OVER-LEARNING FORCES INSTABILITY FOR AT MOST FIVE AGENTS: TRACE BOUNDS FOR $\rho(A - \mathcal{E})$ AND A QUARTIC THRESHOLD

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. In the averaging-plus-learning dynamics $X_{t+1} = AX_t + \mathcal{E}(\bar\sigma - X_t)$ of Popescu and Vaidya, n agents average their neighbours' opinions with the weights of a row-stochastic matrix A while each is pulled towards an external ground truth $\bar\sigma$ at its own *learning rate* $\mathcal{E}(i) \geq 0$. The error obeys $X_{t+1} - \bar\sigma = (A - \mathcal{E})(X_t - \bar\sigma)$, so it dies from every initial condition exactly when $\rho(A - \mathcal{E}) < 1$. Popescu, Syatriadi and Vaidya ask whether a single agent learning at rate at least 3 already destroys that: is $\rho(A - \mathcal{E}) \geq 1$ as soon as $\mathcal{E}(i) \geq 3$ for some i? They settle $n = 1$ and $n = 2$, and settle every n only under an extra hypothesis (a symmetric A, or a Gershgorin isolation condition on the remaining agents). We prove the answer is yes for every $n \leq 5$, for the wider class of row-substochastic A and with no sign condition on the other learning rates. The instruments are the second and the fourth moment of the spectrum: writing $d_i = \mathcal{E}(i) - A(i,i)$, and r_i for the total averaging weight agent i sends to the other agents,

$$\rho(A - \mathcal{E})^2 \geq \frac{1}{n}\sum_i d_i^2, \qquad \operatorname{tr}\big((A - \mathcal{E})^4\big) \geq \sum_i d_i^4 - 4\sum_i \max(d_i, 0)\, r_i,$$

and the two are joined by a constraint peculiar to the problem: a fast learner pays for its outgoing weight out of its own diagonal, so that $\mathcal{E}(i_0) \geq 3$ forces $d_{i_0} \geq 2$ *and* $r_{i_0} \leq d_{i_0} - 2$. Four consequences hold at every n. The extremal function f_n of Popescu, Syatriadi and Vaidya satisfies $f_n(R) \geq (R - 1)^+/\sqrt{n}$, and also $f_n(R) \geq 1$ as soon as $R \geq 3$ and $(R - 1)^4 \geq 5n$, where the threshold they obtain from an $\|\cdot\|_\infty$ Ostrowski–Elsner spectral variation bound is $(n2^n + 3n - 1)/2$ — and their bound is vacuous at $R = 3$ for every $n \geq 2$. If a quarter of the agents learn at rate at least 3, then $\rho(A - \mathcal{E}) \geq 1$; and so it is if the set of agents reachable from one fast learner is closed under the influence relation and has at most four members. We also prove $f_3(3) < 2$, so that the exact formula $f_n(R) = (R - 1)^+$ of Popescu, Syatriadi and Vaidya stops at $n = 2$; the certificate is an explicit quadratic Lyapunov form, the characteristic polynomial being irreducible with a near-double root. Finally we show where the method stops: an explicit 7×7 instance satisfying the hypotheses has both trace powers below n while $\rho(A - \mathcal{E}) \geq 2$, and on it the fourth-moment estimate is an equality. Every theorem and proposition below is machine-checked in Lean 4.

1. Introduction

1.1. Averaging, learning, and the operator $A - \mathcal{E}$.

In the DeGroot model of opinion dynamics each of n agents replaces its opinion by a weighted average of its neighbours', the weights forming a row-stochastic matrix A. Popescu and Vaidya [4] add an external signal: alongside the averaging, agent i is pulled towards a ground truth $\bar\sigma$ at a rate $\mathcal{E}(i) \geq 0$ of its own. With $\mathcal{E} = \operatorname{diag}(\mathcal{E}(1), \dots, \mathcal{E}(n))$ the *averaging-plus-learning* recursion is

$$X_{t+1} \;=\; \underbrace{AX_t}_{\text{averaging}} \;+\; \underbrace{\mathcal{E}(\bar\sigma - X_t)}_{\text{learning}}. \tag{1}$$

Because A is row-stochastic it fixes the constant vector $\bar\sigma$, so the error $X_t - \bar\sigma$ obeys the linear recursion

$$X_{t+1} - \bar\sigma \;=\; (A - \mathcal{E})(X_t - \bar\sigma), \tag{2}$$

and, by the classical characterisation of convergent powers [2, Theorem 5.6.12], the error dies from every initial condition if and only if $\rho(A - \mathcal{E}) < 1$, where ρ denotes the spectral radius, the largest modulus of a complex eigenvalue.

A rate $\mathcal{E}(i) > 1$ is *over-learning*: agent i overshoots the truth at every step rather than approaching it. Enough over-learning must destabilise the network, and the question is how much is enough. Popescu, Syatriadi and Vaidya [5] pose it as follows; we quote their statement verbatim.

Problem 1.1 ([5, Problem 2.20]). "Let A be a row-stochastic matrix of size $n > 0$ and let $\mathcal{E}$ be a diagonal nonnegative matrix of size n. Suppose that there exists an $i \in \{1, 2, \dots, n\}$ such that $\mathcal{E}(i) \geq 3$. Is it true that $\rho(A - \mathcal{E}) \geq 1$?"

1.2. **What is known.** Everything below is from [5, Appendix A], where the problem is restated and attacked, and that appendix is the only part of [5] that bears on it: the body of the paper asks instead when $A - \mathcal{E}$ is *zero-convergent*, and the general results there that force $\rho(A - \mathcal{E}) \geq 1$ assume $A - \mathcal{E} \geq 0$ entrywise, which Problem 1.1 excludes — a fast agent has $(A - \mathcal{E})(i, i) \leq 1 - 3 = -2$, since $A(i, i) \leq 1$.

The constant 3 is not decoration. If A is *symmetric* row-stochastic, then $\mathcal{E}(i) \geq 2$ for a single i already forces $\rho(A - \mathcal{E}) \geq 1$: every eigenvalue of A is at most 1, the least eigenvalue of $-\mathcal{E}$ is at most -2, and Weyl's inequality puts an eigenvalue of $A - \mathcal{E}$ at most -1 [5, Theorem A.2]. Without symmetry the threshold 2 is false, and [5] exhibits the 3×3 pair

$$A = \begin{pmatrix} 1/5 & 0 & 4/5 \\ 1 & 0 & 0 \\ 0 & 1/5 & 4/5 \end{pmatrix}, \qquad \mathcal{E} = \operatorname{diag}\bigl(1/2,\, 0,\, 2\bigr), \tag{3}$$

for which it asserts $\rho(A - \mathcal{E}) < 0.92$. Under a Gershgorin isolation hypothesis — $\mathcal{E}(i) \geq 2$ for one i and $0 \leq \mathcal{E}(j) < 2A(j, j)$ for all $j \neq i$ — the threshold 2 is restored at every n [5, Proposition A.4].

The quantitative attack goes through an extremal function.

Definition 1.2 ([5, Definition A.5])**.** For $n \geq 1$ and $R \geq 0$,

$$f_n(R) \;:=\; \inf_{\mathcal{E}, A}\; \rho(\mathcal{E} - A),$$

the infimum over real diagonal $\mathcal{E} = \operatorname{diag}(\mathcal{E}(1), \dots, \mathcal{E}(n))$ with $\max_i \mathcal{E}(i) = \mathcal{E}(1) = R$ and over row-substochastic $A \in \mathbb{R}^{n \times n}$.

As printed, Definition A.5 imposes no sign condition on $\mathcal{E}$, although the problem it serves assumes $\mathcal{E} \geq 0$; nothing below depends on which reading is intended, since Theorem 3.1 and Corollary 4.3 hold for an arbitrary real diagonal $\mathcal{E}$. Since a row-stochastic matrix is row-substochastic, a lower bound on $f_n(R)$ is a lower bound on $\rho(A - \mathcal{E})$ for every instance of Problem 1.1 with $\max_i \mathcal{E}(i) = R$ [5, Remark A.6], and [5, Problem A.7] expects $f_n(3) \geq 1$ for all n, which it says would resolve Problem 1.1. Two results are proved about f_n. The first is exact and low-dimensional:

$$f_1(R) = f_2(R) = (R - 1)^+ \qquad \text{[5, Theorem A.12]}, \tag{4}$$

which settles Problem 1.1 for $n = 1$ and $n = 2$: an instance of the problem has $R := \max_i \mathcal{E}(i) \geq 3$, so $\rho(A - \mathcal{E}) \geq f_n(R) = R - 1 \geq 2$. The second holds at every n and comes from adapting the Ostrowski–Elsner spectral variation bound [3, 1] to the induced $\|\cdot\|_\infty$ norm:

$$f_n(R) \;\geq\; \Bigl(R - n^{1/n}(2R + 1)^{1 - 1/n}\Bigr)^+ \qquad \text{[5, Theorem A.11]}, \tag{5}$$

from which [5] deduces that $f_n(R) > 1$ for every

$$R \;>\; \bar{R}_n \;:=\; n2^{n-1} + \frac{3n - 1}{2} \;=\; \frac{n2^n + 3n - 1}{2}, \tag{6}$$

together with the simpler sufficient condition $R \geq n(2^{n-1} + 2)$. The threshold (6) is exponential in n, and at the value of R the problem is about it says nothing at all:

Remark 1.3. At $R = 3$ the right-hand side of (5) is 0 for every $n \geq 2$. Indeed $n^{1/n}7^{1-1/n} = 7(n/7)^{1/n} = 7 \exp \varphi(n)$ with $\varphi(n) = \frac{1}{n} \log \frac{n}{7}$, and $\varphi'(n) = n^{-2}\bigl(1 + \log \frac{7}{n}\bigr)$ is positive for $n < 7e$ and negative for $n > 7e$, while $\varphi(n) > 0$ for $n > 7$; so over the reals $n \geq 2$ the minimum of φ is at $n = 2$, giving $n^{1/n}7^{1-1/n} \geq 7\sqrt{2/7} = \sqrt{14} > 3$. At $n = 5$, for instance, (5) reads $f_5(3) \geq (3 - 5^{1/5}7^{4/5})^+ = (-3.54\dots)^+ = 0$. (An elementary calculation, recorded here for orientation; it is not part of the formal development.)

So before the present paper Problem 1.1 was open for every $n \geq 3$ in the absence of a symmetry or isolation hypothesis, and the only general lower bound on $f_n(3)$ was the trivial one.

1.3. **Results.** Write $Z := \mathcal{E} - A$, so that $\rho(Z) = \rho(A - \mathcal{E})$, set $d_i := \mathcal{E}(i) - A(i,i)$ for the diagonal of Z, and let

$$r_i \;:=\; \sum_j A(i,j) - A(i,i)$$

be the total averaging weight agent i sends to the other agents; for a row-substochastic A it lies in $[0,1]$. Two inequalities carry the paper. The first is Theorem 3.1: for entrywise nonnegative A and any real diagonal $\mathcal{E}$,

$$\rho(A - \mathcal{E})^2 \;\geq\; \frac{1}{n}\sum_{i=1}^{n} d_i^2. \tag{7}$$

The second is its fourth-moment companion, Proposition 9.2: if the rows of A sum to at most 1 as well, then

$$\operatorname{tr}\big((A-\mathcal{E})^4\big) \;\geq\; \sum_{i=1}^{n} d_i^4 - 4\sum_{i=1}^{n} \max(d_i, 0)\, r_i.$$

The mechanisms are classical. The off-diagonal contributions to $\operatorname{tr}(Z^2)$ are the products $A(i,j)A(j,i) \geq 0$; the fourth power is expanded and estimated term by term; and $|\operatorname{tr}(Z^m)| \leq n\rho(Z)^m$, the trace being the sum of the m-th powers of the eigenvalues. No novelty is claimed for either of them, nor for the technique of bounding a spectral radius below by a trace power, which is standard. What is new is that they apply to Problem 1.1, that in the regime the problem is about they are far stronger than (5), and that they combine with one elementary constraint peculiar to the problem. A fast learner pays for its outgoing averaging weight out of its own diagonal: if $\mathcal{E}(i_0) \geq 3$ then, A being row-substochastic,

$$d_{i_0} \geq 2 \qquad \text{and} \qquad r_{i_0} \leq d_{i_0} - 2.$$

The second of these is invisible to the second moment, and it is what carries the argument past $n = 4$. Six consequences follow.

- **Problem 1.1 is true for every** $n \leq 5$ (Theorem 9.5), for row-substochastic A and with no sign condition on $\mathcal{E}$. The previous range was $n \leq 2$; the second moment alone reaches $n \leq 4$ (Theorem 4.1).
- $f_n(R) \geq (R-1)^+/\sqrt{n}$, so $f_n(R) > 1$ as soon as $R > 1 + \sqrt{n}$ (Theorem 4.2 and Corollary 4.3); and $f_n(R) \geq 1$ **as soon as** $R \geq \max\big(3, 1 + (5n)^{1/4}\big)$ (Theorem 9.7 and Corollary 9.8), which is the better of the two at every $n \geq 6$. Either replaces the exponential threshold (6): at $n = 10$, $\bar{R}_{10} = 5134.5$ against $1 + \sqrt{10} = 4.162\ldots$ and $1 + 50^{1/4} = 3.659\ldots$.
- **A quarter of the population suffices, at every** n: if k agents all learn at rate at least 3 and $n \leq 4k$, then $\rho(A - \mathcal{E}) \geq 1$ (Theorem 4.5).
- **A small influence set suffices, at every** n: if some set of at most four agents contains a fast learner and no agent of the set puts averaging weight outside it, then $\rho(A - \mathcal{E}) \geq 1$ (Theorem 5.2). The hypothesis constrains the influence graph, not the population size.
- $f_3(3) < 2$ (Theorem 10.2), so the exact formula (4) of [5] does not extend past $n = 2$. The certificate is an explicit quadratic Lyapunov form rather than a factorisation of the characteristic polynomial, which is irreducible over $\mathbb{Q}$ with a near-double root.
- **The two moments stop at** $n = 6$ (Theorem 11.1): an explicit 7×7 instance satisfying the hypotheses of Problem 1.1 has both $\operatorname{tr}((A-\mathcal{E})^2)$ and $\operatorname{tr}((A-\mathcal{E})^4)$ below n, while $\rho(A - \mathcal{E}) \geq 2$ shows it is no counterexample; and on it the fourth-moment estimate is an equality (Proposition 11.2), so the distance between $n \leq 5$ and $n = 7$ is not slack in the estimates.

The third and the fourth are of a shape the tools of [5] do not produce: its general-n results constrain the *slow* agents (symmetry of A, or $\mathcal{E}(j) < 2A(j,j)$ for $j \neq i$), whereas these constrain the fast one and its neighbourhood.

1.4. **Grounds for the novelty claims.** No novelty is claimed for the two trace bounds, nor for the technique of bounding a spectral radius below by a trace power. What is claimed is that Problem 1.1 was open for every $n \geq 3$, that no bound on $f_n(R)$ other than (4) and (5) was in print, and that no value or bound for $f_n(R)$ at any $n \geq 3$ was; those are negative claims about the literature, so the searches behind them deserve to be stated. At the time of writing [5] had one forward citation on the two indexing services we consulted (one reports none; the other reports a single preprint, by two of its three authors), and that preprint cites [5] exactly once, for the averaging-plus-learning setup itself: it contains no occurrence of f_n, of the threshold 3, or of $\rho(A-\mathcal{E})$ — indeed neither the symbol ρ nor the phrase "spectral radius" occurs anywhere in it. A full-text sweep of a local preprint corpus (mathematics and computer science, coverage through 2026-09) for the vocabulary of [5] returned [5] itself, that one citing preprint, and a 2023 preprint on mean-field opinion dynamics whose only match is a bibliography line for [4]. A second sweep of the same corpus, on the intersection of "row-stochastic" with "spectral radius", added one earlier paper of the same line, [6], which uses the notation $A-\mathcal{E}$ for the same operator but works entirely in the stable regime $\mathcal{E}(i) < A(i,i)$ — learning rates below the diagonal, hence below 1 — and says nothing about over-learning, about the threshold 3, or about f_n. Web searches returned [5] for the problem and textbook material for the trace bound. The nearest published body of work is the *grounded Laplacian* line in control theory, which studies $L+D$ for a graph Laplacian L and a nonnegative diagonal D of stubborn-agent weights; it works on the opposite, M-matrix side — weights in $[0,1]$, the smallest eigenvalue, the stable regime — and we found nothing there about learning rates above 1, about $f_n(R)$, or about $\rho(A-\mathcal{E})$ in the over-learning regime. Every one of these negatives should be read as "not found", not as "does not exist".

Section 6 records two explicit 3×3 instances with exact certificates. The first is the pair (3) of [5]: we verify their numerical assertion by factoring the characteristic polynomial, so that the failure of the threshold 2 is now proved rather than computed. The second, $A = I_3$ and $\mathcal{E} = \mathrm{diag}(3,0,0)$, shows that the conclusion $\rho \geq 1$ of Theorem 4.1 cannot be improved past $\rho \geq R-1$. Section 7 reports what a numerical search says about $f_n(3)$, and is careful to say that those values are upper bounds obtained by heuristic minimisation, not theorems. Section 8 says where the second moment stops. Sections 9 to 11 carry the fourth moment: the estimate and the two theorems that come out of it, the certificate behind $f_3(3) < 2$, and the 7×7 instance that fixes the ceiling of the method.

1.5. **Changes from version 1.** This is a revised and extended version. The numbering of the first version is preserved and none of its statements is withdrawn or corrected: Problem 1.1, Definition 1.2, Remark 1.3, Lemma 2.1, Theorem 3.1, Theorems 4.1 and 4.2, Corollary 4.3, Proposition 4.4, Theorem 4.5, Lemma 5.1, Theorem 5.2 and Propositions 6.1 and 6.2 keep both their statements and their numbers. What is new is Sections 9 to 11 — the fourth-moment estimate and the two theorems it yields (Theorems 9.5 and 9.7), the Lyapunov certificate and $f_3(3) < 2$ (Proposition 10.1 and Theorem 10.2), and the 7×7 instance (Theorem 11.1) — together with a rewritten introduction. Two passages of the first version are superseded and have been rewritten where they stood. Section 7 gave $f_3(3) < 2$ as a computation and said that a certificate had not been found; Theorem 10.2 now supplies one. And Section 8 reported the fourth moment as attempted and unsuccessful; the ingredient it was missing was not a scalar inequality but the constraint $r_{i_0} \leq d_{i_0} - 2$ on the fast learner.

2. Notation and preliminaries

Throughout, $n \geq 1$ is an integer and matrices are real and $n \times n$ unless said otherwise. We write $A(i,j)$ for the entry of A in row i and column j, and $\mathcal{E}(i)$ for the i-th diagonal entry of a diagonal matrix $\mathcal{E}$, following [5]. A matrix A is *nonnegative* if $A(i,j) \geq 0$ for all i,j; *row-stochastic* if in addition $\sum_j A(i,j) = 1$ for every i; and *row-substochastic* if in addition $\sum_j A(i,j) \leq 1$ for every i. For $x \in \mathbb{R}$, $x^+ := \max(x,0)$.

The spectrum $\mathrm{spec}(M) \subseteq \mathbb{C}$ of a real matrix M is the spectrum of its complexification, that is, the set of roots of its characteristic polynomial $\det(\mu I - M)$; it is nonempty because $n \geq 1$, and $\rho(M) := \max\{|\mu| : \mu \in \mathrm{spec}(M)\}$. Two remarks on how the statements below are phrased. First, all of them are inequalities $\rho(M) \geq s$ or $\rho(M) < s$, and it is these that the formal development records, in the equivalent forms "some $\mu \in \mathrm{spec}(M)$ has $|\mu| \geq s$" and "every $\mu \in \mathrm{spec}(M)$ has $|\mu| < s$"; the two are negations of one another. Second, $\rho(A-\mathcal{E}) = \rho(\mathcal{E}-A)$, so we move freely between the operator $A-\mathcal{E}$ of (2) and the matrix $Z = \mathcal{E} - A$ whose diagonal $d_i = \mathcal{E}(i) - A(i,i)$ is the quantity that carries the argument.

We use one elementary fact about substochastic matrices repeatedly.

Lemma 2.1. *If A is row-substochastic then $A(i,i) \le 1$ for every i.*

Proof. $A(i,i) \le \sum_j A(i,j) \le 1$, the first inequality because the omitted terms are nonnegative. $\square$

Consequently, for a row-substochastic A an agent with $\mathcal{E}(i) \ge R$ has $d_i = \mathcal{E}(i) - A(i,i) \ge R - 1$. This is the only place where substochasticity is used anywhere below; in particular nothing needs the row sums to be exactly 1, and nothing needs $\mathcal{E} \ge 0$.

3. The trace bound

Theorem 3.1 (Trace bound)**.** *Let $n \ge 1$, let A be a nonnegative $n \times n$ matrix, let $\mathcal{E}$ be any real diagonal matrix, and let $s \in \mathbb{R}$. If every $\mu \in \operatorname{spec}(A - \mathcal{E})$ satisfies $|\mu| < s$, then*

$$\sum_{i=1}^{n} \big(\mathcal{E}(i) - A(i,i)\big)^2 \;<\; n\, s^2.$$

Contrapositively, if $ns^2 \le \sum_i (\mathcal{E}(i) - A(i,i))^2$ then $A - \mathcal{E}$ has an eigenvalue of modulus at least s; and a single agent already suffices, in the sense that $ns^2 \le (\mathcal{E}(i_0) - A(i_0,i_0))^2$ for one index i_0 has the same conclusion. Equivalently,

$$\rho(A - \mathcal{E}) \;\ge\; \Big(\frac{1}{n}\sum_{i=1}^{n}\big(\mathcal{E}(i) - A(i,i)\big)^2\Big)^{1/2}.$$

Proof. Put $Z = \mathcal{E} - A$ and $d_i = \mathcal{E}(i) - A(i,i)$, so $Z^2 = (A - \mathcal{E})^2$ and $Z(i,i) = d_i$, $Z(i,j) = -A(i,j)$ for $i \ne j$. Then

$$\operatorname{tr}(Z^2) \;=\; \sum_i \sum_j Z(i,j)Z(j,i) \;=\; \sum_i d_i^2 + \sum_{i \ne j} A(i,j)A(j,i) \;\ge\; \sum_i d_i^2,$$

the inequality because A is nonnegative: this is the only step that uses anything about A, and it is where the model enters.

For the other side, put $M := A - \mathcal{E}$, so that $Z^2 = M^2$, and work over $\mathbb{C}$. The characteristic polynomial of M^2 has degree n and splits, so its root multiset $\{\lambda_1, \ldots, \lambda_n\}$ has exactly n members and $\operatorname{tr}(M^2) = \sum_k \lambda_k$. Each λ_k lies in $\operatorname{spec}(M^2) = \{\mu^2 : \mu \in \operatorname{spec}(M)\}$, so $\lambda_k = \mu_k^2$ for some $\mu_k \in \operatorname{spec}(M)$. By hypothesis $|\mu_k| < s$ for every k (which forces $s > 0$, the spectrum being nonempty), whence $|\lambda_k| = |\mu_k|^2 < s^2$ and, the multiset being nonempty,

$$\big|\operatorname{tr}(Z^2)\big| \;=\; \big|\operatorname{tr}(M^2)\big| \;\le\; \sum_{k=1}^{n} |\lambda_k| \;<\; n\, s^2.$$

Combining, $\sum_i d_i^2 \le \operatorname{tr}(Z^2) \le |\operatorname{tr}(Z^2)| < ns^2$. The contrapositive is immediate, and the single-agent form follows since $(\mathcal{E}(i_0) - A(i_0,i_0))^2 \le \sum_i d_i^2$. For the last display, apply the contrapositive with $s = (\frac{1}{n}\sum_i d_i^2)^{1/2}$. $\square$

Theorem 3.1 is sharp in its own terms: for a diagonal A the inequality $\operatorname{tr}(Z^2) \ge \sum_i d_i^2$ is an equality, and for Z diagonal with one large entry the factor $1/n$ is exactly the loss incurred by spreading a single large eigenvalue over n of them. Section 6 makes that second point precise.

4. Consequences for the problem

4.1. At most four agents.

Theorem 4.1. *Let $1 \le n \le 4$, let A be row-substochastic, and let $\mathcal{E}$ be any real diagonal matrix such that $\mathcal{E}(i_0) \ge 3$ for some i_0. Then $\rho(A - \mathcal{E}) \ge 1$.*

Proof. By Lemma 2.1, $d_{i_0} = \mathcal{E}(i_0) - A(i_0,i_0) \ge 3 - 1 = 2$, so $d_{i_0}^2 \ge 4 \ge n = n \cdot 1^2$. Apply the single-agent form of Theorem 3.1 with $s = 1$. $\square$

This answers Problem 1.1 affirmatively for $n \le 4$, where the range previously available was $n \le 2$. Three of the hypotheses of Problem 1.1 are not needed: A may be row-substochastic rather than row-stochastic, the remaining learning rates may be negative, and the fast agent's rate enters only through $\mathcal{E}(i_0) \ge 3$, not through $\max_i \mathcal{E}(i)$.

4.2. **A $\sqrt{n}$ threshold.**

Theorem 4.2. *Let $n \ge 1$, let A be row-substochastic, and let $\mathcal{E}$ be any real diagonal matrix such that $\mathcal{E}(i_0) \ge 1 + \sqrt{n}$ for some i_0. Then $\rho(A - \mathcal{E}) \ge 1$.*

Proof. By Lemma 2.1, $d_{i_0} \ge (1 + \sqrt{n}) - 1 = \sqrt{n}$, so $d_{i_0}^2 \ge n = n \cdot 1^2$, and Theorem 3.1 applies with $s = 1$. □

Corollary 4.3. *For every $n \ge 1$ and $R \ge 0$,*

$$f_n(R) \;\ge\; \frac{(R-1)^+}{\sqrt{n}},$$

and consequently $f_n(R) > 1$ for every $R > 1 + \sqrt{n}$.

Proof. For $R \le 1$ there is nothing to prove. For $R > 1$, let $\mathcal{E}$ and A be admissible in Definition 1.2, so that $\mathcal{E}(1) = R$ and A is row-substochastic, hence $A(1,1) \le 1$ and $d_1 \ge R - 1 > 0$. Theorem 3.1 with $s = (R-1)/\sqrt{n}$ and $i_0 = 1$ gives $ns^2 = (R-1)^2 \le d_1^2$ and therefore $\rho(\mathcal{E} - A) = \rho(A - \mathcal{E}) \ge (R-1)/\sqrt{n}$. Take the infimum. □

Corollary 4.3 lowers the threshold above which $f_n(R) > 1$ from $\bar{R}_n = (n2^n + 3n - 1)/2$ to $1 + \sqrt{n}$:

n	1	2	3	4	5	10
$\bar{R}_n$ of (6)	2	6.5	16	37.5	87	5134.5
$1 + \sqrt{n}$	2	2.414	2.732	3	3.236	4.162

The two lower bounds (5) and Corollary 4.3 are not comparable pointwise: as $R \to \infty$ with n fixed, (5) behaves like $R - O(R^{1-1/n})$ and so eventually beats $(R-1)/\sqrt{n}$, whose slope is $1/\sqrt{n}$. The point is the opposite regime. At the bounded values of R that Problem 1.1 is about, (5) is vacuous for every $n \ge 2$ (Remark 1.3) while Corollary 4.3 is not; at $R = 3$ it gives $f_n(3) \ge 2/\sqrt{n}$, which is what Theorem 4.1 is, and which decays only polynomially.

4.3. **Every agent over-learning.** The next statement is elementary — it is one line from Theorem 3.1, and it is also a two-line Gershgorin argument, since every eigenvalue of $\mathcal{E} - A$ lies in a disc centred at d_i of radius $1 - A(i,i)$ and so has real part at least $\mathcal{E}(i) - 1$. We record it because it is the exact "all agents over-learn" companion to the two general-n results of [5], and neither of those covers it: no symmetry and no isolation hypothesis is imposed. No novelty is claimed for it.

Proposition 4.4. *Let $n \ge 1$, let A be row-substochastic, let $c \ge 0$, and let $\mathcal{E}$ be a real diagonal matrix with $\mathcal{E}(i) \ge c + 1$ for every i. Then $\rho(A - \mathcal{E}) \ge c$. In particular, if every agent learns at rate at least 2 then $\rho(A - \mathcal{E}) \ge 1$, at every n.*

Proof. By Lemma 2.1, $d_i \ge (c+1) - 1 = c \ge 0$ for every i, so $\sum_i d_i^2 \ge nc^2$, and Theorem 3.1 applies with $s = c$. □

4.4. **Many fast learners.** Over-learning by a positive fraction of the population forces instability at every n, however large.

Theorem 4.5. *Let $n \ge 1$, let A be row-substochastic, let $\mathcal{E}$ be a real diagonal matrix, and let $S \subseteq \{1, \ldots, n\}$ be a set of agents with $\mathcal{E}(i) \ge 3$ for every $i \in S$. If $n \le 4|S|$, then $\rho(A - \mathcal{E}) \ge 1$.*

Proof. Each $i \in S$ has $d_i \ge 2$ by Lemma 2.1, so $d_i^2 \ge 4$ and $\sum_{i \in S} d_i^2 \ge 4|S| \ge n$. Since the omitted terms d_j^2, $j \notin S$, are nonnegative, $\sum_i d_i^2 \ge n = n \cdot 1^2$, and Theorem 3.1 applies with $s = 1$. □

In words: if a quarter of the agents learn at rate 3, the network cannot forget its error, whatever the averaging weights are and whatever the other rates are. The general-n results of [5] are of the opposite shape — they ask the slow agents to be well-behaved — so this does not follow from them.

5. A small influence set is enough

The hypothesis $n \leq 4$ of Theorem 4.1 is a constraint on the population. It can be replaced by a constraint on the influence graph alone, which is the more natural object: what matters is not how many agents there are but how many the fast learner can reach.

Call a set $S \subseteq \{1, \ldots, n\}$ of agents *out-closed* if no agent of S puts averaging weight outside S, that is, if $A(i,j) = 0$ whenever $i \in S$ and $j \notin S$. Equivalently, S is closed under the influence relation "i listens to j"; the set of agents reachable from a given one is the smallest such set containing it. For an out-closed S the principal submatrix of $A - \mathcal{E}$ on S sits block-triangularly inside $A - \mathcal{E}$, and its spectrum embeds. We state this in the form used below, where S is presented as an injection.

Lemma 5.1 (Spectrum embedding)**.** *Let M be an $n \times n$ matrix and let $\iota : \{1, \ldots, m\} \to \{1, \ldots, n\}$ be injective with out-closed image, in the sense that $M(\iota(a), j) = 0$ for every $a \in \{1, \ldots, m\}$ and every j outside the image of ι. Then every eigenvalue of the principal submatrix $M_\iota := (M(\iota(a), \iota(b)))_{a,b}$ is an eigenvalue of M.*

Proof. It suffices to show that $\det(M_\iota) = 0$ implies $\det(M) = 0$, since restriction commutes with the shift $M \mapsto \mu I - M$: the (a, b) entry of $(\mu I_n - M)_\iota$ is $\mu\delta_{ab} - M(\iota(a), \iota(b))$, because ι is injective, and $\mu I_n - M$ inherits the out-closedness from M (its entries outside the image, on rows in the image, are those of $-M$).

So suppose $\det(M_\iota) = 0$. Then M_ι has a nonzero left null vector $w \in \mathbb{C}^m$. Spread it out by zero: let $v \in \mathbb{C}^n$ have $v(\iota(a)) = w(a)$ and $v(j) = 0$ for j outside the image. Then $v \neq 0$, and for every j,

$$(v^\top M)(j) \;=\; \sum_{i=1}^{n} v(i)M(i,j) \;=\; \sum_{a=1}^{m} w(a)M(\iota(a), j).$$

If $j = \iota(b)$ lies in the image this is the b-th entry of $w^\top M_\iota$, hence 0; if j lies outside the image, every $M(\iota(a), j)$ vanishes by hypothesis, so the sum is 0 again. Thus $v^\top M = 0$ with $v \neq 0$ and $\det(M) = 0$. ☐

Theorem 5.2. *Let $n \geq 1$, let A be row-substochastic, and let $\mathcal{E}$ be a real diagonal matrix. Suppose some out-closed set S of agents satisfies $1 \leq |S| \leq 4$ and contains an agent i_0 with $\mathcal{E}(i_0) \geq 3$. Then $\rho(A - \mathcal{E}) \geq 1$, whatever n is.*

Proof. Write $m = |S|$ and let ι enumerate S. The matrix $A - \mathcal{E}$ is out-closed along ι: for $j \notin S$ and $i \in S$ we have $j \neq i$, so the (i, j) entry of $A - \mathcal{E}$ is $A(i, j) = 0$. Its restriction is $(A - \mathcal{E})_\iota = A_\iota - \mathcal{E}_\iota$, where $\mathcal{E}_\iota$ is the diagonal matrix of the rates of the agents of S, again because ι is injective. Moreover A_ι is row-substochastic: its rows are subsums of rows of A over nonnegative entries. Since $1 \leq m \leq 4$ and $\mathcal{E}_\iota$ has an entry at least 3, Theorem 4.1 gives an eigenvalue of $A_\iota - \mathcal{E}_\iota$ of modulus at least 1, and Lemma 5.1 promotes it to an eigenvalue of $A - \mathcal{E}$. ☐

The block-triangular embedding of Lemma 5.1 is standard, and only the consequence for Problem 1.1 is claimed here. Theorem 5.2 suggests the right general statement is about the strongly connected component of the fast learner in the condensation of the influence graph; the out-closed version is the half of that which needs no condensation.

6. Two exact instances

Both instances below are 3×3, and both are certified by factoring the characteristic polynomial exactly over $\mathbb{Q}$ and bounding the roots of the factors; no floating-point arithmetic and no numerical root-finding enters the certificates.

6.1. The threshold cannot be lowered to 2.

The pair (3) is that of [5], which states $\rho(A - \mathcal{E}) < 0.92$ for it without proof. The following makes that exact. It is the negative control for everything above: it shows the hypothesis $\mathcal{E}(i) \geq 3$ of Problem 1.1 cannot be weakened to $\mathcal{E}(i) \geq 2$, so no theorem of this paper survives that weakening.

Proposition 6.1 (after [5, Appendix A])**.** *For the row-stochastic A and the nonnegative diagonal $\mathcal{E}$ of* (3)*,*

$$\det\big(\mu I - (A-\mathcal{E})\big) = \Big(\mu + \tfrac{4}{5}\Big)\Big(\mu^2 + \tfrac{7}{10}\mu - \tfrac{1}{5}\Big),$$

and every $\mu \in \operatorname{spec}(A-\mathcal{E})$ satisfies $|\mu| < \frac{23}{25} = 0.92$. In particular $\rho(A-\mathcal{E}) < 1$ although $\max_i \mathcal{E}(i) = 2$.

Proof. $A-\mathcal{E}$ is the matrix with rows $(-3/10, 0, 4/5)$, $(1,0,0)$, $(0, 1/5, -6/5)$, and the displayed factorisation is a direct expansion of the 3×3 determinant. If $\mu + 4/5 = 0$ then $|\mu| = 4/5 < 23/25$. Otherwise $\mu^2 + \frac{7}{10}\mu - \frac{1}{5} = 0$; multiplying by 400 gives $(20\mu+7)^2 = 129$, so $|20\mu + 7| = \sqrt{129} < 57/5$ because $129 < (57/5)^2 = 129.96$. Hence $20|\mu| = |(20\mu+7) - 7| \le |20\mu+7| + 7 < \frac{57}{5} + 7 = \frac{92}{5}$, that is, $|\mu| < \frac{92}{100} = \frac{23}{25}$. □

6.2. **The conclusion cannot be strengthened.** Theorem 4.1 concludes $\rho(A-\mathcal{E}) \ge 1$ from $\mathcal{E}(i_0) \ge 3$. Since $f_1(R) = f_2(R) = (R-1)^+$, one might hope for $\rho \ge R-1 = 2$, or more. The following shows that 2 is the ceiling at $n=3$: no conclusion of the form $\rho(A-\mathcal{E}) \ge c$ with $c > R-1$ can be drawn from $\mathcal{E}(i_0) \ge R$ there, because an instance attains $R-1$ exactly.

Proposition 6.2. *For $A = I_3$ and $\mathcal{E} = \operatorname{diag}(3,0,0)$,*

$$\det\big(\mu I - (A-\mathcal{E})\big) = (\mu+2)(\mu-1)^2,$$

so $\operatorname{spec}(A-\mathcal{E}) = \{-2, 1\}$ and $\rho(A-\mathcal{E}) = 2$. In particular $\mathcal{E}(1) = 3$ does not force $\rho(A-\mathcal{E}) \ge \frac{5}{2}$ at $n=3$; and since A is row-stochastic and $\mathcal{E}$ nonnegative, the hypotheses of Theorem 4.1 are satisfiable, so that theorem is not vacuous.

Proof. $A-\mathcal{E} = \operatorname{diag}(-2,1,1)$ and the determinant of a diagonal matrix is the product of its entries. □

7. Numerical evidence about $f_n(3)$

This section reports computations, not theorems. Nothing in it is proved, nothing in it is used anywhere else in this paper, and none of it is part of the formal development.

We minimised $\rho(\mathcal{E} - A)$ over the admissible pairs of Definition 1.2 at $R=3$, by random-restart Nelder–Mead over a softmax parametrisation of the rows of A and a scaled-sigmoid parametrisation of $\mathcal{E}(2), \dots, \mathcal{E}(n)$, evaluating ρ from the characteristic polynomial by the Faddeev–LeVerrier recursion followed by Durand–Kerner root-finding in double precision, and cross-checking against $\|M^{2^{20}}\|_\infty^{2^{-20}}$. The estimator was validated on the pair (3), where it returns $0.9178908\ldots$, in agreement with Proposition 6.1. The smallest values found were

n	2	3	4	5
smallest ρ found	2.000	1.8078	1.7756	1.7835

These are upper bounds for $f_n(3)$ obtained by a heuristic search: they are evidence about the infimum and proofs of nothing. Two things are worth saying about them.

First, $f_3(3) < 2$: the search at $n=3$ found a rational witness,

$$A = \begin{pmatrix} 9/10 & 1/10 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \mathcal{E} = \operatorname{diag}\big(3, \tfrac{5}{4}, \tfrac{5}{4}\big), \tag{8}$$

with $\det(\mu I - (A-\mathcal{E})) = (\mu + \frac{21}{10})(\mu + \frac{5}{4})^2 - \frac{1}{10}$ and $\rho(A-\mathcal{E}) = 1.82559\ldots < 2$. So the exact formula $f_n(R) = (R-1)^+$ of (4) does *not* extend beyond $n=2$. That last conclusion is a theorem and not a computation: Theorem 10.2 below certifies $\rho(A-\mathcal{E}) < 2$ for (8) exactly. The certificate is not a factorisation of the characteristic polynomial — the cubic is irreducible over $\mathbb{Q}$, its two dominant roots are a complex conjugate pair near a double root, and neither the crude bound $\big\|\big(\frac{1}{2}(A-\mathcal{E})\big)^k\big\|_\infty < 1$ for a small power k nor a modulus test on rational factors succeeds — but a quadratic Lyapunov form, which needs no factorisation at all. The value 1.8078 in the table above remains a measurement.

Second, the sequence $f_n(3)$ is non-increasing, so the value at $n=5$ sitting above the value at $n=4$ is search noise rather than a real increase. Indeed $f_{n+1}(R) \le f_n(R)$ for every $R \ge 0$: given an admissible pair $(\mathcal{E}, A)$ of size n, adjoin an $(n+1)$-st agent that does not learn ($\mathcal{E}(n+1) = 0$) and puts all its averaging weight on agent 1. The enlarged pair is admissible, and the enlarged $Z' = \mathcal{E}' - A'$ is block lower triangular with diagonal blocks Z and $[0]$, so $\operatorname{spec}(Z') = \operatorname{spec}(Z) \cup \{0\}$ and $\rho(Z') = \rho(Z)$. Consequently $\inf_n f_n(3) = \lim_n f_n(3)$, and if Problem 1.1 fails at some n it fails at every larger one. The measured values decrease slowly and stay far above 1, which is consistent with the affirmative answer that [5] expects, but the search reached only $n = 5$ and says nothing about the limit.

8. Where the second moment stops

Theorem 3.1 at $s = 1$ needs $n \leq \sum_i d_i^2$, and a lone fast learner at rate exactly 3 contributes exactly 4. So the hypothesis $n \leq 4$ of Theorem 4.1 is not an artefact of the proof: it is where the second moment of the spectrum runs out. Up to that point nothing about the averaging weights has been used beyond their nonnegativity. The next section uses them. The fourth moment sees the closed walks of length three and four in the influence graph, and together with the constraint that a fast learner pays for its outgoing weight out of its own diagonal it reaches $n = 5$. A crude estimate of it does not: bounding the length-three term by Cauchy–Schwarz alone, with $\sum_i d_i^2 < n$ and $(F^3)(i,i) \leq 1$, where F is the off-diagonal part of A as in Section 9, gives $\operatorname{tr}(Z^4) \geq 16 - 4n$, which needs $n < 3.2$ and is worse than the second moment. Section 11 shows that the fourth moment stops in its turn, and where.

8.1. Looking for a counterexample.

A counterexample at some n would settle Problem 1.1 negatively and would be cheap to certify: a single explicit rational pair with $\|(A - \mathcal{E})^k\|_\infty < 1$ for one k suffices. The search reported in Section 7 stopped at $n = 5$ because its cost grows steeply — the runs took 58, 660 and 2084 seconds at $n = 3, 4, 5$ — and a serious hunt in the range $n = 8, \ldots, 16$ wants a compiled inner loop. Given the slow measured decrease of $f_n(3)$ we do not expect one, but the expectation is not evidence.

9. The fourth moment

Split A into its diagonal and its off-diagonal part. Let $D := \operatorname{diag}(d_1, \ldots, d_n)$ with $d_i = \mathcal{E}(i) - A(i,i)$ as before, and let F be the off-diagonal part of A, so that $F(i,i) = 0$ and $F(i,j) = A(i,j)$ for $i \neq j$. Then

$$A - \mathcal{E} = F - D, \qquad Z = \mathcal{E} - A = D - F,$$

and if A is row-substochastic then F is nonnegative with zero diagonal and row sums

$$r_i \;=\; \sum_j F(i,j) \;=\; \sum_j A(i,j) - A(i,i) \;\in\; [0,1],$$

the *outgoing weight* of agent i: what it gives to the others. The fourth power is even, so $\operatorname{tr}((A - \mathcal{E})^4) = \operatorname{tr}(Z^4)$.

Lemma 9.1. *For any two real $n \times n$ matrices D and F,*

$$\operatorname{tr}\big((F - D)^4\big) = \operatorname{tr}(F^4) - 4\operatorname{tr}(DF^3) + 4\operatorname{tr}(D^2F^2) + 2\operatorname{tr}\big((DF)^2\big) - 4\operatorname{tr}(D^3F) + \operatorname{tr}(D^4).$$

Proof. Expand $(F - D)^4$ into its sixteen words in D and F and fold them into cyclic classes by $\operatorname{tr}(abcd) = \operatorname{tr}(dabc)$. The four words with one F give $-4\operatorname{tr}(D^3F)$; the six with two F split into the four cyclic rotations of D^2F^2 and the two of $DFDF$; the four with three F give $-4\operatorname{tr}(DF^3)$; and the two extreme words give $\operatorname{tr}(F^4)$ and $\operatorname{tr}(D^4)$. □

Proposition 9.2. *Let $d \in \mathbb{R}^n$, let $D = \operatorname{diag}(d_1, \ldots, d_n)$, and let F be nonnegative with zero diagonal and row sums $r_i = \sum_j F(i,j) \leq 1$. Then*

$$\operatorname{tr}\big((F - D)^4\big) \;\geq\; \sum_{i=1}^n d_i^4 - 4\sum_{i=1}^n \max(d_i, 0)\, r_i.$$

Proof. Take the six terms of Lemma 9.1 in turn. First $\operatorname{tr}(D^4) = \sum_i d_i^4$ and $\operatorname{tr}(D^3F) = \sum_i d_i^3 F(i,i) = 0$, the diagonal of F being zero. Next $\operatorname{tr}(F^4) \geq 0$, each closed four-walk contributing a product of nonnegative numbers. For the two-F block,

$$4\operatorname{tr}(D^2F^2) + 2\operatorname{tr}\big((DF)^2\big) = \sum_{i,j}\big(4d_i^2 + 2d_id_j\big)F(i,j)F(j,i),$$

whose summand need not be nonnegative; but $F(i,j)F(j,i)$ is symmetric in i and j, so adding the sum to itself with i and j exchanged turns it into $4\sum_{i,j}\big(d_i^2 + d_id_j + d_j^2\big)F(i,j)F(j,i)$, and $d_i^2 + d_id_j + d_j^2 = \big(d_i + \frac{1}{2}d_j\big)^2 + \frac{3}{4}d_j^2 \geq 0$; so the block is nonnegative. Finally the length-three term. Every entry of F is at most its own row sum and hence at most 1, so $(F^2)(j,i) = \sum_k F(j,k)F(k,i) \leq \sum_k F(j,k) \leq 1$ and therefore $0 \leq (F^3)(i,i) = \sum_j F(i,j)(F^2)(j,i) \leq r_i$, whence

$$-4\operatorname{tr}(DF^3) = -4\sum_i d_i\,(F^3)(i,i) \;\geq\; -4\sum_i \max(d_i,0)\,(F^3)(i,i) \;\geq\; -4\sum_i \max(d_i,0)\,r_i. \qquad \square$$

Lemma 9.3. *Let $n \geq 1$, let M be a real $n \times n$ matrix, let $s \in \mathbb{R}$ and let $k \geq 1$. If every $\mu \in \operatorname{spec}(M)$ satisfies $|\mu| < s$, then $\operatorname{tr}(M^k) < n\,s^k$.*

Proof. As in the second half of the proof of Theorem 3.1, with M^k in place of M^2: the characteristic polynomial of M^k splits over $\mathbb{C}$, its root multiset has n members, each of them is μ^k for some $\mu \in \operatorname{spec}(M)$, and $|\mu^k| < s^k$. $\square$

Corollary 9.4. *Let $n \geq 1$, let A be row-substochastic and let $\mathcal{E}$ be any real diagonal matrix. If $\rho(A - \mathcal{E}) < 1$ then*

$$\sum_{i=1}^n d_i^2 < n \qquad \textit{and} \qquad \sum_{i=1}^n d_i^4 - 4\sum_{i=1}^n \max(d_i,0)\,r_i < n.$$

Proof. The first is Theorem 3.1 at $s = 1$. For the second, combine Proposition 9.2, applied to D and F as above, with Lemma 9.3 at $k = 4$ and $s = 1$. $\square$

9.1. **The fast learner pays for its outgoing weight.** The two constraints of Corollary 9.4 pull in opposite directions: a large d_{i_0} helps the second and, through the term $-4\max(d_{i_0},0)r_{i_0}$, hurts it. What saves the argument is that a fast learner cannot have both a large diagonal and a large outgoing weight. Let A be row-substochastic and let $\mathcal{E}(i_0) = R$. Then $A(i_0,i_0) = R - d_{i_0}$, and $A(i_0,i_0) \geq 0$ together with $\sum_j A(i_0,j) \leq 1$ gives

$$(9) \qquad d_{i_0} \;\geq\; R - 1 \qquad \text{and} \qquad r_{i_0} \;=\; \sum_j A(i_0,j) - A(i_0,i_0) \;\leq\; 1 - A(i_0,i_0) \;=\; 1 - R + d_{i_0}.$$

At $R \geq 3$ the second reads $r_{i_0} \leq d_{i_0} - 2$: a fast learner can spend outgoing weight at all only in so far as its diagonal exceeds 2, and it pays for every unit of weight with a unit of diagonal. This is what the second moment cannot see, and it is exactly the term the crude estimate of Section 8 throws away.

9.2. **At most five agents.**

Theorem 9.5. *Let $1 \leq n \leq 5$, let A be row-substochastic, and let $\mathcal{E}$ be any real diagonal matrix such that $\mathcal{E}(i_0) \geq 3$ for some i_0. Then $\rho(A - \mathcal{E}) \geq 1$.*

Proof. Suppose $\rho(A - \mathcal{E}) < 1$, and put

$$x := d_{i_0}, \qquad s := \sum_{i \neq i_0} d_i^2, \qquad \sigma := \sum_{i \neq i_0} \max(d_i, 0).$$

By (9), $x \geq 2$ and $r_{i_0} \leq x - 2$. The first constraint of Corollary 9.4 reads $x^2 + s < n \leq 5$, and $x^2 \geq 4$, so $s < 1$. By Cauchy–Schwarz over the $n-1$ remaining agents, and since $\max(d_i,0)^2 \leq d_i^2$,

$$\sigma^2 \;\leq\; (n-1)\sum_{i \neq i_0} \max(d_i,0)^2 \;\leq\; (n-1)s \;\leq\; 4s \;<\; 4,$$

so $\sigma < 2$. In the second constraint, $\max(x, 0) = x$ and $r_i \le 1$ for $i \ne i_0$, so $\sum_i \max(d_i, 0) r_i \le x(x-2)+\sigma$; discarding the nonnegative terms d_i^4 with $i \ne i_0$,

$$n \;>\; x^4 - 4\bigl(x(x-2)+\sigma\bigr) \;=\; x^4 - 4x^2 + 8x - 4\sigma.$$

Finally $x^4 - 4x^2 + 8x - 16 = (x-2)(x^3 + 2x^2 + 8) \ge 0$ for $x \ge 2$, so $x^4 - 4x^2 + 8x \ge 16$ and $n > 16 - 4\sigma > 16 - 8 = 8$, contradicting $n \le 5$. ☐

The proof is symbolic; no case analysis, no enumeration and no numerical computation enters it, and the same is true of Theorem 9.7 below. As with Theorem 4.1, three hypotheses of Problem 1.1 are not needed: A may be row-substochastic rather than row-stochastic, the other learning rates may be negative, and only $\mathcal{E}(i_0) \ge 3$ is used, not $\max_i \mathcal{E}(i)$. Since the statement is for $n \le 5$ it contains Theorem 4.1 rather than standing beside it. The margin in the last line is wide — 8 against 5 — but it cannot be cashed in for another dimension, because the step $\sigma < 2$ used $n \le 5$: at $n = 6$ the same chain gives only $s < 2$ and $\sigma < \sqrt{10}$, and ends at $n > 16 - 4\sqrt{10} = 3.35\ldots$, which $n = 6$ does not contradict. What can and cannot be done with these two moments is settled in Section 11.

Proposition 9.6. *For $A = I_5$ and $\mathcal{E} = \operatorname{diag}(3, 0, 0, 0, 0)$,*

$$\det\bigl(\mu I - (A - \mathcal{E})\bigr) = (\mu + 2)(\mu - 1)^4,$$

so $\operatorname{spec}(A - \mathcal{E}) = \{-2, 1\}$ and $\rho(A - \mathcal{E}) = 2$. In particular the hypotheses of Theorem 9.5 are satisfiable at $n = 5$ with A row-stochastic and $\mathcal{E}$ nonnegative, so that theorem is not vacuous; and its conclusion cannot be strengthened past $\rho \ge R - 1$ there.

Proof. $A - \mathcal{E} = \operatorname{diag}(-2, 1, 1, 1, 1)$, and the determinant of a diagonal matrix is the product of its entries. ☐

9.3. **A quartic threshold.** The same two constraints, run at a general rate rather than at $R = 3$, replace the square-root threshold of Theorem 4.2 by a quartic one.

Theorem 9.7. *Let $n \ge 1$, let A be row-substochastic, and let $\mathcal{E}$ be a real diagonal matrix with $\mathcal{E}(i_0) = R \ge 3$ for some i_0. If $5n \le (R-1)^4$ — equivalently $R \ge 1 + (5n)^{1/4}$ — then $\rho(A - \mathcal{E}) \ge 1$.*

Proof. Suppose $\rho(A - \mathcal{E}) < 1$ and keep x, s, σ as in the previous proof, so that $x \ge R - 1 \ge 2$ and $r_{i_0} \le x - R + 1$ by (9). The first constraint of Corollary 9.4 gives $x^2 + s < n$, hence $s < n$, and Cauchy–Schwarz gives $\sigma^2 \le (n-1)s < n^2$, hence $\sigma < n$. The second constraint then gives

$$n \;>\; x^4 - 4\bigl(x(x-R+1)+\sigma\bigr) \;=\; x^4 - 4x^2 + 4(R-1)x - 4\sigma \;>\; x^4 - 4x^2 + 4(R-1)x - 4n.$$

Writing $a := R - 1 \ge 2$, the identity

$$x^4 - 4x^2 + 4ax - a^4 = (x-a)\bigl((x+a)(x^2+a^2-4) + 4a\bigr)$$

exhibits its left-hand side as a product of nonnegative factors when $x \ge a \ge 2$, so $x^4 - 4x^2 + 4(R-1)x \ge (R-1)^4 \ge 5n$ and $n > 5n - 4n = n$, which is absurd. ☐

Corollary 9.8. *For every $n \ge 1$ and every $R \ge 3$ with $(R-1)^4 \ge 5n$ — in particular for every $R \ge \max\bigl(3,\, 1 + (5n)^{1/4}\bigr)$ — one has $f_n(R) \ge 1$.*

Proof. Let $\mathcal{E}$ and A be admissible in Definition 1.2, so that $\mathcal{E}(1) = R$ and A is row-substochastic. Theorem 9.7 at $i_0 = 1$ gives $\rho(\mathcal{E} - A) = \rho(A - \mathcal{E}) \ge 1$. Take the infimum. ☐

Since $(5n)^{1/4} < \sqrt{n}$ exactly when $n > 5$, Corollary 9.8 improves on Corollary 4.3 at every $n \ge 6$, agrees with it at $n = 5$, and is subsumed by Theorem 9.5 below that. Against the threshold (6) of [5] both are of a different order:

n	4	5	6	10	100
R_n of (6)	37.5	87	200.5	5134.5	$> 10^{31}$
$1 + \sqrt{n}$	3	3.236	3.449	4.162	11
$1 + (5n)^{1/4}$	3.115	3.236	3.340	3.659	5.729

At $n = 100$, for instance, Theorem 9.7 makes a learning rate of 6 enough, since $5 \cdot 100 = 500 \le 5^4 = 625$, where Theorem 4.2 asks for $1 + \sqrt{100} = 11$. As with (5), the comparison is not pointwise in R: for a fixed n and $R \to \infty$ the bound (5) eventually beats everything here, and the point is again the bounded values of R that Problem 1.1 is about.

Remark 9.9. Corollary 9.8 gives $f_n(R) \ge 1$ at the threshold, not $f_n(R) > 1$. The strict inequality holds above it, by the same chain run at a level $s > 1$: if $\rho(\mathcal{E} - A) < s$ then Corollary 9.4 becomes $\sum_i d_i^2 < ns^2$ and $\sum_i d_i^4 - 4\sum_i \max(d_i, 0) r_i < ns^4$ (Theorem 3.1 and Lemma 9.3 at that s), whence $\sigma < ns$ and, by the same factorisation, $ns^4 > (R-1)^4 - 4ns$. So $\rho(\mathcal{E} - A) \ge s$ whenever $(R-1)^4 \ge n(s^4 + 4s)$, and since $s \mapsto n(s^4 + 4s)$ is continuous with value $5n$ at $s = 1$, every $R > 1 + (5n)^{1/4}$ with $R \ge 3$ has $f_n(R) > 1$. (An elementary calculation, recorded here for orientation; the formal development carries the case $s = 1$.)

10. The extremal function at $n = 3$

Section 7 exhibited the pair (8) and reported $\rho(A - \mathcal{E}) = 1.8256\ldots$ as a computation. We now certify $\rho(A - \mathcal{E}) < 2$ for it. The obstacle is that the characteristic polynomial $\mu^3 + \frac{23}{5}\mu^2 + \frac{109}{16}\mu + \frac{509}{160}$ is irreducible over $\mathbb{Q}$ and its two dominant roots are a complex conjugate pair near a double root, so the device of Section 6 — factor over $\mathbb{Q}$, bound each factor — is unavailable. A quadratic Lyapunov form replaces it and needs no factorisation.

Proposition 10.1. *Let $n \ge 1$, let $s \ge 0$, let M be a real $n \times n$ matrix, and let $q : \mathbb{C}^n \to \mathbb{R}$ be positive off the origin and homogeneous of degree 2 under complex scaling: $q(v) > 0$ for every $v \ne 0$, and $q(\mu v) = |\mu|^2 q(v)$ for every $\mu \in \mathbb{C}$ and $v \in \mathbb{C}^n$. If $q(Mv) < s^2 q(v)$ for every $v \ne 0$, then every $\mu \in \operatorname{spec}(M)$ satisfies $|\mu| < s$.*

Proof. Let $\mu \in \operatorname{spec}(M)$. Then $\mu I - M$ is singular over $\mathbb{C}$, so $Mv = \mu v$ for some $v \ne 0$, and $|\mu|^2 q(v) = q(\mu v) = q(Mv) < s^2 q(v)$ with $q(v) > 0$; divide by $q(v)$. □

Proposition 10.1 is the easy half of the discrete-time Lyapunov, or Stein, criterion — $\rho(M) < 1$ if and only if $P - M^\top P M$ is positive definite for some positive definite P — with q the Hermitian form $v \mapsto \bar{v}^\top P v$ of such a P. The criterion is classical and no novelty is claimed for it, nor for the way the certificate below was produced. Stating it for an abstract q rather than for a matrix P is a convenience: positive definiteness then has to be checked only for the two forms that actually occur, and each of those is visibly a sum of squares.

Theorem 10.2. *Let A and $\mathcal{E}$ be the pair (8). Then every $\mu \in \operatorname{spec}(\mathcal{E} - A)$ satisfies $|\mu| < 2$. Since A is row-stochastic, $\mathcal{E}$ is diagonal and nonnegative and $\max_i \mathcal{E}(i) = \mathcal{E}(1) = 3$, the pair is admissible in Definition 1.2, so*

$$f_3(3) \; < \; 2 \; = \; (3-1)^+,$$

and the exact formula (4) of [5] does not extend to $n = 3$.

Proof. Write $M := A - \mathcal{E}$, so that $Mv = \left(-\frac{21}{10}v_1 + \frac{1}{10}v_2,\ -\frac{5}{4}v_2 + v_3,\ v_1 - \frac{5}{4}v_3\right)$, and let q be the Hermitian form of

$$P = \begin{pmatrix} 128 & -16 & 16 \\ -16 & 6 & -4 \\ 16 & -4 & 8 \end{pmatrix}, \qquad \text{that is,} \qquad q(v) = 2\left|8v_1 - v_2 + v_3\right|^2 + \left|2v_2 - v_3\right|^2 + 5\,|v_3|^2.$$

Then $q(\mu v) = |\mu|^2 q(v)$, and q is positive off the origin because its three linear forms are triangular in (v_1, v_2, v_3): if $q(v) = 0$ then $v_3 = 0$, then $v_2 = 0$, then $v_1 = 0$. One has the identity

$$4q(v) - q(Mv) = \tfrac{1}{4200}\left|168v_1 - 43v_2 + 60v_3\right|^2 + \tfrac{11}{5712}\left|68v_2 + 15v_3\right|^2 + \tfrac{601}{272}|v_3|^2$$

for every $v \in \mathbb{C}^3$ — a polynomial identity in the six real coordinates of v — and its right-hand side is positive off the origin for the same triangularity reason. So $q(Mv) < 4q(v)$ for every $v \ne 0$, and Proposition 10.1 with $s = 2$ gives $|\mu| < 2$ for every $\mu \in \operatorname{spec}(M)$. Finally $(-M)^\top P(-M) = M^\top P M$, so the same P serves for $\mathcal{E} - A = -M$, which is the matrix Definition 1.2 uses. □

The matrix P was found by solving the discrete Lyapunov equation $4P - M^\top PM = 4I$ exactly over $\mathbb{Q}$ and then searching for a nearby integer matrix; the two sums of squares come from the $LDL^\top$ factorisations of P and of $4P - M^\top PM$. None of that search is part of the proof. What the proof uses is the displayed identity, which is an identity between polynomials with rational coefficients and is verified as such.

Proposition 10.3. *For the same pair, $\rho(A - \mathcal{E}) \geq \frac{6}{5}$.*

Proof. Here $d_1 = 3 - \frac{9}{10} = \frac{21}{10}$ and $3 \cdot \left(\frac{6}{5}\right)^2 = \frac{108}{25} \leq \frac{441}{100} = d_1^2$, so the single-agent form of Theorem 3.1 applies with $s = \frac{6}{5}$. $\square$

So (8) is not a counterexample to Problem 1.1: it is an admissible pair with a fast learner whose spectral radius is comfortably above 1, and what it refutes is the formula, not the problem. Combining Corollary 4.3 with Theorem 10.2,

$$\tfrac{2}{\sqrt{3}} = 1.1547\ldots \ \leq\ f_3(3)\ <\ 2,$$

and the measured value 1.8078 of Section 7 lies in that interval. The left end is Corollary 4.3 at $n = 3$, $R = 3$ and holds for every admissible pair; the larger constant 6/5 of Proposition 10.3 is a bound for the single pair (8) and says nothing about the infimum.

11. The ceiling of the two moments

Theorem 9.5 uses exactly two facts about the spectrum, the two constraints of Corollary 9.4. The following instance shows that those two cannot reach $n = 7$, whatever is done with them.

Theorem 11.1. *Let $n = 7$, let $\mathcal{E} = \operatorname{diag}\big(3, \frac{7}{10}, \frac{7}{10}, \frac{7}{10}, \frac{7}{10}, \frac{7}{10}, \frac{7}{10}\big)$, and let A be the matrix with*

$$A(1,1) = 1, \qquad A(2,3) = A(3,4) = A(4,2) = 1, \qquad A(5,6) = A(6,7) = A(7,5) = 1$$

and every other entry 0. Then A is row-stochastic, $\mathcal{E}$ is nonnegative diagonal with $\mathcal{E}(1) = 3$, so the pair satisfies the hypotheses of Problem 1.1; and

$$\operatorname{tr}\big((A-\mathcal{E})^2\big) = \frac{347}{50} = 6.94, \qquad \operatorname{tr}\big((A-\mathcal{E})^4\big) = \frac{3203}{5000} = 0.6406,$$

both below $n = 7$, while $\rho(A - \mathcal{E}) \geq 2$.

Proof. Row 1 of A is e_1 and each other row is a standard basis vector, so A is row-stochastic. Here $d_1 = 3 - 1 = 2$ and $d_i = \frac{7}{10}$ for $i \geq 2$, and the influence graph is a loop at agent 1 together with two vertex-disjoint directed triangles, on $\{2,3,4\}$ and on $\{5,6,7\}$. It has no 2-cycle, so the off-diagonal contributions to $\operatorname{tr}(Z^2)$ all vanish and $\operatorname{tr}(Z^2) = \sum_i d_i^2 = 4 + 6 \cdot \frac{49}{100} = \frac{347}{50}$. For the fourth power, use Lemma 9.1: the absence of 2-cycles kills the two-F block, the absence of a closed four-walk kills $\operatorname{tr}(F^4)$, and $(F^3)(i,i)$ is 1 at each of the six triangle vertices and 0 at agent 1, so

$$\operatorname{tr}(Z^4) = \sum_i d_i^4 - 4\sum_i d_i\,(F^3)(i,i) = 16 + 6 \cdot \tfrac{2401}{10000} - 4 \cdot 6 \cdot \tfrac{7}{10} = \tfrac{3203}{5000}.$$

Both traces are those of $(A-\mathcal{E})^2$ and $(A-\mathcal{E})^4$, the powers being even. Finally agent 1 is absorbing, so the first row of $\mu I - (A - \mathcal{E})$ vanishes identically at $\mu = -2$; hence $-2 \in \operatorname{spec}(A - \mathcal{E})$ and $\rho(A-\mathcal{E}) \geq 2$. $\square$

So no inequality of the form $|\operatorname{tr}((A-\mathcal{E})^m)| \geq n$ with $m \in \{2,4\}$ can be derived from the hypotheses of Problem 1.1 at $n = 7$: here is a pair that satisfies them and violates both. The instance is not a counterexample to the problem — its spectral radius is at least 2 — so what fails is the method and not the expected answer. Padding it with further agents that learn at rate 0 and put all their averaging weight on agent 1 leaves every trace unchanged, since no edge enters a padding agent, so the obstruction persists at every $n \geq 7$; that padding step is on paper and is not part of the formal development.

Two vertex-disjoint triangles are the mechanism. In Lemma 9.1 the only term that can be negative is $-4\operatorname{tr}(DF^3)$, which counts the closed three-walks through each agent weighted by that agent's excess rate, and a directed triangle of weight 1 is what makes it large; $n = 7$ is the first dimension in which two vertex-disjoint triangles fit alongside a fast learner. Nor is the estimate of Proposition 9.2 to blame:

Proposition 11.2. *On the instance of Theorem 11.1 the estimate of Proposition 9.2 is an equality:*

$$\sum_{i=1}^{7} d_i^4 - 4\sum_{i=1}^{7} \max(d_i,0)\, r_i \;=\; \operatorname{tr}\big((A-\mathcal{E})^4\big) \;=\; \frac{3203}{5000}.$$

Proof. Every term the estimate discards vanishes here, as in the proof of Theorem 11.1; all the d_i are positive, so $\max(d_i, 0) = d_i$; and $(F^3)(i,i) = r_i$ for every agent, both sides being 1 at a triangle vertex and 0 at agent 1, whose outgoing weight is $r_1 = 0$. □

The distance between the $n \leq 5$ of Theorem 9.5 and the $n = 7$ of Theorem 11.1 is therefore not slack in the estimates. One dimension is left in the method, $n = 6$, and we have not settled it.

Remark 11.3. What $n = 6$ would need. Run the proof of Theorem 9.5 at $n = 6$ and the step that loses is the Cauchy–Schwarz bound $\sigma \leq \sqrt{(n-1)s}$: it allows all five remaining agents to sit on weight-1 triangles at once, which five vertices cannot do. What would replace it is a bound on the total triangle mass: for F nonnegative with zero diagonal and row sums at most 1, is $\operatorname{tr}(F^3) \leq 3\lfloor n/3 \rfloor$? A hill-climbing search over such F attains $3, 3, 3, 6, 6, 6$ at $n = 3, \ldots, 8$ and finds nothing larger; the value $3\lfloor n/3\rfloor$ is attained by $\lfloor n/3\rfloor$ vertex-disjoint directed triangles of weight 1, and for $n \leq 7$ an exhaustive enumeration of those F whose rows are zero or a standard basis vector returns the same six numbers. All of that is computation and not proof, and none of it is part of the formal development. The natural linear relaxation cannot supply the bound. Give each directed triangle t — a cyclic class (i, j, k) of distinct agents — the weight $w_t = F(i,j)F(j,k)F(k,i)$, so that $\operatorname{tr}(F^3) = 3\sum_t w_t$; the only linear facts available are $\sum_{t\ni i} w_t = (F^3)(i,i) \leq r_i \leq 1$ for each agent i, and maximising $\sum_t w_t$ under those n constraints is fractional triangle matching on the complete graph, whose optimum is exactly $n/3$: summing the n constraints gives $3\sum_t w_t \leq n$, and the assignment giving every triangle the same weight attains it. That yields only $\operatorname{tr}(F^3) \leq n$, which is the crude bound of Section 8 again. The floor has to come from the multiplicative structure, a vertex's outgoing weight being a budget shared by every triangle through it; we have no proof of it, and nothing in this paper depends on one.

12. Verification

Every theorem and proposition stated above, together with Lemmas 2.1, 9.1 and 9.3 and Corollary 9.4, has been stated and proved in Lean 4 against Mathlib, and the proofs have been accepted by the Lean kernel. Outside the formal development are Remark 1.3, Section 7, Section 8, Remark 9.9, the padding argument after Theorem 11.1 and Remark 11.3; each is marked as a computation or as commentary where it stands, and none of them is used to prove anything else. Three statements are recorded in the equivalent form in which they are used rather than verbatim: Corollaries 4.3 and 9.8, because the development does not define the extremal function f_n and carries instead Theorem 3.1 at $s = (R-1)/\sqrt{n}$ and Theorem 9.7, from which the corollaries are read off; and Lemma 5.1, which appears as the singularity statement — $\det(M_\iota) = 0$ forces $\det(M) = 0$ for an out-closed ι — together with the eigenvalue transfer that statement yields. The development is six files, about 1600 lines: the second-moment bound and its consequences, the two explicit 3×3 instances of Section 6, the restriction to an out-closed set, the fourth-moment machinery with Theorems 9.5 and 9.7, the Lyapunov certificate with the $f_3(3)$ witness, and the 7×7 instance. Since Mathlib's spectral radius is the Banach-algebra one and over $\mathbb{R}$ would see only the real spectrum, the statements are phrased over the complexification, in the two dual forms "some eigenvalue has modulus at least s" and "every eigenvalue has modulus below s", which are proved to be negations of one another. The proofs are symbolic throughout: nothing is evaluated by compiled reflection, no proof relies on a decision procedure that computes, and the development contains no occurrence of Lean's compiled-evaluation tactic `native_decide`. So the explicit instances of Sections 6, 10 and 11 are exact rational algebra rather than machine arithmetic, and the searches that produced them — the minimisation of Section 7, the rounding search behind the matrix P of Theorem 10.2, and the structured search that produced the 7×7 instance — lie outside the proofs entirely: each instance is verified from its definition, and nothing would change if the searches had found it by luck. An axiom check covering every declaration of all six modules — 126 declarations as they are written, 170 constants once the auxiliaries the compiler generates from them are counted —

returns for each of them exactly the three axioms of ordinary classical mathematics in Lean (`propext`, `Classical.choice`, `Quot.sound`), and in particular no incomplete proof and no compiled-evaluation axiom. Repository reference to be added at submission.

Acknowledgements

The problem, the extremal function f_n, the counterexample (3) and the constant 0.92 attached to it are all due to Popescu, Syatriadi and Vaidya [5], and the averaging-plus-learning model to Popescu and Vaidya [4]; this paper answers a question they asked, and the question is the harder half.

References

[1] Ludwig Elsner. *An optimal bound for the spectral variation of two matrices*. Linear Algebra and its Applications **71** (1985), 77–80; `doi:10.1016/0024-3795(85)90236-8`.

[2] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*, second edition. Cambridge University Press, Cambridge, 2013 (first edition 1985). The theorem number is that of the second edition, where Theorem 5.6.12 reads: "Let $A \in M_n$. Then $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$."

[3] Alexander Ostrowski. *Mathematische Miszellen. XXVII. Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen*. Jahresbericht der Deutschen Mathematiker-Vereinigung **60** (1957), 40–42.

[4] Ionel Popescu and Tushar Vaidya. *Averaging plus learning models and their asymptotics*. Proceedings of the Royal Society A **479** (2023), no. 2275, Paper No. 20220681; `doi:10.1098/rspa.2022.0681`.

[5] Ionel Popescu, Jeven Syatriadi and Tushar Vaidya. *Anchoring and Mixed-Norm Contractions in Averaging-Learning Dynamics*. arXiv:2602.22627v2 (28 February 2026). Preprint; the arXiv listing showed no journal reference and no later version as of 2026-09-03. Every problem, theorem, definition, proposition and remark number cited here is that of v2.

[6] Tushar Vaidya, Carlos Murguia and Georgios Piliouras. *Learning agents in Black–Scholes financial markets*. Royal Society Open Science **7** (2020), no. 10, article 201188; `doi:10.1098/rsos.201188`. Preprint: arXiv:1704.07597, whose current version v4 (11 July 2020) already carries the published title; the arXiv listing still records the title of version 1, *Learning Agents in Black-Scholes Financial Markets: Consensus Dynamics and Volatility Smiles*, and no journal reference.