besançon, K. Bestuzheva, S. Borst, J. Dionísio, J. Ehls, L. Eifler, M. Ghannam, A. Gleixner, A. Göß, et al. The scip optimization suite 10.0. *arXiv preprint arXiv:2511.18580*, 2025.

A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622:178–210, April 2023. ISSN 0020-0255. doi: 10.1016/j.ins.2022.11.139.

L. Kaufman and P. J. Rousseeuw. Clustering by means of medoids. In Yadolah Dodge (ed.), *Statistical Data Analysis Based on the L_1 Norm and Related Methods*, pp. 405–416, Amsterdam, 1987. North-Holland.

P. Kumar, D. Kumar, and R. K. Agrawal. Fuzzy k-plane clustering method with local spatial information for segmentation of human brain MRI image. *Neural Computing and Applications*, 34 (6):4855–4874, November 2021. ISSN 1433-3058. doi: 10.1007/s00521-021-06677-1.

P. Kumar, R. K. Agrawal, and D. Kumar. Fast and robust spatial fuzzy bounded k-plane clustering method for human brain MRI image segmentation. *Applied Soft Computing*, 133:109939, January 2023. ISSN 1568-4946. doi: 10.1016/j.asoc.2022.109939.

L.-M. Liu, Y.-R. Guo, Z. Wang, Z.-M. Yang, and Y.-H. Shao. k-proximal plane clustering. *International Journal of Machine Learning and Cybernetics*, 8(5):1537–1554, April 2016. ISSN 1868-808X. doi: 10.1007/s13042-016-0526-y.

Y. Ma, A. Y. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, January 2008. ISSN 1095-7200. doi: 10.1137/060655523.

J. MacQueen. Some methods for classification and analysis of multivariate observations. 1967. URL https://api.semanticscholar.org/CorpusID:6278891.

O. L. Mangasarian. Arbitrary-norm separating plane. *Operations Research Letters*, 24(1–2):15–23, February 1999. ISSN 0167-6377. doi: 10.1016/s0167-6377(98)00049-2.

N. Manwani and P. S. Sastry. K-plane regression. *Information Sciences*, 292:39–56, January 2015. ISSN 0020-0255. doi: 10.1016/j.ins.2014.08.058.

N. Megiddo and A. Tamir. On the complexity of locating linear facilities in the plane. *Operations Research Letters*, 1(5):194–197, November 1982. ISSN 0167-6377. doi: 10.1016/0167-6377(82)90039-6.

S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli. K-means and alternative clustering methods in modern power systems. *IEEE Access*, 11:119596–119633, 2023. ISSN 2169-3536. doi: 10.1109/access.2023.3327640.

S. Mirzaei. Hyperspectral image classification using K-plane clustering and kernel principal component analysis. *Multimedia Tools and Applications*, 82(30):47387–47403, May 2023. ISSN 1573-7721. doi: 10.1007/s11042-023-15437-3.

R. T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton, 2015. ISBN 9780691015866. Description based upon print version of record.

M. C. Tsakiris and R. Vidal. Hyperplane clustering via dual principal component pursuit. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3472–3481. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/tsakiris17a.html.

P. Tseng. Nearest q-flat to m points. *Journal of Optimization Theory and Applications*, 105(1):249–252, April 2000. ISSN 1573-2878. doi: 10.1023/a:1004678431677.

D. Wang, C. Ding, and T. Li. *K-Subspace Clustering*, pp. 506–521. Springer Berlin Heidelberg, 2009. ISBN 9783642041747. doi: 10.1007/978-3-642-04174-7_33.

X. Xu, M. Zhong, and C. Guo. A hyperplane clustering algorithm for estimating the mixing matrix in sparse component analysis. *Neural Processing Letters*, July 2017. ISSN 1573-773X. doi: 10.1007/s11063-017-9661-z.

X. Yang, H. Yang, F. Zhang, L. Zhang, X. Fan, Q. Ye, and L. Fu. Piecewise linear regression based on plane clustering. *IEEE Access*, 7:29845–29855, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2902620.

T. Zhang, A. Szlam, and G. Lerman. Median k-flats for hybrid linear modeling with many outliers. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 234–241. IEEE, September 2009. doi: 10.1109/iccvw.2009.5457695.

Y. Zhang, H. Wang, W. Wang, and S. Sanei. K-plane clustering algorithm for analysis dictionary learning. In *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–4. IEEE, September 2013. doi: 10.1109/mlsp.2013.6661910.

A INSTANCE GENERATION

We first construct k random hyperplanes within the unit cube $[0, 1]^n$. A dataset of m points is then produced by repeatedly sampling a hyperplane according to predefined mixture weights, drawing the first $n-1$ coordinates uniformly from $[0, 1]$, and solving for the final coordinate so that the point lies exactly on the selected hyperplane. A small perturbation is optionally added in the direction of the hyperplane's normal vector to simulate noise. This procedure allows precise control over dimensionality and noise level while preserving known ground-truth cluster assignments. All the instances used in Section 5 are generated with the low noise level; Table 3 summarizes them.

Table 3: Synthetic low-noise testbed used in the computational study (64 instances, 7,200 s time limit).

Family	Instances	m	n	k
`LowDim`	16	10–26	2–3	2–3
`HighDim`	24	11–20	2–5	2–5
`SuperHighDim`	24	13–28	5–7	2–3

B Results on the `LowDim` instances

Table 4 reports the results on the `LowDim` instances, which are discussed in Section 5.

Table 4: `LowDim` (16 instances): `MultiNorm` vs. `ISCA`; same layout as Table 1.

				`MultiNorm`			`ISCA`			
Inst.	m	n	k	**Time**	**Nodes**	**Gap**	**Set.**	**Time**	**Nodes**	**Gap**
000	10	2	2	0	103	0.0%	B	1	871	0.0%
004	10	2	3	5	2,595	0.0%	B	10	11,727	0.0%
008	10	3	2	6	2,353	0.0%	A	0	673	0.0%
001	14	2	2	4	1,729	0.0%	A	13	10,141	0.0%
005	14	2	3	6	2,390	0.0%	A	30	28,738	0.0%
009	14	3	2	17	6,931	0.0%	A	38	29,301	0.0%
012	14	3	3	197	208,596	0.0%	B	1659	2,185,055	0.0%
002	18	2	2	2	918	0.0%	A	16	11,782	0.0%
006	18	2	3	33	15,222	0.0%	A	409	177,189	0.0%
010	18	3	2	34	7,163	0.0%	C	85	67,324	0.0%
013	18	3	3	876	809,047	0.0%	C	4070	5,094,067	13.5%
003	22	2	2	2	325	0.0%	A	0	142	0.0%
007	22	2	3	52	18,275	0.0%	A	116	92,600	0.0%
011	22	3	2	56	20,333	0.0%	B	3398	1,900,021	0.0%
014	22	3	3	2150	1,851,507	0.0%	A	7200	5,257,265	79.5%
015	26	3	3	7200	4,107,457	77.0%	A	7200	3,633,336	97.0%
Mean				27.6	12,314	4.8%		77.3	61,676	11.9%

C Construction of the starting simplex

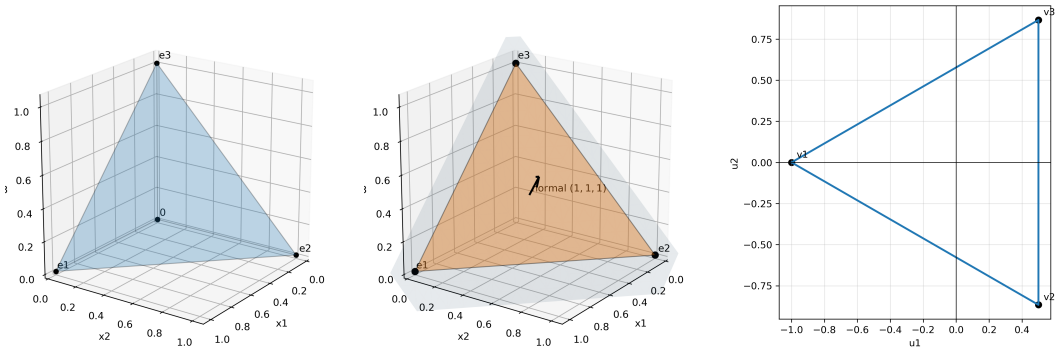
We build the simplex Δ of Section 2 in three steps. First, the vertices $e_1, \dots, e_{n+1}$ of the standard simplex in $\mathbb{R}^{n+1}$ are pairwise equidistant and lie in the affine subspace $\{x \in \mathbb{R}^{n+1} : \mathbf{1}^\top x = 1\}$; their centroid is $c := \frac{1}{n+1}\mathbf{1}$. Second, translating them by $-c$ gives the vectors $e_i - c$, $i \in [n+1]$, which lie in the linear subspace $A_0 := \{x \in \mathbb{R}^{n+1} : \mathbf{1}^\top x = 0\}$. Third, we map them to $\mathbb{R}^n$ with $U^\top$, where $U \in \mathbb{R}^{(n+1)\times n}$ is the Helmert matrix, whose columns form an orthonormal basis of A_0 (its k-th column is $(e_1 + \cdots + e_k - k\, e_{k+1})/\sqrt{k(k+1)}$):

$$U_{ik} = \begin{cases} 1/\sqrt{k(k+1)} & \text{if } i \le k, \\ -k/\sqrt{k(k+1)} & \text{if } i = k+1, \\ 0 & \text{if } i > k+1, \end{cases} \qquad i \in [n+1],\ k \in [n],$$

and we scale the result by $\sqrt{(n+1)/n}$ so that every vertex has unit 2-norm. This gives $v_i = \sqrt{(n+1)/n}\, U^\top (e_i - \frac{1}{n+1}\mathbf{1})$, $i \in [n+1]$, as in Section 2. We record two basic properties of these vertices, denoted by e'_i below, in the following lemma (with $y_i := e_i - c$ in its proof).

Lemma 11. *Let $e'_1, \dots, e'_{n+1} \in \mathbb{R}^n$ be the vertices of the n-simplex centered at the origin obtained by the construction above. Then, for all $i \neq j$,*

$$e'_i \cdot e'_j = -\frac{1}{n} \qquad \textit{and} \qquad \sum_{i=1}^{n+1} e'_i = 0.$$



(a) 3-simplex in $\mathbb{R}^3$ (b) Subspace A (c) Projected 2-simplex

Figure 2: Visualization of the 3-simplex, its subspace A, and the projected origin-centered 2-simplex.

D PROOFS

Lemma 1. *If $0 \in \mathcal{S} \subseteq B$, the simplicial relaxation of k-HC w.r.t. $\mathcal{S}$ admits an optimal solution (w^*, γ^*) with $w^* \in \partial\mathcal{S}$.*

Proof of Lemma 1. Let (w^*, γ^*) be an optimal solution with $w^* \notin \partial\mathcal{S}$. Since $w^* \in \overline{\mathcal{S}}$, w^* lies in the exterior of $\mathcal{S}$. Let $t^* := \max\{t \in [0,1] : tw^* \in \mathcal{S}\}$, which exists as $0 \in \mathcal{S}$ and $\mathcal{S}$ is closed. Then $t^* < 1$ and $\hat{w} := t^* w^* \in \partial\mathcal{S} \subseteq \overline{\mathcal{S}}$. Set $\hat{\gamma} := t^*\gamma^*$. For every $i \in [m]$, $a_i^\top \hat{w} - \hat{\gamma} = t^*(a_i^\top w^* - \gamma^*)$, hence $\sum_{i\in[m]}(a_i^\top \hat{w} - \hat{\gamma})^2 = (t^*)^2 \sum_{i\in[m]}(a_i^\top w^* - \gamma^*)^2 \leq \sum_{i\in[m]}(a_i^\top w^* - \gamma^*)^2$. Therefore, $(\hat{w}, \hat{\gamma})$ is optimal as well. □

Theorem 2. *Let $F^* = \mathrm{conv}\{p_1, \ldots, p_n\}$ be a nondegenerate facet with unit vertices, let $u \in \mathrm{cone}(F^*)$ be a unit vector other than these vertices, and let $\bar{\lambda} := E^{-1}u$ be its conic coordinates w.r.t. F^*. For $i \in [n]$, let $F_i' := \mathrm{conv}(\{u\} \cup \{p_\ell : \ell \neq i\})$, and let $J := \{i \in [n] : \bar{\lambda}_i > 0\}$. Then:*

(i) valid augmentation: *$S' := \mathrm{conv}(F^* \cup \{u\})$ is an n-simplex contained in B, and $\mathrm{conv}(F^* \cup \{0\}) \cup S' = \bigcup_{i\in J} \mathrm{conv}(F_i' \cup \{0\})$;*

(ii) degeneracy: *F_i' is degenerate if and only if $\bar{\lambda}_i = 0$, i.e., if u lies on the facet of $\mathrm{cone}(F^*)$ that does not contain p_i; moreover, $|J| \geq 2$;*

(iii) partition and nesting: *the cones $\mathrm{cone}(F_i')$, $i \in J$, cover $\mathrm{cone}(F^*)$ and have pairwise disjoint interiors, and $U(F_i') \subseteq U(F^*)$ for all $i \in J$;*

(iv) no feasible point is lost: *every $w \in U(F^*)$ with $\|w\|_2 \geq 1$ belongs to $U(F_i')$ for some $i \in J$.*

Proof of Theorem 2. We work in the conic coordinates of F^*, in which p_i is e_i, $\mathrm{cone}(F^*)$ is $\mathbb{R}^n_+$, $\mathrm{conv}(F^* \cup \{0\})$ is $\{\lambda \geq 0 : \mathbf{1}^\top\lambda \leq 1\}$, and $U(F^*)$ is $\{\lambda \geq 0 : \mathbf{1}^\top\lambda \geq 1\}$. Let $Q := \sum_{i\in[n]} \bar{\lambda}_i$. As $u = \sum_{i\in[n]} \bar{\lambda}_i p_i$ with $\bar{\lambda} \geq 0$ and unit p_i, the triangle inequality gives $1 = \|u\|_2 \leq Q$, with equality if and only if a single $\bar{\lambda}_i$ is positive (as the p_i are linearly independent), i.e., if and only if u is a vertex of F^*; hence, $Q > 1$. If $u = w^*/\|w^*\|_2$ with $w^* \in F^*$, then $\mu := E^{-1}w^*$ satisfies $\mu \geq 0$ and $\mathbf{1}^\top\mu = 1$, hence $\bar{\lambda} = \mu/\|w^*\|_2$, which proves the claims stated after the theorem.

(ii) In conic coordinates, the vertices of F_i' are $\bar{\lambda}$ and e_ℓ, $\ell \neq i$, whose determinant is $\pm\bar{\lambda}_i$; hence, they are linearly dependent, i.e., $0 \in \mathrm{aff}(F_i')$ (as they are affinely independent), if and only if $\bar{\lambda}_i = 0$. If $|J| \leq 1$, then $u = \bar{\lambda}_i p_i$ for some i, and $\|u\|_2 = 1$ gives $\bar{\lambda}_i = 1$, i.e., $Q = 1$, a contradiction.

(iii) Let $\lambda \geq 0$, let $i^* \in \arg\min_{i\in J} \lambda_i/\bar{\lambda}_i$, and let $t := \lambda_{i^*}/\bar{\lambda}_{i^*} \geq 0$. Then, $\lambda - t\bar{\lambda} \geq 0$, $(\lambda - t\bar{\lambda})_{i^*} = 0$, and $\lambda = t\bar{\lambda} + \sum_{\ell\neq i^*}(\lambda_\ell - t\bar{\lambda}_\ell)e_\ell \in \mathrm{cone}(F_{i^*}')$. Conversely, the vertices of every F_i' lie in $\mathbb{R}^n_+$. If λ is in the interior of $\mathrm{cone}(F_i')$, i.e., $\lambda = t\bar{\lambda} + \sum_{\ell\neq i} r_\ell e_\ell$ with $t > 0$ and $r_\ell > 0$, then $\lambda_i/\bar{\lambda}_i = t < \lambda_\ell/\bar{\lambda}_\ell$ for all $\ell \in J \setminus \{i\}$, so i is the unique minimizer of the ratio and the interiors are disjoint. A point of $U(F_i')$ reads $t\bar{\lambda} + \sum_{\ell\neq i} r_\ell e_\ell$ with $t, r_\ell \geq 0$ and $t + \sum_{\ell\neq i} r_\ell \geq 1$; its coordinates sum to $tQ + \sum_{\ell\neq i} r_\ell \geq t + \sum_{\ell\neq i} r_\ell \geq 1$, so it lies in $U(F^*)$.

(i) All the vertices of S' lie in the convex set B, and $u \notin \mathrm{aff}(F^*)$ as $\mathbf{1}^\top\bar{\lambda} = Q > 1$, so S' is an n-simplex in B. The inclusion $\supseteq$ holds since the vertices of every $\mathrm{conv}(F_i' \cup \{0\})$ belong to $\mathrm{conv}(F^* \cup \{0, u\}) = \mathrm{conv}(F^* \cup \{0\}) \cup S'$ (the latter equality holds as the segment between 0 and u crosses F^*). For $\subseteq$, take $\lambda = s\bar{\lambda} + \beta$ with $s \geq 0$, $\beta \geq 0$, and $s + \mathbf{1}^\top\beta \leq 1$, which describes $\mathrm{conv}(\{0, e_1, \ldots, e_n, \bar{\lambda}\})$. With i^* and t as above, $t \geq s$ since $\lambda_i \geq s\bar{\lambda}_i$ for all i, and the coefficients of the representation of λ in $\mathrm{cone}(F_{i^*}')$ sum to $t + \mathbf{1}^\top(\lambda - t\bar{\lambda}) = \mathbf{1}^\top\lambda - t(Q-1) \leq \mathbf{1}^\top\lambda - s(Q-1) = s + \mathbf{1}^\top\beta \leq 1$. Hence, $\lambda \in \mathrm{conv}(F_{i^*}' \cup \{0\})$.

(iv) By (iii), $w \in \mathrm{cone}(F_i')$ for some $i \in J$; write $w = \sum_\ell \nu_\ell p_\ell'$ with $\nu \geq 0$, where p_ℓ' are the (unit) vertices of F_i'. Then $1 \leq \|w\|_2 \leq \sum_\ell \nu_\ell \|p_\ell'\|_2 = \mathbf{1}^\top\nu$, i.e., $w \in U(F_i')$. □

Theorem 3. *Let $\mathcal{S}$ be a complex obtained from Δ by finitely many augmentations, and let $\mathcal{F}$ be the set of the nondegenerate facets of its boundary. Then the up-extensions $U(F)$, $F \in \mathcal{F}$, have pairwise disjoint interiors, and their union is the weak complement $\overline{\mathcal{S}}$ of $\mathcal{S}$.*

Proof of Theorem 3. We show by induction on the number of augmentations that the cones $\mathrm{cone}(F)$, $F \in \mathcal{F}$, cover $\mathbb{R}^n$ with pairwise disjoint interiors and that $\mathcal{S} \cap \mathrm{cone}(F) = \mathrm{conv}(F \cup \{0\})$ for

all $F \in \mathcal{F}$. For $\mathcal{S} = \Delta$, this holds as Δ is convex and $0 \in \operatorname{int}\Delta$, so every ray from the origin leaves Δ through exactly one point of its boundary. Consider an augmentation of $F^* \in \mathcal{F}$. By the induction hypothesis, the simplex $S' \subseteq \operatorname{cone}(F^*)$ meets $\mathcal{S}$ only in F^*, as the points of $S' \setminus F^*$ satisfy $\alpha^\top w > 1$, so the new complex is a simplicial complex. By Theorem 2(iii), replacing $\operatorname{cone}(F^*)$ by $\operatorname{cone}(F_i')$, $i \in J$, preserves the cover and the disjointness of the interiors, and, by Theorem 2(i) and the induction hypothesis, $(\mathcal{S} \cup S') \cap \operatorname{cone}(F^*) = \bigcup_{i\in J} \operatorname{conv}(F_i' \cup \{0\})$, which implies $(\mathcal{S} \cup S') \cap \operatorname{cone}(F_i') = \operatorname{conv}(F_i' \cup \{0\})$ by the disjointness of the interiors of the cones. Outside $\operatorname{cone}(F^*)$, nothing changes.

Since $\mathcal{S}$ is a finite union of full-dimensional simplices, $\overline{\mathcal{S}} = \mathbb{R}^n \setminus \operatorname{int}\mathcal{S}$. Let w belong to the interior of $\operatorname{cone}(F)$ for some (unique) $F \in \mathcal{F}$. By the invariant, $w \notin \operatorname{int}\mathcal{S}$ if and only if the conic coordinates of w w.r.t. F sum to at least 1, i.e., if and only if $w \in U(F)$. As both $\overline{\mathcal{S}}$ and $\bigcup_{F\in\mathcal{F}} U(F)$ are closed and the union of the interiors of the cones is dense in $\mathbb{R}^n$, the equality holds everywhere. The interiors of the up-extensions are disjoint as they are contained in the interiors of the corresponding cones. □

Lemma 5. $\Delta^\circ = \operatorname{conv}\{-nv_1, \ldots, -nv_{n+1}\}$. *Hence, the gauge of* Δ *is* $p_\Delta(x) = \max_{i\in[n+1]}\{-n\, v_i^\top x\}$.

Proof of Lemma 5. Since $0 \in \operatorname{int}(\Delta)$, the polar is $\Delta^\circ = \{y \in \mathbb{R}^n : x^\top y \le 1\ \forall x \in \Delta\}$. Because $\Delta = \operatorname{conv}\{v_1, \ldots, v_{n+1}\}$, its support function is $h_\Delta(y) = \max_{i\in[n+1]} v_i^\top y$, hence $\Delta^\circ = \bigcap_{i\in[n+1]}\{y : v_i^\top y \le 1\}$. For each $j \in [n+1]$, let $y = -nv_j$. For $i \ne j$, $v_i^\top y = -n(v_i^\top v_j) = 1$, and $v_j^\top y = -n\|v_j\|_2^2 = -n \le 1$. Thus, $-nv_j \in \Delta^\circ$ and it satisfies at equality the n inequalities with $i \ne j$, whose normals $\{v_i : i \ne j\}$ are linearly independent: $-nv_j$ is a vertex of Δ°. Since Δ° is defined by $n+1$ inequalities any n of which have linearly independent normals, it has at most $n+1$ vertices, which proves $\Delta^\circ = \operatorname{conv}\{-nv_1, \ldots, -nv_{n+1}\}$. The gauge follows from $p_\Delta = h_{\Delta^\circ}$. □

Lemma 6. *For all* $x \in \mathbb{R}^n$, $\|x\|_2 \le p_\Delta(x) \le n\|x\|_2$. *Moreover,* $\min\{\|x\|_2 : p_\Delta(x) = 1\} = 1/n$.

Proof of Lemma 6. By Lemma 5, $\Delta = \bigcap_{i\in[n+1]}\{x : -nv_i^\top x \le 1\}$. Each of these halfspaces has distance $1/n$ from the origin, as $\|v_i\|_2 = 1$; hence, $\frac{1}{n}B \subseteq \Delta \subseteq B$, where the second inclusion holds as the vertices of Δ are unit vectors. For convex sets containing the origin, $K_1 \subseteq K_2$ implies $p_{K_2} \le p_{K_1}$; therefore, $\|x\|_2 = p_B(x) \le p_\Delta(x) \le p_{\frac{1}{n}B}(x) = n\|x\|_2$. The minimum is attained at $x = -\frac{1}{n}v_i$, for which $p_\Delta(x) = 1$ (the maximum in Lemma 5 is attained at i, since $-nv_\ell^\top x = v_\ell^\top v_i \le 1$) and $\|x\|_2 = 1/n$. □

Theorem 7. *For* $K \in \{B, \Delta\}$, *let* $\mathrm{OPT}_K := \min\big\{\sum_{i\in[m]} \min_{j\in[k]}(a_i^\top w_j - \gamma_j)^2 : p_K(w_j) \ge 1\ \forall j \in [k]\big\}$, *where* $p_B = \|\cdot\|_2$. *Then* $\frac{1}{n^2}\mathrm{OPT}_B \le \mathrm{OPT}_\Delta \le \mathrm{OPT}_B$, *and the factor* $\frac{1}{n^2}$ *is tight.*

Proof of Theorem 7. By Lemma 6, $p_\Delta(w) \ge \|w\|_2$, so every solution feasible for OPT_B is feasible for OPT_Δ, and $\mathrm{OPT}_\Delta \le \mathrm{OPT}_B$. Conversely, let $(w_j, \gamma_j)_{j\in[k]}$ be feasible for OPT_Δ. By Lemma 6, $1 \le p_\Delta(w_j) \le n\|w_j\|_2$, i.e., $\|w_j\|_2 \ge 1/n$. Let $t_j := \max\{1, 1/\|w_j\|_2\} \le n$. The solution $(t_j w_j, t_j\gamma_j)_{j\in[k]}$ is feasible for OPT_B, and $(a_i^\top t_j w_j - t_j\gamma_j)^2 = t_j^2(a_i^\top w_j - \gamma_j)^2 \le n^2(a_i^\top w_j - \gamma_j)^2$ for all i, j. Taking the minimum over j, summing over i, and minimizing over the feasible solutions of OPT_Δ gives $\mathrm{OPT}_B \le n^2\,\mathrm{OPT}_\Delta$. For tightness, let $k = 1$ and consider the $2n$ points $\pm e_1, \ldots, \pm e_n$. For any (w, γ), $\sum_{i\in[n]}\big((e_i^\top w - \gamma)^2 + (-e_i^\top w - \gamma)^2\big) = 2\|w\|_2^2 + 2n\gamma^2$, which is minimized with $\gamma = 0$. Hence, $\mathrm{OPT}_B = 2$ and, by Lemma 6, $\mathrm{OPT}_\Delta = 2\min\{\|w\|_2^2 : p_\Delta(w) \ge 1\} = 2/n^2$. □

Proposition 8. *Let* $F = \operatorname{conv}\{p_1, \ldots, p_n\}$ *be a facet with unit vertices. Then every* $w \in U(F)$ *satisfies* $\|w\|_2^2 \ge 1 - \frac{1}{2}h(F)^2$. *Consequently, if* $h(F_j) \le h < \sqrt{2}$ *for the facets* F_j, $j \in [k]$, *of a node, then, for every fixed assignment, the relaxation of the node is within a factor* $1 - \frac{1}{2}h^2$ *of the optimal value of* k*-HC restricted to the region of the node.*

Proof of Proposition 8. Let $w \in U(F)$, and write $w = tz$ with $z = \sum_{i\in[n]} \mu_i p_i \in F$ ($\mu \ge 0$, $\mathbf{1}^\top\mu = 1$) and $t \ge 1$. Since $p_i^\top p_\ell = 1 - \frac{1}{2}\|p_i - p_\ell\|_2^2$ for unit vectors, $\|z\|_2^2 = \sum_{i,\ell\in[n]} \mu_i\mu_\ell p_i^\top p_\ell = 1 - \frac{1}{2}\sum_{i,\ell\in[n]} \mu_i\mu_\ell\|p_i - p_\ell\|_2^2 \ge 1 - \frac{1}{2}h(F)^2$, where we used $\sum_{i,\ell}\mu_i\mu_\ell = 1$. Hence, $\|w\|_2^2 =$

$t^2\|z\|_2^2 \geq 1 - \frac{1}{2}h(F)^2$. For the second claim, let $(w_j, \gamma_j)_{j\in[k]}$ be an optimal solution of the relaxation of the node for a fixed assignment, of value L. For each j with $\|w_j\|_2 < 1$, replace (w_j, γ_j) by $(w_j, \gamma_j)/\|w_j\|_2$: the new w_j is a unit vector in $\mathrm{cone}(F_j)$ and, thus, it belongs to $U(F_j)$ (as in the proof of Theorem 2(iv)), and the residuals of hyperplane j are multiplied by $1/\|w_j\|_2 \leq (1 - \frac{1}{2}h^2)^{-1/2}$. The resulting solution is feasible for k-HC restricted to the region of the node, with value at most $L/(1 - \frac{1}{2}h^2)$. □

Lemma 9. *Let $M \geq D$. For every assignment of the points to the hyperplanes, k-HC restricted to that assignment has an optimal solution with $\|w_j\|_2 = 1$, $|\gamma_j| \leq \max_{i\in[m]} \|a_i\|_2$, and $|w_j^\top a_i - \gamma_j| \leq M$ for all $i \in [m]$, $j \in [k]$. Hence, Formulation* (2) *with $\|w_j\|_2 \geq 1$ in place of $w_j \in U(F_j)$ has the same optimal value as k-HC.*

Proof of Lemma 9. Fix an assignment $C_1, \ldots, C_k$. For a nonempty cluster C_j with $m_j := |C_j|$ and $\bar{a}_j := \frac{1}{m_j}\sum_{\ell\in C_j} a_\ell$, for a fixed w_j the value $\sum_{\ell\in C_j}(a_\ell^\top w_j - \gamma_j)^2$ is minimized by $\gamma_j = \bar{a}_j^\top w_j$; hence, some optimal solution has a unit w_j (by the normalization argument of the introduction) and $\gamma_j = \bar{a}_j^\top w_j$, so that $|\gamma_j| \leq \|\bar{a}_j\|_2 \leq \max_i \|a_i\|_2$ and, for every $i \in [m]$, $|a_i^\top w_j - \gamma_j| = |w_j^\top(a_i - \bar{a}_j)| \leq \|a_i - \bar{a}_j\|_2 \leq \frac{1}{m_j}\sum_{\ell\in C_j}\|a_i - a_\ell\|_2 \leq D \leq M$. For an empty cluster, any unit w_j with $\gamma_j := a_1^\top w_j$ satisfies the same bounds. With such a solution and $d_i := |a_i^\top w_{j(i)} - \gamma_{j(i)}|$, where $j(i)$ is the cluster of a_i, every constraint with $x_{ij} = 0$ is satisfied and $\sum_i d_i^2$ equals the value of the assignment. Conversely, every integer feasible solution of the formulation has $d_i \geq |w_{j(i)}^\top a_i - \gamma_{j(i)}|$, so its value is at least that of its assignment in k-HC. Symmetry breaking is valid as relabeling the clusters by the smallest index of their points gives $x_{ij} = 0$ for $j > i$. □

Proposition 10. *With the notation above,* $\min\{\mathrm{LB}, \mathrm{UB}, \min_{c\in P} L_c\} \leq \mathrm{OPT}$.

Proof of Proposition 10. As in Remark 4, every feasible solution belongs to the region of an active node, of a node pruned by bound, or of a discarded node. In the first case, $\mathrm{LB} \leq \mathrm{OPT}$; in the second, $\mathrm{OPT} = \mathrm{UB}$; in the third, $L_c \leq \mathrm{OPT}$, as L_c is a valid lower bound on the value of the solutions in the region of c. □

Lemma 11. *Let $e'_1, \ldots, e'_{n+1} \in \mathbb{R}^n$ be the vertices of the n-simplex centered at the origin obtained by the construction above. Then, for all $i \neq j$,*

$$e'_i \cdot e'_j = -\frac{1}{n} \qquad \textit{and} \qquad \sum_{i=1}^{n+1} e'_i = 0.$$

Proof of Lemma 11. For $i \neq j$,

$$y_i \cdot y_j = (e_i - c)^\top(e_j - c) = e_i^\top e_j - e_i^\top c - c^\top e_j + c^\top c = 0 - \tfrac{1}{n+1} - \tfrac{1}{n+1} + \tfrac{1}{n+1} = -\tfrac{1}{n+1},$$

and $\|y_i\|_2^2 = y_i \cdot y_i = 1 - \frac{1}{n+1} = \frac{n}{n+1}$. Since U has orthonormal columns spanning A_0, $U^\top$ is an isometry from A_0 onto $\mathbb{R}^n$, hence

$$\tilde{e}_i := U^\top y_i \quad \text{satisfy} \quad \tilde{e}_i \cdot \tilde{e}_j = y_i \cdot y_j \quad \text{and} \quad \|\tilde{e}_i\|_2 = \|y_i\|_2 = \sqrt{\tfrac{n}{n+1}}.$$

After the common normalization $e'_i = \sqrt{\frac{n+1}{n}}\,\tilde{e}_i$, we obtain, for $i \neq j$,

$$e'_i \cdot e'_j = \frac{\tilde{e}_i \cdot \tilde{e}_j}{\|\tilde{e}_i\|_2\,\|\tilde{e}_j\|_2} = \frac{-\frac{1}{n+1}}{\frac{n}{n+1}} = -\frac{1}{n}.$$

Since $\sum_{i=1}^{n+1} y_i = 0$ we have $\sum_{i=1}^{n+1} e'_i = \sqrt{\frac{n+1}{n}}\sum_{i=1}^{n+1}\tilde{e}_i = \sqrt{\frac{n+1}{n}}U^\top\sum_{i=1}^{n+1} y_i = 0$. □