THE CONSTANT $\eta_{2,K} = 1$ IS IMPOSSIBLE: AN EXACT COUNTEREXAMPLE TO THE SINGLE-SHUFFLE INEQUALITY AT $n = 2,\ K = 2,\ d = 4$

KOREA SUPERINTELLIGENCE LABS

Abstract. Yun, Sra and Jadbabaie compare the expected iterate of stochastic gradient descent on a quadratic finite sum under single shuffling, random reshuffling and plain gradient descent, and conjecture that for every number of components n, every epoch count K and every dimension d there is a step-size constant $\eta_{n,K} \in (0, 1]$ for which $(1 - \eta_{n,K})I \preceq A_i \preceq I$ forces $\|W_{\mathrm{ss}}\| \le \|W_{\mathrm{rs}}\| \le \|W_{\mathrm{gd}}\|$. About the first component of their conjecture at $n = 2$ they write: *"Nevertheless, for $n = 2$, we believe that the conjecture itself is likely true for any d and K, with step-size constants $\eta_{2,K} = 1$."* We refute that belief, with two explicit witnesses. For the symmetric positive definite *integer* matrices $A = \mathrm{diag}(1, 10^2, 10^4, 10^6)$ and an explicit B with entries at most 10^7, at $K = 2$ and $d = 4$, we have $\|W_{\mathrm{ss}}\| > \rho^2 \ge \|W_{\mathrm{rs}}\|$ with $\rho = 106130000$; the true ratio is $\|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\| = 1.035608302215\ldots$. Rescaled into the conjecture's own normalisation the pair satisfies $\frac{1}{84375000} I \preceq A', B' \preceq I$ and still violates the inequality. A second, far better conditioned pair — $A = \mathrm{diag}(1, 7^2, 62^2, 624^2)$ and an explicit B with entries of at most ten digits — does the same at $\frac{1}{389765} I \preceq A', B' \preceq I$, a factor 216.5 sharper, with true ratio $1.048908816189\ldots$; and we prove that *no* renormalisation of that pair can reach past $\frac{1}{389376}$, so its bookkeeping is within 0.1% of everything it has to give and what remains is a deficiency of the witness rather than of the accounting. Hence no admissible constant reaches $1 - 2.5656\ldots \times 10^{-6}$, while $\eta = \frac{1}{4}$ is admissible, by the authors' own Theorem 3 together with the $n = 2$ case of the Recht–Ré inequality. The conjecture itself survives at $n = 2$; what dies is the value of the constant, and a stated belief becomes a quantitative question whose answer, the supremum $\eta^*_{2,2}$ of the admissible constants, is bracketed by $\frac{1}{4} \le \eta^*_{2,2} \le 1 - 2.5656\ldots \times 10^{-6}$ — about five and a half orders of magnitude in $1 - \eta$, and no better. Each certificate is four denominator-cleared $LDL^\top$ identities and two Rayleigh quotients, all in exact integer arithmetic, so the refutations are checked by kernel reduction alone, with no compiled evaluation and no floating point anywhere in the proofs. We also record, with credit, the two halves of the $n = 2$ picture that do hold — the reshuffling-versus-gradient-descent inequality for all positive semidefinite A, B at $\eta = 1$, and the single-shuffle inequality at $d = 2$ for every K — and we explain why the counterexample is invisible to the searches that look for it: the violating eigenvalue is the *negative* one, and the largest eigenvalue can never violate the inequality. Every statement below is machine-checked in Lean 4.

1. Introduction

1.1. Three ways to sweep a finite sum. Minimising a finite sum $\frac{1}{n}\sum_{i=1}^n f_i$ by stochastic gradient descent requires a decision that the objective does not make for you: in what order to visit the components. Three orders are standard. *Single shuffling* draws one permutation σ of $[n]$ once and reuses it in every one of the K epochs; *random reshuffling* draws a fresh permutation each epoch; *gradient descent* uses the full gradient, which is the same as visiting the components in any order with the iterate held fixed within an epoch.

On a quadratic finite sum with $f_i(x) = \frac{1}{2}x^\top M_i x - b_i^\top x$ and step size α, one epoch of a scheme multiplies the error of the iterate by a product of the matrices $A_i = I - \alpha M_i$ taken in the visiting order, so comparing the three schemes over K epochs is comparing three products. Writing $\mathbb{S}^d_+$ for the real symmetric positive semidefinite $d \times d$ matrices, $\mathbb{S}^d_{++}$ for the positive definite ones, $\|\cdot\|$ for the spectral norm, $\preceq$ for the Löwner order, and

$$P_\sigma \;=\; A_{\sigma(1)} A_{\sigma(2)} \cdots A_{\sigma(n)},$$

the three matrices of the comparison are

$$W_{\mathrm{ss}} = \frac{1}{n!}\sum_{\sigma\in S_n} P_\sigma^K, \qquad W_{\mathrm{rs}} = \Big(\frac{1}{n!}\sum_{\sigma\in S_n} P_\sigma\Big)^K, \qquad W_{\mathrm{gd}} = \Big(\frac{1}{n}\sum_{i=1}^n A_i\Big)^{nK}. \tag{1}$$

Yun, Sra and Jadbabaie [1] pose the following as a COLT 2021 open problem. We quote it verbatim from their preprint.

> **Conjecture 1.** "For any $n \geq 2$, $K \geq 1$, and $d \geq 1$, there exists a step-size constant $\eta_{n,K} \in (0, 1]$ such that the following statement holds: Suppose $A_1, \ldots, A_n \in \mathbb{S}_+^d$ satisfy $(1 - \eta_{n,K})I \preceq A_i \preceq I$ for all $i \in [n]$. Then, for the three matrices defined as [(1)] the following spectral norm inequalities are satisfied: $\|W_{\mathrm{ss}}\| \leq \|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$."

In words: single-shuffle SGD beats reshuffling SGD, which beats gradient descent. The second inequality, $\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$, is the noncommutative arithmetic–geometric mean inequality of Recht and Ré [3] with a conditioning hypothesis added. Without that hypothesis the Recht–Ré statement holds for $n = 2$ [3] and for $n = 3$ [4, 6] — that attribution is the one [1] themselves give — and it is false at $n = 5$, by the nonconstructive counterexample of Lai and Lim [6]. Two neighbouring statements have their own histories: the symmetrised version was disproved earlier [7], and the expectation-of-norm version, with the norms inside the two averages rather than outside them, is known for products of up to three matrices [5]. The first inequality, $\|W_{\mathrm{ss}}\| \leq \|W_{\mathrm{rs}}\|$, is new with [1].

Binghui Peng [2] moves both components: his Theorem 2 refutes the first inequality at $n = 3$, $K = 2$, $d = 4$ *for every* η, and his Theorem 3 proves the second for every $n \geq 2$ under $(1 - \frac{1}{4n^2+1})I \preceq A_i \preceq I$. He does not treat the single-shuffle inequality at $n = 2$, which is what this paper is about.

1.2. **The case $n = 2$, and a belief.** At $n = 2$ the symmetric group is $\{\mathrm{id}, (12)\}$, so with $A, B \in \mathbb{S}_+^d$ and $N := AB$ — note $BA = N^\top$, since A and B are symmetric — the three matrices of (1) are

$$W_{\mathrm{ss}} = \mathrm{sym}(N^K), \qquad W_{\mathrm{rs}} = (\mathrm{sym}\, N)^K, \qquad W_{\mathrm{gd}} = \big(\tfrac{1}{2}(A + B)\big)^{2K}, \tag{2}$$

where $\mathrm{sym}\, X = \frac{1}{2}(X + X^\top)$. Since $\mathrm{sym}\, N$ is symmetric, $\|(\mathrm{sym}\, N)^K\| = \|\mathrm{sym}\, N\|^K$, and the first inequality of Conjecture 1 at $n = 2$ becomes a *power inequality for the real numerical radius* $w_{\mathbb{R}}(X) := \|\mathrm{sym}\, X\|$:

$$\textbf{(Q)} \qquad w_{\mathbb{R}}(N^K) \;\leq\; w_{\mathbb{R}}(N)^K \quad \text{for } N = AB,\ A, B \in \mathbb{S}_+^d? \tag{3}$$

Both sides of $\|W_{\mathrm{ss}}\| \leq \|W_{\mathrm{rs}}\|$ are homogeneous of the same degree under $A \mapsto sA$ and under $B \mapsto tB$ separately (Lemma 2.1), so the normalisation $A_i \preceq I$ costs nothing and the hypothesis $(1-\eta)I \preceq A_i \preceq I$ is exactly a cap $1/(1 - \eta)$ on the condition numbers of A and B. This is what the authors themselves say: "the actual requirement on A_i's for the first inequality is a restriction that they are PD with condition numbers at most $1/(1 - \eta_{n,K})$" [1]. So $\eta = 1$ at $n = 2$ *is* question (Q) for arbitrary positive semidefinite A and B.

What was known at $n = 2$ before this paper is the following, all of it from [1] except the third line.

regime	statement	source
$d = 2$, all K, $\eta = 1$	$\frac{1}{2}\|(AB)^K + (BA)^K\| \leq \|\frac{1}{2}(AB + BA)\|^K$, $A, B \in \mathbb{S}_+^2$	[1], Thm. 4
all d, $K = 2^m$, $\eta = \frac{1}{2K}$	the same, under $(1 - \frac{1}{2K})I \preceq A, B \preceq I$, $A, B \in \mathbb{S}_{++}^d$	[1], Thm. 3
$\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$, all d, K, $\eta = 1$	$\|\frac{1}{2}(AB + BA)\| \leq \|\frac{1}{2}(A + B)\|^2$, $A, B \in \mathbb{S}_+^d$	[3] at $n = 2$

Immediately after their Theorem 4 the authors write, verbatim — with their two internal cross-references replaced by the numbers they carry in `2103.07079v1`:

> "The restriction of Theorem 4 to $d = 2$ is an artifact of our proof technique that does not extend to $d \geq 3$ (see Section 5.3 for more details). Nevertheless, for $n = 2$, we believe that the conjecture itself is likely true for any d and K, with step-size constants $\eta_{2,K} = 1$."

1.3. **Results.** That belief is false, already at $K = 2$ and $d = 4$, and it is false with an exact integer witness.

- **The refutation** (Theorem 3.1). For $A = \operatorname{diag}(1, 10^2, 10^4, 10^6)$ and the explicit symmetric positive definite integer matrix B of (5), at $K = 2$ and $d = 4$,
$$\|W_{\mathrm{ss}}\| \ > \ \rho^2 = 11263576900000000 \ \geq \ \|W_{\mathrm{rs}}\|, \qquad \rho = 106130000.$$
By Lemma 2.1 this pair may be rescaled into the normalisation $0 \preceq A, B \preceq I$ of Conjecture 1 without changing either side, so the value $\eta_{2,2} = 1$ does not work there.
- **The refutation in the conjecture's own normalisation** (Theorem 4.1). With $A' = 10^{-6}A$ and $B' = 10125000^{-1}B$ one has $\frac{1}{84375000}I \preceq A', B' \preceq I$ — which is the hypothesis of Conjecture 1 at $1 - \eta = \frac{1}{84375000}$ — and the same violation. No hand rescaling is left to the reader.
- **A second, better conditioned witness** (Theorem 4.3). Take $A = \operatorname{diag}(1, 7^2, 62^2, 624^2)$ and the explicit integer matrix B of (6), again at $K = 2$ and $d = 4$. The rescaled pair $A' = A/389376$ and $B' = B/5220000000$ satisfies
$$\tfrac{1}{389765}I \preceq A', B' \preceq I$$
and violates $\|W_{\mathrm{ss}}\| \leq \|W_{\mathrm{rs}}\|$; so Conjecture 1 fails at $n = 2$, $K = 2$, $d = 4$ for every $\eta > 1 - \frac{1}{389765}$, a factor 216.5 sharper. That the improvement is strict, that is the inequality $\frac{1}{84375000} < \frac{1}{389765}$, is itself part of the formal development.
- **The second witness is nearly spent** (Proposition 4.5). Its condition numbers are $\operatorname{cond}(A) = 389376$ exactly and $\operatorname{cond}(B) \leq 389765$, by four more integer certificates; and conversely, every two-sided Löwner sandwich $mI \preceq A \preceq MI$ has $M \geq 389376\,m$ and every $mI \preceq B \preceq MI$ has $M \geq 380791\,m$. Consequently *any* normalisation $(1 - \eta)I \preceq A/s_A,\ B/s_B \preceq I$ of this pair has $1 - \eta \leq \frac{1}{389376}$, and the conditioning cap 389765 actually used is only 0.0999% above the smallest one the pair admits. Whatever is still missing is therefore missing from the witness, not from its bookkeeping.
- **A bracket for the constant** (Proposition 4.8). Consequently every admissible η at $n = 2$, $K = 2$ satisfies
$$\eta \ < \ 1 - \tfrac{1}{389765} \ = \ 1 - 2.5656485\ldots \times 10^{-6},$$
while $\eta = \frac{1}{4}$ is admissible, by Theorem 3 of [1] at $K = 2$ together with Theorem 6.1 below. In terms of $1 - \eta$ the bracket runs from $\frac{3}{4}$ down to 2.566×10^{-6}, that is over about five and a half orders of magnitude, and closing it is the question this paper leaves open.
- **The two halves that hold**, recorded with credit and reproved here. Theorem 6.1: for *all* $A, B \in \mathbb{S}^d_+$ and all $c \geq 0$ with $A + B \preceq 2cI$ one has $-c^2 I \preceq \frac{1}{2}(AB + BA) \preceq c^2 I$, hence $\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$ for every K and d with no conditioning hypothesis; this is Recht and Ré at $n = 2$ [3], cited for exactly this purpose in [1]. Theorem 6.3: at $d = 2$, $w_{\mathbb{R}}(N^K) \leq w_{\mathbb{R}}(N)^K$ for every K and every real 2×2 matrix N with real spectrum; this is Theorem 4 of [1] with the hypothesis $A, B \in \mathbb{S}^2_+$ weakened to a spectral one, and their own route already proves it.
- **Why the witness is hard to find** (Section 5). The violating eigenvalue of W_{ss} is the negative one. The largest eigenvalue can never violate the inequality — $\lambda_{\max}(\operatorname{sym}(N^2)) \leq \|\operatorname{sym} N\|^2$ holds for every square N by a one-line identity — so any search that measures $\lambda_{\max}$ is provably blind to this phenomenon, and a search that maximises the ratio $\|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\|$ sits on a plateau at exactly 1. This is not a metaphor: random sampling and direct maximisation of the ratio both return 1.0 at $d = 4$, where a violation exists.

1.4. **What is not claimed.** Four limits, stated at the start rather than left to be inferred.

First, and most importantly, **Conjecture 1 survives at $n = 2$.** It asks only for the *existence* of some $\eta_{2,K} \in (0, 1]$, and the better of the two witnesses here still needs condition numbers 3.89×10^5 and 3.81×10^5 (the first needs 10^6 and 8.35×10^7); it therefore says nothing against the conjecture as stated. What it refutes is the value $\eta_{2,K} = 1$, which the authors offer as a belief and not as a theorem, and it does so courteously: their two $n = 2$ theorems are untouched, and so is the rest of their paper. It does refute half of the sentence that precedes the belief — the restriction to $d = 2$ in their Theorem 4 is not only an artifact of the proof technique, since at $\eta = 1$ the statement is genuinely false from $d = 4$ on.

Second, nothing here bears on $n \geq 3$. Peng's Theorem 2 (his refutation, at $n = 3$) and his Theorem 3 (the reshuffling-versus-gradient-descent inequality, for every $n \geq 2$) [2] are untouched; the only remark we make about [2] is that the constant $\frac{1}{4n^2+1}$ of its Theorem 3 is not tight at $n = 2$, where by Theorem 6.1 the truth is $\eta = 1$.

Third, Theorems 6.1 and 6.3 are **not new results**. The mathematics of the first is Recht and Ré's [3] and the mathematics of the second is Yun, Sra and Jadbabaie's [1]; what is ours in both is a proof written so that a proof assistant can carry it, which in the second case means avoiding the power inequality for the numerical radius entirely. They appear here because a paper that refutes one half of a picture should say what the other half is.

Fourth, on provenance. Both preprints were re-fetched live from the arXiv API on 3 September 2026, and both are still at v1 with an empty journal-reference field: `2103.07079v1` (submitted 12 March 2021, primary `cs.LG`) and `2607.22620v1` (submitted 16 June 2026, primary `math.OC`). The first of the two does have a conference version, recorded in the bibliography. Every quotation above and below was read out of the LaTeX e-print of [1], not from a secondary source. A forward-citation sweep of [1] on 3 September 2026 returned twelve citing works, of which exactly one responds to the open problem: [9], which proves a weak arithmetic–geometric mean inequality for well-conditioned positive semidefinite matrices and general n — the second inequality of Conjecture 1, not the first, and with no counterexample. A full pass over a local corpus of 369 shards of arXiv source text for the phrases of this problem was completed for this version: 29 of the shards contain the phrase "single-shuffle", in 25 distinct contexts, of which all but one belong to card-based cryptography, to the shuffle model of differential privacy, or to the convergence literature for without-replacement sampling; exactly one file in that corpus states the SS–RS–GD conjecture, and it is [1] itself, [2] sitting in a recent-announcement shard swept separately. A search inside those files for a bound on $\eta_{2,K}$ in either direction returns nothing. A neighbourhood sweep for the noncommutative arithmetic–geometric mean inequality returns [3, 5, 7] and one further paper, [8], which is again on the reshuffling-versus-gradient-descent side of Conjecture 1 and says nothing about single shuffling. So the refutations below are reported as not found in the literature, which is not the same as reporting them as new. The reference list records, for each entry, how far the reading went. Every theorem and section number attributed below to [1] or to [2] was resolved by compiling that paper's own e-print, and is the numbering of the version cited.

2. The objects

2.1. Symmetric and skew parts. For a real $d \times d$ matrix X put

$$\operatorname{sym} X = \tfrac{1}{2}(X + X^\top), \qquad \operatorname{skew} X = \tfrac{1}{2}(X - X^\top),$$

and define the *real numerical radius* $w_{\mathbb{R}}(X) = \|\operatorname{sym} X\|$. For $A, B \in \mathbb{S}^d_+$ and $N = AB$ the identities (2) hold, so question (Q) of (3) is the $n = 2$ case of the first inequality of Conjecture 1 at $\eta = 1$. Recall also that $N = AB$ is similar to $A^{1/2}BA^{1/2} \succeq 0$ when A is positive definite, so such an N has real nonnegative spectrum; this is the only property of the pair (A, B) that Section 6.2 uses.

2.2. Two-sided Löwner bounds. Every statement below is phrased in the Löwner order rather than in a norm. For $c \in \mathbb{R}$ and a real $d \times d$ matrix M write

$$\mathrm{LB}(c, M) \quad :\Longleftrightarrow \quad cI - M \succeq 0 \;\text{ and }\; cI + M \succeq 0,$$

that is, $-cI \preceq M \preceq cI$. For *symmetric* M this says exactly that every eigenvalue of M lies in $[-c, c]$, i.e. that $\|M\| \leq c$, and $\|M\|$ is the least such c. So a statement of the form $\|X\| > \|Y\|$ is written here as: $\mathrm{LB}(c, Y)$ holds for an explicit c, and every c' with $\mathrm{LB}(c', X)$ satisfies $c < c'$. Nothing is lost, and the elementary half of the dictionary,

$$\mathrm{LB}(c, M) \implies |x^\top M x| \leq c\,(x^\top x) \quad \text{for all } x \in \mathbb{R}^d, \tag{4}$$

is what all the certificates below actually use. The matrices W_{ss} and W_{rs} of the counterexample are verified to be symmetric, so for them the Löwner statements really are statements about spectral norms.

2.3. **Two lemmas.** The first lemma is the scale invariance quoted in the introduction; it is the authors' own remark that "scaling a single matrix A_i does not change the inequality $\|W_{\mathrm{ss}}\| \le \|W_{\mathrm{rs}}\|$" [1], in the form the later proofs need.

Lemma 2.1 (scale invariance)**.** *For all real $d \times d$ matrices A, B, all $s, t \in \mathbb{R}$ and all $K \in \mathbb{N}$,*

$$\tfrac{1}{2}\big(((sA)(tB))^K + ((tB)(sA))^K\big) = (st)^K \cdot \tfrac{1}{2}\big((AB)^K + (BA)^K\big),$$

$$\big(\tfrac{1}{2}((sA)(tB) + (tB)(sA))\big)^K = (st)^K \cdot \big(\tfrac{1}{2}(AB + BA)\big)^K.$$

Both sides of $\|W_{\mathrm{ss}}\| \le \|W_{\mathrm{rs}}\|$ therefore carry the same factor $|st|^K$, so a counterexample may be rescaled at will — in particular into the normalisation $0 \preceq A, B \preceq I$ demanded by Conjecture 1.

The second lemma is the workhorse. It says that a two-sided Löwner bound squares, and it is proved without a matrix square root, which is what makes it cheap to verify formally.

Lemma 2.2 (the squaring step)**.** *Let $c \ge 0$ and let D be a real symmetric $d \times d$ matrix with $-cI \preceq D \preceq cI$. Then $D^2 \preceq c^2 I$.*

Proof. Put $P = cI - D$ and $Q = cI + D$, both positive semidefinite by hypothesis and both symmetric. Since P and Q commute with cI one has $QP = PQ = c^2I - D^2$ and $Q + P = 2cI$, whence the identity

$$Q^\top PQ + P^\top QP \;=\; (QP)Q + (PQ)P \;=\; (c^2I - D^2)(Q + P) \;=\; 2c\,(c^2I - D^2).$$

Both summands on the left are positive semidefinite, being congruences of the positive semidefinite P and Q. For $c > 0$ divide by $2c$. For $c = 0$ the hypothesis forces $D = 0$ and the claim is trivial. ☐

No continuous functional calculus and no square root occur here; the only fact used is that $X^\top PX \succeq 0$ whenever $P \succeq 0$.

3. The counterexample

3.1. **The data.** Let

$$A = \begin{pmatrix} 1 & & & \\ & 10^2 & & \\ & & 10^4 & \\ & & & 10^6 \end{pmatrix}, \qquad B = \begin{pmatrix} 10000000 & 1100000 & 10000 & 0 \\ 1100000 & 200000 & -6000 & 100 \\ 10000 & -6000 & 2100 & 110 \\ 0 & 100 & 110 & 10 \end{pmatrix}, \tag{5}$$

both symmetric with integer entries, and take $K = 2$, $d = 4$. Then $AB + BA$ and $(AB)^2 + (BA)^2$ both have even entries, so

$$S := \tfrac{1}{2}(AB + BA) \qquad \text{and} \qquad W_{\mathrm{ss}} = \tfrac{1}{2}\big((AB)^2 + (BA)^2\big)$$

are again integer matrices; explicitly

$$S = 10^3 \begin{pmatrix} 10000 & 55550 & 50005 & 0 \\ 55550 & 20000 & -30300 & 50005 \\ 50005 & -30300 & 21000 & 55550 \\ 0 & 50005 & 55550 & 10000 \end{pmatrix},$$

$$W_{\mathrm{ss}} = 10^9 \begin{pmatrix} 222000 & 1636200 & -1750175 & 11000011 \\ 1636200 & 558000 & -1131200 & -1800180 \\ -1750175 & -1131200 & 599000 & 1691750 \\ 11000011 & -1800180 & 1691750 & 222000 \end{pmatrix},$$

and $W_{\mathrm{rs}} = (\mathrm{sym}(AB))^2 = S^2$. Finally put

$$\rho = 106130000, \qquad \rho^2 = 11263576900000000, \qquad v = (-4, 1, -1, 4)^\top.$$

Theorem 3.1 (the belief $\eta_{2,K} = 1$ is false)**.** *With A, B, ρ as above, A and B are symmetric positive definite,* $\mathrm{LB}(\rho^2, W_{\mathrm{rs}})$ *holds, and every c with* $\mathrm{LB}(c, W_{\mathrm{ss}})$ *satisfies $c > \rho^2$. Since W_{ss} and W_{rs} are symmetric, this says*

$$\|W_{\mathrm{ss}}\| \;>\; \rho^2 \;\geq\; \|W_{\mathrm{rs}}\|.$$

Equivalently, in Rayleigh-quotient form: $|x^\top W_{\mathrm{rs}} x| \leq \rho^2(x^\top x)$ for every $x \in \mathbb{R}^4$, while $\rho^2(v^\top v) < |v^\top W_{\mathrm{ss}} v|$ for the integer vector $v = (-4, 1, -1, 4)^\top$. Rescaling the pair into the normalisation of Conjecture 1 by Lemma 2.1 — Theorem 4.1 does this explicitly — the first inequality of Conjecture 1 therefore fails at $n = 2$, $K = 2$, $d = 4$ for the step-size constant $\eta_{2,2} = 1$.

Proof sketch. The argument is four exact statements over $\mathbb{Z}$ and one application of Lemma 2.2.

(1) $A \succeq 0$ and $B \succeq 0$.
(2) $-\rho I \preceq S \preceq \rho I$.
(3) Hence $W_{\mathrm{rs}} = S^2 \preceq \rho^2 I$ by Lemma 2.2 applied to $D = S$, and $W_{\mathrm{rs}} \succeq 0$ because $W_{\mathrm{rs}} = S^\top S$; so $\mathrm{LB}(\rho^2, W_{\mathrm{rs}})$.
(4) $v^\top W_{\mathrm{ss}} v = -396503392000000000$ and $v^\top v = 34$, so

$$|v^\top W_{\mathrm{ss}} v| = 396503392000000000 \;>\; 382961614600000000 = \rho^2 \cdot 34 = \rho^2(v^\top v),$$

which by the Rayleigh dictionary (4) rules out $\mathrm{LB}(c, W_{\mathrm{ss}})$ for every $c \leq \rho^2$.

Each of the four positive semidefiniteness statements in (1) and (2) is supplied as a *denominator-cleared $LDL^\top$ identity* over the integers: an identity

$$N \cdot M \;=\; U\,\mathrm{diag}(e)\,U^\top, \qquad N \in \mathbb{Z}_{>0}, \quad e \in \mathbb{Z}^4_{\geq 0}, \quad U \in \mathbb{Z}^{4\times 4},$$

which forces $M \succeq 0$ at once, since the right-hand side is a congruence of a nonnegative diagonal matrix and $N > 0$. For $M \in \{A, B, \rho I - S, \rho I + S\}$ the four multipliers N have 1, 5, 21 and 22 digits and the entries of U at most 12 digits.

What rests on computation, and what was searched. The proof rests on exactly six finite integer verifications: the four matrix identities above, and the two scalar evaluations $v^\top W_{\mathrm{ss}} v$ and $v^\top v$ of step (4). Each is a decidable statement about 4×4 integer matrices and is checked by kernel reduction, with no floating point at any point. Two exhaustive searches were run to *produce* the data, and neither is part of the proof: the witness vector v was chosen by sweeping all 13^4 integer vectors with $\|v\|_\infty \leq 6$ and keeping the one maximising $|v^\top W_{\mathrm{ss}} v|/(v^\top v)$, and ρ was chosen by a scan for the smallest round value for which the two $LDL^\top$ certificates of step (2) exist. The pair (A, B) itself came from a numerical optimisation described in Section 5; the proof does not depend on how it was found. ☐

Remark 3.2 (how large the gap is)*.* The theorem records a strict inequality; the same integer data record a quantitative one. From step (4), $\|W_{\mathrm{ss}}\| \geq 396503392000000000/34 = 198251696000000000/17$, and from step (3), $\|W_{\mathrm{rs}}\| \leq \rho^2$, so the certificate itself gives

$$\frac{\|W_{\mathrm{ss}}\|}{\|W_{\mathrm{rs}}\|} \;\geq\; \frac{198251696000000000}{17\rho^2} \;=\; 1.0353606650999325\ldots$$

Outside the formal development, the true ratio was computed by three routes sharing no code — floating point via the singular value decomposition, 120-digit interval arithmetic, and exact rational bisection on the characteristic polynomials — and all three agree on

$$\|W_{\mathrm{ss}}\| = 1.1662522564675913\ldots \times 10^{16}, \qquad \|W_{\mathrm{rs}}\| = 1.1261518992976322\ldots \times 10^{16},$$

$$\|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\| = 1.0356083022147982\ldots,$$

so the machine-checked certificate captures all but 0.70% of the true excess over 1. The spectrum of $N = AB$ is $\{34676.50\ldots, 5288235.95\ldots, 24442640.41\ldots, 31234447.13\ldots\}$, all real and positive, as it must be for a product of two positive definite matrices.

3.2. **Controls.** A refutation should come with evidence that the apparatus producing it does not produce refutations where none exist. Five checks accompany Theorem 3.1, all of them verified in the same development.

Proposition 3.3 (controls). *With A, B, ρ, v as above:*

(i) (the inequality holds where it should) *If A and B are any two commuting real $d \times d$ matrices then $W_{\mathrm{ss}} = W_{\mathrm{rs}}$ for every K; the SS–RS inequality holds with equality. (This is the authors' own remark that Conjecture 1 is true with $\eta_{n,K} = 1$ for commuting A_i.)*

(ii) (the bound on W_{rs} is not vacuous) $\mathrm{LB}(10^{16}, W_{\mathrm{rs}})$ *is false, witnessed by $u = (1, 1, 1, 1)^\top$ with $u^\top W_{\mathrm{rs}} u = 45044456100000000$ and $u^\top u = 4$. Together with step (3) of the proof this brackets $\|W_{\mathrm{rs}}\|$ in $(10^{16}, \rho^2]$, so ρ^2 is not an artificially large bound.*

(iii) (the witness vector matters) *At $w = (1, 0, 0, 0)^\top$ one has $w^\top W_{\mathrm{ss}} w = 222000000000000$, well inside $[-\rho^2, \rho^2]$: the violation is a property of v, not of the shape of the argument.*

(iv) (the witness is interior) *$A \succeq I$, so A is positive definite and the counterexample does not sit on the boundary of the positive semidefinite cone.*

(v) (W_{ss} and W_{rs} are symmetric) *So for both of them the spectral norm really is the least two-sided Löwner bound, and Theorem 3.1 really is the norm inequality it is read as.*

4. The normalised refutations, and how far they reach

4.1. **The first witness, in the conjecture's own normalisation.** Conjecture 1 asks for $(1-\eta)I \preceq A_i \preceq I$. By Lemma 2.1 the pair of Section 3 may be rescaled into that form without changing either side of the inequality. Put

$$A' = 10^{-6} A, \qquad B' = 10125000^{-1} B, \qquad \nu = (10^6 \cdot 10125000)^{-1}, \qquad \varepsilon_0 = \tfrac{1}{84375000},$$

and let $W'_{\mathrm{ss}}, W'_{\mathrm{rs}}$ be the two matrices of (2) formed from A', B' at $K = 2$.

Theorem 4.1 (Conjecture 1 fails at $n = 2$, $K = 2$, $d = 4$ for every $\eta > 1 - \frac{1}{84375000}$). *With the notation above,*

$$\varepsilon_0 I \preceq A' \preceq I, \qquad \varepsilon_0 I \preceq B' \preceq I,$$

which is exactly the hypothesis of Conjecture 1 at $1 - \eta = \varepsilon_0$, while $\mathrm{LB}(\nu^2\rho^2, W'_{\mathrm{rs}})$ holds and every c with $\mathrm{LB}(c, W'_{\mathrm{ss}})$ satisfies $c > \nu^2\rho^2$. That is, $\|W'_{\mathrm{ss}}\| > \nu^2\rho^2 \geq \|W'_{\mathrm{rs}}\|$.

Proof sketch. Lemma 2.1 with $s = 10^{-6}$, $t = 10125000^{-1}$ turns W'_{ss} into $\nu^2 W_{\mathrm{ss}}$ and W'_{rs} into $\nu^2 W_{\mathrm{rs}}$, and a two-sided Löwner bound scales with a nonnegative scalar in both directions, so the last two assertions are Theorem 3.1 multiplied by $\nu^2 > 0$. The four hypotheses $\varepsilon_0 I \preceq A' \preceq I$ and $\varepsilon_0 I \preceq B' \preceq I$ are four more denominator-cleared $LDL^\top$ identities over $\mathbb{Z}$, for the matrices $10^6 I - A$, $675A - 8I$, $10125000I - B$ and $25B - 3I$; the first two are diagonal and trivial, the last two are certificates with multipliers of 9 and 42 digits. The binding constraint is the lower bound on B: the certificate proves $\frac{3}{25} I \preceq B$, and $\frac{3}{25}/10125000 = \varepsilon_0$. □

Remark 4.2 (why $\frac{3}{25}$ and not the true minimum). Outside the formal development, exact rational bisection on the characteristic polynomial of B gives $\lambda_{\min}(B) = 0.121234069352096\ldots$ and $\lambda_{\max}(B) = 10121960.3179198\ldots$, so $\mathrm{cond}(B) = 83491054.7177\ldots$ and this witness in fact rules out every $\eta \geq 1 - 1.1977331\ldots \times 10^{-8}$. Formalising the rounded bound $\frac{3}{25} \leq \lambda_{\min}(B)$ instead of $\frac{60617}{500000} \leq \lambda_{\min}(B)$ shrinks the required integer multiplier from 68 digits to 42 at a cost of one percent in the bound, which is why the machine-checked constant is $\frac{1}{84375000}$ rather than the marginally better one.

4.2. **A better-conditioned witness.** The condition numbers 10^6 and 8.35×10^7 of the pair (5) are exactly what the bound $\eta < 1 - \frac{1}{84375000}$ costs; a pair with smaller condition numbers refutes more. They can be brought down by more than two orders of magnitude while everything stays exact. *In this subsection, and only here, A and B denote the pair* (6) *below, and S, W_{ss}, W_{rs} are the matrices formed from it; Sections 5 and 6 return to* (5). Let

$$A = \mathrm{diag}(1, 7^2, 62^2, 624^2) = \mathrm{diag}(1, 49, 3844, 389376), \tag{6}$$

$$B = \begin{pmatrix} 5074956432 & 841928256 & 7921368 & 0 \\ 841928256 & 280642752 & -13956696 & 74958 \\ 7921368 & -13956696 & 3066336 & 126945 \\ 0 & 74958 & 126945 & 26908 \end{pmatrix},$$

again both symmetric with integer entries, and again $K = 2$, $d = 4$. As before $AB + BA$ and $(AB)^2 + (BA)^2$ have even entries, so $S = \frac{1}{2}(AB + BA)$ and $W_{\mathrm{ss}} = \frac{1}{2}((AB)^2 + (BA)^2)$ are integer matrices, and $W_{\mathrm{rs}} = S^2$; explicitly

$$S = \begin{pmatrix} 5074956432 & 21048206400 & 15228829980 & 0 \\ 21048206400 & 13751494848 & -27166708764 & 14595259575 \\ 15228829980 & -27166708764 & 11786995584 & 24958656450 \\ 0 & 14595259575 & 24958656450 & 10477329408 \end{pmatrix},$$

while W_{ss} has entries of up to 22 digits and is not printed here. Of it we shall use only one Rayleigh quotient and the first diagonal entry,

$$(W_{\mathrm{ss}})_{11} = 607297025958980797744.$$

Put

$$\rho_1 = 42869796155, \qquad \rho_1^2 = 1837819422371252784025, \qquad v = (-8, 4, -4, 8)^\top,$$
$$s_A = 389376, \qquad s_B = 5220000000, \qquad \nu_1 = (s_A s_B)^{-1}, \qquad \varepsilon_1 = \tfrac{1}{389765}.$$

Theorem 4.3 (Conjecture 1 fails at $n = 2$, $K = 2$, $d = 4$ for every $\eta > 1 - \frac{1}{389765}$)**.** *With the data above, A and B are symmetric positive definite,* $\mathrm{LB}(\rho_1^2, W_{\mathrm{rs}})$ *holds, and every c with* $\mathrm{LB}(c, W_{\mathrm{ss}})$ *satisfies $c > \rho_1^2$; since W_{ss} and W_{rs} are symmetric this says*

$$\|W_{\mathrm{ss}}\| \;>\; \rho_1^2 = 1837819422371252784025 \;\geq\; \|W_{\mathrm{rs}}\|.$$

Moreover, writing $A' = A/s_A$ and $B' = B/s_B$ and letting $W'_{\mathrm{ss}}, W'_{\mathrm{rs}}$ be the two matrices of (2) *formed from A', B' at $K = 2$,*

$$\varepsilon_1 I \preceq A' \preceq I, \qquad \varepsilon_1 I \preceq B' \preceq I, \qquad \|W'_{\mathrm{ss}}\| > \nu_1^2 \rho_1^2 \geq \|W'_{\mathrm{rs}}\|,$$

the first two of which are exactly the hypothesis of Conjecture 1 at $1 - \eta = \varepsilon_1$. So the first inequality of Conjecture 1 fails at $n = 2$, $K = 2$, $d = 4$ for every step-size constant

$$\eta \;>\; 1 - \tfrac{1}{389765} \;=\; 1 - 2.5656485\ldots \times 10^{-6}.$$

Finally the improvement over Theorem 4.1 is strict: $\varepsilon_0 = \frac{1}{84375000} < \frac{1}{389765} = \varepsilon_1$, so this witness leaves open a strictly smaller set of step-size constants, by a factor 216.5.

Proof sketch. The shape is the one of Theorems 3.1 and 4.1, with new numbers. For the unnormalised half: $A \succeq 0$, $B \succeq 0$, $-\rho_1 I \preceq S \preceq \rho_1 I$ are four denominator-cleared $LDL^\top$ identities $N \cdot M = U \operatorname{diag}(e)\, U^\top$ over $\mathbb{Z}$, with multipliers N of 1, 11, 62 and 64 digits; Lemma 2.2 applied to $D = S$ turns the third and fourth into $W_{\mathrm{rs}} = S^2 \preceq \rho_1^2 I$, and $W_{\mathrm{rs}} = S^\top S \succeq 0$ gives $\mathrm{LB}(\rho_1^2, W_{\mathrm{rs}})$. Against it,

$$|v^\top W_{\mathrm{ss}} v| = 308248252327926150589440 \;>\; 294051107579400445444000 = \rho_1^2 \cdot 160 = \rho_1^2 (v^\top v),$$

which by (4) excludes $\mathrm{LB}(c, W_{\mathrm{ss}})$ for every $c \leq \rho_1^2$. For the normalised half, four further identities give $I \preceq A \preceq s_A I$ and $\frac{s_B}{389765} I \preceq B \preceq s_B I$ — the two for A are diagonal and trivial, the two for B have multipliers of 37 and 70 digits — and dividing by s_A and s_B turns them into $\varepsilon_1 I \preceq A' \preceq I$ and $\varepsilon_1 I \preceq B' \preceq I$, using $\frac{1}{s_A} = \frac{1}{389376} > \varepsilon_1$ on the A side. Lemma 2.1 with $s = s_A^{-1}$, $t = s_B^{-1}$ then multiplies the first display by $\nu_1^2 > 0$. The last assertion is the comparison of two rationals.

What rests on computation, and what was searched. The proof rests on ten finite integer verifications: the eight matrix identities above and the two scalar evaluations $v^\top W_{\mathrm{ss}} v$ and $v^\top v$; each is a decidable statement about 4×4 integer matrices, reduced by the kernel, with no floating point in the decision path. Four searches were run to *produce* the data and none of them is part of the proof. Two of them are the rounding sweeps of Remark 4.9, of 798 and 1793 candidate roundings, which is how the pair (6) was obtained. The third is the choice of ρ_1: it is the least integer with $-\rho_1 I \preceq S \preceq \rho_1 I$, located by bisection on that integer, and indeed $\|S\| = 42869796154.2358769\ldots$, so $\rho_1 = \lceil \|S\| \rceil$. The fourth is the violating vector, obtained by the same construction as for (5) but over a larger box: it maximises $|v^\top W_{\rm ss} v|/(v^\top v)$ over an exhaustive sweep of the integer vectors with $\|v\|_\infty \le 8$, all 17^4 of them. That vector and the four Rayleigh-quotient vectors used below are emitted from A and B by the exact pipeline described in Section 7. □

Remark 4.4 (how large the gap is, for the second witness). As in Remark 3.2, the certificate itself gives a quantitative statement: $\|W_{\rm ss}\| \ge |v^\top W_{\rm ss} v|/(v^\top v) = 1926551577049538441184$ and $\|W_{\rm rs}\| \le \rho_1^2$, so

$$\frac{\|W_{\rm ss}\|}{\|W_{\rm rs}\|} \ge \frac{1926551577049538441184}{1837819422371252784025} = 1.0482812150084901\ldots$$

Outside the formal development, exact rational bisection on the characteristic polynomials gives

$$\|W_{\rm ss}\| = 1.927704994620283871906 3\ldots \times 10^{21}, \qquad \|W_{\rm rs}\| = 1.8378194223057371814676\ldots \times 10^{21},$$
$$\|W_{\rm ss}\|/\|W_{\rm rs}\| = 1.0489088161892291977\ldots,$$

so the machine-checked certificate captures all but 1.28% of the true excess over 1. The spectrum of $N = AB$ is $\{5.8765613987\ldots\times 10^8,\ 6.1307398003\ldots\times 10^9,\ 1.3533207476\ldots\times 10^{10},\ 2.0839172856\ldots\times 10^{10}\}$, real and positive as it must be. The mechanism of Remark 5.1 is the same here and more pronounced: the eigenvalues of $W_{\rm ss}$ are

$$\{-1.9277\times 10^{21},\ -5.4256\times 10^{20},\ +1.4187\times 10^{21},\ +1.7069\times 10^{21}\},$$

the largest in modulus again the negative one, so that

$$\frac{\lambda_{\max}(W_{\rm ss})}{\lambda_{\max}(W_{\rm rs})} = 0.92875\ldots \qquad \text{against} \qquad \frac{\|W_{\rm ss}\|}{\|W_{\rm rs}\|} = 1.04891\ldots$$

The next proposition is what tells a reader that this particular pair has been used up. It bounds the conditioning of A and B from *both* sides: above by certificates, below by Rayleigh quotients, so that the two together pin down the best normalisation the pair admits at all.

Proposition 4.5 (the conditioning of the second witness, and a matching floor)**.** *With A, B as in* (6)*:*

- (i) (above) $I \preceq A \preceq 389376\, I$ *and* $\frac{5220000000}{389765} I \preceq B \preceq 5220000000\, I$*; equivalently, every Rayleigh quotient* $x^\top Ax/(x^\top x)$ *lies in* $[1, 389376]$ *and every* $x^\top Bx/(x^\top x)$ *in* $[\frac{5220000000}{389765}, 5220000000]$*. Hence* $\operatorname{cond}(A) \le 389376$ *and* $\operatorname{cond}(B) \le 389765$*.*
- (ii) (below) *Every* $m, M \in \mathbb{R}$ *with* $mI \preceq A \preceq MI$ *satisfies* $M \ge 389376\, m$*, and every* m, M *with* $mI \preceq B \preceq MI$ *satisfies* $M \ge 380791\, m$*. With (i),* $\operatorname{cond}(A) = 389376$ *exactly.*
- (iii) (the floor) *Consequently, for all* $s_A, s_B, \varepsilon > 0$*,*

$$\varepsilon I \preceq A/s_A \preceq I \ \ \text{and} \ \ \varepsilon I \preceq B/s_B \preceq I \qquad \Longrightarrow \qquad \varepsilon \le \tfrac{1}{389376}.$$

That is: no *normalisation of this pair into the hypothesis of Conjecture 1 does better than* $1 - \eta = \frac{1}{389376}$*, and the value* $\varepsilon_1 = \frac{1}{389765}$ *used in Theorem 4.3 falls short of that floor by a factor* $\frac{389765}{389376}$*: the conditioning cap* 389765 *is* 0.0999% *above the least one the pair admits,* 389376*.*

Proof sketch. Part (i) is four denominator-cleared $LDL^\top$ identities over $\mathbb{Z}$, for $A - I$, $389376\, I - A$, $5220000000\, I - B$ and $389765\, B - 5220000000\, I$; the first two are diagonal. The Rayleigh reading is (4). Part (ii) is two Rayleigh quotients per matrix and nothing else: for A, which is diagonal, the basis vectors e_1 and e_4, which are exactly extremal — the first gives $m \le A_{11} = 1$ and the second $M \ge A_{44} = 389376$, whence $M \ge 389376 \ge 389376\, m$; for B, the integer vectors

$$(-99603279965,\ 554535898683,\ 4873084956647,\ -5000000000000)^\top$$

and

$$(-7000000000000,\ -11935088769448,\ -74379295969,\ -173244147)^\top,$$

obtained by rounding the two extreme eigenvectors of B computed at 80 digits and keeping the best of about 65 scalings. They realise $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ to full double precision, which is why the constant that comes out is $380791 = \lfloor \operatorname{cond}(B) \rfloor$ and not something weaker. Part (iii) rescales: from $\varepsilon I \preceq A/s_A \preceq I$ one gets $(\varepsilon s_A) I \preceq A \preceq s_A I$, so (ii) gives $s_A \ge 389376\, \varepsilon s_A$ and, s_A being positive,

$389376\,\varepsilon \le 1$. The same step on B gives $380791\,\varepsilon \le 1$, which is weaker. Nothing here is an exhaustive computation except the four identities of (i); the vectors of (ii) were searched for, but the proof only evaluates them. □

Remark 4.6 (why $\frac{1}{389765}$ and not the floor $\frac{1}{389376}$). Exactly as in Remark 4.2, the machine-checked constant is the one with the cheaper certificate rather than the best one. Outside the formal development, exact rational bisection gives $\lambda_{\min}(B) = 13704.38656569963\ldots$, $\lambda_{\max}(B) = 5218514684.26630558\ldots$, hence $\operatorname{cond}(B) = 380791.55599\ldots$, while $\operatorname{cond}(A) = 389376$ exactly because A is diagonal. The binding side is therefore A, and the floor of Proposition 4.5(iii) is $\frac{1}{389376}$; the pair $(s_B, 389765)$ was chosen to keep the two B certificates at 37 and 70 digits. The cost of that choice is under 0.1% in the bound, and Proposition 4.5(iii) is what says it is only that.

Proposition 4.7 (controls for the second witness)**.** *With A, B, ρ_1, s_B as above:*

(i) (the bound on $W_{\rm rs}$ is not vacuous) $\mathrm{LB}(10^{21}, W_{\rm rs})$ *is false, witnessed by* $u = (1,1,1,1)^\top$ *with* $u^\top W_{\rm rs} u = 5322633632322636326814$ *and* $u^\top u = 4$. *With the theorem this brackets* $\|W_{\rm rs}\|$ *in* $(10^{21}, \rho_1^2]$.

(ii) (the witness vector matters) *At* $w = (1,0,0,0)^\top$ *one has* $w^\top W_{\rm ss} w = 60729702595898079744$, *well inside* $[-\rho_1^2, \rho_1^2]$.

(iii) (the constant cannot be rounded up) $\frac{1}{380800} I \preceq B/s_B$ *is* false*, refuted by the same integer vector that realises* $\lambda_{\min}(B)$. *So the certificates of Theorem 4.3 sit near the edge of what this pair supports; Proposition 4.5(iii) is the same statement quantified over all normalisations.*

(iv) (the witness is interior) $A - I \succeq 0$ *and* $389765\,B - 5220000000\,I \succeq 0$, *so A and B are positive definite and the hypothesis of Theorem 4.3 is satisfiable strictly inside the cone.*

The two structural controls of Proposition 3.3, (i) and (v), apply verbatim here and are not repeated.

4.3. **The bracket, and how the second witness was found.**

Proposition 4.8 (a bracket for the step-size constant at $K = 2$)**.** *Call* $\eta \in (0,1]$ admissible *at $n = 2$, $K = 2$ if the implication of Conjecture 1 holds for it in every dimension d. Then*

$$\eta = \tfrac{1}{4} \text{ is admissible}, \qquad \text{and every admissible } \eta \text{ satisfies } \ \eta < 1 - \tfrac{1}{389765} = 1 - 2.5656485\ldots \times 10^{-6}.$$

Writing $\eta^*_{2,2}$ for the supremum of the admissible η, this reads $\frac{1}{4} \le \eta^*_{2,2} \le 1 - \frac{1}{389765}$. The pair (5) *of Section 3 alone gives the weaker upper bound $\eta^*_{2,2} \le 1 - \frac{1}{84375000}$.*

Proof sketch. Admissibility is downward closed: a smaller η is a stronger hypothesis. The upper half is Theorem 4.3, since the normalised pair there satisfies $(1-\eta)I \preceq A', B' \preceq I$ for every $\eta \ge 1 - \varepsilon_1$ and violates the conclusion; the weaker bound is Theorem 4.1 read the same way. The lower half is *not* proved here: it is read off Theorem 3 of [1], which at $K = 2 = 2^1$ reads $(1 - \frac{1}{2K})I = \frac{3}{4} I \preceq A, B \preceq I$ and gives $\|W_{\rm ss}\| \le \|W_{\rm rs}\|$ for $A, B \in \mathbb{S}^d_{++}$, combined with Theorem 6.1 below, which gives $\|W_{\rm rs}\| \le \|W_{\rm gd}\|$ unconditionally. Only the upper half is part of the formal development. □

Remark 4.9 (how (6) was found: the rounding grid, not the homotopy). This is search provenance, outside the formal development; none of it is needed to read Theorem 4.3. The route from (5) to (6) is a homotopy in the conditioning: start at the pair (5), impose a cap on $\max(\log \operatorname{cond} A, \log \operatorname{cond} B)$, minimise the margin m of Section 5 under it, lower the cap, re-minimise, halve the step on failure. In double precision that walks $\max \operatorname{cond}$ from 8.35×10^7 down to 5.42×10^5 in seconds, where two independent 400-restart cold searches reached only 4.1×10^{11} and 5.6×10^8. The float point then has to be rounded back to exact integers — fix the scales c_i to integers, refit, round the remaining matrix to a lattice with denominator q, clear denominators — and it is *there*, not in the homotopy, that the conditioning was being lost. Taking the first (c, q) that certifies gives $\max \operatorname{cond} = 2.53 \times 10^6$: a factor $2525774/541996 = 4.66$ thrown away in the rounding, more than the last two rungs of the homotopy had gained. That first exact pair is $A = \operatorname{diag}(2500, 185761, 9928801, 622702116)$ with a B of 22-digit entries, $\operatorname{cond}(A) = 155675529/625 = 249080.85\ldots$, $\operatorname{cond}(B) = 2525774.2\ldots$, certified ratio 1.0659567 and $LDL^\top$ certificates of 140–155 digits; it gives $\eta^*_{2,2} \le 1 - 3.9591821 \times 10^{-7}$ and is superseded by Theorem 4.3. But each candidate rounding costs only a few exact 4×4 $LDL^\top$ decompositions, so the whole (c, q) grid can simply be priced: sweeping 798 roundings *of the same float point* in 66 seconds, and keeping the one with the smallest exact max cond that still certifies a violation, returns max cond = 507487 — below the float point the homotopy had reached, since the lattice may land in a slightly better spot than its centre — with 11-digit entries in place of 22-digit ones. Restarting the homotopy from that exact pair moves the cap two further rungs, to 3.91×10^5, and a second sweep of 1793 roundings (378 of them certifiable, 421 seconds) at the new point produces (6). A third cycle returns an empty ladder: the margin is -1.0489, too close to the boundary -1 for the cap to move at all, and $\operatorname{cond}(A) = 389376$ and $\operatorname{cond}(B) = 380792$ are now balanced, which is the signature of a local optimum of $\max(\operatorname{cond} A, \operatorname{cond} B)$. Together with Proposition 4.5(iii) — which says the bookkeeping on (6) is within 0.1% of spent — that is why this paper stops where it does. The whole search cost 0.20 CPU-hours. The lesson generalises past this problem: when a floating-point search is converted to an exact witness by rounding to a lattice, the rounding step deserves to be swept exhaustively before the float number is believed to be the frontier, because it is both the cheaper of the two to sweep and, here, the one that was binding.

5. Why the counterexample is invisible to search

Nothing in Section 3 explains how the pair (5) was found, and the answer is worth a section, because the obvious searches provably cannot find it.

Remark 5.1 (the doubling identity, and $\lambda_{\max}$-blindness). For every real square X, writing $S = \operatorname{sym} X$ and $\Omega = \operatorname{skew} X$,

$$\operatorname{sym}(X^2) = \tfrac{1}{2}\big(X^2 + (X^\top)^2\big) = \tfrac{1}{2}\big((S+\Omega)^2 + (S-\Omega)^2\big) = S^2 + \Omega^2,$$

and $\Omega^2 = -\Omega\Omega^\top \preceq 0$ because $\Omega^\top = -\Omega$. Hence $\operatorname{sym}(X^2) \preceq (\operatorname{sym} X)^2$ and in particular

$$\lambda_{\max}\big(\operatorname{sym}(X^2)\big) \;\le\; \|\operatorname{sym} X\|^2 \qquad \text{for every } X. \tag{7}$$

So at $K = 2$ a violation of $w_{\mathbb{R}}(N^2) \le w_{\mathbb{R}}(N)^2$ can only ever come from $\lambda_{\min}(\operatorname{sym}(N^2))$, never from $\lambda_{\max}$. At the witness (5) the eigenvalues of W_{ss} are

$$\{-1.16625 \times 10^{16},\ -5.5389 \times 10^{14},\ +2.5940 \times 10^{15},\ +1.12234 \times 10^{16}\},$$

the largest *in modulus* being the negative one, and consequently

$$\frac{\lambda_{\max}(W_{\mathrm{ss}})}{\lambda_{\max}(W_{\mathrm{rs}})} = 0.996612366804\ldots \qquad \text{while} \qquad \frac{\|W_{\mathrm{ss}}\|}{\|W_{\mathrm{rs}}\|} = 1.035608302215\ldots$$

An optimiser that measures $\lambda_{\max}$ is therefore blind to this phenomenon as a matter of theorem, not of luck. Only identity (7) is elementary mathematics; the four eigenvalues and the two ratios are floating-point computations reported here and are not part of the machine-checked development.

Identity (7) also says that at $K = 2$ question (Q) is *equivalent* to a lower bound on the *margin*

$$m(N) \;=\; \frac{\lambda_{\min}\big(\operatorname{sym}(N^2)\big)}{\|\operatorname{sym} N\|^2}, \qquad \text{namely} \qquad w_{\mathbb{R}}(N^2) \le w_{\mathbb{R}}(N)^2 \iff m(N) \ge -1.$$

Minimising m is what found the witness; maximising the ratio $\|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\|$ does not, because that ratio sits on a plateau at exactly 1.

Remark 5.2 (measurements on the search landscape, outside the formal development). The following are floating-point measurements, reported as such. (i) 20000 random positive semidefinite pairs at $d \le 6$, $K \le 8$ give a maximum of $\|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\|$ of exactly 1.0 — including at $d = 4$, where a violation exists. (ii) Over 200000 random pairs at $d \le 6$, both $\max \lambda_{\max}(W_{\mathrm{ss}})/\lambda_{\max}(W_{\mathrm{rs}})$ and $\max \|W_{\mathrm{ss}}\|/\|W_{\mathrm{rs}}\|$ are 0.999999999999. (iii) The infimum of the margin m over $\{AB : A, B \in \mathbb{S}^d_+\}$ is $-\frac{1}{3}$ at $d = 2$, approached in the limit at $N = \left(\begin{smallmatrix} 1 & 4 \\ 0 & 1 \end{smallmatrix}\right)$, for which $\|\operatorname{sym} N\| = 3$ and $\lambda_{\min}(\operatorname{sym}(N^2)) = -3$ exactly; that N is a Jordan block and so is not itself a product of two positive definite matrices, which is why the value is approached rather than attained. At $d = 3$ some 950 seconds of Nelder–Mead and Powell restarts bottom out at -0.999999998, approaching -1 from above and never crossing, with the near-extremal point at $\log \operatorname{cond}(A) = 59.2$, $\log \operatorname{cond}(B) = 63.8$; at those condition numbers double precision is at its limit and a high-precision re-verification was not completed, so this is two agreeing double-precision measurements and not a settled fact. At $d = 4$ the witness (5) gives $m = -1.0356$, the witness (6) gives $m = -1.0489$, and searches reach -1.273. If the $d = 3$ measurement is right then $d = 4$ is the smallest dimension in which the $\eta = 1$ inequality can fail, and $d = 3$ is the interesting open dimension. Nothing in this remark is claimed as a theorem, and the $d = 3$ entry in particular is not claimed as a fact: it is two agreeing double-precision runs, and it is stated here only to say where the next question is.

Remark 5.3 (an independent search that found nothing). A separate numerical study of the same question, parametrising $\{AB : A, B \in \mathbb{S}^d_{++}\} = \{Q\Lambda Q^{-1} : \Lambda \text{ positive diagonal}\}$ — a parametrisation that is correct, complete, and *contains* the witness (5) — ran roughly 2.4 million differential-evolution evaluations and 1500 Nelder–Mead restarts at $d = 3, 4, 5$ and reported a maximum ratio of 1.0000000000 with no violation. Remark 5.1 explains the discrepancy completely: an objective built on $\lambda_{\max}$ cannot exceed 1, and one built on the ratio sits on the plateau. The same study also failed to rediscover the $d = 3$ gap between the complex and the real numerical radius that [1] record — for the 3×3 matrix printed there, $w_{\mathbb{C}} \approx 3.0004$ against $w_{\mathbb{R}} = 3$, a ratio 1.000144 — which it then confirmed by hand from that matrix: an independent sign that the difficulty is in the landscape and not in the mathematics.

6. The two halves that hold

6.1. **Reshuffling beats gradient descent at $n = 2$, unconditionally.** The second inequality of Conjecture 1 at $n = 2$ needs no conditioning hypothesis at all.

Theorem 6.1 (Recht–Ré at $n = 2$; the proof here is new, the statement is not)**.** *Let $A, B \in \mathbb{S}^d_+$ and let $c \geq 0$ satisfy $A + B \preceq 2cI$. Then*

$$-c^2 I \ \preceq\ \tfrac{1}{2}(AB + BA) \ \preceq\ c^2 I.$$

Taking $c = \|\frac{1}{2}(A + B)\|$ this gives $\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$ for every K and every d, that is, $\eta_{2,K} = 1$ suffices for the second inequality of Conjecture 1. Moreover the constant c^2 is attained, at $A = B = cI$; and the positive semidefiniteness hypothesis cannot be relaxed to $\eta > 1$: for every $\delta \in (0, 1]$ and every $K \geq 1$ the 1×1 pair $A = (\delta)$, $B = (-\delta)$ satisfies $(1 - \eta)I \preceq A, B \preceq I$ with $\eta = 1 + \delta$ and has $W_{\mathrm{gd}} = 0$ while $W_{\mathrm{rs}} = (-\delta^2)^K \neq 0$.

Proof. Put $G = \frac{1}{2}(A + B)$ and $D = \frac{1}{2}(A - B)$, so that $\frac{1}{2}(AB + BA) = G^2 - D^2$. From $0 \preceq A, B$ and $A + B \preceq 2cI$ we get $-cI \preceq G \preceq cI$, and also $-cI \preceq D \preceq cI$, because $cI - D = (cI - G) + B$ and $cI + D = (cI - G) + A$ are each a sum of two positive semidefinite matrices. Lemma 2.2 applied twice gives $G^2 \preceq c^2 I$ and $D^2 \preceq c^2 I$, whence

$$c^2 I - (G^2 - D^2) = (c^2 I - G^2) + D^2 \succeq 0, \qquad c^2 I + (G^2 - D^2) = (c^2 I - D^2) + G^2 \succeq 0.$$

For the two supplements: at $A = B = cI$ one has $\frac{1}{2}(AB + BA) = c^2 I$, so no smaller constant works; and the 1×1 computation is immediate. □

This statement is the $n = 2$ case of the noncommutative arithmetic–geometric mean inequality of Recht and Ré [3], and the authors of [1] cite it for exactly this purpose: “Since $\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$ holds for $n = 2$ and any PSD matrices [3], the theorems prove the entire Conjecture 1 for the special cases considered in them.” What is new here is only the route: Lemma 2.2 replaces the matrix square root that the usual argument takes, which is what makes the whole thing formalisable in a few lines. Recht and Ré’s own argument at $n = 2$ runs instead through the stronger inequality $\|AB\| \leq \|\frac{1}{2}A + \frac{1}{2}B\|^2$, which they attribute to Bhatia and Kittaneh [11] and place in a line of work they trace back to an earlier paper of the same two authors [12]; that attribution was read here in their own preprint.

Remark 6.2. One consequence is worth a sentence. Peng’s Theorem 3 [2] proves $\|W_{\mathrm{rs}}\| \leq \|W_{\mathrm{gd}}\|$ for every $n \geq 2$ under $(1 - \frac{1}{4n^2+1})I \preceq A_i \preceq I$; at $n = 2$ that constant is $\frac{1}{17}$, and Theorem 6.1 shows it is not tight there, since $\eta = 1$ works. Nothing is claimed here about $n \geq 3$.

6.2. **The single-shuffle inequality at $d = 2$, for every K.** At $d = 2$ the inequality that fails at $d = 4$ holds, for every K, and it holds under a hypothesis on the spectrum alone.

Theorem 6.3 ([1], Theorem 4, with the hypothesis weakened)**.** *Let N be a real 2×2 matrix whose two eigenvalues are real, i.e.* $(\operatorname{tr} N)^2 \geq 4 \det N$*, and let $c \geq 0$ satisfy $-cI \preceq \operatorname{sym} N \preceq cI$. Then for every $K \in \mathbb{N}$,*

$$-c^K I \preceq \operatorname{sym}(N^K) \preceq c^K I, \qquad \textit{that is} \qquad w_{\mathbb{R}}(N^K) \leq w_{\mathbb{R}}(N)^K.$$

The hypothesis on the spectrum is not removable: for the rotation $N = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ one has $\operatorname{sym} N = 0$ *and* $\operatorname{sym}(N^2) = -I$*, so $w_{\mathbb{R}}(N^2) = 1 > 0 = w_{\mathbb{R}}(N)^2$.*

Specialised to $N = AB$ with $A, B \in \mathbb{S}^2_+$ — which has real spectrum — this is Theorem 4 of [1]. Their proof runs through the *complex* numerical radius $w(M) = \sup\{|v^* M v| : v \in \mathbb{C}^d, \|v\| \leq 1\}$, the power inequality $w(M^K) \leq w(M)^K$ [10], and their own 2×2 lemma $w(Q\Lambda Q^{-1}) = \frac{1}{2}\|Q\Lambda Q^{-1} + Q^{-\top}\Lambda Q^{\top}\|$ for invertible 2×2 Q and nonnegative diagonal Λ, i.e. $w = w_{\mathbb{R}}$ on that class — a lemma they show to fail at $d = 3$, with the explicit matrix $\left(\begin{smallmatrix} 3 & .05 & .3 \\ -.05 & 3 & -.1 \\ -.3 & .1 & 2 \end{smallmatrix}\right)$, for which $w \approx 3.0004 > 3 = w_{\mathbb{R}}$. For a real 2×2 matrix with real spectrum the numerical range is an ellipse whose foci are the two eigenvalues — the elliptical range theorem [13, 14], for which [15] gives a short modern proof — so its major axis is real and $w = w_{\mathbb{R}}$ there as well; the widening of the hypothesis from $A, B \in \mathbb{S}^2_+$ to "N has real spectrum" is routine once that identification is available, which is why we present the theorem as theirs. For the power inequality we cite [10], the elementary proof that [1] themselves cite; we make no attribution claim beyond that, since the announcement usually credited for the inequality is a meeting abstract we could not verify against a primary record.

Proof sketch. The proof given here uses neither the numerical range nor the power inequality, both of which are expensive to formalise; it replaces them by a two-term linear recursion, and uses no square roots and no binomial coefficients. Write $m = \frac{1}{2}\operatorname{tr} N$ and $r^2 = m^2 - \det N$, which is ≥ 0 exactly by the real-spectrum hypothesis. Cayley–Hamilton at 2×2 gives $N^K = p_K N + (f_K - m p_K)I$ where

$$f_{K+1} = m f_K + r^2 p_K, \qquad p_{K+1} = f_K + m p_K, \qquad f_0 = 1, \; p_0 = 0.$$

Applying sym, the matrix $\operatorname{sym}(N^K) = p_K \operatorname{sym} N + (f_K - m p_K)I$ has half-trace f_K, and its off-diagonal entry and its diagonal gap are p_K times those of $\operatorname{sym} N$. Writing $\sigma = c - |m|$ and using Sylvester's criterion at 2×2 in both directions, the whole statement reduces to the scalar inequality

$$f_K + \sigma p_K \leq (m + \sigma)^K \qquad \text{for } m, \sigma \geq 0, \; 0 \leq r^2 \leq \sigma^2,$$

whose induction step is one line: $f_{K+1} + \sigma p_{K+1} = (m + \sigma) f_K + (r^2 + m\sigma) p_K \leq (m + \sigma)(f_K + \sigma p_K)$. Negative m is handled by the parity $f_K(-m) = (-1)^K f_K(m)$, $p_K(-m) = (-1)^{K+1} p_K(m)$. Nothing here is an exhaustive computation. □

Remark 6.4 (one step deliberately left out). The specialisation of Theorem 6.3 to $N = AB$ needs the fact that a product of two positive semidefinite matrices has real spectrum. That fact is standard, but it is *not* part of the machine-checked development: at 2×2 it is the polynomial inequality $(ap + 2bq + cr)^2 \geq 4(ac - b^2)(pr - q^2)$ for $\left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right), \left(\begin{smallmatrix} p & q \\ q & r \end{smallmatrix}\right) \succeq 0$, which is true but tight along a curve (equality at $a = c = p = r = 1$, $b = -q$), so the only elementary proof we found goes through $(\sqrt{ap} + \sqrt{cr})^2 \geq 4|bq|$ and needs square roots. The formal statement therefore carries $(\operatorname{tr} N)^2 \geq 4 \det N$ as a hypothesis rather than deriving it.

Changes from version 1 of this paper. Nothing stated in version 1 is withdrawn. Every lemma, theorem and proposition of that version stands here with its statement unchanged, with a single exception: Proposition 4.8, whose upper bound moves from $1 - \frac{1}{84375000}$ to $1 - \frac{1}{389765}$. Version 1's bound is true and is the weaker one; it is still recorded in the statement, and the theorem that gives it, Theorem 4.1, is unchanged. Among the remarks, one is rewritten and one gains a clause. The rewritten one is Remark 4.9, which in version 1 reported a better-conditioned witness as a computation outside the formal development and closed by saying that a further one had since been certified. That further witness is now Theorem 4.3; the intermediate one is superseded and is kept in the remark only as a step of the search, which is all that remark now carries. The clause is in Remark 5.2, recording the margin of the second witness. What is new: Section 4.2 in its entirety — Theorem 4.3, Propositions 4.5 and 4.7, Remarks 4.4 and 4.6 — together with the two items about it in Section 1, the corresponding sentences of the abstract, the paragraph of Section 7 on the pipeline that generates the new module, and one bibliography entry, [8]. Two reports are strengthened rather than corrected: the corpus statement in Section 1, where version 1 reported a pass over recent preprints and this version reports the completed pass over all 369 shards of the corpus; and the counts in Section 7, which grow with the fourth module and are stated under version 1's counting convention.

7. Verification

Every statement of Sections 2–6 that is asserted as a lemma, proposition or theorem — Lemmas 2.1 and 2.2, Theorems 3.1, 4.1, 4.3, 6.1, 6.3, Propositions 3.3, 4.5 and 4.7, and the upper half of Proposition 4.8 — has been formally verified in Lean 4 against *Mathlib* [16, 17]. The development is four modules and 1770 lines carrying 140 theorems and 40 definitions; it contains no `sorry` and declares no axiom by hand. A print of the axiom set of each of the 32 headline declarations — 20 for the first witness and the two halves that hold, 12 for the second witness — returns `propext`, `Classical.choice` and `Quot.sound` and nothing else — in particular no `sorryAx` and, unlike many certificate-carrying developments, **no compiled-evaluation axiom**: every integer certificate is reduced by the Lean kernel itself through `decide`, so the counterexamples rest on kernel reduction alone and on no trusted compiled code. Two engineering points are worth recording because they shaped the certificates. First, the same $LDL^\top$ identities over $\mathbb{Q}$ are *not* kernel-reducible: `decide` on a 4×4 rational matrix equation stalls on the normalisation inside rational addition, at any recursion depth. Clearing denominators into $\mathbb{Z}$ removes the problem entirely, which is why every certificate in this paper is an integer identity $N \cdot M = U \operatorname{diag}(e) U^\top$ with $N > 0$ and $e \geq 0$. Second, bounding $S = \frac{1}{2}(AB + BA)$ and recovering $W_{\text{rs}} = S^2 \preceq \rho^2 I$ from Lemma 2.2, rather than bounding W_{rs} directly, whose entries are of the order of the squares of those of S, shrinks the certificate multipliers by an order of magnitude; the largest integer appearing in the first three modules has 42 digits, and in the module for the second witness 78.

Every numeral in the module for the second witness is a function of the fourteen integers of (6) — the four diagonal entries of A and the ten upper-triangular entries of B: two scripts and a fixed template, which carries the prose and nothing else, take that pair to the complete module, computing S, W_{ss}, the least integer ρ_1, the violating vector, the $LDL^\top$ certificates, the normalisation $(s_B, 389765)$ chosen for certificate size, the two eigenvector-derived vectors of Proposition 4.5(ii) and the controls of Proposition 4.7. Run again in a clean directory from those integers alone, they reproduce the committed module byte for byte. The same pipeline was first run on the pair (5) and reproduced that witness's ρ, certificates and condition numbers digit for digit before it was used on anything new.

Several things quoted above are deliberately outside the formal development, and are labelled as such where they appear: the numerical values of Remarks 3.2, 4.2, 4.4, 4.6, 4.9, 5.1, 5.2 and 5.3 — ratios, eigenvalues, condition numbers, and the search statistics — and the lower half $\eta = \frac{1}{4}$ of Proposition 4.8, which is quoted from Theorem 3 of [1] and not reproved. Separately, for both witnesses the certificate literals were parsed back out of the Lean source and recomputed from A and B alone in exact integer arithmetic by an independent script — 60 such checks for the second witness — which closes the one failure mode a kernel-checked certificate cannot catch by itself: a typographical error that is *consistent* between a matrix and its certificate. All such checks pass.

The source of the development is currently held in a private repository: repository reference to be added at submission.

On the reference list. Each entry records how far the reading went. The five classical records [11, 12, 13, 14, 15] are cited for attribution only; each was verified at Crossref on 3 September 2026, and none of those papers was read here.

References

[1] C. Yun, S. Sra and A. Jadbabaie, *Can single-shuffle SGD be better than reshuffling SGD and GD?*, arXiv:2103.07079v1 [cs.LG], 12 March 2021; published as *Open Problem: Can Single-Shuffle SGD be Better than Reshuffling SGD and*

GD?, in: Proceedings of the Thirty Fourth Conference on Learning Theory, Proceedings of Machine Learning Research **134**, PMLR, 2021, 4653–4658. Read here in full from the arXiv LaTeX e-print of v1; all quotations above are from that source, Conjecture 1 and the two $n = 2$ theorems from its statement files and the belief sentence from the paragraph following Theorem 4. Every theorem, lemma and section number attributed to it above is v1's, obtained by compiling that e-print. arXiv record checked live on 3 September 2026: still one version, journal-reference field empty; the proceedings record above, including the changed title, was read off the PMLR volume page the same day.

[2] B. Peng, *A resolution of the SS–RS–GD inequalities*, arXiv:2607.22620v1 [math.OC], 16 June 2026. Its Theorem 2 refutes the first inequality of Conjecture 1 at $n = 3$, $K = 2$, $d = 4$ for every η, and its Theorem 3 proves the second for every $n \geq 2$ under $(1 - \frac{1}{4n^2+1})I \preceq A_i \preceq I$; it does not treat the $n = 2$ single-shuffle inequality. The two theorem numbers are v1's, obtained by compiling the e-print — the paper's theorem counter is shared with the conjecture it quotes. Checked live on 3 September 2026: still one version, no journal reference.

[3] B. Recht and C. Ré, *Toward a noncommutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences*, in: Proceedings of the 25th Annual Conference on Learning Theory, Proceedings of Machine Learning Research **23**, PMLR, 2012, 11.1–11.24. The arXiv preprint carries a different title: *Beneath the valley of the non-commutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences*, arXiv:1202.4184. Volume, page range and booktitle read off the PMLR proceedings entry on 3 September 2026. Of the preprint, the $n = 2$ section was read here, for the statement used above and for its attribution to [11, 12]; the rest was not.

[4] T. Zhang, *A note on the non-commutative arithmetic-geometric mean inequality*, arXiv:1411.5058v3; published, under the changed title *A note on the matrix arithmetic-geometric mean inequality*, in Electron. J. Linear Algebra **34** (2018), 283–287, doi:10.13001/1081-3810.3555. Proves the Recht–Ré inequality for products of three matrices when the number of matrices is a multiple of three, which includes the case $n = 3$ used above. Title, authors and version verified against the arXiv record, and the journal record against Crossref, on 3 September 2026; only the abstract was read here.

[5] A. Israel, F. Krahmer and R. Ward, *An arithmetic-geometric mean inequality for products of three matrices*, arXiv:1411.0333v4; Linear Algebra Appl. **488** (2016), 1–12, doi:10.1016/j.laa.2015.09.013. Its main theorem proves, for products of up to three matrices, arbitrarily many matrices and every unitarily invariant norm, the *expectation-of-norm* variant of the conjecture — the norms inside the two averages — which the paper itself distinguishes from the norm-of-expectation form used here, and which is how [1] classify it. Title, authors and journal record verified on 3 September 2026; the abstract and the statements of the two conjectures and of the main theorem were read here, the proofs were not.

[6] Z. Lai and L.-H. Lim, *Recht–Ré noncommutative arithmetic-geometric mean conjecture is false*, arXiv:2006.01510; in: Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research **119**, PMLR, 2020, 5608–5617. The counterexample is nonconstructive and is obtained at $n = 5$; the same paper proves the conjecture for products of two and of three matrices with arbitrarily many matrices, extending [4]. The arXiv record's own journal reference reads "PMLR 108", which is wrong — volume 108 is a different conference — and the volume, pages and booktitle above were read off the PMLR proceedings entry on 3 September 2026, in agreement with dblp. The abstract and the statement of the $m \leq 3$ results were read here.

[7] W. Albar, M. Junge and M. Zhao, *On the symmetrized arithmetic-geometric mean inequality for operators*, arXiv:1803.02435. Shows that the constant in the symmetrised inequality of [3] exceeds 1, disproving its most optimistic form, and predates [6]. Authors and identifier verified against the arXiv record on 3 September 2026; the title is the e-print's own, the arXiv metadata record carrying a typographical error in the last word. Only the abstract was read here.

[8] W. Albar, M. Junge and M. Zhao, *Noncommutative versions of the arithmetic-geometric mean inequality*, arXiv:1703.00546v1 [math.OA], 1 March 2017. Returned, together with [3, 5, 7], by a sweep of the arXiv abstracts for the noncommutative arithmetic–geometric mean inequality carried out for this version; by its abstract it proves arithmetic–geometric mean inequalities with a constant independent of the dimension for general operators, together with an order version under an additional hypothesis. That is the neighbourhood of the second inequality of Conjecture 1, of Theorem 6.1 and of [9], and it says nothing about single shuffling. Identifier, title, author list and version read off the arXiv record on 3 September 2026, where the journal-reference field is empty; only the abstract was read here.

[9] X. Liu, *A note on arithmetic-geometric mean inequality for well-conditioned matrices*, in: 2025 59th Annual Conference on Information Sciences and Systems (CISS), IEEE, 2025, 1–3, doi:10.1109/CISS64860.2025.10944733. The only work found that responds to the open problem of [1]: it establishes a *weak* matrix arithmetic-geometric mean inequality for positive semidefinite matrices of small condition number — that is, the second inequality of Conjecture 1 for general n, in the neighbourhood of Theorem 3 of [2] and of Theorem 6.1 above — and does not treat the $n = 2$ single-shuffle inequality. Record verified at Crossref and abstract fetched 3 September 2026; there is no arXiv version.

[10] C. Pearcy, *An elementary proof of the power inequality for the numerical radius*, Michigan Math. J. **13** (1966), no. 3, 289–291, doi:10.1307/mmj/1031732779. The proof of the power inequality that [1] cite. Record verified against the publisher's page and zbMATH on 3 September 2026; the paper was not read here.

[11] R. Bhatia and F. Kittaneh, *Notes on matrix arithmetic–geometric mean inequalities*, Linear Algebra Appl. **308** (2000), no. 1–3, 203–211, doi:10.1016/S0024-3795(00)00048-3. Cited only as the source to which [3] attribute $\|AB\| \leq \|\frac{1}{2}A + \frac{1}{2}B\|^2$.

[12] R. Bhatia and F. Kittaneh, *On the singular values of a product of operators*, SIAM J. Matrix Anal. Appl. **11** (1990), no. 2, 272–277, doi:10.1137/0611018. Cited only as the paper [3] name as the start of that line of work.

[13] O. Toeplitz, *Das algebraische Analogon zu einem Satze von Fejér*, Math. Z. **2** (1918), no. 1–2, 187–197, doi:10.1007/BF01212904.

[14] F. D. Murnaghan, *On the field of values of a square matrix*, Proc. Natl. Acad. Sci. USA **18** (1932), no. 3, 246–248, doi:10.1073/pnas.18.3.246.

[15] C.-K. Li, *A simple proof of the elliptical range theorem*, Proc. Amer. Math. Soc. **124** (1996), no. 7, 1985–1986, doi:10.1090/S0002-9939-96-03307-2.

[16] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, 2021, 625–635, doi:10.1007/978-3-030-79876-5_37.

[17] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381, doi:10.1145/3372885.3373824.