Posterior-Cut Preconditioning for Exact Learning of k-Term DNFs

J. R. Landers

May 2026 draft

Abstract

Exact learning with membership and equivalence queries can be viewed as posterior collapse: the surviving hypotheses must become functionally identical on the Boolean cube. This note develops a DNF-specific preconditioning framework based on positive equivalence counterexamples. A positive witness and its one-bit neighborhood define a local fiber that certifies active literals; under an isolation condition the witness recovers and peels an entire term. If peeling fails, a peel-or-overlap dichotomy forces rescued one-bit boundary structure, yielding a boundary-rescue compression statement. The framework also explains equivalence queries as global posterior cuts and interprets the recent Alman–Nadimpalli–Patel–Servedio exact learner as constructing compact central hypotheses in an induced feature geometry. The main certified consequence is the instance-dependent bound

$$Q(f) \le q(n+1) + \mathrm{poly}(n)2^{\widetilde{O}(\sqrt{k-q})}$$

whenever q terms are peeled before invoking the residual learner. Numerical probes show that local peeling removes many terms in sparse and moderately dense random DNFs, while the obstruction regime exhibits high boundary-rescue overlap, suggesting a concrete route toward residual compression.

1 Introduction

Let $\mathcal{X} = \{0,1\}^n$ and let H_k be the class of Boolean functions representable as a DNF with at most k terms. In Angluin's membership/equivalence query model [1], the learner may ask labels $f(x)$ and may propose a hypothesis g, receiving either acceptance or a counterexample. The long line of exact DNF learning work includes the classical $\mathrm{poly}(n, 2^k)$ algorithm of Blum and Rudich [4], monotone-theory and related exact query algorithms [5, 6, 12], and limitations for proper or strongly proper DNF hypotheses [8, 9]. Alman, Nadimpalli, Patel, and Servedio recently obtained the first improvement in the k dependence, learning k-term DNFs in $\mathrm{poly}(n)2^{\widetilde{O}(\sqrt{k})}$ time using membership and equivalence queries [2]. Recent work has also revisited DNF learning under membership-query PAC and random-walk models [3], reinforcing the role of local movement on the Boolean cube.

This paper asks what these algorithms are doing geometrically. A transcript defines a version space $V_t \subseteq H_k$. A membership query chooses a point and asks for a local split of V_t. An equivalence query chooses a full hypothesis; if rejected, the oracle supplies a point at which the proposal is wrong, thereby cutting the posterior by the corresponding disagreement slice. The ideal proposal is the posterior-majority function, but it need not have a compact DNF representation. Thus the central problem is representational: construct hypotheses that are near the posterior center in an induced geometry.

The contribution here is a certified preconditioning mechanism for this central-hypothesis problem. Positive counterexamples are structural for DNFs. One-bit perturbations of a positive witness certify active literals; if the witness isolates a residual term, the term is exactly recovered and peeled. If the term cannot be peeled, then its private region must have one-bit boundary points rescued by other terms. This gives both an exact data-dependent preprocessing bound and a structural description of the obstruction.

2 Posterior Cuts

After a transcript τ_t, define

$$V_t = \{h \in H_k : \ h \text{ is consistent with } \tau_t\}.$$

With any full-support prior, the noiseless posterior is supported exactly on V_t. For $x \in \mathcal{X}$, let $p_t(x) = \mathbb{P}_{h\sim P_t}[h(x) = 1]$ and

$$D_t = \{x : \exists h, g \in V_t \text{ with } h(x) \neq g(x)\}.$$

Exact learning is posterior collapse: $D_t = \varnothing$, equivalently $p_t(x) \in \{0,1\}$ for all x.

A membership query at x partitions the version space into two label classes. The best guaranteed local split is

$$\alpha(V_t) = \max_x \min\{P_t[h(x)=0], P_t[h(x)=1]\}.$$

Equivalence queries give a complementary global cut. For a proposal g, define

$$\Lambda_t(g) = \max_{x \in \mathcal{X}} P_t\{h \in V_t : h(x) \neq g(x)\}.$$

If $\Lambda_t(g) \leq 1-\beta$, then every rejected equivalence query to g leaves posterior mass at most $1-\beta$.

Lemma 1 (Majority halving)**.** *Let* $g_t^{\mathrm{Maj}}(x) = \mathbf{1}[p_t(x) \geq 1/2]$*. Then* $\Lambda_t(g_t^{\mathrm{Maj}}) \leq 1/2$*.*

Proof. At each x, the posterior mass disagreeing with the majority label is at most $1/2$. Taking the maximum over possible counterexamples proves the claim. □

The lemma is information-theoretic. For structured classes, the obstacle is constructing a compact enough surrogate. This is closely aligned with online mistake-bound learning and attribute-efficient methods such as Winnow [14], and with PTF representations for DNF [10]. The same viewpoint clarifies why proper equivalence query learning can be substantially harder than unrestricted equivalence queries [8, 9]: the best central proposal need not live in the syntactic class used to describe the target.

3 Positive Witness Fibers

DNFs are asymmetric: a positive example says that at least one term fired. For $x \in \mathcal{X}$, define the agreeing literal

$$\lambda_i^x = \begin{cases} z_i, & x_i = 1, \\ \neg z_i, & x_i = 0. \end{cases}$$

Definition 2 (Local kill set)**.** *For a positive point* x^+*, define*

$$K_f(x^+) = \{i \in [n] : f(x^+ \oplus e_i) = 0\}.$$

Lemma 3 (Killed neighbors certify active literals)**.** *Let* $f = T_1 \vee \cdots \vee T_m$ *and suppose* $f(x^+) = 1$*. If* $f(x^+ \oplus e_i) = 0$*, then every term active at* x^+ *contains* $\lambda_i^{x^+}$*.*

Proof. If an active term did not contain the agreeing literal in coordinate i, then flipping coordinate i would leave that term satisfied, forcing $f(x^+ \oplus e_i) = 1$. □

Thus a positive witness and its one-bit neighborhood define a fiber inside the version space. If $|K_f(x^+)| = s$, active conjunctions satisfied by x^+ are reduced by a factor 2^{-s}: the killed coordinates are mandatory literals and the remaining coordinates are optional.

4 Certified Peeling

Definition 4 (Isolated witness)**.** *Write* $f = T \vee R$*, where* $R \in H_{k-1}$*. A point* x^+ *is isolated for* T *if*

$$T(x^+) = 1, \qquad R(x^+) = 0, \qquad R(x^+ \oplus e_i) = 0 \ \ \forall i \in \mathrm{var}(T).$$

Theorem 5 (One-term recovery)**.** *If* x^+ *is isolated for* T*, then after querying all* $x^+ \oplus e_i$*,*

$$K_f(x^+) = \mathrm{var}(T), \qquad T = \bigwedge_{i \in K_f(x^+)} \lambda_i^{x^+}.$$

Proof. If $i \notin \mathrm{var}(T)$, flipping i leaves T true, so the label remains positive. If $i \in \mathrm{var}(T)$, flipping i kills T and isolation prevents any residual term from rescuing the point. The recovered literals are exactly the agreeing literals. □

The preconditioner maintains a recovered DNF G_q with no false positives and asks $G_q \overset{?}{=} f$. Any rejected equivalence query gives a positive witness outside the peeled terms. If isolated, one local star peels another term.

Corollary 6 (Residual ANPS bound)**.** *If q terms are certified and peeled, then*

$$Q_{\text{pre}}(n,k) \leq q(n+1) + Q(n,k-q).$$

Using the ANPS learner [2] for the residual,

$$Q_{\text{pre}}(n,k) \leq q(n+1) + \text{poly}(n)2^{\widetilde{O}(\sqrt{k-q})}.$$

This is not a worst-case improvement over ANPS. It is a certified, instance-dependent reduction of the parameter that drives the augmented-PTF degree, Winnow weight bound, stem-search probability, and feature count.

5 Peel-or-Overlap Geometry

Let $f = T_1 \vee \cdots \vee T_m$. For a term T_j, set $R_j = \bigvee_{\ell \neq j} T_\ell$ and

$$P_j = \{x : T_j(x) = 1,\ R_j(x) = 0\}.$$

Theorem 7 (Peel-or-overlap dichotomy)**.** *For every term T_j, exactly one of the following holds:*

1. *$P_j = \varnothing$, so T_j is redundant in this representation.*
2. *There exists $x \in P_j$ such that $R_j(x \oplus e_i) = 0$ for all $i \in \text{var}(T_j)$.*
3. *$P_j \neq \varnothing$, T_j is not locally peelable, and every $x \in P_j$ has a rescued one-bit boundary:*

$$\exists i \in \text{var}(T_j) \quad R_j(x \oplus e_i) = 1.$$

Moreover, for $r_j = |\,\text{var}(T_j)|$, some coordinate rescues at least $|P_j|/r_j$ private points.

Proof. If P_j is empty, the first case holds. Otherwise, either a private point has no rescued boundary in the variables of T_j, giving a peel witness, or every private point has a rescued boundary. The final statement follows by covering P_j by the r_j rescue sets and applying the pigeonhole principle. □

Failure to peel is therefore not lack of structure; it is certified overlap.

6 Boundary-Rescue Compression

Represent each term as a subcube $C_j = \{x : T_j(x) = 1\}$ and let $S_j = \text{var}(T_j)$. Define private density $\rho_j = |P_j|/|C_j|$.

Theorem 8 (Nearby rescuer)**.** *If T_j is nonredundant and not locally peelable, then there exist $i^\star \in S_j$ and $\ell \neq j$ such that, with $a_{j\ell} = |S_\ell \setminus S_j|$,*

$$a_{j\ell} \leq \log_2\left(\frac{r_j(k-1)}{\rho_j}\right).$$

Proof. By the dichotomy and pigeonhole, some coordinate-rescuer pair $(i^\star, \ell)$ rescues at least $|P_j|/(r_j(k-1))$ private points. On the larger cube C_j, the constraints introduced by variables in $S_\ell \setminus S_j$ cut measure by $2^{-a_{j\ell}}$. Hence $|P_j|/(r_j(k-1)) \leq 2^{-a_{j\ell}}|C_j|$, giving the bound. □

If all terms have width at most w and $\rho_j \geq \rho$, then every unpeelable nonredundant term has a rescuer with

$$|S_\ell \setminus S_j| \leq L = \left\lceil \log_2\left(\frac{w(k-1)}{\rho}\right) \right\rceil.$$

This defines a boundary-rescue graph. A cluster with one anchor and rescue radius L has rough description length

$$\log N_w + (m-1)\log B(n,w,L), \quad B(n,w,L) = w2^w \sum_{a=0}^{L} \binom{n}{a} 2^a,$$

instead of $m \log N_w$, where $N_w = \sum_{a=0}^{w} \binom{n}{a} 2^a$.

7 Certified Overlap Modules

The boundary-rescue statement is a compression certificate, not by itself a complete learner. One setting in which overlap compression becomes an exact learner is a query-separable module decomposition.

Definition 9 (Query-separable overlap decomposition)**.** *Let $f = G_q \vee R$, where G_q is the peeled DNF. A residual decomposition $R = M_1 \vee \cdots \vee M_c$ is query-separable with local width ℓ if each module M_a depends only on a variable set Y_a with $|Y_a| \le \ell$, and for every a and every assignment $y \in \{0,1\}^{Y_a}$ there is a constructible point $x(a,y) \in \mathcal{X}$ such that*

$$x(a,y)_{Y_a} = y, \qquad G_q(x(a,y)) = 0, \qquad M_b(x(a,y)) = 0 \quad \forall b \neq a.$$

Theorem 10 (Exact learning from certified modules)**.** *If a preconditioning phase peels q terms and certifies a query-separable residual decomposition into c modules of local width at most ℓ, then f is exactly learnable using*

$$q(n+1) + c2^\ell + 1$$

queries after the certificates are available.

Proof. For each module M_a, query $x(a,y)$ for all $y \in \{0,1\}^{Y_a}$. At this point the peeled terms and all other modules are off, so the observed label is exactly $M_a(y)$. Thus the full truth table of each module is recovered using at most 2^ℓ membership queries. The final hypothesis is $G_q \vee \widehat{M}_1 \vee \cdots \vee \widehat{M}_c$, and one equivalence query verifies equality. □

This theorem is deliberately conditional. It identifies the missing structural step needed to turn boundary-rescue compression into a worst-case improvement: prove that high rescue-rate residuals organize into query-separable modules, or into another compact representation from which central equivalence hypotheses can be constructed.

8 Relation to ANPS

ANPS maintains eligible pairs $(T', R_{T'})$, where T' is a candidate stem and $R_{T'}$ is an auxiliary variable set. Its feature space contains augmented monomials

$$T' \cdot \prod_{i \in S} x_i, \qquad S \subseteq R_{T'}, \ |S| \le d_{\max}.$$

If the maintained family is fully expressive, every target term has a valid stem whose missing literals lie in its $R_{T'}$. Then a Chebyshev construction expresses the DNF as an augmented PTF with

$$d_{\max} = O(\sqrt{k \log k}), \qquad W_{\max} = 2^{O(\sqrt{k}\log^2 k)}.$$

Winnow2 over the induced feature space learns the corresponding threshold function. In posterior-cut language, positive counterexamples grow stems, negative counterexamples grow auxiliary variables, and Winnow mistakes are equivalence-counterexample cuts in the induced geometry. The relevant-variable subroutine uses noise-operator ideas related to junta testing [13, 7]; in the present language, that subroutine prevents long residual terms from inflating the local feature set.

Certified peeling should be performed before this phase when possible. A peeled term is not merely a discovered stem; it is removed from the residual target. Consequently the ANPS parameters can be instantiated with $s = k - q$.

More explicitly, after q certified peels the residual augmented-PTF degree and weight may be bounded as

$$d_s = O(\sqrt{s \log s}), \qquad W_s = 2^{O(\sqrt{s} \log^2 s)}, \qquad s = k - q.$$

The feature-count term also changes from roughly

$$|\mathcal{F}| \binom{O(k^2 \log k)}{\leq d_k} \quad \text{to} \quad |\mathcal{F}_{\mathrm{res}}| \binom{O(s^2 \log s)}{\leq d_s},$$

up to the usual $\widetilde{O}(\cdot)$ factors. This is the precise sense in which peeling lowers the parameter driving the enhanced-feature phase.

9 Transcript Programs

The same mechanisms can be expressed at the level of adaptive query programs. Let

$$\sigma = (a_1, \ldots, a_T), \qquad a_t \in \{\mathrm{MQ}(x_t), \mathrm{EQ}(g_t)\},$$

where x_t and g_t may depend on the previous transcript. Define the retention ratio

$$r_t = \frac{P_0(V_{t+1})}{P_0(V_t)}, \qquad c_t = 1 - r_t.$$

For a membership query, r_t is the posterior mass of the returned label side. For a rejected equivalence query, r_t is the posterior mass of the counterexample slice. A sequence should be evaluated not only by $\prod_t r_t$, but by the structural state it leaves behind:

$$\left(\prod_{t=1}^{T} r_t, \quad \text{certificates accumulated by } \sigma \right).$$

Positive-witness stars are valuable because they score well in both coordinates: they often leave a small posterior fiber, and they certify literals, peels, or rescued boundaries. This explains why a purely greedy one-step hybrid can underperform a more structural transcript. The immediate cut may be smaller, but the induced residual geometry may be better for constructing central hypotheses.

10 Numerical Probes

The experiments are geometric probes, not asymptotic evidence. Scripts and the HTML companion note are included with the source.

All exact finite-version-space experiments represent Boolean functions as truth-table integers. For $n = 4$, all non-contradictory conjunctions are enumerated, all disjunctions of at most k such conjunctions are formed, and duplicates are removed by truth table. This gives the exact finite class H_k as a set of functions, rather than syntactic formulas. Membership cuts, posterior-majority equivalence cuts, and legal-H_k center proposals are then computed exactly from the current version space.

The syntactic DNF experiments sample fixed-width terms without replacement. For each sampled target, the full Boolean cube is enumerated to identify active terms, private regions, locally isolated witnesses, and rescued one-bit boundaries. This brute-force step is intentionally not proposed as an algorithm; it is a microscope for the geometry predicted by the theorems.

10.1 Exact finite version spaces

For $n = 4$, distinct truth-table functions in H_k were enumerated.

k	$\|H_k\|$	MQ cut	Maj-EQ cut	best H_k retained
1	82	.195	.805	.195
2	1,886	.320	.680	.320
3	15,358	.418	.582	.418
4	44,262	.472	.528	.472
5	61,294	.495	.505	.495

Majority equivalence proposals invert the local skew: small membership minorities become strong counterexample cuts.

For $n = 4, k = 3$, over 80 sampled targets with adversarial counterexamples:

policy	mean queries	mean cut/query
greedy MQ	15.44	.477
posterior-majority EQ	9.56	.583
best legal H_k EQ	10.49	.557
one-step hybrid	11.10	.555

The positive-witness star $\mathrm{EQ}(0) \to x^+$ followed by all $x^+ \oplus e_i$ left mean retained mass .0258 for a random positive witness and .0234 for the best positive witness in the sampled $n = 4, k = 3$ runs.

The representability bottleneck in the central-hypothesis view was also visible in the traces. Across 936 posterior states visited by majority-EQ runs for $n = 4, k = 3$, the exact posterior-majority truth table itself belonged to H_k in approximately 75.2% of states. When it did not, the best legal H_k center was often close: the mean retained-mass gap between the legal center and the true posterior majority was .008, although the largest observed gap was .167. This supports the interpretation that centrality is information theoretically easy but representationally delicate.

10.2 Sequential peeling

Random fixed-width DNFs were sampled, and a finite-cube routine repeatedly peeled locally isolated terms. The table reports 120 trials per row.

n	k	w	mean q	mean s	$\sqrt{k} - E\sqrt{s}$	rescue
10	16	4	15.57	0.42	2.903	.041
10	32	4	26.22	5.78	3.655	.214
12	32	4	31.27	0.73	4.489	.067
12	48	4	32.86	15.14	3.859	.315
12	64	4	2.31	61.69	0.213	.755
14	64	5	63.68	0.32	6.930	.027

A phase curve at $n = 12, w = 4$ shows the transition. Full peel rates fall from 92.5% to 0% across the tested range, while residual rescue rates rise from .019 to .758. This matches the dichotomy: the obstruction to peeling appears as boundary-rescue overlap.

k	mean peeled	median s	full peel	stuck start	rescue
16	15.85	0	92.5%	0.0%	.019
24	23.64	0	83.8%	0.0%	.043
32	30.99	0	72.5%	0.0%	.078
40	39.01	0	68.8%	0.0%	.087
48	32.59	2	37.5%	2.5%	.297
56	7.50	55	3.8%	35.0%	.666
64	1.00	64	0.0%	62.5%	.758

This table suggests two regimes. In the first, isolated private witnesses are abundant and the residual parameter collapses before the enhanced-feature phase. In the second, private mass is surrounded by rescuers, so the next improvement must exploit overlap rather than search harder for isolated witnesses.

11 Artifacts and Reproducibility

The accompanying workspace contains three kinds of artifacts. First, the two source notes develop the posterior-collapse and central-equivalence viewpoints separately. Second, the scripts `cut_sequence_experiments.py` and `anps_witness_preconditioning_experiments.py` generate the finite version-space and random syntactic DNF probes reported above. Third, the HTML notes provide an interactive reading layer with the same numerical tables.

The exact $n = 4$ enumeration is deterministic once n and k are fixed. Terms are represented by non-contradictory literal patterns, truth tables are encoded as integers, and syntactic duplicates are removed. The random syntactic probes use explicit seeds and fixed-width terms sampled without replacement. Because the syntactic probes enumerate the entire Boolean cube, they should be read as measurements of geometry rather than scalable algorithms.

The paper makes three separations that are useful for checking future work:

- *certification versus discovery*: the peeling theorem is a certified reduction once an isolated witness is found, while finding many such witnesses efficiently remains a separate algorithmic question;
- *peeling versus overlap*: failure of the local certificate is not discarded, but recorded as rescued boundary structure;
- *information versus representation*: posterior-majority centers give ideal cuts, but exact learning needs compact hypotheses that can be submitted to an equivalence oracle.

These separations are the intended interface with ANPS. Any future improvement can target one of them: better witness discovery, stronger overlap decomposition, or more compact central surrogates over the induced residual geometry.

12 Discussion

The present results do not improve the worst-case ANPS bound. They give:

- a certified preconditioning theorem that lowers the residual term parameter when isolated witnesses are found;
- a structural obstruction theorem showing that failure to peel forces rescued boundary overlap;
- a compression lemma for overlap-heavy residuals under width and private density assumptions;
- numerical evidence that the peelable and overlap-heavy regimes are visible in finite DNF geometries.

The next target is a representation theorem of the form

$$\text{few peelable terms} \Longrightarrow \text{compact overlap modules or compact central surrogates.}$$

Such a theorem would convert the overlap side of the dichotomy into a true worst-case improvement. Until then, posterior-cut preconditioning gives a rigorous instance-dependent refinement and a concrete language for attacking the remaining residual geometry.

References

[1] D. Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, 1988.

[2] J. Alman, S. Nadimpalli, S. Patel, and R. A. Servedio. Faster exact learning of k-term DNFs with membership and equivalence queries. *arXiv:2507.20336*, 2025.

[3] J. Alman, S. Nadimpalli, S. Patel, and R. A. Servedio. DNF learning via locally mixing random walks. *arXiv:2505.18839*, 2025.

[4] A. Blum and S. Rudich. Fast learning of k-term DNF formulas with queries. *Journal of Computer and System Sciences*, 51(3):367–373, 1995. Preliminary version in STOC 1992.

[5] N. H. Bshouty. Exact learning Boolean functions via the monotone theory. *Information and Computation*, 123(1):146–153, 1995.

[6] N. H. Bshouty. A subexponential exact learning algorithm for DNF using equivalence queries. *Information Processing Letters*, 59(1):37–39, 1996.

[7] N. H. Bshouty. Almost optimal distribution-free junta testing. *arXiv:1901.00717*, 2019.

[8] L. Hellerstein and V. Raghavan. Exact learning of DNF formulas using DNF hypotheses. *Journal of Computer and System Sciences*, 70(4):435–470, 2005.

[9] L. Hellerstein, D. Kletenik, L. Sellie, and R. A. Servedio. Tight bounds on proper equivalence query learning of DNF. In *Proceedings of COLT*, volume 23 of JMLR Proceedings, pages 31.1–31.18, 2012.

[10] A. R. Klivans and R. A. Servedio. Learning DNF in time $2^{\widetilde{O}(n^{1/3})}$. *Journal of Computer and System Sciences*, 68(2):303–318, 2004.

[11] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, 1993.

[12] E. Kushilevitz. A simple algorithm for learning $O(\log n)$-term DNF. *Information Processing Letters*, 61(6):289–292, 1997.

[13] Z. Liu, X. Chen, R. A. Servedio, Y. Sheng, and J. Xie. Distribution-free junta testing. *ACM Transactions on Algorithms*, 15(1):1–23, 2018.

[14] N. Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1988.

[15] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.