1id1 Preprint | 20 November 2026
1id1:core:20261120-0001@v1 Version 1
https://1id1.net/id/core/20261120-0001/v1/

Supplementary Material for Learning Orthant Comparability

Wuming Pan PANWUMING@SCU.EDU.CN
College of Computer Science, Sichuan University
Chengdu 610065, China

Abstract

We provide closure-learning implementation details, reproducible finite checks, the precise Lean verification scope, and full proofs of the random-instance identification results. Joint antichain estimates give the limiting no-missed-direction probability $\exp(-e^{-c})$ when $\binom{N}{2}2^{1-d} - (d-1)\ln 2 \to c$. Explicit witnesses certify minimum covering sizes in dimensions two through six and upper bounds in seven and eight. The program checks these witnesses, closure and compression; finite checks supplement, but do not replace, proofs.

Keywords: comparability learning, improper learning, sample compression, signed graphs, PAC-Bayes bounds

1 Implementation and reproducible checks

We implement $g_P(z) = \mathbf{1}[\mathcal{V}(P) \subseteq C(z)]$ from at most $d-1$ positive examples without auxiliary bits, using support–sign inputs. Dense pairs incur an additional coordinate-comparison cost.

Compression and reconstruction. Initialise parity union–find with d isolated coordinates. Process positives in sample order: for support T, choose $r \in T$ and impose $\sigma_r\sigma_i = \alpha_r\alpha_i$ for $i \in T \setminus \{r\}$. Retain the example iff an edge merges components. Under realizability, nonmerging edges impose no new parity condition; each retained example reduces the component count, giving at most $d-1$. Reconstruction builds their graph and stores a satisfying signing a_i and component IDs $t(i)$.

Closure and Gibbs-mass queries. For support T and signs α_i, scan $(t(i), a_i\alpha_i)$. Conflicting signs within a component make the query inconsistent as an added positive: return closure and Gibbs mass zero. Otherwise, if q components occur, the mass is $2^{-(q-1)}$ and closure is one iff $q = 1$. Timestamped component arrays give $O(|T|)$ queries after one graph preprocessing pass. For $q = 2$ the mass is $1/2$, so majority voting must break ties negatively to equal closure.

Boundary cases. With no positives, closure accepts exactly support-one pairs, which are always comparable and cannot be negative in a realizable sample. A full-support positive fixes one projective direction. In dimension two, opposite parallel signed edges form a negative two-cycle, attaining the two-example inconsistency bound. Isolated coordinates remain free variables in consistency and counting and cannot be deleted in a parsimonious reduction.

Scope of the counting reduction. A signed NAE clause on three distinct variables is represented by a negative support-three example, with one coordinate per variable and

no extra coordinates. The resulting map from assignments to raw directions is bijective, including unused variables. The standard counting dichotomy (Creignou and Hermann, 1996) gives hardness under polynomial-time Turing reductions. For clarity, allowing repeated variables in its input language does not undermine restriction to distinct-variable clauses: a clause with opposite literals of one variable is tautological; one with only one repeated literal reduces to a disequality of two signed literals. Replace such a disequality $a \neq b$ by $\mathrm{NAE}(a, b, z)$ and $\mathrm{NAE}(a, b, -z)$, with a fresh variable z. There are exactly two extensions when $a \neq b$, and none otherwise, so the known factor 2 per replacement is divided out in the Turing reduction. A clause consisting of three identical literals is unsatisfiable. After this preliminary restriction, the clause-to-pair map itself is parsimonious. Counts of projective directions are half the counts of raw directions; the main counting theorem uses raw ones.

Verification program. The standard-library Python program `verify_claims.py` enumerates nonzero support–sign patterns modulo reversal and every positive version space for $2 \leq d \leq 5$, including the empty sample. For every query it checks graph-based closure and Gibbs mass by enumerating directions. It also checks version-space preservation by at most $d - 1$ greedy positives and cardinality halving at strict nonempty positive intersections. Enumerating unordered pairs verifies every displayed covering witness. These are certificate checks, not exhaustive searches over point sets or empirical evaluations of learning applications.

Lean verification: coverage and exclusions. In Definition 2.1 of the main paper, Lean models the order on feature vectors, not its pullback to a separate set of records. The supplied `LOC` source package pins Lean and Mathlib to version 4.35.0. It verifies the sign representation, the affine positive version space, and positive compression including its version-space-preserving tightness example. For main-paper Theorem 4.2, it verifies part (a) and the upper and deterministic lower mistake bounds in part (c). For Theorem 4.3, it verifies a fixed map that selects at most $d - 1$ positive examples and whose closure fits every realizable mixed sample, with no auxiliary information. The map uses `Classical.choose`: its existence and correctness, not an efficient executable selection algorithm, are formalised. Theorem 4.2(b) and all implementation running-time claims remain outside the development.

The Lean model uses ordered pairs, including equal endpoints, and finite sets of positive constraints. Symmetry of comparability permits the mathematical restriction to the paper's unordered, distinct-endpoint pairs; a separate quotient-type interface is not formalised. The lower-bound constructions use two ordered levels per coordinate. The additional Theorem 4.4 verification covers only finitely supported distributions; its confidence corollary covers the regime where the displayed threshold is at most one. General-distribution reduction, proper-learning hardness, and the later probability-threshold results are not formally verified. Rebuilding and the supplied axiom audit find no proof placeholders or axioms beyond `propext`, `Classical.choice`, and `Quot.sound`.

2 Setting and notation

We use the notation of the main paper. $\Sigma_d = \{-1, 1\}^d$ is the set of directions, $\preceq_\sigma$ the coordinate-reversed product order, and h_σ the comparability concept. For $x \neq y \in \mathbb{R}^d$ with no tied coordinate, $\chi(x, y) = (\mathrm{sgn}(y_i - x_i))_i \in \Sigma_d$, and x, y are comparable under σ iff

1id1:core:20261120-0001@v1 1id1.net

$\chi(x,y) = \pm\sigma$. Directions σ and $-\sigma$ define the same concept; we write $[\sigma]$ for the projective class and $\mathbb{P}\Sigma_d$ for the set of 2^{d-1} classes. For a finite $A \subseteq \mathbb{R}^d$,

$$\mathrm{Ant}(A) = \{\sigma \in \Sigma_d : A \text{ is a } \preceq_\sigma \text{-antichain}\} = \Sigma_d \setminus \bigcup_{\{x,y\}\in\binom{A}{2}} C(x,y),$$

where $C(x,y) = \{\sigma : x, y \text{ comparable under } \sigma\}$. Under the general-position assumption $C(x,y) = \{\pm\chi(x,y)\}$, so $\mathrm{Ant}(A)$ is closed under negation and $|\mathrm{Ant}(A)|/2$ is the number of projective classes not hit by any pair of A.

Learning reading. A negative example $\{x,y\}$ (incomparable pair) eliminates the single class $[\chi(x,y)]$. If $\mathcal{N}$ is a set of negative examples, the version space they define is $\Sigma_d \setminus \bigcup_{n\in\mathcal{N}} C(n)$. In particular $\mathrm{Ant}(A) \neq \emptyset$ iff the *all-negative* labelling of $\binom{A}{2}$ is realizable by $\mathcal{C}_d$.

Random model. $X^{(1)}, X^{(2)}, \ldots$ are independent random points in $\mathbb{R}^d$ whose coordinates are independent with continuous laws, identically distributed across points within each coordinate (but allowed to differ between coordinates). $A_N = \{X^{(1)}, \ldots, X^{(N)}\}$, $d = d(N) \to \infty$. For $\sigma \in \Sigma_d$,

$$K_\sigma = \#\{\pi \in \tbinom{[N]}{2} : \pi \text{ comparable under } \sigma\}, \qquad \lambda_N = \mathbb{E}K_\sigma = \binom{N}{2} 2^{1-d},$$

and

$$Z_N = \tfrac{1}{2}|\mathrm{Ant}(A_N)| = \#\{[\sigma] \in \mathbb{P}\Sigma_d : K_\sigma = 0\},$$
$$\nu_N := 2^{d-1} e^{-\lambda_N}, \qquad c_N = \lambda_N - (d-1)\ln 2,$$

so that $\nu_N = e^{-c_N}$. Finally

$$\eta_N = N^3 3^{-d} + N\,2^{-d} + \lambda_N 2^{-d}$$

collects the error terms. Whenever $\lambda_N = O(d)$, we have $N^2 2^{-d} = O(d)$, which forces $\eta_N \to 0$: indeed $N^3 3^{-d} = (N^2 2^{-d})^{3/2}(2^{3/2}/3)^d = O(d^{3/2})\,(2^{3/2}/3)^d$.

3 Tools

3.1 Suen's inequality

Let $\{I_i\}_{i\in\mathcal{I}}$ be indicators with a *dependency graph* Γ: whenever $A, B \subseteq \mathcal{I}$ are disjoint with no edge of Γ between them, $\{I_i\}_{i\in A}$ and $\{I_i\}_{i\in B}$ are independent. Write $i \sim j$ for adjacency, $p_i = \mathbb{E}I_i$, $S = \sum_i I_i$, and

$$\mu = \sum_i p_i, \quad \Delta = \sum_{\{i,j\}:\, i\sim j} \mathbb{E}(I_i I_j), \quad \Delta' = \sum_{\{i,j\}:\, i\sim j} p_i p_j, \quad \delta = \max_i \sum_{j\sim i} p_j, \quad \varepsilon = \max_i p_i,$$

where the sums defining Δ, Δ' run over unordered adjacent pairs.

Lemma 3.1 (Suen's inequality, Janson's form (Suen, 1990; Janson, 1998)) *With the notation above, if $\varepsilon \le \frac{1}{2}$ then*

$$\big(1 - \Delta' e^{4\delta}\exp(\Delta e^{4\delta})\big)\prod_i (1-p_i) \;\le\; \mathbb{P}(S=0) \;\le\; \exp(-\mu + \Delta e^{2\delta}).$$

1id1:core:20261120-0001@v1 1id1.net

Proof The upper bound is Theorem 2 of Janson (1998). For the lower bound, Theorem 8 there states

$$\mathbb{P}(S=0) \geq \bigl(1-\Delta_0^* \exp(\Delta^*)\bigr) \prod_i (1-p_i),$$

where

$$\Delta^* = \sum_{\{i,j\}:i\sim j} \mathbb{E}(I_i I_j) \prod_{k\sim\{i,j\}} (1-p_k)^{-1}, \qquad \Delta_0^* = \sum_{\{i,j\}:i\sim j} p_i p_j \prod_{k\sim\{i,j\}} (1-p_k)^{-1}.$$

Here $k \sim \{i,j\}$ means that k is adjacent to at least one endpoint. Since $-\ln(1-p_k) \leq p_k/(1-\varepsilon)$, each product is at most $\exp(2\delta/(1-\varepsilon)) \leq e^{4\delta}$. Thus $\Delta^* \leq \Delta e^{4\delta}$ and $\Delta_0^* \leq \Delta' e^{4\delta}$, giving the displayed lower bound. It remains valid when its prefactor is negative, but is useful here only when the error quantities are small. ∎

If Δ, Δ' and δ are small, and $\sum_i p_i^2$ is small, this gives $\mathbb{P}(S=0) = e^{-\mu}(1 + O(\Delta + \Delta' + \sum_i p_i^2))$ with bounded factors depending on δ. Unlike an additive total-variation estimate, such relative control is useful when $e^{-\mu}$ is exponentially small.

3.2 Pair correlations

Fix distinct projective classes $[\sigma_1], \ldots, [\sigma_k]$, with $k \geq 1$. For each pair π, let I_π indicate comparability under at least one of these classes. The events "comparable under σ_j" are disjoint for distinct classes, so $p_\pi = \mathbb{E} I_\pi = k\, 2^{1-d}$. Two pairs are adjacent in the dependency graph iff they share a point.

Lemma 3.2 (Joint comparability of adjacent pairs) *Let $\pi = \{u, v\}$ and $\pi' = \{u, w\}$ share the point u. Then*

$$\mathbb{P}(\pi \text{ comparable under } \sigma,\ \pi' \text{ comparable under } \sigma') \ \leq\ 4 \cdot 3^{-d}$$

for all $\sigma, \sigma' \in \Sigma_d$. The probability equals $2 \cdot 3^{-d} + 2 \cdot 6^{-d}$ when $\sigma' = \pm\sigma$. Consequently $\mathbb{E}(I_\pi I_{\pi'}) \leq 4k^2\, 3^{-d}$.

Proof Let $s_i = \operatorname{sgn}(X_i^{(v)} - X_i^{(u)})$ and $t_i = \operatorname{sgn}(X_i^{(w)} - X_i^{(u)})$. Within coordinate i the three values are exchangeable and distinct a.s., so the six rankings are equiprobable and $\mathbb{P}(s_i = a, t_i = b) = \frac{1}{3}$ if $a = b$ and $\frac{1}{6}$ if $a \neq b$. Coordinates are independent. The event in question is $s = \tau\sigma,\ t = \tau'\sigma'$ for some signs τ, τ'; for fixed (τ, τ') its probability is $\prod_i \mathbb{P}(s_i = \tau\sigma_i, t_i = \tau'\sigma_i') \leq 3^{-d}$, and there are four sign choices. If $\sigma' = \pm\sigma$ the coordinates all agree or all disagree, giving $2 \cdot 3^{-d} + 2 \cdot 6^{-d}$. The last claim sums over the k^2 choices of (σ_j, σ_l). ∎

3.3 Why the standard additive estimate is insufficient here

For one fixed class put $p = 2^{1-d}$, $M = \binom{N}{2}$ and $q = 2 \cdot 3^{-d} + 2 \cdot 6^{-d}$. With B_π the pairs meeting π, including π itself, the convention of Arratia et al. (1989, Theorem 1) gives

$$b_1 = M(2N-3)p^2, \qquad b_2 = M\, 2(N-2)q, \qquad b_3 = 0.$$

1id1:core:20261120-0001@v1 1id1.net

The b_2 sum excludes π, unlike b_1. The same convention defines b_3 by conditioning on the indicators outside B_π; they depend only on other sample points. Thus a total-variation bound is $2(1 \wedge \lambda_N^{-1})(b_1 + b_2)$. This suffices at bounded intensity but not for the relative estimate needed below.

Proposition 3.3 (Scale mismatch and the variance identity) *For the fixed-direction pair count,*

$$\operatorname{Var} K_\sigma - \mathbb{E}K_\sigma = -Mp^2 + M\,2(N-2)(q-p^2) = (2+o(1))N^3 3^{-d}$$

as $N, d \to \infty$. At $\lambda_N = (d-1)\ln 2 + O(1)$, the displayed Chen–Stein bound, divided by $e^{-\lambda_N}$, is of order

$$d^{1/2}\left(\frac{2^{5/2}}{3}\right)^d \longrightarrow \infty.$$

Proof The diagonal variance terms are $Mp(1-p)$, not Mp. Only ordered pairs of distinct indicators sharing an endpoint contribute covariance, and each contributes $q - p^2$. This proves the identity. Its leading term is $2N^3 3^{-d}$; the remaining terms have negligible ratios as $N, d \to \infty$. At the critical scale, $N^2 = \Theta(d2^d)$, $b_2 = \Theta(N^3 3^{-d})$, $b_1 = o(b_2)$, and $\lambda_N = \Theta(d)$. Multiplication by λ_N^{-1} and division by $e^{-\lambda_N} = \Theta(2^{-d})$ give the asserted order. ∎

This is a comparison of a particular *upper bound* with the scale of the desired probability, not a lower bound on the actual total-variation error, and not an impossibility result for all Chen–Stein methods. We use factorial moments of the missed-class count. They require joint zero probabilities, each of which is a zero-event for pair indicators with a local dependency graph. No assertion about a complete or minimal dependency graph on the missed-class indicators themselves is needed.

4 Poisson estimates with multiplicative error

Theorem 4.1 (Joint antichain probabilities) *There are absolute constants $C, c_0 > 0$ such that for every $k \geq 1$, every k distinct classes $[\sigma_1], \ldots, [\sigma_k]$, and all N, d with $k\,2^{1-d} \leq \frac{1}{2}$ and $k^2\eta_N \leq c_0$,*

$$\mathbb{P}(K_{\sigma_1} = \cdots = K_{\sigma_k} = 0) = e^{-k\lambda_N}(1+\theta), \qquad |\theta| \leq C\,k^2\eta_N.$$

Proof With I_π as in Section 3, $S = \sum_\pi I_\pi = \sum_j K_{\sigma_j}$ and the event is $\{S = 0\}$. We compute the quantities of Lemma 3.1. There are $\binom{N}{2}$ pairs, $p_\pi = k2^{1-d} = \varepsilon$, $\mu = k\lambda_N$. Each pair is adjacent to $2(N-2)$ others, so $\delta = 2(N-2)k2^{1-d} \leq 4kN2^{-d}$, and the number of unordered adjacent pairs is $\binom{N}{2}(N-2) \leq N^3/2$. Hence, by Lemma 3.2,

$$\Delta \leq \tfrac{1}{2}N^3 \cdot 4k^2 3^{-d} = 2k^2 N^3 3^{-d}, \qquad \Delta' \leq \tfrac{1}{2}N^3 \cdot 4k^2 4^{-d} \leq 2k^2 N^3 3^{-d}.$$

Finally, for $x \leq \frac{1}{2}$, $-x - x^2 \leq \ln(1-x) \leq -x$, so

$$e^{-k\lambda_N - k\lambda_N\varepsilon} \leq \prod_\pi (1-p_\pi) \leq e^{-k\lambda_N}, \qquad k\lambda_N\varepsilon = 2k^2\lambda_N 2^{-d}.$$

1id1:core:20261120-0001@v1 1id1.net

Under the hypothesis $k^2\eta_N \le c_0$ with c_0 small, all of $\Delta, \Delta', \delta,\ k\lambda_N\varepsilon$ are $O(k^2\eta_N)$ and bounded by a small constant, and Lemma 3.1 gives $e^{-k\lambda_N}(1 - O(k^2\eta_N)) \le \mathbb{P}(S = 0) \le e^{-k\lambda_N}(1 + O(k^2\eta_N))$. ∎

Corollary 4.2 *In the regime $\eta_N \to 0$: $\mathbb{E}Z_N = \nu_N(1+O(\eta_N))$, and for each fixed k the k-th factorial moment satisfies*

$$\mathbb{E}[Z_N(Z_N-1)\cdots(Z_N-k+1)] = (2^{d-1})_k\, e^{-k\lambda_N}\,(1+O(k^2\eta_N)) = \nu_N^k(1+O(k^2\eta_N+k^2 2^{1-d})),$$

where $(m)_k = m(m-1)\cdots(m-k+1)$.

Proof $Z_N = \sum_{[\sigma]} \mathbf{1}[K_\sigma = 0]$ and the k-th factorial moment is the sum of $\mathbb{P}(K_{\sigma_1} = \cdots = K_{\sigma_k} = 0)$ over ordered k-tuples of distinct classes; apply Theorem 4.1 and $(2^{d-1})_k = 2^{k(d-1)}(1 + O(k^2 2^{1-d}))$. ∎

5 The antichain-direction threshold

Theorem 5.1 (Antichain-direction threshold) *Let $d = d(N) \to \infty$ and $c_N = \lambda_N - (d-1)\ln 2$.*

(a) *If $c_N \to +\infty$, then $\mathbb{P}(\mathrm{Ant}(A_N) = \emptyset) \to 1$.*

(b) *If $c_N \to -\infty$, then $\mathbb{P}(\mathrm{Ant}(A_N) \neq \emptyset) \to 1$; more precisely $Z_N/\nu_N \to 1$ in probability, where $\nu_N = 2^{d-1}e^{-\lambda_N} \to \infty$.*

(c) *If $c_N \to c \in \mathbb{R}$, then Z_N converges in distribution to $\mathrm{Poisson}(e^{-c})$; in particular $\mathbb{P}(\mathrm{Ant}(A_N) = \emptyset) \to \exp(-e^{-c})$.*

Proof (b) and (c). In both cases $\lambda_N \le (d-1)\ln 2 + O(1)$, so $N^2 2^{-d} = O(d)$ and $\eta_N \to 0$. For (c), $\nu_N = e^{-c_N} \to e^{-c}$, and by Corollary 4.2 every factorial moment of Z_N converges to e^{-kc}, the k-th factorial moment of $\mathrm{Poisson}(e^{-c})$. The Poisson law is determined by its moments, so $Z_N \to \mathrm{Poisson}(e^{-c})$ in distribution by the method of moments; $\mathbb{P}(Z_N = 0) \to \exp(-e^{-c})$. For (b), Corollary 4.2 with $k = 1, 2$ gives $\mathbb{E}Z_N = \nu_N(1+o(1))$ and $\mathbb{E}[Z_N(Z_N-1)] = \nu_N^2(1+o(1))$, hence

$$\mathrm{Var}\, Z_N = \mathbb{E}Z_N + \mathbb{E}[Z_N(Z_N-1)] - (\mathbb{E}Z_N)^2 = \nu_N(1+o(1)) + o(\nu_N^2),$$

and Chebyshev's inequality gives concentration around $\mathbb{E}Z_N$. Since $\mathbb{E}Z_N/\nu_N \to 1$ and $\nu_N = e^{-c_N} \to \infty$, this also gives $Z_N/\nu_N \to 1$.

(a). Adding points only removes antichain directions. Choose $\omega_N \to \infty$ with $\omega_N \le \min\{c_N/2, \sqrt{d}\}$, and let $N' \le N$ be the least integer satisfying $\lambda_{N'} \ge (d-1)\ln 2 + \omega_N$. Minimality gives $N' = O(\sqrt{d2^d})$; hence the last increment is $(N'-1)2^{1-d} = O(\sqrt{d}\,2^{-d/2}) = o(1)$. Thus $c_{N'} = \omega_N + o(1)$, $\lambda_{N'} = O(d)$, and $\eta_{N'} \to 0$. By Markov's inequality and Corollary 4.2,

$$\mathbb{P}(Z_N \ge 1) \le \mathbb{E}Z_{N'} = e^{-\omega_N + o(1)} \longrightarrow 0.$$

∎

1id1:core:20261120-0001@v1 1id1.net

Remark 5.2 (Location versus critical-window precision) *The condition for the limiting profile is precisely*

$$\lambda_N - (d-1)\ln 2 \longrightarrow c, \quad \textit{equivalently} \quad \binom{N}{2} = 2^{d-1}\big((d-1)\ln 2 + c + o(1)\big).$$

It implies $d = 2\log_2 N - \log_2 \ln N - 1 + o(1)$*, describing the location relative to the fixed-direction scale* $2\log_2 N$*. The converse is false: an* $o(1)$ *perturbation in* d *can change the intensity by an unbounded amount. Therefore the coarse dimension formula is not a parameterisation of the limiting critical window. Since* d *is integer-valued, finite-*c *limits are understood along sequences of integer pairs* (N, d) *satisfying the precise intensity condition. The probability of no missed class tends to* $\exp(-e^{-c})$*; the probability of an antichain direction tends to its complement.*

6 Consequences for learning from negative feedback

Fix a hidden direction σ^* and suppose the learner receives the labels $h_{\sigma^*}(\pi)$ of all $\binom{N}{2}$ pairs $\pi \in \binom{A_N}{2}$. Let $\mathcal{V}_N^-$ be the version space determined by the *negative* examples alone and $\mathcal{V}_N$ the version space determined by all examples.

Lemma 6.1 *Almost surely,* $\frac{1}{2}|\mathcal{V}_N^-| = 1 + Z_N - \mathbf{1}[K_{\sigma^*} = 0]$.

Proof A pair π is negative iff $[\chi(\pi)] \neq [\sigma^*]$, and it then eliminates exactly the class $[\chi(\pi)]$. Hence the classes eliminated by negatives are exactly the classes other than $[\sigma^*]$ hit by some pair, and $\frac{1}{2}|\mathcal{V}_N^-| = 1 + \#\{[\sigma] \neq [\sigma^*] : K_\sigma = 0\}$. ∎

Theorem 6.2 (Negative feedback alone) *Let* $d \to \infty$ *and* $c_N = \lambda_N - (d-1)\ln 2$.

(a) *If* $c_N \to +\infty$*, the negative examples alone identify* σ^* *(that is,* $\mathcal{V}_N^- = \{\pm\sigma^*\}$*) with probability* $\to 1$.

(b) *If* $c_N \to c \in \mathbb{R}$*, then* $\frac{1}{2}|\mathcal{V}_N^-| - 1 \to \mathrm{Poisson}(e^{-c})$ *in distribution, and* $\mathbb{P}(\mathcal{V}_N^- = \{\pm\sigma^*\}) \to \exp(-e^{-c})$.

(c) *If* $c_N \to -\infty$*, then* $|\mathcal{V}_N^-|/(2\nu_N) \to 1$ *in probability with* $\nu_N \to \infty$*; in particular the negatives alone identify* σ^* *with probability* $\to 0$.

Consequently, negative feedback alone identifies the direction with probability $1 - o(1)$ *iff* $\binom{N}{2} \geq 2^{d-1}\big((d-1)\ln 2 + \omega(1)\big)$*, a factor* $(d-1)\ln 2$ *above the counting lower bound* $\binom{N}{2} \geq 2^{d-1} - 1$.

Proof By Lemma 6.1, $\frac{1}{2}|\mathcal{V}_N^-| - 1 = Z_N - \mathbf{1}[K_{\sigma^*} = 0]$. In (a), $Z_N = 0$ with probability tending to one, and the identity then forces a unique remaining target class. In (b), $\lambda_N = (d-1)\ln 2 + O(1)$, so $\eta_N \to 0$ and $\mathbb{P}(K_{\sigma^*} = 0) = e^{-\lambda_N}(1 + o(1)) \to 0$ by Theorem 4.1; hence the same identity and Theorem 5.1(c) give (b). In (c), $\nu_N \to \infty$ and $Z_N/\nu_N \to 1$ in probability by Theorem 5.1(b), while $0 \leq \mathbf{1}[K_{\sigma^*} = 0] \leq 1$; so $|\mathcal{V}_N^-|/(2\nu_N) \to 1$ in probability,

1id1:core:20261120-0001@v1 1id1.net

whether or not λ_N stays bounded, and $\mathbb{P}(\mathcal{V}_N^- = \{\pm\sigma^*\}) = \mathbb{P}(Z_N = \mathbf{1}[K_{\sigma^*} = 0]) \le \mathbb{P}(Z_N \le 1) \to 0$. For necessity in the consequently clause, if c_N does not tend to $+\infty$, take a subsequence on which it is bounded above, and then a further subsequence converging to a finite value or to $-\infty$. Cases (b) and (c) rule out identification probability tending to one along that subsequence. The counting bound is the observation that each negative removes one of the $2^{d-1} - 1$ classes other than $[\sigma^*]$. ∎

The distinction is between one positive full-support pair, which determines the target class, and negatives that eliminate observed non-target classes. The exact version-space identity above uses N random instances and their dependent pairs. It is not a formula for an independent stream of a prescribed number of negative examples.

Corollary 6.3 (Identification theorem for all λ) *In the random-instance setting above, suppose $\lambda_N \to \lambda \in [0, \infty)$ and let $\mathcal{I}_N$ be the event that $\mathcal{V}_N = \{\pm\sigma^*\}$. Then $\mathbb{P}(\mathcal{I}_N) \to 1 - e^{-\lambda}$ for every $\lambda \in [0, \infty)$. If $\lambda_N \to \infty$, then $\mathbb{P}(\mathcal{I}_N) \to 1$.*

Proof $\mathcal{I}_N = \{K_{\sigma^*} \ge 1\} \cup \{K_{\sigma^*} = 0,\ \mathcal{V}_N^- = \{\pm\sigma^*\}\}$, a disjoint union, because a positive example identifies $\pm\sigma^*$ in general position while with no positive examples $\mathcal{V}_N = \mathcal{V}_N^-$. By Lemma 6.1 the second event is $\{K_{\sigma^*} = 0,\ Z_N = 1\}$. If $\lambda_N \to \lambda < \infty$, then $c_N \to -\infty$ and $\mathbb{P}(Z_N = 1) \le \mathbb{P}(Z_N \le \nu_N/2) \to 0$ by Theorem 5.1(b), so $\mathbb{P}(\mathcal{I}_N) = \mathbb{P}(K_{\sigma^*} \ge 1) + o(1) \to 1 - e^{-\lambda}$ by the bounded-intensity Poisson theorem in the main paper. If $\lambda_N \to \infty$, fix $T > 0$ and choose the least $N_T \le N$ with $\lambda_{N_T} \ge T$. Consecutive increments tend to zero, so $\lambda_{N_T} \to T$. Monotonicity and the bounded-intensity limit give $\limsup \mathbb{P}(K_{\sigma^*} = 0) \le e^{-T}$. Let $T \to \infty$. ∎

The thinning step is needed because the bounded-intensity Poisson theorem cannot itself be invoked with $\lambda = \infty$.

Corollary 6.4 (Vacuous monotone bias: concentration) *Let $F_N = |\operatorname{Ant}(A_N)|/2^d$ be the fraction of directions under which monotonicity imposes no constraint on the sample.*

(a) *If $\lambda_N \to \lambda \in [0, \infty)$, then $F_N \to e^{-\lambda}$ in probability.*

(b) *If $\lambda_N \to \infty$ with $c_N \to -\infty$, then $2^{d-1} F_N / \nu_N \to 1$ in probability.*

(c) *If $c_N \to c$, then $2^{d-1} F_N \to \text{Poisson}(e^{-c})$.*

(d) *If $c_N \to +\infty$, then $F_N = 0$ with probability $\to 1$.*

Proof $2^{d-1} F_N = Z_N$; apply Theorem 5.1, noting that $\lambda_N \to \lambda < \infty$ implies $c_N \to -\infty$ and $\nu_N = 2^{d-1} e^{-\lambda_N}$, so $Z_N / \nu_N \to 1$ means $F_N / e^{-\lambda_N} \to 1$. ∎

1id1:core:20261120-0001@v1 1id1.net

7 Small-dimensional witnesses and a general upper bound

The main paper remarks that for A in general position $|\operatorname{Ant}(A)| \geq 2^d - 2\binom{N}{2}$, so $\binom{N}{2} < 2^{d-1}$ guarantees an antichain direction. Define

$$N_0(d) = \min\Big\{N : \tbinom{N}{2} \geq 2^{d-1}\Big\},$$

$$N^*(d) = \min\{N : \exists A \subseteq \mathbb{R}^d \text{ in general position}, |A| = N,\ \operatorname{Ant}(A) = \emptyset\}.$$

The union bound says exactly $N^*(d) \geq N_0(d)$. Since $\operatorname{Ant}(A)$ depends only on the d coordinate rankings of A, we may take $A = \{(\pi_1(u), \ldots, \pi_d(u)) : u \in [N]\}$ for permutations $\pi_1, \ldots, \pi_d$ of $[N]$; relabelling the points lets us fix $\pi_1 = \mathrm{id}$. Then $\operatorname{Ant}(A) = \emptyset$ iff the $\binom{N}{2}$ sign vectors $(\operatorname{sgn}(\pi_i(v) - \pi_i(u)))_i$, normalised to have first coordinate $+$, exhaust $\{+\} \times \{-,+\}^{d-1}$.

Proposition 7.1 (Exact and near-exact tightness) *$N^*(d) = N_0(d)$ for $2 \leq d \leq 6$, namely $(N^*(2), \ldots, N^*(6)) = (3, 4, 5, 7, 9)$. Moreover $N^*(7) \in \{12, 13\}$ and $N^*(8) \in \{17, 18\}$.*

Proof Lower bounds are the union bound: $N_0(d) = 3, 4, 5, 7, 9, 12, 17$ for $d = 2, \ldots, 8$ ($\binom{3}{2} = 3 \geq 2$, $\binom{4}{2} = 6 \geq 4$, $\binom{5}{2} = 10 \geq 8$, $\binom{7}{2} = 21 \geq 16$, $\binom{9}{2} = 36 \geq 32$, $\binom{12}{2} = 66 \geq 64$, $\binom{17}{2} = 136 \geq 128$). Upper bounds are exhibited by the permutation tuples below. For each tuple the program enumerates every unordered pair and checks that all 2^{d-1} normalised sign patterns occur. The union bound proves minimality for $d \leq 6$; verification of a witness alone does not prove minimality in dimensions seven and eight. No claim about an exhaustive search over all tuples is required. ∎

Table 1: Verified covering witnesses. The lower bounds use only pair counting; coverage is checked by the supplied program.

d	$N_0(d)$	Witness size N	Classes covered
2	3	3	2
3	4	4	4
4	5	5	8
5	7	7	16
6	9	9	32
7	12	13	64
8	17	18	128

The following tuples use indices $u = 0, \ldots, N-1$, with $\pi_1(u) = u$. Each line lists one permutation in the order $\pi_2, \ldots, \pi_d$. Commas separate entries, including multi-digit ones.

Dimension 2, $N = 3$.

(0, 2, 1)

Dimension 3, $N = 4$.

(0, 2, 3, 1)
(0, 3, 1, 2)

1id1:core:20261120-0001@v1 1id1.net

Dimension 4**,** $N = 5$**.**

(0, 2, 4, 3, 1)
(0, 4, 2, 1, 3)
(1, 2, 4, 0, 3)

Dimension 5**,** $N = 7$**.**

(3, 2, 1, 4, 5, 6, 0)
(0, 6, 5, 1, 2, 4, 3)
(0, 4, 5, 6, 3, 1, 2)
(3, 1, 6, 4, 0, 5, 2)

Dimension 6**,** $N = 9$**.**

(4, 2, 6, 3, 0, 8, 7, 5, 1)
(4, 6, 0, 8, 3, 2, 5, 7, 1)
(2, 1, 5, 8, 6, 4, 7, 0, 3)
(3, 4, 5, 2, 8, 0, 7, 6, 1)
(0, 8, 7, 5, 2, 1, 3, 6, 4)

Dimension 7**,** $N = 13$**.**

(3, 11, 5, 6, 0, 2, 8, 9, 10, 12, 1, 7, 4)
(10, 4, 9, 1, 8, 2, 5, 3, 6, 7, 0, 12, 11)
(2, 9, 10, 11, 7, 6, 12, 0, 1, 3, 4, 5, 8)
(8, 11, 0, 7, 10, 6, 1, 2, 3, 12, 5, 4, 9)
(2, 7, 8, 1, 9, 12, 5, 6, 3, 10, 4, 11, 0)
(8, 5, 6, 0, 7, 4, 12, 1, 10, 9, 11, 3, 2)

Dimension 8**,** $N = 18$**.**

(7, 16, 11, 12, 3, 8, 1, 17, 2, 4, 10, 9, 13, 15, 14, 6, 5, 0)
(8, 12, 5, 13, 16, 1, 4, 11, 0, 15, 10, 3, 9, 2, 6, 17, 14, 7)
(0, 8, 15, 3, 13, 14, 1, 10, 7, 12, 17, 6, 9, 11, 2, 4, 5, 16)
(7, 12, 3, 10, 15, 16, 11, 1, 8, 2, 9, 0, 6, 17, 14, 13, 5, 4)
(7, 17, 0, 10, 8, 15, 3, 12, 14, 9, 2, 16, 1, 6, 4, 11, 5, 13)
(15, 13, 8, 6, 4, 3, 1, 2, 17, 10, 16, 5, 7, 11, 12, 14, 0, 9)
(9, 16, 10, 1, 5, 7, 11, 13, 12, 8, 3, 2, 17, 14, 0, 15, 4, 6)

Proposition 7.2 (General asymptotic upper bound) *As* $d \to \infty$*,*

$$2^{d/2} < N_0(d) \le N^*(d) \le (1 + o(1))\sqrt{2^d (d-1)\ln 2},$$

and consequently

$$\frac{N^*(d)}{N_0(d)} \le (1 + o(1))\sqrt{(d-1)\ln 2}.$$

Proof The first inequality follows from $\binom{N}{2} < N^2/2$. For the upper bound take

$$N = \left\lceil \sqrt{2^d((d-1)\ln 2 + \omega_d)} \right\rceil, \qquad \omega_d \to \infty, \quad \omega_d = o(d).$$

1id1:core:20261120-0001@v1 1id1.net

Then $c_N \to \infty$, and Theorem 5.1(a) shows that a random N-point sample (which is in general position almost surely) satisfies $\mathrm{Ant}(A_N) = \emptyset$ with positive probability, so such an A exists. Finally $N_0(d) \geq \sqrt{2^d}(1 - o(1))$. ∎

Remark 7.3 (Limits of the covering bounds) *The lower bound on the minimum covering* sample size *is attained for $d \leq 6$. This does not assert equality in the numerical inequality $|\mathrm{Ant}(A)| \geq 2^d - 2\binom{N}{2}$ for every displayed witness, nor sharpness of the sample-size lower bound in all dimensions. The remaining worst-case gap is at most $(1 + o(1))\sqrt{(d-1)\ln 2}$ in N. The random critical scale exceeds the counting scale by this factor in N, or by order d in the number of pairs.*

8 Remaining probabilistic and combinatorial questions

The first unresolved minimum-size cases are $d = 7, N = 12$ and $d = 8, N = 17$. A covering witness at either size would settle the corresponding case, whereas nonexistence requires a proof beyond a failed finite search.

A separate problem is the joint law of the many hit classes beyond the fixed-dimensional marginal estimates proved here. We do not postulate a Poisson point process in total variation for the entire growing class set. For models with ties, negative constraints can have support three, so random signed NAE satisfiability becomes relevant. The general-position negative version space is the complement of a list of classes and is easy to describe, but is *not* generally affine: for $d = 3$, a single negative full-support pair leaves six raw directions, which is not a power of two.

References

Richard Arratia, Larry Goldstein, and Louis Gordon. Two moments suffice for Poisson approximations: The Chen–Stein method. *The Annals of Probability*, 17(1):9–25, 1989. doi: 10.1214/aop/1176991491.

Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Information and Computation*, 125(1):1–12, 1996. doi: 10.1006/inco.1996.0016.

Svante Janson. New versions of Suen's correlation inequality. *Random Structures & Algorithms*, 13(3–4):467–483, 1998.

W.-C. Stephen Suen. A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph. *Random Structures & Algorithms*, 1(2):231–242, 1990.

1id1:core:20261120-0001@v1 1id1.net