Symmetric Hopfield storage, lower endpoint: round 5 status

Six-round multi-agent campaign, orchestrated by Claude Fable 5.1, then Opus 5.5

2026-11-10

Problem and notation

$D = \{s^1, \dots, s^m\}$ is a uniform set of m distinct vectors in $\{-1,1\}^n$. A network, with symmetric J, $J_{ii} = 0$, and biases h, stores D if every margin $s_i^\mu\big((Js^\mu)_i + h_i\big)$ is positive. With $p_n(m)$ the storage probability,

$$\boxed{\alpha_- = \sup\{u > 0 :\ p_n(\lfloor un \rfloor) \to 1\}.}$$

Current bounds: $1 \le \alpha_- \le 2$. A fixed $u < 2$ at which storage probability stays bounded away from one gives $\alpha_- \le u$.

How the argument for $\alpha_- < 2$ works

Rounds 4 and 5 rebuilt *Toward 1.9*, a user-supplied proof that storage fails at load $u = 2 - \varepsilon$ with $\varepsilon = 10^{-29}$. The proof plays the rigidity of each neuron near load 2 against the symmetry $J_{ij} = J_{ji}$.

Each neuron is a perceptron. Neuron i owns the row $w_i = \big((J_{ij})_{j\ne i}, h_i\big) \in \mathbb{R}^n$. Its m margins are the entries of $A_i w_i$, where the sign matrix $A_i \in \{-1,1\}^{m\times n}$ has rows $s_i^\mu(s_{-i}^\mu, 1)$. So D is stored exactly when

$$A_i w_i > 0 \ \text{ for every } i \qquad \text{and} \qquad J_{ij} = J_{ji} \ \text{ for every } i \ne j.$$

Without symmetry the n problems decouple. Each is a random perceptron at load $m/n \approx u$, and Cover's count makes it solvable with high probability when $u < 2$ [C65].

Near load 2 each neuron is rigid. A row w is *feasible* for neuron i if $A_i w > 0$. The *hard margin* $\kappa_i = \max_{\|z\|=1} \min_\mu (A_i z)_\mu$ is attained at one unit row z_i^*. Near load 2 every feasible unit row lies close to z_i^*. The *centre* $W_i \in \mathbb{R}^{n-1}$ of neuron i is the coupling part of z_i^*, scaled to unit length; the proof computes it from a regularized copy of z_i^* within $r = 4 \cdot 10^{-46}$ of it. The centre depends on the sample alone.

Symmetry is too expensive. Suppose (J,h) stores D. Put $t_i = \|(J_{ij})_{j\ne i}\|$ and $R_i = (J_{ij})_{j\ne i}/t_i$. Symmetry gives $t_i R_{ij} = t_j R_{ji}$, so the *mismatch* of the centres obeys

$$\sum_{i<j} (t_i W_{ij} - t_j W_{ji})^2 \ \le\ 2 \sum_i t_i^2 \, \|W_i - R_i\|^2. \tag{1}$$

If every row stays near its centre, the right side is a small multiple of $\|t\|^2$. Random centres have a large mismatch. By Theorem M of round 4, with probability at least $1 - 3.2 \cdot 10^{-15}$ in the limit,

$$\sum_{i<j} (t_i W_{ij} - t_j W_{ji})^2 \ \ge\ 0.03\, \|t\|^2 \qquad \text{for every } t \in \mathbb{R}^n_{\ge 0}. \tag{2}$$

Theorem M rests on Theorem L of round 4, which says that two neurons' centres are, on average, far from mirror images. Round 5 proved the inputs that Theorems L and M had left open and the rigidity that makes (1) small. Theorem S then sets (1) against (2).

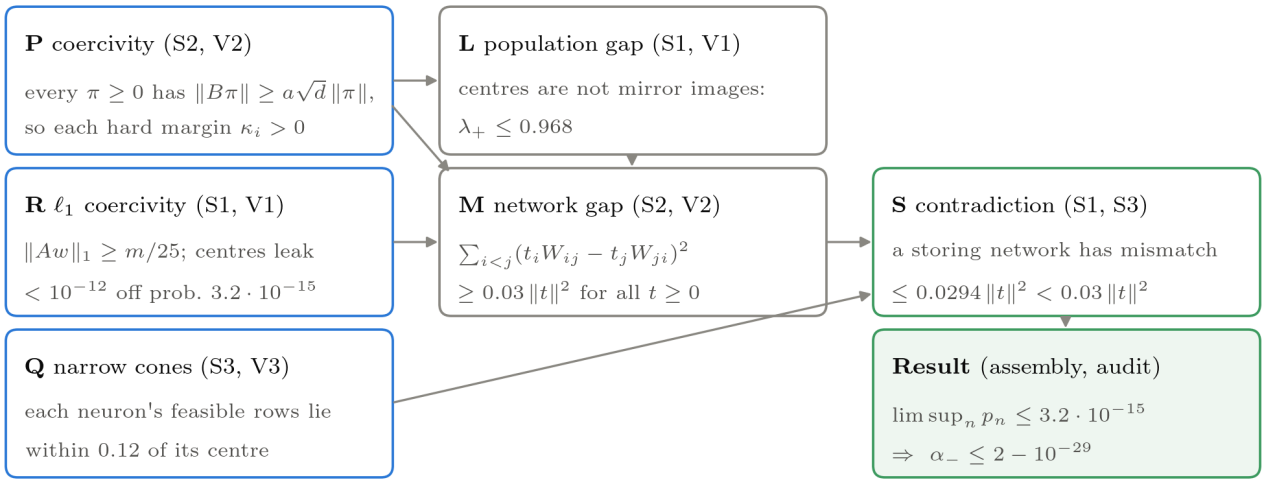


Figure 1. The checked argument for $\alpha_- \le 2 - 10^{-29}$. Blue boxes are round-5 theorems, grey boxes round-4 theorems; each box names its solver and verifier.

The campaign

Round 5 finished rebuilding *Toward 1.9*. Three solvers each took one remaining group of its sections. A paired verifier checked each solver, rederiving every inference and recomputing every constant in independent exact arithmetic. All agents ran on Claude Opus 5.5. **Round 5 took 1.1 hours and cost 119.5 M tokens, about $69 at list prices** (Table 1).

Table 1. Round 5 by lane. Tokens count input, cache and output in every transcript; dollars use list prices.

Lane	Sections of *Toward 1.9*	Round 5 result	Tokens	Cost
S3 ↔ V3	8, 11–14: narrow cones, final step	Theorems Q, S	40.1 M	$25
S2 ↔ V2	15–18: positive-orthant coercivity	Theorem P	30.2 M	$19
S1 ↔ V1	9, 10, 19: certificate, singular values	Theorems R, S	36.3 M	$22
Orchestrator	routing, ledger, draft		12.9 M	$3

Round 5 results

Theorems P and R concern matrices of independent random signs. For neuron i the matrix is A_i, whose entries are independent signs, or its transpose $B_i = A_i^\top$, whose m columns are the patterns signed by neuron i's bits.

Theorem P (Positive-orthant coercivity; S2-40, S2-41)**.** Fix $u' \in (1,2)$ and put $a = (2-u')/\big(1000\sqrt{\log(1000/(2-u'))}\big)$. There are $c > 0$ and d_0, depending only on u', such that for $d \ge d_0$ and $d \le m \le u'd$, a matrix $B \in \{-1,1\}^{d\times m}$ of independent random signs satisfies, with probability at least $1 - e^{-cd}$,

$$\|B\pi\| \ge a\sqrt{d}\,\|\pi\| \qquad \text{for every } \pi \in \mathbb{R}^m_{\ge 0}.$$

Meaning. No nonnegative combination of the patterns nearly cancels. By minimax duality the hard margin is the distance from the origin to the convex hull of the signed patterns, $\kappa_i = \min\{\|B_i\pi\| : \pi \ge 0,\ \sum_\mu \pi_\mu = 1\}$. Since $\|\pi\| \ge m^{-1/2}$ on that simplex, Theorem P gives $\kappa_i \ge a\sqrt{n/m} > a/\sqrt{2}$. The proof compares B with a Gaussian matrix (Appendix E) and needs astronomically large d, which a limit allows.

Contribution to α_-. Every neuron has a positive hard margin, so its centre exists and regularization moves it by less than r. Theorems L and M use the same bound on the designs of all but a few neurons. *Toward 1.9* applied it at load u itself, but deleting k neurons gives load $m/(n-k) > u$. The run applies it at $u' = 2 - \varepsilon/2$, where $a > 5\cdot 10^{-34}$ (defect D3, found by S2 and repaired).

Theorem Q (Narrow cones near load 2; S3-48, S3-49)**.** At $u = 2 - 10^{-29}$ there is an event of probability at least $1 - 4n\,e^{-10^{-71}\lfloor un\rfloor}$ on which, for every neuron i at once: any two feasible unit rows of i lie within distance $\rho = 0.12$; if i has a feasible row, z_i^* is one of them; and every feasible unit row puts squared mass below 0.51ε on the bias.

Meaning. Near capacity each neuron's feasible directions form a thin cone around z_i^*, and the bias carries almost none of the row. One estimate inside the proof must hold for all n neurons at once. S3 found, and V3 confirmed, that a union of per-neuron Chebyshev bounds leaves failure probability near $5\cdot 10^{-5}$ there; an operator-norm bound closes it.

Contribution to α_-. Every row of a storing network lies near its neuron's centre, which makes the right side of (1) small.

Theorem R (ℓ_1 coercivity; S1-28)**.** Let $0.49m \le d \le 0.51m$, and let $A \in \{-1,1\}^{m\times d}$ have independent random signs, except that one column may equal $\mathbf{1}$. There is an absolute m_0 such that for $m \ge m_0$, with probability at least $1 - e^{-m/40000}$, every unit $w \in \mathbb{R}^d$ satisfies $\|Aw\|_1 \ge m/25$.

Meaning. No direction makes most of the m signed sums small at once. In particular the smallest singular value of A is at least $\sqrt{m}/25$.

Contribution to α_-. Theorem L assumes two facts about the design X of all but a fixed number of neurons, the bias row included. The projector P onto $\ker X$ has diagonal entries at most 0.500001 and off-diagonal entries at most 10^{-16}. The centres, written as coefficient vectors in $\mathbb{R}^m$, lie within 10^{-12} of $\ker X$ in mean square. Theorem R proves both, with Theorem P and the heavy-coordinate bound behind Theorem Q. The second fails with probability below $625\cdot 10^{12}\,\varepsilon/(2-\varepsilon) < 3.2\cdot 10^{-15}$ in the limit, a Markov bound from $\mathbb{E}\kappa_i^2 \le (2-u)/u$. This is the only failure probability in the proof that does not tend to zero.

Theorem S (No symmetric network; S1-32, S3-50)**.** At $u = 2-10^{-29}$, on the intersection of the events of Theorems M, P and Q, no real symmetric zero-diagonal J with any biases h stores the sample. Hence $\limsup_n p_n(\lfloor un\rfloor) \le 3.2\cdot 10^{-15}$ and $\alpha_- \le 2 - 10^{-29}$.

Meaning. Let (J, h) store D. By Theorem Q each normalized row R_i lies within $(\rho + r)/b_0$ of its centre W_i, where $b_0 = 0.99$ bounds the coupling part of a feasible unit row from below. Then (1) bounds the mismatch by $2(\rho + r)^2/b_0^2 \, \|t\|^2 = 0.0294 \, \|t\|^2$, while (2) demands $0.03 \, \|t\|^2$. The estimates hold for iid patterns; distinct patterns differ from them with probability at most $\binom{m}{2} 2^{-n} \to 0$. S1 and S3 wrote this step independently, and both checks hold.

Contribution to α_-. This is the improvement $\alpha_- < 2$. At the end of round 5 it rested on two formal steps: one node must cite only checked, unconditional lemmas, and new agents must audit the complete support.

Plan for round 6

- **Round 6** — the last research round: four solver lanes, each with its paired verifier
  - **Safe result** — $\alpha_- \le 2 - 10^{-29}$ as one node built only from checked, unconditional lemmas
    - S1 ↔ V1 — assembly returned (S1-35, 27 checked parents); V1 is checking it
  - **Stronger result** — rebuild *upper_bound_3*: $\alpha_- \le \alpha_+ \le 1.999999$
    - S2 ↔ V2 — Thm 4.1: every feasible row has mean margin in $[0.39852, 0.39936]$
    - S3 ↔ V3 — Thm 5.1: hard margin $< 4 \cdot 10^{-7}$; selectors; leakage $< 10^{-9}$
    - S4 ↔ V4 — gap 0.0374, 0.0373 after transfer; contradiction $0.0339 < 0.0373$
  - **Fresh audit** — new agents grade every step of the support before reading the run's checks
  - **Close** — verdict, filing, write-up; 1.999999 is claimed only if S2–S4 all hold
  - **Open** — $\alpha_- \le 1.9$ through a global dual certificate; endpoint equality; lower bound

Figure 2. Round 6 plan. Four solver lanes, each with its paired verifier; the grey line is not pursued.

A second reference. Before round 6 the user supplied *upper_bound_3*, a continuation of *Toward 1.9* that claims

$$p_n\big(\lfloor (2 - 10^{-6})n \rfloor\big) \to 0, \qquad \text{hence} \qquad \alpha_- \le \alpha_+ \le 1.999999.$$

The orchestrator read all fourteen pages. The document keeps the passage from Theorem L to Theorem M that rounds 4 and 5 checked, and it replaces the two estimates that forced the load so close to 2.

- **Mean margins (its Theorem 4.1).** At load 1.999999, every feasible unit row w of a neuron has mean margin $\mu(w) = \frac{1}{m} \sum_\mu (A_i w)_\mu$ in $[L, U] = [0.39852, 0.39936]$. For feasible unit w_1, w_2, the direction of $w_1 + w_2$ is feasible, so $2L \le \mu(w_1 + w_2) \le U \|w_1 + w_2\|$, and then $\|w_1 - w_2\|^2 \le 4(1 - L^2/U^2) = 0.0168$. Every feasible cone has diameter below 0.13, which replaces Theorem Q.
- **Hard margin (its Theorem 5.1).** Each hard margin satisfies $\kappa_i < 4 \cdot 10^{-7}$ except with probability e^{-cm}. This pointwise bound replaces the mean bound $\mathbb{E}\kappa_i^2 \le (2 - u)/u$. With it, the centres leak less than 10^{-9} out of $\ker X$ for all but a vanishing fraction of labels, so the failure probability tends to zero instead of stopping at $3.2 \cdot 10^{-15}$.
- **Closing.** The good labels give $(\sqrt{3/80} - \sqrt{10^{-9}})^2 > 0.03748$, and a positive determinant absorbs the rest, so the population gap rises to 0.0374. The transfer gives 0.0373 in (2), while a storing network would have mismatch at most $2(0.13 + 10^{-9})^2/0.999^2 = 0.0339$.

Lanes. Round 6 runs four solvers. S1 writes the safe result: $\alpha_- \le 2 - 10^{-29}$ as one node whose 27 parents are all checked and unconditional. S1 has returned this node (S1-35), and V1 is checking it. S2 rebuilds Theorem 4.1 with the extensions it needs: replacement on domains with a tilt, one-sided concentration without convexity, and Gaussian comparison at the $1/m$ scale. S3 rebuilds Theorem 5.1 and the selectors and leakage bound at the new parameters. S4 rebuilds the gap 0.0374, its transfer at load 1.999999, and the contradiction $0.0339 < 0.0373$. The verifiers V2–V4 check each lane as it returns. The final verdict claims $\alpha_- \le 1.999999$ only if all three new lanes hold, and otherwise reports $\alpha_- \le 2 - 10^{-29}$. Fresh agents then audit the complete support. The document's own next target, a global dual certificate at load 1.9, stays open.

Appendix E. Gaussian comparison and sign matrices

Theorems P and R reduce statements about random sign matrices to Gaussian ones, where exact comparison tools exist.

Gordon's inequality. Let $G \in \mathbb{R}^{m\times d}$, $g \in \mathbb{R}^m$ and $h \in \mathbb{R}^d$ have independent standard Gaussian entries, let $S \subset \mathbb{R}^d$ and $T \subset \mathbb{R}^m$ be compact, and let ψ be continuous. Then for every real c,

$$\Pr\Big(\min_{x\in S}\max_{y\in T}\big(y^\top Gx + \psi(x,y)\big) < c\Big) \le 2\Pr\Big(\min_{x\in S}\max_{y\in T}\big(\|x\|g^\top y - \|y\|h^\top x + \psi(x,y)\big) \le c\Big)$$

[G85; TOH15, Thm 3]. Taking T a unit sphere turns the left side into a smallest restricted singular value, and the right side into a problem in two vectors that can be computed. Stojnic [S13] used it to prove Gardner's capacity 2 for the spherical perceptron.

From Gaussians to signs. Replace the entries of B by Gaussians one at a time and track the change of a smooth test function; each swap costs $O(\|\text{third derivative}\|_\infty)$, uniformly in the other entries [Ch06]. Oymak and Tropp prove universality of such restricted singular values for general iid entries [OT17]. *Toward 1.9* uses an elementary head–tail version: coordinates with large coefficients are deleted, then the light coordinates are swapped. *upper_bound_3* keeps the heavy coordinates inside the comparison instead, which its Theorem 4.1 needs.

Ordinary singular values. For m/d bounded above 1, the smallest singular value of an $m \times d$ sign matrix is of order $\sqrt{m}$ with exponentially small failure probability [RV09]; Theorem R is an ℓ_1 version with an explicit constant.

Literature. Gordon's comparison descends from Slepian's lemma; its use for perceptron capacity began with Stojnic [S13]. Gardner [G88] computed the capacity 2 by the replica method, and Cover [C65] proved it by counting.

References

[C65] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* EC-14 (1965), 326–334.

[Ch06] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.* 34 (2006), 2061–2076.

[G85] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.* 50 (1985), 265–289.

[G88] E. Gardner. The space of interactions in neural network models. *J. Phys. A* 21 (1988), 257–270.

[OT17] S. Oymak and J. A. Tropp. Universality laws for randomized dimension reduction, with applications. *Inf. Inference* 7 (2018), 337–446.

[RV09] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* 62 (2009), 1707–1739.

[S13] M. Stojnic. Another look at the Gardner problem. arXiv:1306.3979 (2013).

[TOH15] C. Thrampoulidis, S. Oymak and B. Hassibi. Regularized linear regression: a precise analysis of the estimation error. *Proc. COLT* (2015), 1683–1709.