

# Empirical Investigation of an Open Conjecture: Can Single-Shuffle SGD be Better than Reshuffling SGD and GD?

Agentic NL→Lean 4 Pipeline  
Job #20

April 26, 2026

## Abstract

This report documents the empirical investigation of an open mathematical conjecture that could not be formally proved or disproved in Lean 4 with Mathlib. Numerical experiments were conducted to gather evidence for or against the conjecture. The empirical verdict is: **Empirically Supported**. The conjecture remains formally open.

## 1 Conjecture Statement

### Conjecture 1.

*Can Single-Shuffle SGD be Better than Reshuffling SGD and GD?*

Let  $n \geq 2$ ,  $K \geq 1$ , and  $d \geq 1$ . Let  $S_n$  denote the set of all permutations of  $\{1, \dots, n\}$ . For real symmetric  $d \times d$  matrices  $X, Y$ , write  $X \preceq Y$  for the Loewner order (i.e.,  $Y - X$  is positive semidefinite), and let  $\|\cdot\|_2$  be the spectral (operator) norm.

Given symmetric matrices  $A_1, \dots, A_n$ , define for each  $S_n$  the without-replacement product  $P_{S_n} := \prod_{i=1}^n A_{\{i\}} = A_{\{n\}} \cdots A_{\{1\}}$ .

Define:

$$\begin{aligned} W_{SS} &:= (1/n!) \sum_{S_n} (P_{S_n})^K && \text{(single-shuffle)} \\ W_{RS} &:= ((1/n!) \sum_{S_n} P_{S_n})^K && \text{(random reshuffling)} \\ W_{GD} &:= ((1/n) \sum_{i=1}^n A_i)^{nK} && \text{(gradient descent)} \end{aligned}$$

**CONJECTURE:** For every  $n \geq 2$  and  $K \geq 1$  there exists a constant  $c_{n,K} \in (0, 1]$  such that, whenever  $(1 - c_{n,K})I \preceq A_i \preceq I$  for all  $i \in \{1, \dots, n\}$ , one has:

$$\|W_{SS}\|_2 \leq \|W_{RS}\|_2 \leq \|W_{GD}\|_2$$

## 2 Status

**Formal Status:** OPEN — no Lean 4 proof or disproof was found.

**Empirical Verdict:** **Empirically Supported**

The pipeline attempted formal verification in Lean 4 with Mathlib but was unable to produce a compiling proof or disproof. Empirical testing was then conducted to gather numerical evidence.

### 3 Basic Empirical Testing

The following output was produced by the basic numerical experiment:

```
=== EXPERIMENT PLAN ===

We investigate the conjecture that, for any  $n \geq 2$ ,  $K \geq 1$ , there is a
threshold  $\eta_{\{n,K\}}$  in  $(0,1]$  so that whenever  $(1-\eta) I \leq A_i \leq I$  (
  symmetric)
for all  $i$  in  $\{1, \dots, n\}$ :

   $\|W_{SS}\|_2 \leq \|W_{RS}\|_2 \leq \|W_{GD}\|_2$ ,

with the three operators defined as:
   $W_{SS} := (1/n!) \sum_{\sigma} P_{\sigma}^K$ 
   $W_{RS} := ((1/n!) \sum_{\sigma} P_{\sigma})^K$ 
   $W_{GD} := ((1/n) \sum_i A_i)^{\{nK\}}$ 
and  $P_{\sigma} = A_{\{\sigma(n)\}} * \dots * A_{\{\sigma(1)\}}$ .

Experimental strategy (multiple angles):

(E1) Random sampling: For many  $(n,K,d,\eta)$  configurations, sample
  symmetric
  matrices  $A_i$  with eigenvalues in  $[1-\eta, 1]$ . Compute  $W_{SS}$ ,  $W_{RS}$ ,  $W_{GD}$ 
  exactly (all  $n!$  permutations). Record:
     $gap_{SS_{RS}} = \|W_{RS}\| - \|W_{SS}\|$  (should be  $\geq 0$ )
     $gap_{RS_{GD}} = \|W_{GD}\| - \|W_{RS}\|$  (should be  $\geq 0$ )

(E2) Threshold search: For each  $(n,K)$ , scan  $\eta$  from small to large.
  Record the empirical success rate as a function of  $\eta$  to see whether
  a threshold  $\eta_{\{n,K\}}$  in  $(0,1]$  exists.

(E3) Adversarial / counterexample search: Use many random trials with
  *small*  $\eta$  (where conjecture should hold) and also *large*  $\eta$ 
  (where it may fail). Also a local optimization (random restarts +
  simulated-annealing style perturbation) trying to MINIMIZE
   $(gap_{SS_{RS}}, gap_{RS_{GD}})$  to break the conjecture at small  $\eta$ .

(E4) Asymptotic check: Vary  $d$  and  $K$  to see that violations (if any) remain
  absent in small- $\eta$  regime as dimension and power grow.

(E5) Known-easy baseline: Commuting (simultaneously diagonalizable)
  matrices --  $W_{SS} = W_{RS} = W_{GD}$ . Used as a sanity test.

We report counts, worst-case gaps, and plots. The conjecture only promises
*some*  $\eta_{\{n,K\}}$ ; so a "success" is that at  $\eta$  sufficiently small the
inequalities hold in all trials.

--- (E5) Sanity check: commuting (diagonal) matrices ---
 $\|W_{SS}\|=0.68791396$   $\|W_{RS}\|=0.68791396$   $\|W_{GD}\|=0.69526845$ 
max pairwise diff =  $7.35e-03$ 

--- (E1) Random sampling sweep over  $(n,K,d,\eta)$  ---
```

```

Total E1 random trials: 59400
  n  K  d   eta    T   min gap SR   min gap RG  viol_SR  viol_RG
  2  1  2  0.02   800   0.000e+00   1.915e-09     0         0
  2  1  2  0.10   800   0.000e+00   5.991e-07     0         0
  2  1  2  0.25   800   0.000e+00   2.153e-06     0         0
  2  1  2  0.50   800   0.000e+00   2.519e-06     0         0
  2  1  2  0.90   800   0.000e+00   2.234e-05     0         0
  2  1  3  0.02   800   0.000e+00   2.274e-08     0         0
  2  1  3  0.10   800   0.000e+00   2.200e-06     0         0
  2  1  3  0.25   800   0.000e+00   2.655e-06     0         0
  2  1  3  0.50   800   0.000e+00   2.966e-05     0         0
  2  1  3  0.90   800   0.000e+00   1.958e-05     0         0
  2  1  5  0.02   800   0.000e+00   9.425e-08     0         0
  2  1  5  0.10   800   0.000e+00   6.268e-06     0         0
  2  1  5  0.25   800   0.000e+00   5.229e-06     0         0
  2  1  5  0.50   800   0.000e+00   1.261e-04     0         0
  2  1  5  0.90   800   0.000e+00   2.085e-04     0         0
  2  2  2  0.02   800   0.000e+00   5.992e-09     0         0
  2  2  2  0.10   800   6.661e-16   1.221e-06     0         0
  2  2  2  0.25   800   6.661e-16   2.932e-07     0         0
  2  2  2  0.50   800   6.914e-14   5.137e-06     0         0
  2  2  2  0.90   800   4.309e-13   1.813e-07     0         0
  2  2  3  0.02   800   6.994e-15   8.903e-08     0         0
  2  2  3  0.10   800   4.571e-12   2.276e-06     0         0
  2  2  3  0.25   800   5.207e-11   9.177e-06     0         0
  2  2  3  0.50   800   8.265e-10   1.842e-05     0         0
  2  2  3  0.90   800   5.291e-09   2.950e-05     0         0
  2  2  5  0.02   800   8.539e-13   6.915e-07     0         0
  2  2  5  0.10   800   4.154e-10   1.053e-05     0         0
  2  2  5  0.25   800   1.660e-08   5.307e-05     0         0
  2  2  5  0.50   800   8.440e-08   1.014e-04     0         0
  2  2  5  0.90   800   1
... [truncated]

```

## 4 Experiment Code (Basic)

```

"""
Empirical test for the COLT 2021 open conjecture:
 $\|W_{SS}\|_2 \leq \|W_{RS}\|_2 \leq \|W_{GD}\|_2$ 
for symmetric  $A_i$  satisfying  $(1 - \eta) I \leq A_i \leq I$ , with  $\eta$  small enough.
"""

import math
import itertools
import random
import numpy as np
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
from numpy.linalg import matrix_power

```

```

rng = np.random.default_rng(20260421)
random.seed(20260421)

# -----
print("===_EXPERIMENT_PLAN_===")
print("""
We investigate the conjecture that, for any  $n \geq 2$ ,  $K \geq 1$ , there is a
threshold  $\eta_{\{n,K\}}$  in  $(0,1]$  so that whenever  $(1-\eta) I \leq A_i \leq I$  (
    symmetric)
for all  $i$  in  $\{1, \dots, n\}$ :

     $\|W_{SS}\|_2 \leq \|W_{RS}\|_2 \leq \|W_{GD}\|_2$ ,

with the three operators defined as:
     $W_{SS} := (1/n!) \sum_{\sigma} P_{\sigma}^{-K}$ 
     $W_{RS} := ((1/n!) \sum_{\sigma} P_{\sigma})^{-K}$ 
     $W_{GD} := ((1/n) \sum_i A_i)^{-\{nK\}}$ 
and  $P_{\sigma} = A_{\{\sigma(n)\}} * \dots * A_{\{\sigma(1)\}}$ .

Experimental strategy (multiple angles):

(E1) Random sampling: For many  $(n,K,d,\eta)$  configurations, sample
    symmetric
    matrices  $A_i$  with eigenvalues in  $[1-\eta, 1]$ . Compute  $W_{SS}$ ,  $W_{RS}$ ,  $W_{GD}$ 
    exactly (all  $n!$  permutations). Record:
         $gap_{SS_{RS}} = \|W_{RS}\| - \|W_{SS}\|$  (should be  $\geq 0$ )
         $gap_{RS_{GD}} = \|W_{GD}\| - \|W_{RS}\|$  (should be  $\geq 0$ )

(E2) Threshold search: For each  $(n,K)$ , scan  $\eta$  from small to large.
    Record the empirical success rate as a function of  $\eta$  to see whether
    a threshold  $\eta_{\{n,K\}}$  in  $(0,1]$  exists.

(E3) Adversarial / counterexample search: Use many random trials with
    *small*  $\eta$  (where conjecture should hold) and also *large*  $\eta$ 
    (where it may fail). Also a local optimization (random restarts +
    simulated-annealing style perturbation) trying to MINIMIZE
     $(gap_{SS_{RS}}, gap_{RS_{GD}})$  to break the conjecture at small  $\eta$ .

(E4) Asymptotic check: Vary  $d$  and  $K$  to see that violations (if any) remain
    absent in small- $\eta$  regime as dimension and power grow.

(E5) Known-easy baseline: Commuting (simultaneously diagonalizable)
    matrices --  $W_{SS} = W_{RS} = W_{GD}$ . Used as a sanity test.

We report counts, worst-case gaps, and plots. The conjecture only promises
*some*  $\eta_{\{n,K\}}$ ; so a "success" is that at  $\eta$  sufficiently small the
inequalities hold in all trials.
""")

# -----
# Helpers
# -----
def random_sym_in_interval(d, eta, rng):
    """Symmetric  $d \times d$  with all eigenvalues in  $[1-\eta, 1]$ ."""

```

```

# random orthogonal via QR
G = rng.standard_normal((d, d))
Q, R = np.linalg.qr(G)
# fix sign to make Q uniform on O(d)
Q = Q * np.sign(np.diag(R))
evals = 1.0 - eta * rng.random(d) # in [1-eta, 1]
return (Q * evals) @ Q.T

def spectral_norm(M):
# M is complex-or-real; we use 2-norm (largest singular value)
return np.linalg.norm(M, ord=2)

def compute_W(A_list, K):
"""Return ||W_SS||, ||W_RS||, ||W_GD||."""
n = len(A_list)
d = A_list[0].shape[0]
# all permutations
perms = list(itertools.permutations(range(n)))
# P_sigma = A_{sigma(n)} ... A_{sigma(1)}; i.e. multiply left-to-right
# using sigma(1), then sigma(2), ... so result = A_{sigma(n)} @ ... @ A_{sigma(1)}.
# Equivalent: start with I, then for idx in sigma left-to-right do M =
# A[idx] @ M
P_sum = np.zeros((d, d))
PK_sum = np.zeros((d, d))
for sigma in perms:
M = np.eye(d)
for idx in sigma:
M = A_list[idx] @ M
P_sum += M
PK_sum += np.linalg.matrix_power(M, K)
invfact = 1.0 / math.factorial(n)
W_SS = invfact * PK_sum
W_RS = np.linalg.matrix_power(invfact * P_sum, K)
Abar = sum(A_list) / n
W_GD = np.linalg.matrix_power(Abar, n * K)
return spectral_norm(W_SS), spectral_norm(W_RS), spectral_norm(W_GD)

# -----
# (E5) Sanity check on commuting matrices
# -----
print("\n---(E5)Sanitycheck:commuting(diagonal)matrices---")
d_sanity = 4
n_sanity = 3
K_sanity = 2
eta_sanity = 0.3
# all diagonal -> commute -> all three should be equal
A_list = []
for _ in range(n_sanity):
diag = 1.0 - eta_sanity * rng.random(d_sanity)
A_list.append(np.diag(diag))
ss, rs, gd = compute_W(A_list, K_sanity)
print(f"||W_SS||={ss:.8f}||W_RS||={rs:.8f}||W_GD||={gd:.8f}")
print(f"maxpairwise_diff={max(abs(ss-rs),abs(rs-gd),abs(ss-gd)):.2e}")

```

```

# -----
# (E1) Random sampling across many configs
# -----
print("\n--- (E1) Random sampling sweep over (n,K,d,eta) ---")
configs = []
for n in (2, 3, 4):
    for K in (1, 2, 3):
        for d in (2, 3, 5):
            for eta in (0.02, 0.1, 0.25, 0.5, 0.9):
                configs.append((n, K, d, eta))

# number of trials per config
# ... [truncated]

```

## 5 Conclusion

The conjecture remains formally open. Numerical experiments **support** the conjecture — no counterexamples were found across all tested parameter ranges. Further investigation (both formal and empirical) is warranted.