Escaping Nesterov flows and iterates below critical damping

Dev D. Goyal

30 October 2026; revised 1 November 2026

Abstract

We study zero-velocity solutions of $X''+\frac{r}{t}X'+\nabla F(X)=0$, the continuous model of Nesterov's accelerated gradient method introduced by Su, Boyd and Candès [1], and the corresponding Nesterov iteration with momentum $k/(k+r)$, for convex F with a globally 1-Lipschitz gradient and an attained minimum. For $r \geq 3$ classical energy arguments bound all trajectories and iterates, while Jang and Ryu [8] record the subcritical behavior of trajectories as open. We show that the threshold $r = 3$ is sharp. On $\ell^2(\mathbb{N};\mathbb{R}^2)$, for every $0 < r < 3$ there is such an objective whose flow from rest has $\|X(t)\| \to \infty$, and another whose unit-step iterates have $\|X_k\| \to \infty$. In the plane, the explicit potential $f(x,y) = h(\sqrt{x^2+y^2}) - x$, with h the Huber function, has an escaping trajectory from $(2,1)$ for every $0 < r \leq 2$, while every such trajectory is bounded for $r > 2$. Whether boundedness fails in finite dimension for $2 < r < 3$ remains open. The Hilbert-space and discrete thresholds and the planar escape for $1 < r \leq 2$ are formalized in Lean 4; the remaining planar statements and the existence of solutions for $r \leq 1$ have conventional proofs.

Contents

1 Introduction **2**

2 Statements **3**

3 The planar valley **4**
3.1 The potential 4
3.2 Invariants of the flow 4
3.3 Escape for $1 < r \leq 2$ 4
3.4 Escape for $0 < r \leq 1$ 5
3.5 Boundedness for $r > 2$ 5

4 Global solutions from rest **5**

5 Hilbert space: escape for $1 < r < 3$ **5**
5.1 Moreau valleys 5
5.2 Finite excursions 6
5.3 One objective on ℓ^2 6

6 The positive side and low damping in continuous time **6**

7 The discrete iteration **7**
7.1 Finite-horizon approximation 7
7.2 Planar excursions and permanence 7
7.3 One objective on ℓ^2 7
7.4 The positive side and low damping 7

8 Formal verification **8**

9 Related work and novelty **8**

10 Open questions **9**

11 Research process and AI disclosure **9**

1 Introduction

Su, Boyd and Candès modeled Nesterov's accelerated gradient method by

$$X''(t) + \frac{r}{t}X'(t) + \nabla F(X(t)) = 0, \qquad t > 0, \qquad X'(0) = 0, \tag{1.1}$$

with $r = 3$ [1]. For a convex F with a minimizer, the energy $t^2(F(X) - \min F) + \frac{1}{2}\|tX' + (r-1)(X-z)\|^2$ is nonincreasing when $r \geq 3$, which bounds every trajectory. For $r > 3$ trajectories converge weakly [2, 3]; at $r = 3$ point convergence was proved recently [8, 9]. For $r < 3$ objective values still decrease, at the rate $t^{-2r/3}$ [5, 6], but nothing in this literature bounds the trajectory. Jang and Ryu record the subcritical behavior of trajectories as open in the version of their paper we consulted [8].

We show that the classical threshold $r = 3$ is sharp for boundedness, in three senses.

1. *Continuous time on Hilbert space* (Theorem 2.1). On $H = \ell^2(\mathbb{N};\mathbb{R}^2)$, universal boundedness of zero-velocity solutions holds exactly when $r \geq 3$.
2. *An explicit planar potential* (Theorem 2.2). For $f(x,y) = h(R) - x$, $R = \sqrt{x^2+y^2}$, with the Huber function h, the solution from $(2,1)$ at rest escapes to infinity exactly when $0 < r \leq 2$.
3. *The discrete iteration* (Theorem 2.4). For $X_{k+1} = Y_k - \nabla F(Y_k)$, $Y_{k+1} = X_{k+1} + \frac{k}{k+r}(X_{k+1} - X_k)$, universal boundedness on $\ell^2(\mathbb{N};\mathbb{R}^2)$ holds exactly when $r \geq 3$.

For $0 < r \leq 1$ escape is elementary, since the free motion $X' = ct^{-r}$ already has unbounded displacement. The substantive ranges are $1 < r \leq 2$ in the plane and $1 < r < 3$ in Hilbert space. The Hilbert-space objectives are sums of rescaled two-dimensional valleys, each producing a longer and longer finite excursion; a single planar valley of this type cannot reach $r > 2$ (Remark 3.1). Whether a finite-dimensional objective has an unbounded zero-velocity trajectory for some $2 < r < 3$ is the main open question. In every example the set of minimizers is unbounded, which is necessary in finite dimension.

The positive directions for $r \geq 3$ are classical and the ranges $0 < r \leq 1$ are elementary; new arguments are needed only for the negative directions with $1 < r < 3$ and for the planar statement with $1 < r \leq 2$ (Section 2). Section 9 compares the results with the literature. The continuous-time Hilbert-space threshold, the discrete threshold and the planar escape for $1 < r \leq 2$, including actual gradients, global solutions from rest and norm divergence, are formalized in Lean 4. The planar escape of the same potential for $0 < r \leq 1$, its boundedness for $r > 2$ and the existence of solutions for $r \leq 1$ rest on the conventional proofs given here (Section 8).

2 Statements

Throughout, an *actual solution* of (1.1) on a real Hilbert space H is a pair of continuous maps $X, V : [0,\infty) \to H$ such that $X' = V$ on $[0,\infty)$, with a genuine derivative at $t = 0$, and $V'(t) = -\frac{r}{t}V(t) - \nabla F(X(t))$ for $t > 0$, with $V(0) = 0$. The objectives are convex, everywhere finite and Fréchet differentiable, with a globally 1-Lipschitz gradient.

Theorem 2.1 (Hilbert-space threshold)**.** *Let $H = \ell^2(\mathbb{N};\mathbb{R}^2)$ and $r > 0$. Every actual solution for every convex F with a globally 1-Lipschitz gradient and an attained minimum is bounded if and only if $r \geq 3$. More precisely:*

(a) if $r \geq 3$ and z minimizes F, then $\|X(t) - z\| \leq \|X(0) - z\|$ for all $t \geq 0$, on every real Hilbert space;

(b) if $1 < r < 3$, there is a convex $F \geq 0$ with $F(0) = 0$, a 1-Lipschitz gradient bounded by 2, and an actual solution with $X(0)_n = 2^{-n}(2,1)/16$ and $\|X(t)\| \to \infty$;

(c) if $0 < r \leq 1$, there is such an F with an attained minimum and an actual solution with $\|X(t)\| \to \infty$.

Theorem 2.2 (An explicit planar valley)**.** *Let $h(s) = (s^2+1)/2$ for $0 \leq s \leq 1$, $h(s) = s$ for $s \geq 1$, and $f(x,y) = h(R) - x$ with $R = \sqrt{x^2+y^2}$. Then f is convex, $\nabla f(z) = z/\max\{1, \|z\|\} - (1,0)$ is 1-Lipschitz and bounded by 2, $f \geq 0$, and the minimizers of f form the ray $\{(x,0) : x \geq 1\}$. For every $r > 0$ there is an actual solution of (1.1) with $X(0) = (2,1)$, and for every actual solution with these data:*

(a) if $0 < r \leq 2$, then the first coordinate and the norm of $X(t)$ tend to $+\infty$;

(b) if $r > 2$, then X is bounded and its first coordinate converges.

Corollary 2.3. *In $\mathbb{R}^2$, universal boundedness of zero-velocity solutions fails for every $0 < r \leq 2$.*

For the discrete statement let F be convex with a globally 1-Lipschitz gradient, $r > 0$ and $b \in H$, and define

$$X_0 = Y_0 = b, \qquad X_{k+1} = Y_k - \nabla F(Y_k), \qquad Y_{k+1} = X_{k+1} + \frac{k}{k+r}(X_{k+1} - X_k). \tag{2.1}$$

In the index $n = k+1$ the momentum is $(n-1)/(n+r-1)$, the rational family of Chambolle and Dossal [4]; at $r = 3$ it is the classical $(n-1)/(n+2)$. It is not the square-root recursion used in FISTA.

Theorem 2.4 (Discrete threshold)**.** *Let $H = \ell^2(\mathbb{N};\mathbb{R}^2)$ and $r > 0$. The iterates (2.1) are bounded for every convex F with a globally 1-Lipschitz gradient and an attained minimum, and every b, if and only if $r \geq 3$. If $r \geq 3$ and z minimizes F, then $\|X_k - z\| \leq \|b - z\|$ for all k on every real Hilbert space. If $0 < r < 3$, there are $F \geq 0$ attaining 0, with a 1-Lipschitz gradient bounded by 2, and b with $\|X_k\| \to \infty$; for $1 < r < 3$ one may take $F(0) = 0$ and $b_n = 2^{-n}(2,1)/16$.*

Only the negative directions for $1 < r < 3$ in Theorems 2.1 and 2.4, and the planar statement for $1 < r \leq 2$, require new arguments. The positive directions are classical, and the ranges $0 < r \leq 1$ are elementary.

3 The planar valley

3.1 The potential

Let P be the Euclidean projection onto the closed unit ball. Then $\nabla f(z) = P(z) - (1,0)$, which is 1-Lipschitz because P is nonexpansive, and $\|\nabla f\| \le 2$. The function $z \mapsto h(\|z\|)$ is convex because h is convex and nondecreasing on $[0,\infty)$. Inside the unit ball $f(x,y) = \frac{1}{2}((x-1)^2 + y^2)$, which vanishes only at $(1,0)$; outside, $f = R - x \ge 0$, with equality exactly when $y = 0$ and $x \ge 1$.

3.2 Invariants of the flow

Let $X = (x,y)$ be an actual solution from $(2,1)$ at rest, write $v = x'$, $w = y'$ and $V = R - x \ge 0$. The first coordinate of ∇f is nonpositive, so $(t^r v)' = t^r(1 - x/\max\{1,R\}) \ge 0$; since $t^r v(t) \to 0$ as $t \downarrow 0$, $v \ge 0$ and $x \ge 2$, hence $R \ge 2$. The trajectory therefore stays in the region where

$$v' = -\frac{r}{t}v + \frac{V}{R}, \qquad w' = -\frac{r}{t}w - \frac{y}{R}. \tag{3.1}$$

The energy $A = \frac{1}{2}(v^2 + w^2) + V$ satisfies $A' = -\frac{r}{t}(v^2 + w^2) \le 0$, so with $W = \frac{1}{2}w^2 + V$,

$$0 \le W \le A \le A(0) = \sqrt{5} - 2 < 1. \tag{3.2}$$

Finally $y^2 = R^2 - x^2 = V(R + x)$.

3.3 Escape for $1 < r \le 2$

Define the completed-square energy

$$E = \tfrac{1}{2}(tw + (r-1)y)^2 + t^2 V \;\ge\; t^2 V.$$

From (3.1), $(tw + (r-1)y)' = -ty/R$ and $V' = (-vV + yw)/R$, so

$$E' = t\Big(2V - (r-1)\frac{y^2}{R}\Big) - t^2\frac{vV}{R} = \frac{tV}{R}\big((4-2r)x + (3-r)V\big) - t^2\frac{vV}{R}, \tag{3.3}$$

using $y^2 = V(R+x)$ and $R = x + V$. For $1 < r \le 2$ the bracket is nonnegative, and since $R \ge 2$ and $E \ge t^2 V$,

$$(e^{x/2}E)' = e^{x/2}\Big(E' + \frac{v}{2}E\Big) \ge e^{x/2}\, t^2 v V\Big(\frac{1}{2} - \frac{1}{R}\Big) \ge 0.$$

Hence $J = e^{x/2}E$ is nondecreasing and $J(t) \ge J(0) = e(r-1)^2/2 > 0$.

Suppose now that $x \le M$ for all t, with $M \ge 2$. Then $R \le M + 1$ and $y^2 \le (2M+1)V$ by (3.2). For $t \ge 1$,

$$E \le t^2 w^2 + (r-1)^2 y^2 + t^2 V \le (2M+3)t^2 W,$$

while $E \ge J(0)e^{-M/2}$. Hence $W \ge c/t^2$ for $t \ge 1$, with $c = J(0)e^{-M/2}/(2M+3)$. Put $H = x + y^2/(8(M+1))$. By (3.1),

$$H'' + \frac{r}{t}H' = \frac{V}{R} + \frac{w^2 - y^2/R}{4(M+1)} \ge \frac{V}{M+1} + \frac{w^2 - 2V}{4(M+1)} \ge \frac{W}{2(M+1)} \ge \frac{c'}{t^2},$$

with $c' = c/(2(M+1))$, using $y^2/R \le 2V$. Thus $(t^r H')' \ge c' t^{r-2}$ on $[1,\infty)$, so $t^r H'(t) \ge H'(1) + c'(t^{r-1} - 1)/(r-1)$ and $H'(t) \ge c'/(2(r-1)t)$ for all large t, because $r > 1$. Then H grows at least like a multiple of $\log t$, while $H \le M + 1$ under the hypothesis. This contradiction shows that x is unbounded; being nondecreasing, $x \to \infty$, and $\|X\| \ge x$.

3.4 Escape for $0 < r \le 1$

Here $t^r v(t) = \int_0^t s^r V/R\, ds$. Since $V > 0$ near $t = 0$, $\kappa = t_0^r v(t_0) > 0$ for some $t_0 > 0$, and monotonicity of $t^r v$ gives $v(t) \ge \kappa t^{-r}$ for $t \ge t_0$. As $\int^\infty t^{-r} dt = \infty$ for $r \le 1$, $x \to \infty$.

3.5 Boundedness for $r > 2$

Suppose $r > 2$ and $x \to \infty$; recall that x is nondecreasing. Since $V < 1$, the bracket in (3.3) satisfies $(4-2r)x + (3-r)V \le -(r-2)x$ once $(r-2)x \ge 3-r$, and $x/R \ge 2/3$ for $x \ge 2$. Thus $E' \le -\frac{2}{3}(r-2)tV$ for all large t, and since $E \ge 0$, $\int^\infty tV\, dt < \infty$. On the other hand, by Fubini,

$$\int_0^\infty v\, dt = \int_0^\infty t^{-r} \int_0^t s^r \frac{V}{R}\, ds\, dt = \frac{1}{r-1} \int_0^\infty s \frac{V}{R}\, ds < \infty,$$

so x is bounded, a contradiction. Hence x converges, and $y^2 \le V(2x+1)$ is bounded.

Remark 3.1 (Limits of valley potentials). The argument for $r \le 2$ uses that the valley floor is exactly the ray $y = 0$ with $f = R - x$ outside the unit ball, which makes $\frac{y^2}{R} \le 2V$. A valley $\|(x,y)\|_p - x$ with $p < 2$ would give room for $2 < r < 3$, as in Section 5, but its gradient is not Lipschitz along the floor, and every $C^{1,1}$ smoothing makes the floor locally quadratic again. In simulations run during the internal AI referee checks, a fixed smoothing produced only finite excursions for $r = 2.4$. This is a heuristic, not a theorem. In finite dimension a bounded set of minimizers forces bounded trajectories, since $F(X(t)) \le F(X(0))$ and sublevel sets are bounded, so any finite-dimensional example for $2 < r < 3$ needs an unbounded minimizer set.

4 Global solutions from rest

Let G be a globally 1-Lipschitz map on a Banach space with $\|G\| \le 2$, and fix X_0. For $r > 0$ and $0 \le s \le t$ let

$$K_r(t,s) = \int_s^t \Big(\frac{s}{\sigma}\Big)^r d\sigma = \begin{cases} \dfrac{s - t^{1-r} s^r}{r-1}, & r \ne 1, \\ s \log(t/s), & r = 1, \end{cases}$$

so that $0 \le K_r(t,s) \le t - s$. A continuous solution of $X(t) = X_0 - \int_0^t K_r(t,s) G(X(s))\, ds$ is an actual solution of (1.1) with $\nabla F = G$: differentiation gives $V(t) = X'(t) = -t^{-r} \int_0^t s^r G(X(s))\, ds$ and $V' = -\frac{r}{t} V - G(X)$ for $t > 0$, while $\|V(t)\| \le 2t$ and $\|X(t) - X_0\| \le t^2$ give continuity at 0 and the derivative $X'(0) = 0$. On bounded continuous functions b on $[0,\infty)$, the map

$$(\mathcal{T} b)(t) = -e^{-2t} \int_0^t K_r(t,s)\, G\big(X_0 + e^{2s} b(s)\big)\, ds$$

satisfies $\|\mathcal{T} b - \mathcal{T} c\|_\infty \le \frac{1}{4} \|b - c\|_\infty$, because $\sup_t e^{-2t} \int_0^t (t-s) e^{2s} ds \le \frac{1}{4}$. Its fixed point gives $X(t) = X_0 + e^{2t} b(t)$. The Lean development formalizes this construction for $r > 1$; the cases $r \le 1$ use the same estimate with the kernel above.

5 Hilbert space: escape for $1 < r < 3$

5.1 Moreau valleys

Choose an even integer $m \ge 2$ with $p = m/(m-1)$ satisfying $p(r-1) \le 2$; this is possible because $r < 3$. Let $C_m = \{(h-1, b) : h^m + b^m \le 1\}$, a compact convex set containing 0, and for $\varepsilon > 0$ let

$$F_\varepsilon(z) = \max_{g \in C_m} \Big(\langle g, z \rangle - \frac{\varepsilon}{2} \|g\|^2 \Big), \qquad G_\varepsilon(z) = P_{C_m}(z/\varepsilon).$$

Then F_ε is convex, $\nabla F_\varepsilon = G_\varepsilon$ is ε^{-1}-Lipschitz, $F_\varepsilon \geq 0 = F_\varepsilon(0)$, $(G_\varepsilon)_1 \leq 0$, $\|G_\varepsilon\| \leq \sqrt{5}$, and $F_\varepsilon(2,1) \leq 1$; F_ε is the Moreau envelope of $\|\cdot\|_p - x$. Write $z = (x,y)$, $G_\varepsilon(z) = (-a,b)$ and suppose $x \geq 2$. With $(\xi,\eta) = z - \varepsilon G_\varepsilon(z)$, $R = \|(\xi,\eta)\|_p$ and $V = R - \xi$, the projection identities give $1 - a = (\xi/R)^{p-1}$, $b = \operatorname{sgn}(\eta)(|\eta|/R)^{p-1}$ and $F_\varepsilon(z) = V + \frac{\varepsilon}{2}(a^2 + b^2)$. Elementary convexity estimates then yield

$$a \leq \frac{F_\varepsilon}{2}, \qquad yb \leq pF_\varepsilon + \varepsilon, \qquad a \geq \frac{p-1}{B}(F_\varepsilon - \varepsilon), \qquad y^2 \leq 4BF_\varepsilon + 2\varepsilon^2, \tag{5.1}$$

whenever $0 < \varepsilon \leq 1$, $2 \leq x \leq M$, $F_\varepsilon \leq 1$ and $B = M + 2$.

5.2 Finite excursions

Proposition 5.1. *Let $1 < r < 3$ and p as above. For every $M \geq 2$ there is $\varepsilon \in (0,1]$ such that every actual solution for G_ε from $(2,1)$ at rest has $x(t) > M$ for some t, and then for all later times.*

Sketch. As in Section 3, x is nondecreasing and $W = \frac{1}{2}w^2 + F_\varepsilon \leq 1$. Assume $x \leq M$ throughout. With $E = \frac{1}{2}(tw + (r-1)y)^2 + t^2 F_\varepsilon$ one has the exact identity $E' = t(2F_\varepsilon - (r-1)yb) - t^2 va$, and (5.1) with $p(r-1) \leq 2$ gives $(e^{x/2}E)' \geq -(r-1)\varepsilon t e^{M/2}$. For ε small relative to a horizon T this keeps $e^{x/2}E$ above half its initial value on $[0,T]$, which yields $W \geq c/t^2$ on $[1,T]$. For $H = x + ky^2$ with $k = (p-1)/(4pB)$ the identity $H'' + \frac{r}{t}H' = a + 2k(w^2 - yb)$ and (5.1) give $H'' + \frac{r}{t}H' \geq d/t^2$ on $[1,T]$. Then $L = tH' + (r-1)H$ satisfies $L' \geq d/t$, so $tH' \geq 1$ on $[S,T]$ for explicit $S < T$, and $H(T) \geq B + 1$, contradicting $H \leq B$. All constants depend only on r, p and M, and are chosen before the trajectory. □

The explicit constants are those of the formal development: $J_* = e(r-1)^2/2$, $e_0 = J_*/(2e^{M/2})$, $C = 2 + 4B(r-1)^2$, $c = e_0/(2C)$, $K = (p-1)/(2B)$, $d = Kc/2$, $S = \exp(((r-1)B+3)/d)$, $T = Se^{B+1}$ and $\varepsilon = \min\{1, J_*/((r-1)e^{M/2}T^2), e_0/(4(r-1)^2), c/(6T^2)\}$.

5.3 One objective on ℓ^2

Let $a_n = 2^{-n}/16$ and $M_n = 2 + (n+1)/a_n$, and choose ε_n by Proposition 5.1. Define

$$\Phi(z) = \sum_{n \geq 0} a_n^2 \varepsilon_n^2 F_{\varepsilon_n}(z_n/a_n), \qquad \nabla\Phi(z) = \big(a_n \varepsilon_n^2 G_{\varepsilon_n}(z_n/a_n)\big)_{n \geq 0}.$$

Each block gradient is ε_n-Lipschitz with norm at most $3a_n$, so the series converges, Φ is convex with $\Phi \geq 0 = \Phi(0)$, the gradient is 1-Lipschitz with norm at most $3/8$, and the block descent inequalities sum to $|\Phi(z+h) - \Phi(z) - \langle \nabla\Phi(z), h\rangle| \leq \frac{1}{2}\|h\|^2$. Section 4 gives an actual solution from $X_n(0) = a_n(2,1)$. The rescaling $Q_n(s) = a_n^{-1} X_n(s/\varepsilon_n)$ turns block n into an actual solution for G_{ε_n} from $(2,1)$; the factor ε_n^2 in Φ makes the time change exact. By Proposition 5.1 there is s_n with $(Q_n)_1(s) > M_n$ for all $s \geq s_n$, so $\|X(t)\| \geq a_n M_n > n+1$ for all $t \geq s_n/\varepsilon_n$. Hence $\|X(t)\| \to \infty$ along the whole time axis.

6 The positive side and low damping in continuous time

Let $r \geq 3$, let z minimize F and $e = X - z$. The energy $\mathcal{E} = t^2(F(X) - F(z)) + \frac{1}{2}\|tX' + (r-1)e\|^2$ satisfies $\mathcal{E}' = 2t(F(X) - F(z)) - (r-1)t\langle e, \nabla F(X)\rangle \leq (3-r)t(F(X) - F(z)) \leq 0$ by convexity. With $\mathcal{E}(0) = \frac{1}{2}(r-1)^2\|e(0)\|^2$ and $h = \|e\|^2$, discarding nonnegative terms gives $\frac{d}{dt}[t^{r-1}(h - h(0))] \leq 0$, so $\|X(t) - z\| \leq \|X(0) - z\|$. No Lipschitz assumption is needed once a solution is given.

For $0 < r \le 1$, set $\beta = \frac{r+1}{2(r+3)}$, $b = 1-\beta$ and $v_1 = \frac{4}{r+3}$, and let $q(t) = t^2 - \beta t^4$ for $t \le 1$ and $q(t) = b + v_1 \int_1^t \sigma^{-r} d\sigma$ for $t \ge 1$. With $a(x) = 2(r+1)\big(1 - \frac{1-\sqrt{1-4\beta c(x)}}{2\beta}\big)$, $c(x) = \min\{\max\{x,0\}, b\}$, and $f(x) = \int_x^b a$, the function f is convex, $C^{1,1}$ and nonnegative with minimum on $[b,\infty)$, and q solves $q'' + \frac{r}{t}q' = -f'(q)$ with $q(0) = q'(0) = 0$ and $q \to \infty$. Rescaling to a 1-Lipschitz gradient and lifting along a unit vector gives Theorem 2.1(c) on any Hilbert space, including $\mathbb{R}^2$.

7 The discrete iteration

7.1 Finite-horizon approximation

For $h > 0$ and a 1-Lipschitz G consider $X_{k+1} = Y_k - h^2 G(Y_k)$, $Y_{k+1} = X_{k+1} + \beta_k(X_{k+1} - X_k)$ with $\beta_k = k/(k+r)$. Put $z_k = X_{k+1}$, $w_k = h^{-1}(X_{k+1} - X_k)$ and $\tau_k = (k+r)h$. Then exactly

$$Y_{k+1} = z_k + h\beta_k w_k, \qquad w_{k+1} = \beta_k w_k - hG(Y_{k+1}), \qquad z_{k+1} = z_k + hw_{k+1}, \qquad \beta_k = 1 - \frac{rh}{\tau_k}.$$

The last identity is the damping r/t at time τ_k without approximation. Let (Q,U) be an actual solution with $Q(0) = b$, $\|U\| \le B$ on $[0,T]$, and with acceleration continuous on $[0,T]$. Uniform continuity gives one-step defects at most hc for small h, and with $e_k = \|z_k - Q(\tau_k)\| + \|w_k - U(\tau_k)\|$ one obtains $e_{k+1} \le (1+3h)e_k + h(2hB+3c)$, hence $e_N \le e^{3T}(e_0 + T(2hB+3c))$ when $Nh \le T$. Since $e_0 \to 0$ as $h \to 0$, for every $\eta > 0$ there are $h \in (0,1]$ and N with $(N+r)h = T$ and $\|X_{N+1} - Q(T)\| \le \eta$. The acceleration of the Volterra solution of Section 4 extends continuously to $t = 0$, with value $-G(b)/(r+1)$.

7.2 Planar excursions and permanence

Normalize $\widehat{G}_\varepsilon = \frac{\varepsilon^2}{4} G_\varepsilon$, which is 1-Lipschitz for $\varepsilon \le 1$. Its solution from $(2,1)$ at rest, rescaled in time by $\delta = \varepsilon/2$, is a solution for G_ε, so Proposition 5.1 with $L = M+2$ gives a time T with $q_1(T) > M+2$. The approximation with $\eta = 1$ yields a discrete orbit with $(X_{N+1})_1 > M$. The crossing is permanent: with $d_k = (X_{k+1} - X_k)_1$, $d_0 = -h^2(\widehat{G}_\varepsilon)_1(b) \ge 0$ and $d_{k+1} = \beta_k d_k - h^2(\widehat{G}_\varepsilon)_1(Y_{k+1}) \ge 0$, so the first coordinate of the iterates is nondecreasing. This is exact for the discrete orbit and needs no long-time approximation.

7.3 One objective on ℓ^2

With a_n and M_n as before, choose $\varepsilon_n, h_n \in (0,1]$ by the planar discrete excursion, and set

$$\Phi(z) = \sum_n a_n^2 \frac{h_n^2 \varepsilon_n^2}{4} F_{\varepsilon_n}(z_n/a_n), \qquad \nabla\Phi(z) = \Big(a_n \frac{h_n^2\varepsilon_n^2}{4} G_{\varepsilon_n}(z_n/a_n)\Big)_n.$$

As in Section 5, Φ is convex, $\Phi \ge 0 = \Phi(0)$, and $\nabla\Phi$ is 1-Lipschitz with norm at most $3/8$. The unit-step iteration (2.1) from $b_n = a_n(2,1)$ satisfies $(X_k)_n = a_n X_k^{(n)}$ exactly, where $X^{(n)}$ is the planar iteration with step h_n and gradient $\widehat{G}_{\varepsilon_n}$. Each block crosses M_n permanently, so $\|X_k\| \ge a_n M_n > n+1$ for all large k, and $\|X_k\| \to \infty$.

7.4 The positive side and low damping

Let $r \ge 3$, z a minimizer, $t_k = (k+r-1)/(r-1)$, $Z_k = t_k Y_k - (t_k - 1)X_k$ and $\mathcal{E}_k = (t_k^2 - t_k)(F(X_k) - F(z)) + \frac{1}{2}\|Z_k - z\|^2$. The descent lemma at Y_k, combined with convexity at X_k and at z with weights

$t_k - 1$ and 1, together with $t_{k+1}\beta_k = t_k - 1$ and $t_{k+1}^2 - t_{k+1} \le t_k^2$ for $r \ge 3$, gives $\mathcal{E}_{k+1} \le \mathcal{E}_k$. Since $t_0 = 1$ and $Z_0 = b$, $\|Z_k - z\| \le \|b - z\|$, and $X_{k+1} - z$ is a convex combination of $Z_{k+1} - z$ and $X_k - z$, so $\|X_k - z\| \le \|b - z\|$ by induction. For $0 < r \le 1$ take $f(x) = \int_2^x \min\{\max\{t-2, -1\}, 0\}dt$ on $\mathbb{R}$ and $X_0 = 0$: then $X_1 = 1$, $X_2 = 2$, and $X_{k+2} - X_{k+1} \ge \frac{1}{k+1}$, so $X_k \to \infty$; lift along a unit vector.

8 Formal verification

The Lean project in `lean/` pins Lean 4.19.0 (commit `6caaee842e94`) and Mathlib revision `c44e0c8ee63ca166450922a373c7409c5d26b00b`. It merges three earlier Lean developments (Section 11), whose modules are included unchanged, with new modules for the planar generalization. The principal declarations are:

- `NesterovSharpThreshold.exists_subcritical_norm_divergent_flow` and `NesterovSharpThreshold.sharp_boundedness_threshold` (Theorem 2.1(b) and the equivalence for $r > 1$), `NesterovSharpThreshold.sharp_boundedness_threshold_all_positive` and `NesterovSharpThreshold.exists_positive_subcritical_norm_divergent_flow` (all $r > 0$), and `NesterovHighDamping.bounded_distance` (part (a), on every real Hilbert space). Universal boundedness is the definition `NesterovSharpThreshold.UniversalBoundedness`, which quantifies over actual gradients, 1-Lipschitz bounds, attained minima and actual solutions on $\ell^2(\mathbb{N}; \mathbb{R}^2)$.
- `DiscreteNesterov.sharp_discrete_boundedness_threshold`, `DiscreteNesterov.exists_subcritical_discrete_norm_divergence` and `DiscreteNesterov.exists_discrete_norm_divergence_below_three` (Theorem 2.4); the recursion is `DiscreteNesterov.nagState`.
- `NesterovPlanar.planar_valley_escape` and `NesterovPlanar.every_planar_trajectory_escapes` (Theorem 2.2(a) for $1 < r \le 2$, including the potential's formula, convexity, gradient, Lipschitz bound, zero set, existence of a solution and escape of every solution), and `NesterovPlanar.planar_escape_all_low_damping` (Corollary 2.3; for $0 < r \le 1$ it uses the lifted scalar example of Section 6).

Each statement uses the actual Fréchet gradient (`HasGradientAt`), `LipschitzWith 1`, `ConvexOn` on the whole space, genuine derivatives (`HasDerivAt`) with the equation asserted for $t > 0$ and the derivative at 0 taken two-sided after constant extension to negative times, and `Tendsto` of the norm to infinity. All headline theorems depend only on `propext`, `Classical.choice` and `Quot.sound`. A fresh replay on 30 October 2026, compiling all 56 modules from source with `-trust=0` against a Mathlib checkout built at the pinned revision, passed with every audited axiom closure standard; its record is in `verification/`.

Formal scope: Theorem 2.2(a) for $0 < r \le 1$ and part (b), existence of solutions for $r \le 1$ by the kernel of Section 4, and Remark 3.1 are conventional arguments. Everything else in Theorems 2.1, 2.2 and 2.4 is formalized.

9 Related work and novelty

Su, Boyd and Candès introduced (1.1) and the energy argument for $r \ge 3$ [1]. Weak convergence for $r > 3$ is due to Attouch, Chbani, Peypouquet and Redont [2] and May [3], and for the discrete family to Chambolle and Dossal [4]. Point convergence at $r = 3$ was proved by Jang and Ryu [8] and, in Hilbert space and in discrete time, by Boţ, Fadili and Nguyen [9]. For $r < 3$, Attouch,

Chbani and Riahi [5] and Apidopoulos, Aujol and Dossal [6] prove the rate $O(t^{-2r/3})$ for objective values, which is compatible with our escaping trajectories because the minimizer sets are unbounded. Cabot, Engler and Gadat [7] study asymptotically small dissipation; the free-motion mechanism behind the case $r \leq 1$ is classical. Ryu [10] constructed a convergent orbit of infinite length at $r = 3$, a different pathology.

We found no earlier result showing unbounded zero-velocity trajectories or iterates with an attained minimum for $1 < r < 3$, either in the plane for $1 < r \leq 2$ or in Hilbert space. This is a provisional assessment based on the sources we could inspect. The full texts of several theses could not be accessed during the literature comparison, and arXiv itself could not be reached while this manuscript was prepared, so version-specific statements about arXiv preprints rely on earlier search records and results.

10 Open questions

Does every convex F on $\mathbb{R}^d$ with a Lipschitz gradient and an attained minimum have bounded zero-velocity trajectories for $2 < r < 3$? Does the discrete iteration admit a finite-dimensional unbounded orbit for some $1 < r < 3$? What happens when the minimizer set is bounded in Hilbert space? The constructions here say nothing about these cases.

11 Research process and AI disclosure

This is fully autonomous agentic research. Three uploaded research packages contributed: a continuous-time Hilbert-space threshold, an escaping planar trajectory at damping $r = 2$, and a discrete-time threshold for the Nesterov iteration. Their research, proofs, Lean formalizations and earlier notes were produced by autonomous AI research agents in earlier sessions whose records do not name the model or system used.

Opus 5.5 by Anthropic used with Claude Code assessed the three packages for this release and coordinated independent AI referee checks of the first two. It proposed and proved the extension of the planar example from $r = 2$ to every $0 < r \leq 2$; the boundedness of the same planar flows for $r > 2$ was found during those AI referee checks. It generalized the Lean proof of the planar example to every $1 < r \leq 2$, merged the three Lean developments into one project, ran the fresh Lean replay described in Section 8, checked the discrete transfer argument, and wrote this combined manuscript. The editorial revision of 1 November 2026, also made with Opus 5.5 by Anthropic used with Claude Code, changed the presentation only, not the mathematics or the formalization.

Dev D. Goyal is responsible for publication. The work has not been externally peer reviewed; no expert validation is claimed, AI referee checks are not human peer review, and the novelty assessment is provisional. Lean verification establishes only the formal statements described in Section 8; Sections 9 and 10 state the remaining limits.

References

[1] W. Su, S. Boyd and E. J. Candès, A differential equation for modeling Nesterov's accelerated gradient method: theory and insights, *J. Mach. Learn. Res.* 17 (2016), no. 153, 1–43.

[2] H. Attouch, Z. Chbani, J. Peypouquet and P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, *Math. Program.* 168 (2018), 123–175.

[3] R. May, Asymptotic for a second-order evolution equation with convex potential and vanishing damping term, *Turkish J. Math.* 41 (2017), 681–685.

[4] A. Chambolle and C. Dossal, On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm", *J. Optim. Theory Appl.* 166 (2015), 968–982. doi:10.1007/s10957-015-0746-4.

[5] H. Attouch, Z. Chbani and H. Riahi, Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$, *ESAIM Control Optim. Calc. Var.* 25 (2019), article 2. doi:10.1051/cocv/2017083.

[6] V. Apidopoulos, J.-F. Aujol and C. Dossal, Convergence rate of inertial forward–backward algorithm beyond Nesterov's rule, *Math. Program.* 180 (2020), 137–156. doi:10.1007/s10107-018-1350-9.

[7] A. Cabot, H. Engler and S. Gadat, On the long time behavior of second order differential equations with asymptotically small dissipation, *Trans. Amer. Math. Soc.* 361 (2009), 5983–6017.

[8] U. Jang and E. K. Ryu, Point convergence of Nesterov's accelerated gradient method: an AI-assisted proof, preprint (2026), version 2. arXiv:2510.23513v2.

[9] R. I. Boţ, J. Fadili and D.-K. Nguyen, The iterates of Nesterov's accelerated algorithm converge in the critical regimes, preprint (2025). arXiv:2510.22715.

[10] E. K. Ryu, Nesterov flow may travel infinitely long to converge to a minimizer, preprint (2026). arXiv:2604.06651.