Empirical Investigation of an Open Conjecture: Anytime Convergence Rate of Gradient Descent

Agentic NL→Lean 4 Pipeline
Job #46

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: **Inconclusive**. The conjecture remains formally open.

1 Conjecture Statement

Conjecture 1.

Anytime Convergence Rate of Gradient Descent

Let

: →

f:R
d→
R be a convex differentiable function with

L-Lipschitz gradient ("

L-smooth"), i.e.,

(

)-

(

) -

```
f(x)-f(y)  -Lxy for all

,

x,y. Assume

f attains its minimum, and fix an initialization

0



x
0


R
d
 with some minimizer *



arg

min


x*

argminf and value *


=

( *



)
f*

=f(x*

). Consider vanilla gradient descent with a predetermined (oblivious)
    stepsize sequence
(
```

)

0
(
t

)
t0

(possibly depending on

L but not on the stopping time

T):

+
1
= -

(

)
,

=
0
,
1
,
2
,…

.
x
t+1

=x
t-

```
t


f(x
t


),t…=0,1,2,.

The classical worst-case guarantee for suitable constant stepsizes gives an
    anytime bound of order

(


)- *










0- *





2
/

f(x
T


)-f*

CLx
0-


x*



2
/T holding for all




TN and all
```

```
L-smooth convex

f, with a universal constant

C.

Unsolved Problem

Do stepsizes alone yield a strictly faster worst-case anytime rate on the
    actual iterate


x
T


? Equivalently, does there exist a stepsize sequence
(


)


0
(
t


)
t0


 and an exponent

>
1
>1 such that for some universal constant

<∞

C∞<, for every dimension

d, every

L-smooth convex

f attaining its minimum, every initialization

0
x
0


, every choice of minimizer *
```

```
arg

min


x*

argminf, and every



TN,


(


)- *









0- *




2


?
f(x
T


)-f*

C
T

Lx
0-



x*
```

2

?

More generally, what is the best function

(

)
r(T) for which there exists an oblivious stepsize schedule such that

(

)- *

0- *

2

(

)
f(x
T

)-f*

CLx
0-

x*

2
r(T) holds simultaneously for all

T and all

```
L-smooth convex

f?

Solution Claims

Accepted claims are public. Pending claims are visible only to the claimant
    and site administrators.
```

2 Status

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

Empirical Verdict: Inconclusive

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 ===

Conjecture (COLT 2024 open problem, anytime convergence rate of GD):
    There exists an OBLIVIOUS stepsize sequence (eta_t)_{t>=0} (depending on
        L
    but not on the stopping time T) and an exponent alpha > 1 such that, for
        a
    universal constant C, for EVERY L-smooth convex f with a minimum, every
       x0,
    every minimizer x*, and every T in N:
                  f(x_T) - f* <= C * L * ||x0 - x*||^2 / T^alpha .

We empirically probe this from MULTIPLE angles:

  Test 1 (canonical worst case):  Nesterov's tridiagonal hard quadratic.
        This instance saturates the Omega(1/T) lower bound for plain GD with
        constant stepsize, so it is a sharp test for any candidate schedule.

  Test 2 (oblivious schedule families):  constant 1/L, polynomial decay
        eta_t = 1/(L (t+1)^beta) for several beta, and several "long-step"
        schedules that insert larger steps at indices 2^k - 1 (the natural
        anytime extension of the silver-stepsize idea of Altschuler-Parrilo
           &
        Grimmer-Shu-Wang). The long-step schedules are the only known way to
        break the 1/T barrier.

  Test 3 (horizon-tuned benchmark):  for each T, run a horizon-dependent
        (NON-oblivious) silver-style schedule of length T. This is NOT what
        the conjecture asks for, but provides an upper-envelope benchmark.
```

```
  Test 4 (robustness):  random quadratics with varied condition numbers and
        seeds, to check whether the empirical exponent on Nesterov's
           instance
        carries over to other smooth convex problems.

  Test 5 (counterexample search):  for each oblivious schedule, examine
     whether
        sup_T  T^alpha * (f(x_T)-f*) / (L*R^2)  is bounded for alpha = 1.05.
        If, for every tested schedule, this ratio increases with T, we have
        empirical evidence AGAINST the conjecture (no oblivious schedule
           among
        the tested family exhibits alpha > 1).

  Test 6 (asymptotic exponent):  log-log linear regression of normalized
        suboptimality against T to extract empirical alpha.

--- Setup ---
L = 1.0,   d = 600,   ||x0-x*||^2 = R^2 = 199.8336,   f* = -0.1248
T grid: [1 2 3 4] ... [170 208 256] (T_max = 256)

--- Empirical convergence exponent on Nesterov hard quadratic ---
Schedule                      alpha_emp        C*max    subopt(T_max)/(LR^2)
constant 1/L                     0.4763       0.0005               3.011e-05
decay beta=0.10                  0.4320       0.0004               3.799e-05
decay beta=0.25                  0.3681       0.0004               5.308e-05
decay beta=0.50                  0.2697       0.0004               8.875e-05
longstep f=1.5 anytime           0.4613       0.0004               2.987e-05
longstep f=2.0 anytime           0.4487       0.0004               2.964e-05
longstep f=2.4 anytime           0.4400       0.0004               2.945e-05
growing-longstep anyt.           0.4577       0.0004               2.811e-05
horizon-silver (NON-oblivious)     0.5122   (benchmark; not anytime)

--- Robustness: random quadratics (kappa in {50, 200, 1000}, 4 seeds each)
    ---
  constant 1/L                 alpha mean=1.051  range=[0.120, 2.514]  (n
     =12)
  decay beta=0.10              alpha mean=0.778  range=[0.092, 1.749]  (n
     =12)
  decay beta=0.25              alpha mean=0.514  range=[0.066, 1.113]  (n
     =12)
  decay beta=0.50              alpha mean=0.286  range=[0.042, 0.663]  (n
     =12)
  longstep f=1.5 anytime       alpha mean=1.053  range=[0.119, 2.543]  (n
     =12)
  longstep f=2.0 anytime       alpha mean=1.056  range=[0.118, 2.573]  (n
     =12)
  longstep f=2.4 anytime       alpha mean=1.061  range=[0.118, 2.600]  (n
     =12)
  growing-longstep anyt.       alpha mean=1.143  range=[0.127, 2.820]  (n
     =12)

--- Counterexample probe: is sup_T T^alpha * (f(x_T)-f*)/(L R^2) bounded?
    ---
```

```
Using alpha = 1.05 on Nesterov hard quadratic.
Schedule                          C(small T)   C(large T)     growth
constant 1/L                       3.899e-04    1.017e-02      26.09x
decay beta=0.10
... [truncated]
```

4 Advanced Empirical Testing

A research-grade experiment was designed with nonlinear analysis, parameter sweeps, and convergence testing. Output:

```
=== ADVANCED EXPERIMENT PLAN ===

We test whether OBLIVIOUS, ANYTIME stepsize schedules for vanilla gradient
    descent
on L-smooth convex functions can achieve worst-case f(x_T) - f^* <= C L R^2
    / T^alpha
with alpha > 1, beyond the classical 1/T rate.

Beyond the basic experiment we add:
 (1) Multiple problem classes:
      (a) Nesterov tridiagonal hard quadratic at d in {25,100,300} (
          canonical worst
           case for first-order methods, dimension-resolution convergence
               test).
      (b) Drori 1D Huber-like NON-quadratic (the actual analytic worst case
          for
           constant 1/L GD via Drori-Teboulle PEP).
      (c) Symmetric Log-Sum-Exp -- highly nonlinear, non-quadratic, smooth
          convex.
 (2) Modern advanced schedules:
      - Constant 1/L (classical baseline)
      - Polynomial decay (t+1)^(-beta) for beta in {0.25, 0.5}
      - Long-step periodic (Grimmer-Shu-Wang style 7-period pattern)
      - Anytime nested-silver (periodic block of Altschuler-Parrilo silver)
      - Silver-horizon (NON-anytime benchmark; theoretical alpha=log2(1+sqrt
          (2))~1.272)
 (3) Convergence test across dimensions d in {25, 100, 300} for Nesterov to
    assess
     dimension-free behavior; rates should be d-independent for true worst-
         case.
 (4) Lyapunov / "energy" monitoring: Drori-Teboulle Lyapunov
        V_t = t*(f(x_t)-f^*) + (L/2)||x_t-x^*||^2
     should be monotone non-increasing under any schedule that proves the 1/
         T rate.
     We track its relative drift -- positive drift signals breakdown of the
         proof.
 (5) Counterexample probe: for each schedule and alpha in {1.0, 1.1, 1.272},
     compute max over T of T^alpha * (f(x_T)-f*)/(LR^2) up to T = 2048,
         comparing
     small-T and large-T windows: unbounded growth refutes alpha at that
         exponent.
```

```
 (6) Comparison with Drori-Teboulle theoretical worst-case 1/(4T+2) for
     constant-1/L
      GD and with the silver schedule's theoretical T^{-log2(1+sqrt(2))}.

Expected behavior:
 - If conjecture TRUE: at least one ANYTIME oblivious schedule shows
     alpha_emp > 1
   robustly across dimensions and bounded T^alpha*(f-f*) up to large T.
 - If conjecture FALSE: only horizon-tuned silver achieves alpha > 1; all
     anytime
   schedules show alpha <= 1 with T*(f-f*) bounded but T^{1+eps}*(f-f*)
       growing.


Solver: vanilla gradient descent (forward Euler).
  T_MAX = 2048, problem set = ['Nesterov-d25', 'Nesterov-d100', 'Nesterov-
      d300', 'Drori-Huber-1D', 'LSE-d80']
  schedules = ['const-1/L', 'decay-0.25', 'decay-0.50', 'long-step-period',
      'nested-silver-p4', 'silver-horizon*']
  schedule 'const-1/L': eta_avg = 1.0000, eta_max = 1.0000, eta_min = 1.0000
  schedule 'decay-0.25': eta_avg = 0.1978, eta_max = 1.0000, eta_min =
      0.1487
  schedule 'decay-0.50': eta_avg = 0.0435, eta_max = 1.0000, eta_min =
      0.0221
  schedule 'long-step-period': eta_avg = 1.7246, eta_max = 2.4142, eta_min =
       1.4142
  schedule 'nested-silver-p4': eta_avg = 2.6124, eta_max = 15.0711, eta_min
      = 1.4142
  schedule 'silver-horizon*': eta_avg = 11.2129, eta_max = 6726.9999,
      eta_min = 1.4142

--- Empirical convergence exponent alpha_emp (fit on T in [T_MAX/4, T_MAX])
   ---
Problem            Schedule                alpha_emp   subopt(T_MAX)/(LR^2)
Nesterov-d25       const-1/L                  8.3521             3.7339e-10
Nesterov-d25       decay-0.25                 1.4446             6.1031e-05
Nesterov-d25       decay-0.50                 0.4175             7.0277e-04
Nesterov-d25       long-step-period          14.4242             7.0399e-15
Nesterov-d25       nested-silver-p4          20.0288            -5.0940e-18
Nesterov-d25       silver-horizon*           12.8578            -5.0940e-18
Nesterov-d100      const-1/L                  0.8643             2.9128e-05
Nesterov-d100      decay-0.25                 0.4689             1.1201e-04
Nesterov-d100      decay-0.50                 0.2826             2.8081e-04
Nesterov-d100      long-step-period           1.1273             1.3589e-05
Nesterov-d100      nested-silver-p4           1.5248             5.6014
... [truncated]
```

5 Experiment Code (Basic)

```
import numpy as np
import matplotlib
```

```
matplotlib.use("Agg")
import matplotlib.pyplot as plt
from scipy.stats import linregress

print("=== EXPERIMENT PLAN ===")
print("""
Conjecture (COLT 2024 open problem, anytime convergence rate of GD):
    There exists an OBLIVIOUS stepsize sequence (eta_t)_{t>=0} (depending on L
    but not on the stopping time T) and an exponent alpha > 1 such that, for a
    universal constant C, for EVERY L-smooth convex f with a minimum, every x0,
    every minimizer x*, and every T in N:
                f(x_T) - f* <= C * L * ||x0 - x*||^2 / T^alpha .

We empirically probe this from MULTIPLE angles:

  Test 1 (canonical worst case):  Nesterov's tridiagonal hard quadratic.
        This instance saturates the Omega(1/T) lower bound for plain GD with
        constant stepsize, so it is a sharp test for any candidate schedule.

  Test 2 (oblivious schedule families):  constant 1/L, polynomial decay
        eta_t = 1/(L (t+1)^beta) for several beta, and several "long-step"
        schedules that insert larger steps at indices 2^k - 1 (the natural
        anytime extension of the silver-stepsize idea of Altschuler-Parrilo &
        Grimmer-Shu-Wang). The long-step schedules are the only known way to
        break the 1/T barrier.

  Test 3 (horizon-tuned benchmark):  for each T, run a horizon-dependent
        (NON-oblivious) silver-style schedule of length T. This is NOT what
        the conjecture asks for, but provides an upper-envelope benchmark.

  Test 4 (robustness):  random quadratics with varied condition numbers and
        seeds, to check whether the empirical exponent on Nesterov's instance
        carries over to other smooth convex problems.

  Test 5 (counterexample search):  for each oblivious schedule, examine whether
        sup_T  T^alpha * (f(x_T)-f*) / (L*R^2)  is bounded for alpha = 1.05.
        If, for every tested schedule, this ratio increases with T, we have
        empirical evidence AGAINST the conjecture (no oblivious schedule among
        the tested family exhibits alpha > 1).

  Test 6 (asymptotic exponent):  log-log linear regression of normalized
        suboptimality against T to extract empirical alpha.
""")

# ---------- Setup ----------
rng = np.random.default_rng(0)
L = 1.0
```

```
T_max = 256
T_grid = np.unique(np.round(np.logspace(0, np.log10(T_max), 28)).astype(int)
    )

# ---------- Test functions ----------
def nesterov_hard(d):
    """Canonical worst-case quadratic: A=(L/4)*tridiag(-1,2,-1), b=(L/4)e_1.
        """
    A = (L/4.0) * (2*np.eye(d) - np.eye(d, k=1) - np.eye(d, k=-1))
    b = np.zeros(d); b[0] = L/4.0
    x_star = np.linalg.solve(A, b)
    f_star = 0.5*x_star @ A @ x_star - b @ x_star
    x0 = np.zeros(d)
    R2 = float(np.sum((x0 - x_star)**2))
    return A, b, x0, x_star, f_star, R2

def random_quadratic(d, kappa=200.0, seed=0):
    rr = np.random.default_rng(seed)
    Q, _ = np.linalg.qr(rr.standard_normal((d, d)))
    eigs = np.linspace(L/kappa, L, d)
    A = Q @ np.diag(eigs) @ Q.T
    A = 0.5*(A + A.T)
    b = rr.standard_normal(d)
    x_star = np.linalg.solve(A, b)
    f_star = 0.5*x_star @ A @ x_star - b @ x_star
    x0 = np.zeros(d)
    R2 = float(np.sum((x0 - x_star)**2))
    return A, b, x0, x_star, f_star, R2

def run_gd(A, b, x0, eta_seq, fmax=1e30):
    T = len(eta_seq)
    x = x0.copy()
    fs = np.empty(T+1)
    fs[0] = 0.5*x @ A @ x - b @ x
    for t in range(T):
        g = A @ x - b
        x = x - eta_seq[t]*g
        v = 0.5*x @ A @ x - b @ x
        if not np.isfinite(v) or abs(v) > fmax:
            fs[t+1:] = fmax
            return fs
        fs[t+1] = v
    return fs

# ---------- Oblivious (anytime) schedules ----------
def sched_const(T): return np.full(T, 1.0/L)
def sched_decay(T, beta): return 1.0/(L*np.arange(1, T+1)**beta)

def sched_longstep_anytime(T, factor):
    """At indices 2^k - 1, insert a long step of size factor/L; else 1/L."""
    eta = np.full(T, 1.0/L)
    k = 1
    while (1 << k) - 1 < T:
        eta[(1 << k) - 1] = factor/L
```

```
        k += 1
    return eta

def sched_growing_longstep(T):
    """Grow long-step factor with k: eta_{2^k-1} = (1 + 2^{(k-1)/2})/L."""
    eta = np.full(T, 1.0/L)
    k = 1
    while (1 << k) - 1 < T:
        eta[(1 << k) - 1] = (1.0 + 2.0**((k-1)/2.0))/L
        k += 1
    return eta

# Horizon-tuned silver-style schedule (NOT oblivious; benchmark only)
def silver_horizon(T):
    if T <= 0: return np.array([])
    K = int(np.ceil(np.log2(T+1)))
    def build(k):
        if k == 0: return [1.0]
        s = build(k-1)
        rho = 1.0 + 2.0**((k-1)/2.0)
        return s + [rho] + s
    full = build(K)
    return np.array(full[:T]) / L

# ---------- Test 1-2: schedules on Nesterov's hard quadratic ----------
d_test = max(2*T_max + 5, 600)
A, b, x0, xs, f_star, R2 = nesterov_hard(d_test)
print(f"--- Setup ---")
print(f"L = {L},   d = {d_test},   ||x0-x*||^2 = R^2 = {R2:.4f},   f* = {
   f_star:.4f}")
print(f"T grid: {T_grid[:4]} ... {T_grid[-3:]} (T_max = {T_max})")

schedules = {

# ... [truncated]
```

6 Experiment Code (Advanced)

```
import numpy as np
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
import math
from itertools import product

print("=== ADVANCED EXPERIMENT PLAN ===")
print("""
We test whether OBLIVIOUS, ANYTIME stepsize schedules for vanilla gradient
   descent
on L-smooth convex functions can achieve worst-case f(x_T) - f^* <= C L R^2
   / T^alpha
```

```
with alpha > 1, beyond the classical 1/T rate.

Beyond the basic experiment we add:
 (1) Multiple problem classes:
      (a) Nesterov tridiagonal hard quadratic at d in {25,100,300} (
          canonical worst
           case for first-order methods, dimension-resolution convergence
                test).
      (b) Drori 1D Huber-like NON-quadratic (the actual analytic worst case
          for
           constant 1/L GD via Drori-Teboulle PEP).
      (c) Symmetric Log-Sum-Exp -- highly nonlinear, non-quadratic, smooth
          convex.
 (2) Modern advanced schedules:
      - Constant 1/L (classical baseline)
      - Polynomial decay (t+1)^(-beta) for beta in {0.25, 0.5}
      - Long-step periodic (Grimmer-Shu-Wang style 7-period pattern)
      - Anytime nested-silver (periodic block of Altschuler-Parrilo silver)
      - Silver-horizon (NON-anytime benchmark; theoretical alpha=log2(1+sqrt
          (2))~1.272)
 (3) Convergence test across dimensions d in {25, 100, 300} for Nesterov to
     assess
     dimension-free behavior; rates should be d-independent for true worst-
          case.
 (4) Lyapunov / "energy" monitoring: Drori-Teboulle Lyapunov
         V_t = t*(f(x_t)-f^*) + (L/2)||x_t-x^*||^2
     should be monotone non-increasing under any schedule that proves the 1/
         T rate.
     We track its relative drift -- positive drift signals breakdown of the
         proof.
 (5) Counterexample probe: for each schedule and alpha in {1.0, 1.1, 1.272},
     compute max over T of T^alpha * (f(x_T)-f*)/(LR^2) up to T = 2048,
         comparing
     small-T and large-T windows: unbounded growth refutes alpha at that
         exponent.
 (6) Comparison with Drori-Teboulle theoretical worst-case 1/(4T+2) for
     constant-1/L
     GD and with the silver schedule's theoretical T^{-log2(1+sqrt(2))}.

Expected behavior:
 - If conjecture TRUE: at least one ANYTIME oblivious schedule shows
     alpha_emp > 1
   robustly across dimensions and bounded T^alpha*(f-f*) up to large T.
 - If conjecture FALSE: only horizon-tuned silver achieves alpha > 1; all
     anytime
   schedules show alpha <= 1 with T*(f-f*) bounded but T^{1+eps}*(f-f*)
       growing.
""")

SQRT2 = np.sqrt(2.0)
RHO = 1.0 + SQRT2  # silver base

# =================== Problems ===================
def make_nesterov(d, L=1.0):
```

```python
    A = np.diag(2.0*np.ones(d)) + np.diag(-1.0*np.ones(d-1), 1) + np.diag
        (-1.0*np.ones(d-1), -1)
    M = (L/4.0) * A
    e1 = np.zeros(d); e1[0] = 1.0
    b = (L/4.0) * e1
    x_star = np.linalg.solve(M, b)
    f_star = 0.5 * x_star @ M @ x_star - b @ x_star
    x0 = np.zeros(d)
    R = np.linalg.norm(x0 - x_star)
    f = lambda x: 0.5 * x @ (M @ x) - b @ x
    g = lambda x: M @ x - b
    return dict(name=f"Nesterov-d{d}", f=f, g=g, x0=x0.copy(),
                x_star=x_star, f_star=f_star, L=L, R=R)

def make_drori_huber(L=1.0, R=10.0):
    def f(x):
        a = abs(x[0])
        return 0.5*L*x[0]**2 if a <= 1.0 else L*a - 0.5*L
    def g(x):
        if abs(x[0]) <= 1.0: return L * x.copy()
        return np.array([L * np.sign(x[0])])
    x0 = np.array([float(R)])
    x_star = np.array([0.0])
    return dict(name="Drori-Huber-1D", f=f, g=g, x0=x0, x_star=x_star,
                f_star=0.0, L=L, R=R)

def make_lse(d, L=1.0, seed=1):
    rng = np.random.default_rng(seed)
    beta = L  # f is at most beta-smooth
    def f(x):
        z = beta * x
        m = z.max()
        return (m + np.log(np.mean(np.exp(z - m))))/beta - np.mean(x)
    def g(x):
        z = beta * x
        m = z.max()
        e = np.exp(z - m)
        p = e / e.sum()
        return p - 1.0/len(x)
    x_star = np.zeros(d)
    f_star = 0.0
    x0 = rng.standard_normal(d) * 1.0
    R = np.linalg.norm(x0 - x_star)
    return dict(name=f"LSE-d{d}", f=f, g=g, x0=x0, x_star=x_star,
                f_star=f_star, L=L, R=R)

# =================== Schedules ===================
def sched_constant(T, L=1.0):
    return np.ones(T) / L

def sched_decay(T, L=1.0, beta=0.25):
    return (1.0/L) * (1.0 + np.arange(T))**(-beta)

def sched_long_step_periodic(T, L=1.0):
```

```python
    pattern = np.array([SQRT2, 2.0, SQRT2, 1.0+SQRT2, SQRT2, 2.0, SQRT2])
    n = T // len(pattern) + 1
    return np.tile(pattern, n)[:T] / L

def sched_silver_horizon(T, L=1.0):
    """Altschuler-Parrilo silver schedule, recursive: silver(p)=silver(p-1)
        +[1+rho^(p-1)]+silver(p-1)
    Length 2^(p+1)-1. Horizon-tuned (NOT anytime)."""
    h = [SQRT2]
    p = 0
    while len(h) < T:
        p += 1
        h = h + [1.0 + RHO**(p-1)] + h
    return np.array(h[:T]) / L

def sched_nested_silver(T, L=1.0, p_block=4):
    """Periodic ANYTIME schedule: repeat silver(p_block) of length 2^(
        p_block+1)-1."""
    block = sched_silver_horizon(2**(p_block+1) - 1, L=1.0)
    n = T // len(block) + 1
    return np.tile(b
# ... [truncated]
```

7 Conclusion

The conjecture remains formally open. Numerical experiments were **inconclusive** — neither strong support nor clear counterexamples were found. Further investigation (both formal and empirical) is warranted.