THE UPPER HALF OF THE WEIGHTED DATA-SELECTION GAP: $F_w(d,n) = 1 + (2d-n)/d$ FOR $2n \geq 3d$

KOREA SUPERINTELLIGENCE LABS

Abstract. Hanneke, Moran, Shlimovich and Yehudayoff determined the worst-case risk ratio $F_w(d,n)$ of the minimum-norm ERM under a weighted selection of n points of a dataset in $\mathbb{R}^d \times \mathbb{R}$ outside the window $d < n < 2d$, and asked for its value inside that window (Question 2 of their COLT 2025 open problem). Zhang has since proved the lower bound $F_w(d, d+k) \geq 1 + \Gamma_{d,k}$ for an explicit harmonic $\Gamma_{d,k}$, conjectured equality throughout the window, settled the cells $n = 2d-1$, $(3,4)$ and $(4,5)$, and named $(d,n) = (4,6)$ in print as the smallest open cell. We prove that conjecture on the upper half of the window: $F_w(d,n) = 1 + (2d-n)/d$ for every $d \geq 1$ and every n with $\lceil 3d/2 \rceil \leq n \leq 2d$, so in particular $F_w(4,6) = 3/2$. The proof is a case split on a minimum-cardinality positively spanning subset of the whitened gradients — exact recovery, a circuit interface, a penalty anchor and a rigidity branch — closed by an induction that we isolate as a purely arithmetic recursion. That recursion, the lower-bound witness, the collapse $\Gamma_{d,k} = (d-k)/d$ on $2k \geq d$ and the recursion's exact reach for $d \leq 30$ are formalised in Lean 4 and checked by its kernel; the geometric branch inequalities are proved here as ordinary mathematics and are not formalised. The restriction $2n \geq 3d$ is not cosmetic: over the whole window the formula is false, since $F_w(4,5) = 2$. The argument uses neither Steinitz's interior theorem nor volume sampling.

1. Introduction

1.1. **Weighted data selection for the minimum-norm ERM.** A *dataset* is a finite multiset $D = \{(x_i, y_i)\}_{i=1}^N$ in $\mathbb{R}^d \times \mathbb{R}$. With the squared loss $\ell_{(x,y)}(w) = (\langle w, x\rangle - y)^2$ we write $L_D(w) = \frac{1}{N}\sum_i \ell_{(x_i,y_i)}(w)$ and $L_D^* = \min_w L_D(w)$. The learner is the *minimum-norm ERM* A^*: for a nonnegatively weighted objective F, the set $\operatorname{argmin} F$ is a translate of a linear subspace, and $A^*(F)$ is its unique element of smallest Euclidean norm. A *weighted selection with budget* n means choosing $z_1, \ldots, z_n$ in D and F in $\operatorname{conv}(\ell_{z_1}, \ldots, \ell_{z_n})$; we follow the formulation of [1], §2.1. Selections are counted by *effective support*: a repeated point merges its weights, and a point given weight 0 counts as unselected. Then

$$L_D^*(n; A^*) = \inf_{\text{selections}} L_D\big(A^*(F)\big), \qquad F_w(d,n) = \sup_D \frac{L_D^*(n;A^*)}{L_D^*}, \tag{1}$$

with the conventions $0/0 = 1$ and $c/0 = \infty$ for $c > 0$, again as in [1]: a supremum over datasets against an infimum over selections, the adversary picking the data, the selector optimal and seeing all of D, the learner fixed.

Date: September 3, 2026.

1.2. **What is known, and the open window.** Hanneke, Moran, Shlimovich and Yehudayoff [2] computed $F_w(d,n)$ outside one window. Their Theorem 8 reads, verbatim: "Let A^* denote the min-norm ERM. Then, $\sup_D L^*_D(n;A^*)/L^*_D = 1$ if $n \geq 2d$, $d+1$ if $n = d$, ∞ if $n < d$, where D ranges over all finite datasets and $L^*_D = \inf_{w\in W} L_D(w)$ denotes the optimal loss." (Numbers cited from [2] are those of its v2 PDF; the experimental HTML rendering of the same version runs a single counter through the definitions and calls this statement Theorem 9.) The endpoint $n \geq 2d$ is obtained through Steinitz's interior refinement of Carathéodory's theorem and the endpoint $n = d$ through volume sampling, so the two ends of the window have different geometries and nothing interpolates between them for free. Immediately after the theorem, in their Section 4, they record that $d < n < 2d$ remains open, with only the bound $O(d)$ established there. The same authors pose this as Question 2 of their COLT 2025 open problem [3], adding, verbatim: "Theorem 1 implies that this quantity lies in the interval $(1, d+1]$. In the case $n = 2d-1$, we believe we have an argument showing that the value is exactly $1 + 1/d$, though the argument is somewhat ad hoc and does not clearly extend to other values of n."

Zhang [1] introduced an explicit candidate value. Write $n = d+k$ with $1 \leq k \leq d-1$; for $s \in \{k+1,\dots,d\}$ write $d = q_s s + a_s$ with $q_s \geq 1$ and $0 \leq a_s < s$, and set

$$C_{d,s} = \left(\frac{s-a_s}{q_s} + \frac{a_s}{q_s+1}\right)^{-1}, \qquad \Gamma_{d,k} = \max_{k+1\leq s\leq d} (s-k)\, C_{d,s}; \tag{2}$$

equivalently $1/C_{d,s}$ is the minimum of $\sum_{j=1}^{s} 1/r_j$ over positive integer partitions $r_1 + \cdots + r_s = d$, attained at the balanced partition. Reference [1] proves $F_w(d, d+k) \geq 1 + \Gamma_{d,k}$ for $1 \leq k \leq d-1$ (its Theorem 13), the exact values $F_w(d, 2d-1) = 1+1/d$ (Corollary 11), $F_w(3,4) = 5/3$ (Theorem 27) and $F_w(4,5) = 2$ (Theorem 47), and conjectures equality throughout the window (its Conjecture 54). Its Discussion says, verbatim: "The smallest open cell is $(d,n) = (4,6)$, with conjectured value $\frac{3}{2}$."

1.3. **Results.**

Theorem 1 (Theorem A)**.** *For every $d \geq 1$ and every n with $\lceil 3d/2\rceil \leq n \leq 2d$,*

$$F_w(d,n) = 1 + \frac{2d-n}{d}.$$

Writing $n = 2d-j$, the range is $d \geq 2j$ and the value is $1 + j/d$. The instance $j = 1$ recovers $F_w(d, 2d-1) = 1+1/d$ for every $d \geq 2$ in three lines. The smallest instance not already in the literature is the cell [1] names as open, $F_w(4,6) = \frac{3}{2}$; the next new ones are $F_w(5,8) = 7/5$, $F_w(6,9) = 3/2$ and $F_w(6,10) = 4/3$. Three things must be said about the scope.

The value is not ours. On $2n \geq 3d$ the quantity $1 + (2d-n)/d$ is exactly the value conjectured in [1], because $\Gamma_{d,k} = (d-k)/d$ whenever $2k \geq d$ (Theorem 24). What is new is the proof, not the question and not the shape of the answer; the lower bound and even its extremal dataset are published as well (Section 3).

The range is load-bearing. Over the whole window the formula is false: at $(4,5)$ it predicts $7/4$, while [1] proves $F_w(4,5) = 2$; the cell $(3,4)$, also outside the range, happens to agree. Our method stops exactly at $2n = 3d$ for an exhibited reason: the arithmetic step driving the induction fails at precisely $(3,4)$ and $(4,5)$

(Proposition 19), and the recursion computed to $d \le 30$ settles the cells with $2n \ge 3d$ and no others (Theorem 26).

Two halves, verified differently. The upper bound splits into a geometric reduction, producing a system of inequalities for the extremal excess risk, and an induction closing them. The induction is isolated here as a statement about an arbitrary function $P\colon \mathbb{N}\times\mathbb{N}\to\mathbb{Q}$ satisfying those inequalities (Theorem 17); it, the lower-bound optimisation (Theorem 6), the collapse of Γ (Theorem 24) and the reach of the recursion (Theorem 26) are formalised in Lean 4 and checked by its kernel. The geometric reduction — Propositions 9–15 — is proved here as ordinary mathematics and is *not* formalised; Section 10 says why.

Two by-products. The argument uses neither Steinitz's interior theorem nor volume sampling: wherever the former extracts at most $2m$ gradients with 0 interior to their hull, a minimum-cardinality positively spanning subset does the same job by Lemmas 4 and 5, and the recursion with the volume-sampling base case deleted settles exactly the same cells (Theorem 26). And the maximum in (2) is attained at $s = d$ whenever $2k \ge d$ (Theorem 24), and empirically down to $k \ge \lceil d/3\rceil$ — a wider band than Theorem 1 reaches, so $\lceil d/3\rceil \le k < d/2$ is where a sharper circuit branch would pay.

1.4. **Concurrent work.** This open problem has attracted several unrefereed single-author notes in a short period, and priority is delicate. Besides [1] (28 August 2026) and its vector-valued companion [7], which does not touch the scalar window, there are Zenodo deposits by Huang [4] (2 August 2026), proving $F_w(3,4) = 5/3$ (its Theorem 1.1) and $F_w(d, 2d-1) = 1 + 1/d$ (its Theorem 1.2); an anonymous one [5] (11 August 2026), proving $F_w(d, 2d-1) = 1 + 1/d$ and the same harmonic lower bound and showing that bound exact on rank-additive circuit decompositions, while stating that the general open curve needs a reduction it does not have; and two by Yang [6], on $d = 2$, $n = 3$ lower bounds only. We are not aware of an upper bound at $(4, 6)$, or anywhere in the window beyond $n = 2d - 1$, in any of them.

2. Preliminaries

2.1. **Whitening.** Two degenerate situations go first. If $L_D^* = 0$ the ratio in (1) is 1 by the convention $0/0 = 1$, d independent features already determining the predictor. If the features span only a subspace of dimension $\rho < d$, the problem is isometric to a ρ-dimensional one with the same budget, predictions and the minimum-norm output both living in that span, and $(2\rho - n)_+/\rho \le (2d - n)_+/d$ since $m \mapsto 2 - n/m$ increases, so the d-dimensional bound covers it.

So assume $L_D^* > 0$ and full feature rank, and put $H = \frac{1}{N}\sum_i x_i x_i^\top$ (positive definite), $w^* = \operatorname{argmin} L_D$ and $r_i = y_i - \langle w^*, x_i\rangle$. The change of variables $\xi_i = H^{-1/2}x_i$, $\eta_i = r_i/\sqrt{L_D^*}$ and $q = H^{1/2}(w - w^*)/\sqrt{L_D^*}$ is Lemma 1 of [1]; it is elementary and we use it as stated there. It produces

$$(3)\qquad \frac{1}{N}\sum_i \xi_i\xi_i^\top = I_d, \quad \frac{1}{N}\sum_i \eta_i\xi_i = 0, \quad \frac{1}{N}\sum_i \eta_i^2 = 1, \quad \frac{L_D(w)}{L_D^*} = 1 + \|q\|^2.$$

Definition 2. A family $(\xi_i, \eta_i)_{i=1}^N$ in $\mathbb{R}^m \times \mathbb{R}$ satisfying the first three identities of (3), with I_m in place of I_d, is a *whitened system* in $\mathbb{R}^m$; its *gradients* are $g_i = \eta_i\xi_i$, and $\sum_i g_i = 0$. We write E for the residual energy $\frac{1}{N}\sum \eta_i^2$ of a sub-family, normalised to 1 for the system itself.

Definition 3. For $m \geq 1$ and an integer $b \geq 0$ let $\Psi_m(b) = \sup\inf\|\hat{q}\|^2$, the supremum over whitened systems in $\mathbb{R}^m$ and the infimum over selections of at most b of their points with strictly positive weights whose objective is strictly convex, $\hat{q}$ being that objective's unique minimiser. In *deficit coordinates* put $P(m,j) = \Psi_m(2m-j)$.

Restricting the inner infimum to strictly convex selections only increases it, and such an objective has a unique minimiser, so the tie rule defining A^* never acts on the selections counted in Definition 3. With (3) and §2.1 this gives

$$F_w(d,n) \;\leq\; 1 + \Psi_d(n), \tag{4}$$

and Ψ is homogeneous in the residual energy: a family satisfying the first two identities of (3) with $\frac{1}{N}\sum \eta_i^2 = E$ has excess at most $E \cdot \Psi_m(b)$, by rescaling η.

2.2. **Positive spanning sets.** For finite $S \subset \mathbb{R}^m \setminus \{0\}$ let $\mathrm{pos}(S)$ be the set of nonnegative combinations of S; S *positively spans* $\mathbb{R}^m$ when $\mathrm{pos}(S) = \mathbb{R}^m$, and a *positive basis* is an inclusion-minimal such S. We use two classical facts.

Lemma 4 (P1)**.** *If S positively spans $\mathbb{R}^m$ there are $\lambda_s > 0$ with $\sum_{s\in S}\lambda_s s = 0$.*

Proof. Write $-s = \sum_{t\in S}\beta_{st}t$ with $\beta_{st} \geq 0$ for each s, and sum over s; every coefficient of the resulting relation is at least 1. □

Lemma 5 (P2)**.** *A positive basis of $\mathbb{R}^m$ has at least $m+1$ and at most $2m$ elements, and one with exactly $2m$ elements is $\{a_1, -l_1a_1, \ldots, a_m, -l_m a_m\}$ for a linear basis $a_1, \ldots, a_m$ of $\mathbb{R}^m$ and scalars $l_j > 0$.*

Lemma 5 we do not prove; it is classical, and its attribution deserves a word. Davis [8] introduced positively spanning sets and positive bases and set out to classify them; the bounds $m+1 \leq |S| \leq 2m$ are credited to that paper throughout the literature, and McKinney [9], who continues Davis's investigation in infinite dimensions, omits the finite-dimensional proofs as "easy extensions of those … which appear in" [8]. Audet [10] gives a short proof of the upper bound $2m$ — of that bound only, as its abstract states. For the size-$2m$ classification, attributions diverge: [1] credits it, with the bounds, to [8] and [9] and states it as its Lemma 4, while [13] states it as Theorem 5.1 of its version v2 and credits it to Regis [11] and to the monograph of Audet and Hare [12]. We were unable to consult [8] — it is paywalled and absent from every source reachable here — so we assert no priority; nothing below depends on the answer.

Throughout, B denotes a *minimum-cardinality* positively spanning subset of the nonzero gradients of a whitened system — minimum cardinality, not merely inclusion-minimal, though inclusion-minimality is then automatic and is what Proposition 12 uses. A *positive circuit* of size c is a minimal set of c gradients carrying a strictly positive vanishing relation, equivalently a support-minimal nonnegative vector in the kernel of the matrix of gradients.

3. The lower bound

The lower half of Theorem 1 is known, and so are its witness and the notation for it. The bound $F_w(d, d+k) \geq 1 + \Gamma_{d,k}$ is Theorem 13 of [1], which with Theorem 24 below is exactly $1 + (2d-n)/d$ on $2n \geq 3d$, and it is proved independently as Theorem 7 of [5], whose Ψ is the Γ of (2) and whose following discussion reads "Once $m = d$ binds, the excess is $(d-k)/d$". The family $D_{d,c}$ below is displayed under that name, for every d and every $c \geq 1$, in the proof of Theorem 10 of [1], together with the risk identity we rederive; its Corollary 11 exhibits $D_{2,2}$. The general-n evaluation below is the case $s = d$, all parts equal to 1, of Proposition 12 of [1], whose labels carry a large shift M where ours carry the fixed $c = 2$. Equation (15) of [5] is the member $c = 1$; Example 2 of [2] is the same cross-polytope with per-coordinate targets 2 and -1, that is $c = 1/3$ with all labels scaled by $3/2$. Contributed here are only the machine-checked verification and the observation that integer labels independent of n, with the minimum-norm tie rule, replace those sources' shift argument.

Fix an integer $c \geq 1$ and let $D_{d,c} = \{(e_j, c+1), (-e_j, 1-c) : j = 1, \dots, d\}$, so $N = 2d$. Since $(t-c-1)^2 + (t-c+1)^2 = 2(t-c)^2 + 2$ for all real t,

$$L_{D_{d,c}}(w) = \frac{1}{2d} \sum_{j=1}^{d} \left[(w_j - c - 1)^2 + (w_j - c + 1)^2 \right] = 1 + \frac{1}{d} \|w - c\mathbf{1}\|^2,$$

so $w^* = c\mathbf{1}$, $L^*_{D_{d,c}} = 1$, and d times the excess ratio of an output w is $\sum_j (w_j - c)^2$. The formal development fixes $c = 2$, the dataset $\{(e_j, 3), (-e_j, -1)\}$. The objective of any selection separates across coordinates, and coordinate j sees only its own two points. If both are selected, the minimiser there is $((c+1)b_1 + (c-1)b_2)/(b_1 + b_2)$, which sweeps $(c-1, c+1)$, so c is reachable and the coordinate costs 0. If exactly one is selected, the minimiser is that point's label $c \pm 1$, uniquely, and the coordinate costs 1. If neither is, F does not depend on w_j, so $\operatorname{argmin} F$ is a translate of the w_j-axis and the minimum-norm tie rule returns $w_j = 0$, at cost c^2. As $\operatorname{argmin} F$ is a product across coordinates its minimum-norm element is the coordinatewise one, so the cases are exhaustive and independent, and what remains is a finite optimisation.

Theorem 6. *Let d, n, m be natural numbers with $d \leq n \leq 2d$ and $m \geq 2$. Call a* pattern *a vector $p \in \{0,1,2\}^d$, recording how many of the two points of each coordinate are retained, and set*

$$\operatorname{slots}(p) = \sum_{j=1}^{d} p_j, \qquad \operatorname{cost}_m(p) = \sum_{j=1}^{d} \gamma_m(p_j), \quad \gamma_m(2) = 0,\ \gamma_m(1) = 1,\ \gamma_m(0) = m.$$

Then $\operatorname{cost}_m(p) \geq 2d - n$ for every pattern with $\operatorname{slots}(p) \leq n$, with equality for the pattern having $2d - n$ entries equal to 1 and $n - d$ equal to 2. Hence $\min\{\operatorname{cost}_m(p) : \operatorname{slots}(p) \leq n\} = 2d - n$ exactly.

Proof. The content is the per-coordinate inequality $\gamma_m(t) + t \geq 2$ for $t \in \{0,1,2\}$, which holds by inspection and uses $m \geq 2$ at $t = 0$. Summing, $\operatorname{cost}_m(p) + \operatorname{slots}(p) \geq 2d$, so $\operatorname{slots}(p) \leq n$ forces $\operatorname{cost}_m(p) \geq 2d - n$. The stated pattern is legitimate since $d \leq n \leq 2d$ makes $2d - n$ and $n - d$ nonnegative with sum d, and it has slots $= n$, $\operatorname{cost}_m = 2d - n$. □

Corollary 7 (known; [1] Prop. 12 at $s = d$ and Thm. 13, [5] Thm. 7)**.** *For every $d \leq n \leq 2d$ one has $L^*_{D_{d,2}}(n; A^*)/L^*_{D_{d,2}} = 1 + (2d-n)/d$, and therefore $F_w(d,n) \geq 1 + (2d-n)/d$. In particular $F_w(4,6) \geq 3/2$.*

Proof. The reduction above expresses d times the excess ratio of a selection as $\operatorname{cost}_{c^2}$ of its pattern, with slots counting the retained points; apply Theorem 6 with $m = c^2 = 4$. □

Remark 8. The hypothesis $m \geq 2$ is sharp: for $c = 1$, the dataset $\{(e_j, 2), (-e_j, 0)\}$, the pattern missing one coordinate entirely and completing the rest has slots $= 2d-2$ and cost 1, certifying only $1 + 1/d$ at $n = 2d - 2$.

4. The branch inequalities

Nothing in this section is part of the machine-checked development; the results are proved as ordinary mathematics, and Section 10 records what a formal proof would need. Their conclusion is the case split of Proposition 15, the exact hypothesis of the kernel-checked Theorem 17.

Fix a whitened system in $\mathbb{R}^d$, let $G = \{i : g_i \neq 0\}$, and note that $g_i \neq 0$ forces $\eta_i \neq 0$, so ξ_i and g_i span the same ray. We use throughout the *exact-recovery certificate*: if a set S carries weights $\alpha_i > 0$ with $\sum_{i \in S} \alpha_i g_i = 0$ and $\operatorname{span}\{\xi_i : i \in S\} = \mathbb{R}^m$, then $q = 0$ is the unique minimiser of its objective, so that selection has excess 0.

Proposition 9 ((B), exact recovery and the base case)**.** *$\Psi_m(b) = 0$ for every $b \geq 2m$; equivalently $P(m, 0) = 0$. Moreover, if the nonzero gradients span $\mathbb{R}^d$ and B is a minimum-cardinality positively spanning subset of them with $|B| \leq n$, then that system contributes 0 to $\Psi_d(n)$.*

Proof. For the second assertion, Lemma 4 supplies $\alpha > 0$ with $\sum_{i \in B} \alpha_i g_i = 0$ and $\operatorname{span}\{\xi_i : i \in B\} = \operatorname{span}(B) = \mathbb{R}^d$; apply the certificate. For the first, let $V = \operatorname{span}\{g_i : i \in G\}$ and $\rho = \dim V$; a point with $\xi_i \notin V$ has $g_i = 0$ and $\xi_i \neq 0$, hence $\eta_i = 0$, and we call it *null*. Take $B \subseteq G$ of minimum cardinality positively spanning V inside V, so $|B| \leq 2\rho$ by Lemma 5, with its strictly positive vanishing relation, and adjoin $m - \rho$ null points whose features complete a basis of $\mathbb{R}^m$ modulo V — they exist because the ξ_i span $\mathbb{R}^m$ and every ξ_i outside V is null. At $q = 0$ the gradient vanishes and the Hessian is positive definite, so the excess is 0 on $|B| + (m - \rho) \leq m + \rho \leq 2m \leq b$ points. The case $G = \varnothing$ is included: then $\rho = 0$ and m points suffice. □

Remark 10. The nonzero gradients need *not* span $\mathbb{R}^d$, so the second assertion above is a genuine restriction: the whitened system $\big((\sqrt{2}, 0), \sqrt{2}\big)$, $\big((-\sqrt{2}, 0), \sqrt{2}\big)$, $\big((0, \sqrt{2}), 0\big)$, $\big((0, -\sqrt{2}), 0\big)$ in $\mathbb{R}^2$ has nonzero gradients $(2, 0)$, $(-2, 0)$, spanning a line. Such systems fall to the anchor branch, Proposition 13, since a point with $\xi_i \notin \operatorname{span}\{g_i\}$ is null; the first assertion was proved so as not to assume spanning.

Proposition 11 ((C), the circuit interface)**.** *Suppose the gradients contain a positive circuit C of size c, $3 \leq c \leq d$, spanning a subspace P with $\dim P = c - 1$. Put $m = d - c + 1$, $\zeta_i = \operatorname{Proj}_{P^\perp} \xi_i$, $A = \{i : \zeta_i \neq 0\}$ and $E(P) = \frac{1}{N} \sum_{i \in A} \eta_i^2 \leq 1$. Then, for $n \geq d + 1$, that system contributes at most $E(P)\Psi_m(n - c) \leq \Psi_m(n - c)$ to $\Psi_d(n)$.*

Proof. First, $\sum_{i \in A} \zeta_i \zeta_i^\top = N I_{P^\perp}$ and $\sum_{i \in A} \eta_i \zeta_i = \operatorname{Proj}(\sum_i \eta_i \xi_i) = 0$, the terms with $\zeta_i = 0$ contributing to neither; so with $M = |A|$ the rescaling $\zeta \mapsto \sqrt{M/N}\,\zeta$, $\eta \mapsto \sqrt{M/(NE(P))}\,\eta$ turns $(\zeta_i, \eta_i)_{i \in A}$ into a whitened system on $P^\perp \cong \mathbb{R}^m$. If $E(P) = 0$, every $i \in A$ has $\eta_i = 0$, and selecting C with its circuit relation together with m points of A whose ζ_i are independent makes the gradient vanish at $q = 0$: excess 0 on $c + m = d + 1 \leq n$ points.

Assume $E(P) > 0$ and fix $\varepsilon > 0$. Choose a strictly convex selection $S_2 \subseteq A$ with $|S_2| \leq n - c$ and weights $\beta_i > 0$ whose projected objective on $P^\perp$ has unique minimiser u with $\|u\|^2 \le E(P)(\Psi_m(n-c)+\varepsilon)$; the factor $E(P)$ is what the rescaling gives, residuals transforming by $\langle u', \zeta_i' \rangle - \eta_i' = \sqrt{M/(NE)}\,(\sqrt{E}\langle u', \zeta_i \rangle - \eta_i)$. Such an S_2 exists with $|S_2| \ge m$, which is what spanning $P^\perp$ needs, because $n - c \ge d+1-c = m$. Regard u as a point of $\mathbb{R}^d$ orthogonal to P and consider $C \cup S_2$. For $i \in S_2$ the residual at u is $e_i = \langle u, \zeta_i \rangle - \eta_i$, so those points contribute $2\sum \beta_i e_i \xi_i$ to the gradient, whose $P^\perp$-component $2\sum \beta_i e_i \zeta_i$ vanishes by projected stationarity: the contribution is a vector $w \in P$. For $i \in C$ we have $\xi_i \in P$, hence $\langle u, \xi_i \rangle = 0$, and those points contribute $-2\sum_{i \in C} \gamma_i g_i$. Since C is a positive circuit spanning P, any $c-1$ of its vectors form a basis of P and adding a large multiple of the circuit relation makes any representation strictly positive, so $\operatorname{pos}(C) = P$ and some $\gamma_i > 0$ have $\sum_{i \in C} \gamma_i g_i = w$. The gradient at u then vanishes, and the Hessian $\sum_C \gamma_i \xi_i \xi_i^\top + \sum_{S_2} \beta_i \xi_i \xi_i^\top$ is positive definite because $\operatorname{span}\{\xi_i : i \in C\} = P$ while the S_2 features project onto a spanning set of $P^\perp$; so u is the unique minimiser and A^* returns it. Finally $i \in C$ implies $\zeta_i = 0$, hence $i \notin A \supseteq S_2$, so the blocks are disjoint and $|C \cup S_2| = c + |S_2| \le n$. Let $\varepsilon \to 0$. ☐

Proposition 12 ((D), a large circuit exists)**.** *Let B be a minimum-cardinality positively spanning subset of the nonzero gradients with $|B| < 2d$. Then either $|B| \le d+1$, or some positive circuit inside B has size c with $3 \le c \le d$.*

Proof. Let K be the kernel of the matrix with columns B. Every $\lambda \ge 0$ in K is a finite nonnegative combination of support-minimal nonzero elements of $K \cap \mathbb{R}^{|B|}_{\ge 0}$ with supports inside $\operatorname{supp}(\lambda)$: induct on $|\operatorname{supp}\lambda|$, pick such a μ with $\operatorname{supp}\mu \subseteq \operatorname{supp}\lambda$, put $t = \min_{i \in \operatorname{supp}\mu} \lambda_i/\mu_i$, and note that $\lambda - t\mu \ge 0$ lies in K with strictly smaller support. Support-minimal elements are exactly positive circuits, and by Lemma 4 some strictly positive λ lies in K, so the circuits in its decomposition cover B. Size 1 is impossible as $g_i \ne 0$; and if every circuit had size 2, the antipodal pairs would be pairwise disjoint — two sharing an element would put two gradients on one ray, the redundant one deletable against minimality — so B would lie on $|B|/2$ lines and $\operatorname{span}(B) = \mathbb{R}^d$ would force $|B| \ge 2d$. Hence some circuit has size $c \ge 3$. A circuit of size $d+1$ has rank d and so positively spans $\mathbb{R}^d$, giving $|B| \le d+1$; outside that case $c \le d$. ☐

Proposition 13 ((E1), the penalty anchor)**.** *If some point has $\eta_{i_0} = 0$ and $\xi_{i_0} \ne 0$, then that system contributes at most $\Psi_{d-1}(n-1)$ to $\Psi_d(n)$.*

Proof. Project onto $W = \xi_{i_0}^\perp$. Then $P_W I_d P_W = I_W$, and the normal equation and the residual energy are untouched because $\eta_{i_0} = 0$, so the projected family is a whitened system on $W \cong \mathbb{R}^{d-1}$ of energy 1 — not E, unlike Proposition 11. Take a strictly convex projected selection S_2 with $|S_2| \le n-1$ and minimiser $u \in W$, and for $\lambda > 0$ set $G_\lambda = f + \lambda h$ with $f(q) = \sum_{i \in S_2} \beta_i(\langle q, \xi_i \rangle - \eta_i)^2$ and $h(q) = \langle q, \xi_{i_0} \rangle^2$: an objective on at most n points, whose Hessian is positive definite for every $\lambda > 0$, since a vector it kills lies in W and is orthogonal to all the projected features. Let q_λ be its minimiser. From $\lambda h(q_\lambda) \le G_\lambda(q_\lambda) \le f(u)$ we get $h(q_\lambda) \to 0$; the images $(\langle q_\lambda, \xi_i \rangle)_i$ over $S_2 \cup \{i_0\}$ are bounded and $q \mapsto (\langle q, \xi_i \rangle)_i$ is injective, so (q_λ) is bounded and any limit point lies in W with f-value at most $f(u)$, hence equals u by uniqueness. As $L_D^*(n; A^*)$ is an infimum, this bounds the contribution by $\|u\|^2$ with no limit of datasets. ☐

Proposition 14 ((E2), rigidity)**.** *Suppose $|B| = 2d$ and no point has $\eta_i = 0$, $\xi_i \ne 0$. Then, for $d \le n \le 2d$, that system contributes at most $(2d-n)/d$ to $\Psi_d(n)$.*

Proof. Every nonzero gradient lies on one of d lines. By Lemma 5, after positive rescaling $B = \{\pm a_1, \dots, \pm a_d\}$ for a basis (a_j). Let $g = \sum_{j \in S} c_j a_j$ be a nonzero gradient with all $c_j \neq 0$ and $|S| \geq 2$. Put $o_j = \operatorname{sgn}(c_j) a_j$, $r_j = -o_j$ and $T = \{g\} \cup \{r_j : j \in S\} \cup \{\pm a_j : j \notin S\}$. Then $|c_j| o_j = g + \sum_{h \in S, h \neq j} |c_h| r_h$, so every $\pm a_j$ lies in $\operatorname{pos}(T)$ and $\operatorname{pos}(T) = \mathbb{R}^d$, while $|T| = 2d - |S| + 1 \leq 2d - 1$, contradicting the minimum cardinality of B. Hence $|S| = 1$.

The lines are orthonormal. Every nonzero feature is by hypothesis the feature of a nonzero gradient, hence lies on one of the d lines; write $\xi_i = t_i u_j$ for $i \in I_j$ with u_j a unit vector, and $Q_j = \frac{1}{N} \sum_{I_j} t_i^2$. The covariance identity reads $I = \sum_j Q_j u_j u_j^\top$, i.e. $AA^\top = I$ for the square matrix A with columns $\sqrt{Q_j} u_j$; hence $A^\top A = I$, so the u_j are orthonormal and $Q_j = 1$.

The selection. Let $R_j = \frac{1}{N} \sum_{I_j} \eta_i^2$, so $\sum_j R_j \leq 1$ (points with $\xi_i = 0$ may carry residual energy). Projecting the normal equation onto u_k gives $\sum_{i \in I_k} t_i \eta_i = 0$ with every term nonzero, so each line carries a canceling antipodal pair; and R_j is the average of $(\eta_i/t_i)^2$ against weights t_i^2/N summing to $Q_j = 1$, so each line carries a point with $(\eta_i/t_i)^2 \leq R_j$. Put a canceling pair — weighted so that $\beta_i t_i \eta_i + \beta_{i'} t_{i'} \eta_{i'} = 0$, making that coordinate of the minimiser 0 — on $n - d$ lines, and the cheapest single point on each of the remaining $2d-n$, chosen by Lemma 20: that is $2(n-d)+(2d-n) = n$ points, the objective separates across the orthonormal coordinates, its Hessian is positive definite, and the excess is at most $(2d-n)/d$. $\square$

Proposition 15 (the case split)**.** $P(m, 0) = 0$ *for every* m*, and for all* d, j *with* $1 \leq j \leq d - 1$,

$$P(d,j) \ \leq\ \max\Big\{ \ \frac{j}{d}, \quad P(d-1, j-1), \quad \max_{3 \leq c \leq d} P(d-c+1,\, j+2-c) \ \Big\}.$$

Proof. The first assertion is Proposition 9. For the second put $n = 2d - j$, so $d+1 \leq n \leq 2d-1$, and fix a whitened system in $\mathbb{R}^d$. If some point has $\eta_i = 0$ and $\xi_i \neq 0$, Proposition 13 gives the second term; by Remark 10 this covers the case where the nonzero gradients fail to span $\mathbb{R}^d$. Otherwise every point with $\xi_i \neq 0$ has $\eta_i \neq 0$, hence $g_i \neq 0$ and $\xi_i \in \operatorname{span}\{g_i\}$, so the nonzero gradients span $\mathbb{R}^d$; adding a large multiple of $\sum_i g_i = 0$ shows they positively span it, and a minimum-cardinality positively spanning B exists with $d + 1 \leq |B| \leq 2d$ by Lemma 5. If $|B| \leq n$, Proposition 9 gives excess 0, below every term. If $n < |B| < 2d$ then $|B| > n \geq d+1$, so Proposition 12 supplies a positive circuit of size $3 \leq c \leq d$ and Proposition 11 gives $\Psi_{d-c+1}(n-c) = P(d-c+1, j+2-c)$, since $2(d-c+1)-(n-c) = j+2-c$. If $|B| = 2d$, Proposition 14 gives j/d. The reduction to full feature rank of §2.1 is available because $n \geq \lceil 3d/2 \rceil$ forces $n > 3\rho/2$ for every $\rho < d$. $\square$

5. The recursion closes on the upper half

This section is the arithmetic core, and is machine-checked. Nothing in it mentions datasets: it concerns an arbitrary rational-valued function of two natural number arguments satisfying the inequalities produced by Section 4. Subtraction of natural numbers is truncated at 0, as in the formal development; that matters only in $j + 2 - c$, which is 0 when $c > j + 2$, where the base case applies.

Definition 16. A function $P \colon \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ *satisfies the branch bounds* when (i) $P(m, 0) = 0$ for every m, and (ii) for all d, j with $1 \leq j$ and $j + 1 \leq d$ and every

$x \in \mathbb{Q}$: if $j/d \le x$, $P(d-1, j-1) \le x$, and $P(d-c+1, j+2-c) \le x$ for every c with $3 \le c \le d$, then $P(d, j) \le x$.

Theorem 17. *Let $P\colon \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ satisfy the branch bounds. Then*

$$P(d,j) \;\le\; \frac{j}{d} \qquad \text{whenever} \qquad 2j \le d.$$

Proof. Strong induction on j. For $j = 0$ the value is 0 by (i). Let $j \ge 1$ and $2j \le d$, so $d \ge 2$ and $j+1 \le d$; apply (ii) with $x = j/d$. The rigidity term is j/d itself. For the anchor, $d-1 \ge 2j-1 \ge 2(j-1)$, so induction gives $P(d-1, j-1) \le (j-1)/(d-1)$, and $(j-1)/(d-1) \le j/d$ because $d(j-1) \le j(d-1) \iff j \le d$. For the circuit term at $3 \le c \le d$: if $c \ge j+2$ then $j+2-c = 0$ and the term is 0 by (i); otherwise $c \le j+1$, and with $m = d-c+1$, $j' = j+2-c \ge 1$ one has $m \ge 2j'$, since $d-c+1 \ge 2(j-c+2) \iff d+c \ge 2j+3$, which follows from $d \ge 2j$ and $c \ge 3$, so induction gives $P(m, j') \le (j+2-c)/(d-c+1)$, which Lemma 18 bounds by j/d. In each application $j' < j$, so the induction is well founded. □

Lemma 18. *Let c, d, j be natural numbers with $3 \le c$, $c \le j+1$, $c \le d$ and $2j \le d$. Then $(j+2-c)d \le j(d-c+1)$, equivalently $(j+2-c)/(d-c+1) \le j/d$.*

Proof. The inequality rearranges to $c(d-j) \ge 2d-j$. Since $c \ge 3$ it suffices that $3(d-j) \ge 2d-j$, i.e. $d \ge 2j$. No monotonicity in c is needed. □

Proposition 19 (sharpness, and the exact failure)**.** *The function $(m,j) \mapsto j/m$ satisfies the branch bounds, so the conclusion of Theorem 17 cannot be replaced by any smaller bound; the zero function satisfies them too, so the hypotheses are consistent. The inequality of Lemma 18 fails at $(d,j,c) = (3,2,3)$ and at $(4,3,3)$ — that is, at the cells $F_w(3,4)$ and $F_w(4,5)$, the two* [1] *settles by hand — and holds again at $(4,2,3)$, the first cell on the boundary $2j = d$.*

Proof. For j/m the base case is $0/m = 0$ and clause (ii) is trivial, its first hypothesis being already $j/d \le x$; for the zero function both clauses are immediate. The numerical claims are $3 > 2$, $8 > 6$ and $4 \le 4$ in $(j+2-c)d$ against $j(d-c+1)$. □

Lemma 20. *Let $k \le d$ and let $R\colon \{1, \ldots, d\} \to \mathbb{Q}$ be arbitrary. Then some T of size k satisfies $d\sum_{i \in T} R_i \le k\sum_{i=1}^{d} R_i$. Consequently, if $d \le n \le 2d$ and $\sum_i R_i \le 1$, some T of size $2d-n$ satisfies $d\sum_{i\in T} R_i \le 2d-n$. The bound is attained: for $R \equiv 1/d$ with $d \ge 1$, every T of size k has $d\sum_{i\in T} R_i = k$.*

Proof. For $k = 0$ both sides vanish. For $k \ge 1$ choose T of size k minimising $\sum_T R$ and let $i_0 \in T$ maximise R on T. For $j \notin T$ the set $(T \setminus \{i_0\}) \cup \{j\}$ again has size k, so minimality gives $R_{i_0} \le R_j$. Hence $\sum_T R \le kR_{i_0}$ and $(d-k)R_{i_0} \le \sum_{T^c} R$; multiplying the first by $d-k \ge 0$ and the second by $k \ge 0$ chains them into $(d-k)\sum_T R \le k\sum_{T^c} R$, which with $\sum_T R + \sum_{T^c} R = \sum R$ is the claim. Nonnegativity of R is nowhere used: the k smallest of d reals have average at most the overall average. □

6. Theorem A

Theorem 21 (Theorem A, restated)**.** *For every $d \ge 1$ and every n with $\lceil 3d/2 \rceil \le n \le 2d$, $F_w(d,n) = 1 + (2d-n)/d$.*

Proof. The inequality $\geq$ is Corollary 7, whose hypothesis $d \leq n \leq 2d$ is implied. For $\leq$, Proposition 15 says that $P(m,j) = \Psi_m(2m-j)$ satisfies the branch bounds of Definition 16, and $2n \geq 3d$ is $2j \leq d$ for $j = 2d - n$, so Theorem 17 gives $\Psi_d(n) \leq (2d-n)/d$; now apply (4). □

Corollary 22. *$F_w(4,6) = 3/2$, the exact value at the cell* [1] *names in print as the smallest open one.*

Remark 23. Why is $(4,6)$ easier than $(4,5)$, one point away? At $n = 6$ the branch that makes $(4,5)$ hard — the rectangle, a positive basis of $\mathbb{R}^4$ of size 6 — evaporates: six positively spanning gradients carry a strictly positive vanishing relation by Lemma 4, their features span, and Proposition 9 recovers the optimum on $|B| = 6 = n$ points, leaving $|B| \in \{7,8\}$ to the circuit and rigidity branches. In Proposition 19 that is the difference between $(d,j,c) = (4,3,3)$, where the arithmetic step fails, and $(4,2,3)$, where it holds. Theorem 21 and Corollary 22, like all of Section 4, are *not* machine-checked; what is, is their arithmetic half (Theorem 17 with Lemmas 18 and 20) and their lower half (Theorem 6).

7. The conjectured value on the upper half

Theorem 21 confirms the conjecture of [1] on its range because of the following evaluation of (2). It is elementary, and it is here so that the bridge between the two formulas is a checked statement rather than an assertion. Write $W(d,s) = 1/C_{d,s}$; with $d = qs + a$, $0 \leq a < s$, $q \geq 1$,

$$W(d,s) = \frac{s-a}{q} + \frac{a}{q+1} = \frac{s(q+1)-a}{q(q+1)}, \qquad \text{so} \qquad C_{d,s} = \frac{q(q+1)}{s(q+1)-a}.$$

Theorem 24. *Let d, k, s be natural numbers with $k+1 \leq s \leq d$ and $d \leq 2k$. Then $(s-k)C_{d,s} \leq (d-k)/d$, with equality at $s = d$. Consequently $\Gamma_{d,k} = (d-k)/d$ whenever $2k \geq d$, so on the range $\lceil 3d/2 \rceil \leq n \leq 2d$ the value conjectured in* [1] *is exactly the value of Theorem 21.*

Proof. Two steps. First an exact identity, valid whenever $d = qs + a$:

$$d\big(s(q+1)-a\big) = s^2 q(q+1) + a(s-a),$$

which says $W(d,s) \geq s^2/d$ with defect exactly $a(s-a)/(dq(q+1))$; it is the arithmetic–harmonic mean inequality for the balanced partition, made exact. Second, for $k+1 \leq s \leq d$ and $d \leq 2k$,

$$d^2(s-k) \ \leq \ (d-k)s^2,$$

because the difference factors as $(d-s)\big(k(d+s)-ds\big)$ with $d-s \geq 0$, and $2k(d+s) \geq d(d+s) = d^2 + ds \geq 2ds$ using $d \geq s$. Multiplying the second by $q(q+1) \geq 0$ and inserting the first,

$$d \cdot d(s-k)q(q+1) \ \leq \ (d-k)s^2 q(q+1) \ \leq \ (d-k)\, d\big(s(q+1)-a\big),$$

and dividing by $d > 0$ gives $d(s-k)q(q+1) \leq (d-k)(s(q+1)-a)$, the claim cross-multiplied. At $s = d$ one has $q = 1$, $a = 0$, $W(d,d) = d$ and $C_{d,d} = 1/d$, so that term is exactly $(d-k)/d$. □

Remark 25. The hypothesis $2k \geq d$ is what this route needs: $s \mapsto ds/(d+s)$ increases to $d/2$ at $s = d$, which is the condition $ds \leq k(d+s)$ above. It is not needed at every cell: at $(d,k) = (4,1)$ the $s = 2$ term equals $1 > 3/4$ and the identity genuinely fails, whereas at $(3,1)$ it holds although $2k < d$, and exhaustive evaluation for $d \le 60$ found no violation for $k \ge \lceil d/3 \rceil$. Note also that $\Gamma_{4,2} = 1/2$, so the value conjectured in [1] at the target cell $(4,6)$ is $3/2$ — the same number Theorem 21 proves — while $\Gamma_{4,1} = 1$, matching $F_w(4,5) = 2$ one cell outside the range.

8. The exact reach of the recursion

Is the range $2n \ge 3d$ a limitation of the induction of Theorem 17 or of the recursion itself? Computing the recursion decides this. Let

$$U_d(n) = 0 \quad (n \ge 2d), \qquad U_d(n) = \max\Big\{\frac{2d-n}{d},\, U_{d-1}(n-1),\, \max_{3\le c\le d} U_{d-c+1}(n-c)\Big\}$$

for $d < n < 2d$; in one variant with $U_m(m) = m$, the value suggested by $F_w(m,m) = m+1$, and in a second with $U_m(m)$ left undefined, so that a branch reaching the diagonal returns no bound and the undefined value propagates through the maxima. In the second variant, iterating Proposition 15 from its base case gives $\Psi_d(n) \le U_d(n)$ at every cell where $U_d(n)$ is defined. Call a cell *settled* when $U_d(n) = (2d-n)/d$.

Theorem 26. *For every $d \le 30$ and every n with $d < n < 2d$: $U_d(n) = (2d-n)/d$ exactly when $2n \ge 3d$, and $U_d(n)$ is strictly larger, or undefined, when $2n < 3d$. The same holds with the base case $U_m(m) = m$ deleted, and the two settled sets coincide. In the range $d \le 30$ there are 225 cells with $2n \ge 3d$ and 210 with $2n < 3d$, so neither half is vacuous. Spot values: $U_4(6) = 1/2$; $U_d(2d-1) = 1/d$ at $d = 3,4,5,8$; and $U_3(4) = 1 > 2/3$, $U_4(5) = 2 > 1$, $U_5(7) = 1 > 3/5$, these three becoming undefined when $U_m(m) = m$ is deleted.*

Theorem 26 rests on exhaustive computation and is the only statement here that is bounded: it says nothing about $d > 30$. The table is evaluated inside Lean by compiled evaluation (Section 10) and independently by an exact-rational implementation in Python, which agrees cell for cell. Three consequences: the frontier $n \ge \lceil 3d/2 \rceil$ is the recursion's own, not an artefact of the induction; the first two cells lost are $(3,4)$ and $(4,5)$, precisely the two that [1] settles by hand through classifications of small positive bases of $\mathbb{R}^3$ and $\mathbb{R}^4$, and the first lost cell at which $1 + (2d-n)/d$ is still the conjectured value is $(5,7)$, where $\Gamma_{5,2} = 3/5$ — the lost cell $(5,6)$ before it has $\Gamma_{5,1} = 6/5$, so the formula is not even conjectured there; and the volume-sampling base case is never reached, so no volume sampling and no determinantal point processes are needed — which, with Steinitz replaced by Lemmas 4 and 5, is why nothing from the two known endpoints of the window enters the proof.

9. Computations that are not theorems

Nothing here is a theorem or part of the formal development. It is reported because it justified attempting the proof, because it is a negative control on Theorem 21 at the smallest new cell, and because two of the computations independently confirm published results. All of it is exact rational arithmetic, single-threaded, and cost about seven CPU-minutes apart from one sweep noted below.

The instrument. A whitened system with all $\eta_i > 0$ is represented projectively by pairs (d_i, s_i) with $s_i > 0$, $\sum s_i = 1$, $\sum s_i d_i = 0$ and $M = \sum s_i d_i d_i^\top$ positive definite, so that $\|q\|^2 = p^\top M p$ for $p = M^{-1/2}q$ and everything stays rational. The solver enumerates minimum-norm interpolants, full-rank supports of size at most n and the sign vectors realised by the kernel, returning the least value whose sign pattern matches exactly, which is the optimality condition for a strictly positively weighted objective; that value is therefore *achievable*, hence an upper bound on the infimum — the safe direction for a refutation search. It is only that: the infimum in Definition 3 need not be attained, and where it is not the solver returns something strictly larger. It was validated against exact normal-equation solves at random positive weights on random supports, which never went below it.

Confirmations and audits. The exact budget-energy value $\Phi(r, k)$ of the orthogonal block systems of [1], namely $\max_{|J|\geq k+1}(|J| - k)\big/\sum_{j\in J} 1/r_j$, was recomputed from the definition (1) rather than from that paper's proof, on 49 cells over 17 dimension vectors, with 0 mismatches; and the printed table of $1 + \Gamma_{d,k}$ for $d \leq 6$ was reproduced from (2), 15 cells and 0 mismatches. Branch (C) was then checked on 48 circuit/block cells — no violation, and every one an equality, so it is not a lossy bound that happens to survive — and branch (E2) on 63 axis cells from 21 systems with independent line scalings, where in all 63 the solver's value equals that of the selection the branch constructs. The dataset of Section 3 evaluates to $1+(2d-n)/d$ in every cell computable for $d \leq 4$, and at $(4, 5)$ to $7/4 < 2$, correctly, the extremal system there being the $(2, 2)$ block system.

A refutation sweep at $(4, 6)$. Random systems are useless here: almost all have $|B| = 5$, hence a full-rank positive circuit, hence excess 0. We therefore generated from the two hard branches directly — $|B| = 8$ from four lines $\pm u_j$ over six integer bases, $|B| = 7$ from a central triangle with two leaves over 36 parameter choices, and the orthogonal block systems with dimension vectors $(2, 1, 1)$, $(1, 1, 1, 1)$, $(2, 2)$ and the plain rectangle — each perturbed by one or two extra gradients. Over 2126 such systems, plus 59 random controls and 706 seconds, the worst achievable excess was 0 at $|B| = 6$, $5/14$ at $|B| = 7$ and $5/11$ at $|B| = 8$; none exceeded $1/2$, and since the solver returns an achievable value that is the conclusive direction. This is a negative control, not evidence: it searches a targeted family, not all datasets. Likewise the three obstruction configurations of [1] — the rectangle of its Proposition 48, the (R) counterexample of its Proposition 52, and the threshold family of its Proposition 53 at $\varepsilon = 1/100$, $1/10$, $1/5$; the last two with the masses printed there, the rectangle with balanced masses of our choice, since that proposition fixes only the directions — all have achievable excess exactly 0 at $n = 6$, witnessed by the two-triangle support, which is Proposition 9 with $|B| = 6 = n$. Nothing is claimed about them at $n = 5$: there the solver returns only its least exactly attained value, and on the (R) instance that value exceeds the ratio 2452861/2000000 which Proposition 52 of [1] itself attains with an explicit five-point objective.

10. Verification

Theorems 6, 17, 26 and 24, Lemmas 18 and 20 and Proposition 19 are formalised in Lean 4 [14] and checked by the Lean kernel; Mathlib [15] is used by two of the four modules. The development is 863 lines and 71 declarations, compiling in 4.9, 5.6, 61 and 31 seconds at peaks of 1.57, 1.55, 6.76 and 6.78 GB — the four modules in the order `LowerWitness`, `Table`, `Recursion`, `Gamma`, measured on a loaded machine and hence upper bounds. The toolchain is `leanprover/lean4:v4.32.0` and Mathlib is taken at revision `81a5d257c8e410db227a6665ed08f64fea08e997`.

There is no `sorry` anywhere. Theorem 17 and everything it rests on, including Lemmas 18 and 20 and Proposition 19, depend on no axioms beyond `propext`, `Classical.choice` and `Quot.sound` and use no compiled evaluation. The same holds for Theorems 6 and 24, the former needing only `propext` and `Quot.sound`. Theorem 26 is the exception and the only place where `native_decide` is used: each of its assertions carries one compiled-evaluation axiom, so for that statement, and only for it, the Lean compiler and runtime are part of the trusted base. Negative controls are formalised beside the theorems: that $3/2$ at $(4, 6)$ is attained and that the same dataset certifies no more, and $1+1/4$ at $n = 7$; that $m \geq 2$ is load-bearing (Remark 8); that the branch bounds are consistent and that j/m satisfies them; the two circuit failures and the boundary case of Proposition 19; the attainment in Lemma 20; the cell counts and the unsettled cells of Theorem 26; and that $2k \geq d$ is neither vacuous nor removable in Theorem 24, with $\Gamma_{4,2} = 1/2$. Not formalised are Sections 4 and 9, and with them Theorem 21 and Corollary 22; the interface between the halves is Definition 16, stated as a hypothesis rather than hidden inside a definition for exactly that reason. The repository is private at the time of writing; repository reference to be added at submission.

Two gaps in Mathlib keep Section 4 out of the kernel, and both look worth closing on their own account. There is no positive-spanning or positive-basis API: at the revision above, nothing named `positiveSpanning` or `PositiveBasis` exists, nor do the phrases "positive basis" or "positively span". The right definition is available — `PointedCone.hull R s = ⊤`, the `hull` renamed from `PointedCone.span` on 22 March 2026 — and Lemma 4 is two lines from it, Lemma 5 has a one-page proof in [10], and the circuit decomposition of Proposition 12 needs no Minkowski–Weyl theory. The second gap is larger: there is no least-squares or minimum-norm bridge. Mathlib has no `Matrix.solve`, no least-squares construction, no normal equations and no Moore–Penrose pseudoinverse, and `Matrix.PosDef.sqrt`, needed by the whitening of §2.1, was deprecated and then deleted in favour of `CFC.sqrt` with `Matrix.PosSemidef.inv_sqrt`, whose order instances are scoped. Formalising Section 4 is a project on the scale of a Mathlib contribution rather than of this paper.

Acknowledgement of provenance

The problem is not ours. The quantity $F_w(d, n)$, Theorem 8 and the open window $d < n < 2d$ are due to Hanneke, Moran, Shlimovich and Yehudayoff [2, 3]. The candidate value $1 + \Gamma_{d,k}$, the definition (2), the lower bound with the dataset of Section 3, the cells $n = 2d-1$, $(3, 4)$, $(4, 5)$, the block model $\Phi(r, k)$, the obstruction instances of Section 9 and the conjecture proved here on the upper half are all due to Zhang [1]; the endpoint $n = 2d - 1$ was obtained independently by Huang [4] and by the anonymous deposit [5], and the harmonic lower bound independently by the latter. Offered here are the proof of that conjecture on $2n \geq 3d$, the branch decomposition and the recursion closing it, the evaluation $\Gamma_{d,k} = (d - k)/d$ on $2k \geq d$, the recursion's exact reach, and the machine-checked verification of those four.

References

[1] G. Zhang, *Exact Risk Ratios for Weighted Data Selection in Linear Regression*, 33 pages, `arXiv:2608.28007v1` (28 August 2026); cs.LG primary, math.ST cross-list. Re-checked against

the arXiv API on 3 September 2026: v1 is still the only version and carries no journal reference and no DOI, so every number cited above refers to it.

[2] S. Hanneke, S. Moran, A. Shlimovich and A. Yehudayoff, *Data Selection for ERMs*, `arXiv:2504.14572v2` (25 April 2025; v1 20 April 2025). All numbers cited above are those of the v2 PDF, in which the linear-regression result is Theorem 8 and the cross-polytope dataset of Section 3 is Example 2; the experimental HTML rendering of the same version numbers them 9 and 10.

[3] S. Hanneke, S. Moran, A. Shlimovich and A. Yehudayoff, *Open Problem: Data Selection for Regression Tasks*, Proceedings of Machine Learning Research **291** (COLT 2025), 6225–6229; `proceedings.mlr.press/v291/hanneke25e.html`.

[4] M. Huang, *Exact Data Selection for Scalar Linear Regression*, Zenodo deposit, `doi:10.5281/zenodo.21754480` (2 August 2026); unrefereed. Confirmed against the Zenodo API on 3 September 2026; it contains neither the cell $(4, 6)$ nor the formula $1 + (2d - n)/d$.

[5] Anonymous, *One Short of Steinitz: Weighted Regression Selection*, Zenodo deposit, `doi:10.5281/zenodo.21881535` (11 August 2026); unrefereed; a second deposit of the same text at `doi:10.5281/zenodo.22061101` (22 August 2026). Its Theorem 7 is the harmonic lower bound. The deposit's creator field reads "Anonymous" on the Zenodo record, as checked against the Zenodo API on 3 September 2026, and we cite it accordingly.

[6] Z. Yang, *Partial Results on Four Open Problems in Discrete Geometry, Learning Theory, and Graph Theory*, Zenodo deposit, `doi:10.5281/zenodo.22136407` (28 August 2026), whose third direction is this problem, and *A continuous* $3/2$ *lower-bound family for closed-simplex weighted data selection in* $d = 2$, $n = 3$, `doi:10.5281/zenodo.22220409` (1 September 2026); unrefereed, both confined to $d = 2$, $n = 3$ lower bounds, the first describing itself as "preliminary research records, not published papers", and both stating that the regime $d < n < 2d$ remains open.

[7] G. Zhang, *Exact Recovery Thresholds for Weighted Data Selection in Vector-Valued Linear Regression*, `arXiv:2608.30254v1` (31 August 2026); the vector-valued variant, Question 4 of [3], which does not touch the scalar window.

[8] C. Davis, *Theory of positive linear dependence*, Amer. J. Math. **76** (1954), no. 4, 733–746, `doi:10.2307/2372648` (Zbl 0058.25201); not consulted directly here.

[9] R. L. McKinney, *Positive bases for linear spaces*, Trans. Amer. Math. Soc. **103** (1962), no. 1, 131–148, `doi:10.1090/s0002-9947-1962-0147879-x`.

[10] C. Audet, *A short proof on the cardinality of maximal positive bases*, Optim. Lett. **5** (2011), no. 1, 191–194, `doi:10.1007/s11590-010-0229-3`.

[11] R. G. Regis, *On the properties of positive spanning sets and positive bases*, Optim. Eng. **17** (2016), no. 1, 229–262, `doi:10.1007/s11081-015-9286-x`.

[12] C. Audet and W. Hare, *Derivative-Free and Blackbox Optimization*, Springer Series in Operations Research and Financial Engineering, Springer, 2017, `doi:10.1007/978-3-319-68913-5`; cited by [13] for the size-$2m$ classification, and not consulted directly here.

[13] D. Cornaz, S. Kerleau and C. W. Royer, *A characterization of positive spanning sets with ties to strongly connected digraphs*, `arXiv:2411.08994v2` (15 April 2025; v1 13 November 2024). The theorem number quoted above is that of v2, whose LaTeX source we compiled; that paper credits the size-$2m$ classification to [12] and [11], and the size bounds to [12] and [8].

[14] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, Automated Deduction — CADE-28, LNCS **12699**, Springer, 2021, 625–635.

[15] The mathlib Community, *The Lean mathematical library*, Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381.