Exact Inexact-Gradient Polyak–Łojasiewicz Noise Floors, Accelerated Nesterov Lyapunov Stability, Ball-Completion Reflected Resolvent Contraction, and Weighted Frank–Wolfe Duality Certificates

Lluis Eriksson
Independent Researcher, Östersund, Sweden
editor@airr.science

November 29, 2026

Abstract

We establish four self-contained, non-asymptotic quantitative theorems in first-order optimization and monotone operator splitting with explicit, exception-free constants. First, after giving a five-line derivation of Baillon–Haddad co-coercivity, we analyze perturbed gradient descent $x_{k+1} = x_k - \eta(\nabla f(x_k) + e_k)$ on L-smooth (not necessarily convex) μ-Polyak–Łojasiewicz functions, proving a step-size-independent steady-state noise floor $\varepsilon^2/(2\mu)$ for every $\eta \in (0, 1/L]$ (attained with exact equality on isotropic quadratics under constant adversarial error) together with an explicit inexact contraction envelope across the entire stability range $\eta \in (0, 2/L)$. Second, for L-smooth μ-strongly convex objectives under adversarial gradient errors $\|e_k\|_2 \le \varepsilon$, we establish an eight-line Lyapunov stability proof for Nesterov's accelerated gradient method yielding $\Phi_K \le (1 - \frac{1}{2}\sqrt{\mu/L})^K \Phi_0 + \frac{2}{\mu}\varepsilon^2$—retaining accelerated $\sqrt{L/\mu}$ decay at only four times the unaccelerated noise floor while recovering exact $(1 - \sqrt{\mu/L})^K \Phi_0$ convergence when $\varepsilon = 0$. Third, via an exact ball-completion identity $\|(R_{\gamma\nabla g}x - R_{\gamma\nabla g}y) - c(\gamma)(x - y)\|_2 \le \mathrm{rad}(\gamma)\|x - y\|_2$ for the reflected resolvent $R_{\gamma\nabla g} = 2(I + \gamma\nabla g)^{-1} - I$, we prove sharp non-asymptotic linear contraction factors $\rho_{\mathrm{DR}}(\gamma) = \max(|1 - \gamma\mu|/(1 + \gamma\mu), |1 - \gamma L|/(1 + \gamma L))$ for Peaceman–Rachford and $(1 + \rho_{\mathrm{DR}}(\gamma))/2$ for Douglas–Rachford splitting on composite objectives $g + h$ with arbitrary proper nonsmooth closed convex h. Fourth, pairing Bregman mirror descent's fixed-horizon and explicit anytime rates with a three-line $(k + 2)$-weighted telescoping proof for open-loop Frank–Wolfe ($\gamma_k = 2/(k+2)$), we prove an initialization-free primal–dual gap certificate $\min_{\lfloor K/2 \rfloor \le k < K} g_{\mathrm{FW}}(x_k) \le \frac{2(1+\ln 2)\, C_f}{\lceil K/2 \rceil}$ for every $K \ge 2$. All four theorems are verified across 538 deterministic numerical checks.

1 Introduction and Relation to Prior Work

Quantitative first-order optimization and operator-splitting algorithms rely on explicit non-asymptotic contraction factors, noise-stability ceilings, and computable duality-gap certificates [8, 2]. Four key themes dominate modern analysis: (i) non-convex linear convergence under the Polyak–Łojasiewicz (PL) inequality [7] and sensitivity to inexact gradients; (ii) Lyapunov potential proofs of Nesterov acceleration [8, 1] and its quantitative error accumulation under inexact oracles [4]; (iii) sharp contraction rates of reflected resolvents in Peaceman–Rachford and Douglas–Rachford splitting [5]; and (iv) Bregman mirror descent [3] alongside affine-invariant primal–dual gap certificates for Frank–Wolfe conditional gradient methods [6].

Here we unify all four pillars through elementary, completely self-contained, finite-horizon proofs valid without asymptotic $O(\cdot)$ constants: exact Baillon–Haddad co-coercivity and sharp

inexact-gradient μ-PL noise floors across $(0, 2/L)$ (Theorem 2.1); an explicit Lyapunov bound for inexact strongly convex Nesterov acceleration (Theorem 3.1); a four-line ball-completion derivation of sharp Peaceman–Rachford and Douglas–Rachford contraction factors (Theorem 4.1); and fixed-horizon/anytime mirror descent paired with an initialization-free $(k+2)$-weighted Frank–Wolfe duality gap bound (Theorem 5.1).

2 Baillon–Haddad Co-Coercivity and Inexact Polyak–Łojasiewicz Envelopes

A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth ($L > 0$) if $\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x - y\|_2$ for all $x, y \in \mathbb{R}^d$, which implies the quadratic descent upper bound $f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2$. It satisfies the μ-Polyak–Łojasiewicz (μ-PL) condition ($\mu > 0$) if $f^* := \inf_{x\in\mathbb{R}^d} f(x) > -\infty$ and $\frac{1}{2}\|\nabla f(x)\|_2^2 \ge \mu(f(x) - f^*)$ for all $x \in \mathbb{R}^d$. Note that for any nonconstant L-smooth μ-PL function, $\mu \le L$ holds automatically (picking x with $f(x) > f^*$ and testing $y = x - \frac{1}{L}\nabla f(x)$ gives $f(x) - f^* \ge \frac{1}{2L}\|\nabla f(x)\|_2^2 \ge \frac{\mu}{L}(f(x) - f^*)$); in Theorem 2.1(b) we assume $0 < \mu \le L$ explicitly so that constant functions are covered without exception.

Theorem 2.1 (Baillon–Haddad Co-Coercivity and Sharp Inexact-Gradient μ-PL Envelopes)**.** (a) (Baillon–Haddad Co-Coercivity.) *If $f : \mathbb{R}^d \to \mathbb{R}$ is convex and L-smooth ($L > 0$), then for all $x, y \in \mathbb{R}^d$,*

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle \ge \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2, \qquad \langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2. \tag{1}$$

(b) (Sharp Step-Independent Inexact μ-PL Noise Floor and Full $(0, 2/L)$ Stability Envelope.) *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth and μ-PL with $0 < \mu \le L$ (not necessarily convex), and consider perturbed gradient descent $x_{k+1} = x_k - \eta(\nabla f(x_k) + e_k)$ with arbitrary gradient errors $e_k \in \mathbb{R}^d$.*

(i) *For every step size $\eta \in (0, 1/L]$ and every $K \ge 0$,*

$$f(x_K) - f^* \le (1 - \mu\eta)^K (f(x_0) - f^*) + \frac{\eta}{2}\sum_{k=0}^{K-1}(1 - \mu\eta)^{K-1-k}\|e_k\|_2^2. \tag{2}$$

In particular, whenever $\|e_k\|_2 \le \varepsilon$ for all $k < K$, $f(x_K) - f^* \le (1-\mu\eta)^K(f(x_0) - f^*) + \frac{\varepsilon^2}{2\mu}$, and the steady-state noise floor $\frac{\varepsilon^2}{2\mu}$ is sharp (attained on $f(x) = \frac{\mu}{2}\|x\|_2^2$ under constant adversarial error $e_k \equiv -\varepsilon u$, $\|u\|_2 = 1$).

(ii) *Across the full stability range $\eta \in (0, 2/L)$ and for any parameter $\theta \in (0, 1)$, setting $\rho_{\eta,\theta} := 1 - \mu\eta(2 - L\eta)(1 - \theta) \in (0, 1)$,*

$$f(x_{k+1}) - f^* \le \rho_{\eta,\theta}(f(x_k) - f^*) + \frac{\eta}{2}\left(L\eta + \frac{(1 - L\eta)^2}{\theta(2 - L\eta)}\right)\|e_k\|_2^2. \tag{3}$$

Proof. Part (a). Fix $x \in \mathbb{R}^d$ and define $\phi_x(z) := f(z) - \langle \nabla f(x), z\rangle$. Then ϕ_x is convex, L-smooth, and satisfies $\nabla\phi_x(x) = 0$, so x minimizes ϕ_x globally. Evaluating L-smoothness of ϕ_x at $z = y - \frac{1}{L}\nabla\phi_x(y)$ yields $\phi_x(x) \le \phi_x(y - \frac{1}{L}\nabla\phi_x(y)) \le \phi_x(y) - \frac{1}{2L}\|\nabla\phi_x(y)\|_2^2$, which rearranges to the first inequality in (1). Adding the symmetric inequality with x and y swapped yields the second.

Part (b). Let $g_k := \nabla f(x_k)$. By L-smoothness along $x_{k+1} - x_k = -\eta(g_k + e_k)$,

$$f(x_{k+1}) - f(x_k) \le -\eta\left(1 - \frac{L\eta}{2}\right)\|g_k\|_2^2 - \eta(1 - L\eta)\langle g_k, e_k\rangle + \frac{L\eta^2}{2}\|e_k\|_2^2. \tag{4}$$

For (i), when $\eta \in (0, 1/L]$ we have $1 - L\eta \geq 0$, so $-\eta(1 - L\eta)\langle g_k, e_k\rangle \leq \frac{\eta(1-L\eta)}{2}(\|g_k\|_2^2 + \|e_k\|_2^2)$. Substituting into (4) simplifies both coefficients identically to $f(x_{k+1}) - f(x_k) \leq -\frac{\eta}{2}\|g_k\|_2^2 + \frac{\eta}{2}\|e_k\|_2^2$. Applying μ-PL ($\frac{1}{2}\|g_k\|_2^2 \geq \mu(f(x_k) - f^*)$) yields $f(x_{k+1}) - f^* \leq (1 - \mu\eta)(f(x_k) - f^*) + \frac{\eta}{2}\|e_k\|_2^2$, and unrolling from $k = 0$ to $K - 1$ proves (2) [7]. When $\|e_k\|_2 \leq \varepsilon$, $\frac{\eta}{2}\sum_{j=0}^{K-1}(1 - \mu\eta)^j\varepsilon^2 \leq \frac{\eta}{2} \cdot \frac{\varepsilon^2}{\mu\eta} = \frac{\varepsilon^2}{2\mu}$ (with $0^0 := 1$ when $\mu = L$ and $\eta = 1/L$). For sharpness, on $f(x) = \frac{\mu}{2}\|x\|_2^2$ with $e_k \equiv -\varepsilon u$ ($\|u\|_2 = 1$), $x_{k+1} = (1 - \mu\eta)x_k + \eta\varepsilon u \to \frac{\varepsilon}{\mu}u$ as $k \to \infty$, where $f(x_\infty) - f^* = \frac{\varepsilon^2}{2\mu}$. For (ii), when $\eta \in (0, 2/L)$ and $\theta \in (0, 1)$, apply Young's inequality $|\langle g_k, e_k\rangle| \leq \frac{\tau}{2}\|g_k\|_2^2 + \frac{1}{2\tau}\|e_k\|_2^2$ with $\tau := \frac{\theta(2-L\eta)}{|1-L\eta|}$ (or drop the cross term directly when $L\eta = 1$) in (4), obtaining $f(x_{k+1}) - f(x_k) \leq -\frac{\eta(2-L\eta)(1-\theta)}{2}\|g_k\|_2^2 + \frac{\eta}{2}(L\eta + \frac{(1-L\eta)^2}{\theta(2-L\eta)})\|e_k\|_2^2$; since $\eta(2 - L\eta) \leq 1/L$ and $\mu \leq L$, $\mu\eta(2 - L\eta)(1 - \theta) \in (0, 1)$, and applying μ-PL proves (3). □

3 Inexact Strongly Convex Nesterov Acceleration via Lyapunov Potential

Let $f : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and μ-strongly convex ($0 < \mu \leq L$) with unique minimizer x^* and $f^* := f(x^*)$. Set $\alpha := \sqrt{\mu/L} \in (0, 1]$, initialize $z_0 := x_0 \in \mathbb{R}^d$, and consider Nesterov's constant-momentum scheme [8] driven by adversarial inexact gradients $g_k + e_k$ ($g_k := \nabla f(y_k)$):

$$y_k := \frac{x_k + \alpha z_k}{1 + \alpha}, \qquad x_{k+1} := y_k - \frac{1}{L}(g_k + e_k), \qquad z_{k+1} := (1 - \alpha)z_k + \alpha y_k - \frac{\alpha}{\mu}(g_k + e_k). \quad (5)$$

Define the Lyapunov potential $\Phi_k := f(x_k) - f^* + \frac{\mu}{2}\|z_k - x^*\|_2^2$.

Theorem 3.1 (Exact and Adversarial Inexact Strongly Convex Nesterov Stability)**.** *For every L-smooth, μ-strongly convex $f : \mathbb{R}^d \to \mathbb{R}$ ($0 < \mu \leq L$) and arbitrary gradient errors $\{e_k\}_{k\geq 0} \subset \mathbb{R}^d$:*

(a) (Exact Oracle Rate.) *When $e_k \equiv 0$, $\Phi_K \leq \left(1 - \sqrt{\mu/L}\right)^K \Phi_0$ for all $K \geq 0$.*

(b) (Non-Asymptotic Adversarial Inexact Stability.) *For every $K \geq 0$,*

$$\Phi_K \leq \left(1 - \frac{1}{2}\sqrt{\frac{\mu}{L}}\right)^K \Phi_0 + \frac{1}{\sqrt{\mu L}} \sum_{k=0}^{K-1} \left(1 - \frac{1}{2}\sqrt{\frac{\mu}{L}}\right)^{K-1-k} \|e_k\|_2^2. \quad (6)$$

In particular, whenever $\|e_k\|_2 \leq \varepsilon$ for all $k < K$,

$$f(x_K) - f^* \leq \Phi_K \leq \left(1 - \frac{1}{2}\sqrt{\frac{\mu}{L}}\right)^K \Phi_0 + \frac{2}{\mu}\,\varepsilon^2. \quad (7)$$

Proof. Fix $k \geq 0$ and define the exact-gradient reference updates $\tilde{x}_{k+1} := y_k - \frac{1}{L}g_k$ and $\tilde{z}_{k+1} := (1 - \alpha)z_k + \alpha y_k - \frac{\alpha}{\mu}g_k$. Set $A_k := f(y_k) - f^* - \frac{1}{2L}\|g_k\|_2^2$; by L-smoothness at y_k, $A_k \geq f(\tilde{x}_{k+1}) - f^* \geq 0$. We first verify the exact-gradient Lyapunov contraction $A_k + \frac{\mu}{2}\|\tilde{z}_{k+1} - x^*\|_2^2 \leq (1-\alpha)\Phi_k$ [1]. Using the identity $\|(1-\alpha)a+\alpha b\|_2^2 = (1-\alpha)\|a\|_2^2+\alpha\|b\|_2^2-\alpha(1-\alpha)\|a-b\|_2^2$, $\frac{\alpha^2}{2\mu} = \frac{1}{2L}$, and $\alpha(z_k-y_k) = y_k-x_k$ (which implies $(1-\alpha)z_k+\alpha y_k-x^* = (y_k-x^*)+\frac{1-\alpha}{\alpha}(y_k-x_k)$),

$$\begin{aligned}\frac{\mu}{2}\|\tilde{z}_{k+1} - x^*\|_2^2 \leq{}& \frac{\mu(1-\alpha)}{2}\|z_k - x^*\|_2^2 + \frac{\mu\alpha}{2}\|y_k - x^*\|_2^2 \\ &- \alpha\langle g_k, y_k - x^*\rangle - (1-\alpha)\langle g_k, y_k - x_k\rangle + \frac{1}{2L}\|g_k\|_2^2.\end{aligned}$$

Adding $A_k = f(y_k) - f^* - \frac{1}{2L}\|g_k\|_2^2$ cancels $\pm\frac{1}{2L}\|g_k\|_2^2$ identically. Applying μ-strong convexity $\alpha(f(y_k)-f^*-\langle g_k, y_k-x^*\rangle+\frac{\mu}{2}\|y_k-x^*\|_2^2) \leq 0$ and convexity $(1-\alpha)(f(y_k)-f(x_k)-\langle g_k, y_k-x_k\rangle) \leq 0$ yields

$$A_k + \frac{\mu}{2}\|\tilde{z}_{k+1} - x^*\|_2^2 \leq (1-\alpha)\Phi_k, \quad (8)$$

which immediately proves Part (a) when $e_k \equiv 0$.

For Part (b), evaluating L-smoothness at $x_{k+1} = y_k - \frac{1}{L}(g_k + e_k)$ (where the cross term $1 - L \cdot (1/L) = 0$ vanishes identically) gives $f(x_{k+1}) - f^* \le A_k + \frac{1}{2L}\|e_k\|_2^2$. If $\alpha = 1$ ($\mu = L$), (8) forces $A_k = 0$ and $\tilde{z}_{k+1} = x^*$, so $z_{k+1} - x^* = -\frac{1}{L}e_k$ and $\Phi_{k+1} \le \frac{1}{2L}\|e_k\|_2^2 + \frac{L}{2L^2}\|e_k\|_2^2 = \frac{1}{L}\|e_k\|_2^2 \le (1 - \frac{1}{2})\Phi_k + \frac{1}{\sqrt{\mu L}}\|e_k\|_2^2$. If $\alpha \in (0, 1)$, write $z_{k+1} - x^* = (\tilde{z}_{k+1} - x^*) - \frac{\alpha}{\mu}e_k$ and apply $\|u-v\|_2^2 \le (1+\delta)\|u\|_2^2 + (1+\delta^{-1})\|v\|_2^2$ with $\delta := \frac{\alpha}{2(1-\alpha)} > 0$. Since $A_k \ge 0$, $(1+\delta)(1-\alpha) = 1 - \frac{\alpha}{2}$, and $\frac{\alpha^2}{2\mu} = \frac{1}{2L}$,

$$\Phi_{k+1} \le (1+\delta)\left(A_k + \frac{\mu}{2}\|\tilde{z}_{k+1} - x^*\|_2^2\right) + \left(1 + 1 + \frac{1}{\delta}\right)\frac{1}{2L}\|e_k\|_2^2 \le \left(1 - \frac{\alpha}{2}\right)\Phi_k + \frac{1}{\alpha L}\|e_k\|_2^2,$$

using $2 + \delta^{-1} = 2 + \frac{2(1-\alpha)}{\alpha} = \frac{2}{\alpha}$ and $\frac{1}{\alpha L} = \frac{1}{\sqrt{\mu L}}$. Unrolling from $k = 0$ to $K - 1$ gives (6), and summing the geometric series $\sum_{j=0}^{K-1}(1 - \alpha/2)^j < 2/\alpha$ yields $\frac{2}{\alpha\sqrt{\mu L}} = \frac{2}{\mu}$ in (7). □

4 Ball-Completion Reflected Resolvent and Composite Splitting Contraction

For any proper closed convex function $h : \mathbb{R}^d \to (-\infty, +\infty]$ and step size $\gamma > 0$, define its proximal resolvent $J_{\gamma\partial h}(x) := \mathrm{prox}_{\gamma h}(x) = \arg\min_{p\in\mathbb{R}^d}\{h(p) + \frac{1}{2\gamma}\|p - x\|_2^2\} = (I + \gamma\partial h)^{-1}(x)$ and reflected resolvent $R_{\gamma\partial h} := 2J_{\gamma\partial h} - I$ [2]. For composite minimization $\min_{x\in\mathbb{R}^d}\{g(x) + h(x)\}$, define the Peaceman–Rachford and Douglas–Rachford splitting operators

$$T_{\mathrm{PR}} := R_{\gamma\partial h} \circ R_{\gamma\nabla g}, \qquad T_{\mathrm{DR}} := \frac{1}{2}(I + R_{\gamma\partial h} \circ R_{\gamma\nabla g}).$$

Theorem 4.1 (Ball-Completion Reflected Resolvent and Sharp Composite Splitting Contraction)**.** *Let $g : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and μ-strongly convex ($0 < \mu \le L$), let $h : \mathbb{R}^d \to (-\infty, +\infty]$ be any proper closed convex function (possibly nonsmooth), and fix any $\gamma > 0$. Set*

$$c(\gamma) := \frac{1 - \gamma^2\mu L}{(1+\gamma\mu)(1+\gamma L)}, \qquad \mathrm{rad}(\gamma) := \frac{\gamma(L-\mu)}{(1+\gamma\mu)(1+\gamma L)}, \qquad \rho_{\mathrm{DR}}(\gamma) := \max\left(\frac{|1-\gamma\mu|}{1+\gamma\mu}, \frac{|1-\gamma L|}{1+\gamma L}\right). \tag{9}$$

(a) (Exact Reflected-Resolvent Ball Inclusion.) *For all $x, y \in \mathbb{R}^d$,*

$$\left\|(R_{\gamma\nabla g}x - R_{\gamma\nabla g}y) - c(\gamma)(x - y)\right\|_2 \le \mathrm{rad}(\gamma)\,\|x - y\|_2, \tag{10}$$

and consequently $\|R_{\gamma\nabla g}x - R_{\gamma\nabla g}y\|_2 \le \rho_{\mathrm{DR}}(\gamma)\|x - y\|_2 < \|x - y\|_2$ for $x \ne y$, minimized at $\gamma^* = (\mu L)^{-1/2}$ with $\rho_{\mathrm{DR}}(\gamma^*) = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$.

(b) (Sharp Non-Asymptotic Peaceman–Rachford and Douglas–Rachford Contraction.) *For all $x, y \in \mathbb{R}^d$, $R_{\gamma\partial h}$ is nonexpansive ($\|R_{\gamma\partial h}x - R_{\gamma\partial h}y\|_2 \le \|x - y\|_2$), and*

$$\|T_{\mathrm{PR}}x - T_{\mathrm{PR}}y\|_2 \le \rho_{\mathrm{DR}}(\gamma)\,\|x - y\|_2, \qquad \|T_{\mathrm{DR}}x - T_{\mathrm{DR}}y\|_2 \le \frac{1 + \rho_{\mathrm{DR}}(\gamma)}{2}\,\|x - y\|_2. \tag{11}$$

Both constants are sharp (attained for $h \equiv 0$, $g(x) = \frac{\mu}{2}\|x\|_2^2$ when $\gamma \le (\mu L)^{-1/2}$, and for $h = \iota_{\{0\}}$, $g(x) = \frac{L}{2}\|x\|_2^2$ when $\gamma \ge (\mu L)^{-1/2}$).

Proof. Part (a). Put $p := J_{\gamma\nabla g}(x)$, $q := J_{\gamma\nabla g}(y)$, $u := x - p = \gamma\nabla g(p)$, $v := y - q = \gamma\nabla g(q)$, $s := x - y = (p - q) + (u - v)$, and $r := R_{\gamma\nabla g}x - R_{\gamma\nabla g}y = (p - q) - (u - v)$, so $p - q = \frac{s+r}{2}$ and $u - v = \frac{s-r}{2}$. If $\mu = L$, then $\nabla g(p) - \nabla g(q) = \mu(p - q)$, so $u - v = \gamma\mu(p - q)$ and $r = \frac{1-\gamma\mu}{1+\gamma\mu}s = c(\gamma)s$ with $\mathrm{rad}(\gamma) = 0$, verifying (10) with equality. If $\mu < L$, $\tilde{g}(z) := g(z) - \frac{\mu}{2}\|z\|_2^2$

is $(L-\mu)$-smooth and convex; applying (1) to $\tilde{g}$ yields $\langle(\nabla g(p)-\nabla g(q))-\mu(p-q), p-q\rangle \geq \frac{1}{L-\mu}\|(\nabla g(p)-\nabla g(q))-\mu(p-q)\|_2^2$, which rearranges identically to

$$\gamma(L+\mu)\langle u-v, p-q\rangle \geq \gamma^2\mu L\|p-q\|_2^2 + \|u-v\|_2^2.$$

Substituting $p-q=\frac{s+r}{2}$ and $u-v=\frac{s-r}{2}$ and multiplying by 4 gives

$$(1+\gamma\mu)(1+\gamma L)\|r\|_2^2 - 2(1-\gamma^2\mu L)\langle r, s\rangle + (1-\gamma\mu)(1-\gamma L)\|s\|_2^2 \leq 0.$$

Dividing by $(1+\gamma\mu)(1+\gamma L)$ and noting the exact identity $c(\gamma)^2 - \mathrm{rad}(\gamma)^2 = \frac{(1-\gamma\mu)(1-\gamma L)}{(1+\gamma\mu)(1+\gamma L)}$ completes the square directly to $\|r-c(\gamma)s\|_2^2 \leq \mathrm{rad}(\gamma)^2\|s\|_2^2$, proving (10). By the triangle inequality, $\|r\|_2 \leq (|c(\gamma)|+\mathrm{rad}(\gamma))\|s\|_2$. When $\gamma^2\mu L \leq 1$, $c(\gamma)\geq 0$ and $c(\gamma)+\mathrm{rad}(\gamma) = \frac{(1-\gamma\mu)(1+\gamma L)}{(1+\gamma\mu)(1+\gamma L)} = \frac{1-\gamma\mu}{1+\gamma\mu} = \rho_{\mathrm{DR}}(\gamma)$; when $\gamma^2\mu L \geq 1$, $c(\gamma)\leq 0$ and $-c(\gamma)+\mathrm{rad}(\gamma) = \frac{(\gamma L-1)(1+\gamma\mu)}{(1+\gamma\mu)(1+\gamma L)} = \frac{\gamma L-1}{\gamma L+1} = \rho_{\mathrm{DR}}(\gamma)$ [5]. Balancing both branches yields $\gamma^* = (\mu L)^{-1/2}$.

Part (b). For $p_1 = J_{\gamma\partial h}(x_1)$, $p_2 = J_{\gamma\partial h}(x_2)$, optimality gives $w_i := x_i - p_i \in \gamma\partial h(p_i)$, so monotonicity of ∂h implies $\langle w_1 - w_2, p_1 - p_2\rangle \geq 0$. Writing $R_{\gamma\partial h}x_i = p_i - w_i$, we have $\|R_{\gamma\partial h}x_1 - R_{\gamma\partial h}x_2\|_2^2 = \|(p_1-p_2)-(w_1-w_2)\|_2^2 = \|(p_1-p_2)+(w_1-w_2)\|_2^2 - 4\langle w_1-w_2, p_1-p_2\rangle \leq \|x_1-x_2\|_2^2$, proving nonexpansiveness of $R_{\gamma\partial h}$. Composing with Part (a) gives $\|T_{\mathrm{PR}}x - T_{\mathrm{PR}}y\|_2 \leq \rho_{\mathrm{DR}}(\gamma)\|x-y\|_2$, and the triangle inequality on $T_{\mathrm{DR}} = \frac{1}{2}(I+T_{\mathrm{PR}})$ yields (11). For sharpness, when $\gamma \leq (\mu L)^{-1/2}$ choose $h\equiv 0$ ($R_{\gamma\partial h} = I$) and $g(x) = \frac{\mu}{2}\|x\|_2^2$: then $T_{\mathrm{PR}} = \frac{1-\gamma\mu}{1+\gamma\mu}I$ and $T_{\mathrm{DR}} = \frac{1}{1+\gamma\mu}I = \frac{1+\rho_{\mathrm{DR}}(\gamma)}{2}I$; when $\gamma \geq (\mu L)^{-1/2}$ choose $h = \iota_{\{0\}}$ ($J_{\gamma\partial h} = 0$, $R_{\gamma\partial h} = -I$) and $g(x) = \frac{L}{2}\|x\|_2^2$: then $T_{\mathrm{PR}} = \frac{\gamma L-1}{\gamma L+1}I$ and $T_{\mathrm{DR}} = \frac{\gamma L}{\gamma L+1}I = \frac{1+\rho_{\mathrm{DR}}(\gamma)}{2}I$, attaining exact equality in (11) across all $\gamma > 0$. □

5 Bregman Mirror Descent and Initialization-Free Frank–Wolfe Gap

Let $\mathcal{X} \subset \mathbb{R}^d$ be a nonempty closed convex set.

Theorem 5.1 (Bregman Mirror Descent and Weighted Frank–Wolfe Duality Gap Certificate). (a) (Finite-Horizon and Anytime Bregman Mirror Descent.) *Let ψ be differentiable and 1-strongly convex with respect to a norm $\|\cdot\|$ on $\mathcal{X}$ (dual norm $\|\cdot\|_*$), with Bregman divergence $D_\psi(x,y) := \psi(x)-\psi(y)-\langle\nabla\psi(y), x-y\rangle \geq \frac{1}{2}\|x-y\|^2$. Let $f:\mathcal{X}\to\mathbb{R}$ be convex with subgradients $g_k \in \partial f(x_k)$ obeying $\|g_k\|_* \leq G$, and iterate $x_{k+1} = \arg\min_{x\in\mathcal{X}}\{\eta_k\langle g_k, x\rangle + D_\psi(x, x_k)\}$ with $\eta_k > 0$ (assuming first-order optimality holds at each step). Then for any $x^* \in \mathcal{X}$ with $D_0 := D_\psi(x^*, x_0)$ and any $K \geq 1$, setting $\bar{x}_K := \left(\sum_{k=0}^{K-1}\eta_k\right)^{-1}\sum_{k=0}^{K-1}\eta_k x_k$,*

$$f(\bar{x}_K) - f(x^*) \leq \frac{D_0 + \frac{1}{2}\sum_{k=0}^{K-1}\eta_k^2\|g_k\|_*^2}{\sum_{k=0}^{K-1}\eta_k}. \tag{12}$$

In particular, when $D_0 > 0$ and $G > 0$, $\eta_k \equiv \frac{\sqrt{2D_0}}{G\sqrt{K}}$ yields $f(\bar{x}_K) - f(x^*) \leq G\sqrt{2D_0/K}$, while the anytime schedule $\eta_k = \frac{\sqrt{2D_0}}{G\sqrt{k+1}}$ yields $f(\bar{x}_K) - f(x^*) \leq \frac{G\sqrt{2D_0}\,(2+\ln K)}{4(\sqrt{K+1}-1)}$ for every $K \geq 1$.

(b) (Initialization-Free $(k+2)$-Weighted Frank–Wolfe Primal–Dual Gap Certificate.) *Let $\mathcal{X} \subset \mathbb{R}^d$ be compact convex and f be convex and L-smooth on $\mathcal{X}$ with curvature constant $C_f := \sup_{x,s\in\mathcal{X},\,\gamma\in(0,1]}\frac{2}{\gamma^2}(f(x+\gamma(s-x))-f(x)-\gamma\langle\nabla f(x), s-x\rangle) \leq L\,\mathrm{diam}_2(\mathcal{X})^2$. Run open-loop Frank–Wolfe: $s_k \in \arg\min_{s\in\mathcal{X}}\langle\nabla f(x_k), s\rangle$, $x_{k+1} = (1-\gamma_k)x_k + \gamma_k s_k$ with $\gamma_k := \frac{2}{k+2}$, and let $g_{\mathrm{FW}}(x_k) := \langle\nabla f(x_k), x_k - s_k\rangle \geq f(x_k) - f^*$ be the computable primal–dual gap. Then for every $K \geq 1$, $f(x_K) - f^* \leq \frac{2C_f}{K+2}$, and for every $K \geq 2$,*

$$\min_{\lfloor K/2\rfloor \leq k < K} g_{\mathrm{FW}}(x_k) \leq \frac{2(1+\ln 2)\,C_f}{\lceil K/2\rceil} \leq \frac{4(1+\ln 2)\,C_f}{K} \approx \frac{6.7726\,C_f}{K}, \tag{13}$$

independently of the initial suboptimality $f(x_0) - f^*$.

Proof. Part (a). First-order optimality of x_{k+1} gives $\langle \eta_k g_k + \nabla\psi(x_{k+1}) - \nabla\psi(x_k), x^* - x_{k+1}\rangle \geq 0$, hence $\eta_k\langle g_k, x_k - x^*\rangle \leq \eta_k\langle g_k, x_k - x_{k+1}\rangle + \langle \nabla\psi(x_{k+1}) - \nabla\psi(x_k), x^* - x_{k+1}\rangle$. By the exact Bregman three-point identity, the second term equals $D_\psi(x^*, x_k) - D_\psi(x^*, x_{k+1}) - D_\psi(x_{k+1}, x_k)$. Using $D_\psi(x_{k+1}, x_k) \geq \frac{1}{2}\|x_{k+1} - x_k\|^2$ and $\eta_k\langle g_k, x_k - x_{k+1}\rangle - \frac{1}{2}\|x_{k+1} - x_k\|^2 \leq \frac{1}{2}\eta_k^2\|g_k\|_*^2$ gives $\eta_k(f(x_k) - f(x^*)) \leq D_\psi(x^*, x_k) - D_\psi(x^*, x_{k+1}) + \frac{1}{2}\eta_k^2\|g_k\|_*^2$. Summing $k = 0, \ldots, K-1$ and applying Jensen's inequality proves (12) [3]. For $D_0 > 0$ and $G > 0$, substituting $\eta_k \equiv \frac{\sqrt{2D_0}}{G\sqrt{K}}$ gives $G\sqrt{2D_0/K}$; for $\eta_k = \frac{\sqrt{2D_0}}{G\sqrt{k+1}}$, integral comparison gives $\sum_{k=0}^{K-1}(k+1)^{-1/2} \geq \int_1^{K+1} t^{-1/2}\,dt = 2(\sqrt{K+1} - 1)$ and $\sum_{k=0}^{K-1}(k+1)^{-1} \leq 1 + \ln K$, yielding the anytime rate.

Part (b). Let $h_k := f(x_k) - f^* \geq 0$ and $g_k := g_{\mathrm{FW}}(x_k) \geq h_k$ (by convexity of f). By definition of C_f, every step obeys

$$h_{k+1} \leq h_k - \gamma_k g_k + \frac{\gamma_k^2}{2}C_f \leq (1 - \gamma_k)h_k + \frac{\gamma_k^2}{2}C_f. \tag{14}$$

At $k = 0$, $\gamma_0 = 1$ gives $h_1 \leq \frac{1}{2}C_f \leq \frac{2C_f}{3}$; inductively, if $h_k \leq \frac{2C_f}{k+2}$ $(k \geq 1)$, then (14) gives $h_{k+1} \leq \frac{k}{k+2}\frac{2C_f}{k+2} + \frac{2C_f}{(k+2)^2} = \frac{2(k+1)}{(k+2)^2}C_f \leq \frac{2C_f}{k+3}$ since $(k+1)(k+3) = (k+2)^2 - 1 < (k+2)^2$ [6]. Now multiply the first inequality of (14) by $(k+2)$ and substitute $(k+2)\gamma_k = 2$ alongside $h_k \leq g_k$:

$$(k+2)h_{k+1} - (k+1)h_k = h_k + (k+2)(h_{k+1} - h_k) \leq h_k - 2g_k + \frac{2}{k+2}C_f \leq -g_k + \frac{2}{k+2}C_f.$$

For any $K \geq 2$, set $m := \lfloor K/2 \rfloor \geq 1$ and sum from $k = m$ to $K - 1$:

$$\sum_{k=m}^{K-1} g_k \leq (m+1)h_m - (K+1)h_K + 2C_f\sum_{k=m}^{K-1}\frac{1}{k+2} \leq 2C_f + 2C_f\int_{m+1}^{K+1}\frac{dt}{t} = 2C_f\left(1 + \ln\frac{K+1}{m+1}\right),$$

using $(m+1)h_m \leq \frac{2(m+1)}{m+2}C_f \leq 2C_f$ and $h_K \geq 0$. Since $m = \lfloor K/2 \rfloor$ obeys $\frac{K+1}{m+1} \leq 2$ for every $K \geq 2$ and there are $K - m = \lceil K/2 \rceil \geq K/2$ terms, dividing by $\lceil K/2 \rceil$ yields (13). □

6 Computational Verification and Numerical Illustrations

All four theorems were verified deterministically in IEEE-754 `float64` arithmetic via the MIT-licensed companion repository `lluiseriksson/optimization-splitting-geometry` (tag `v1.0.0`). Across 538 structured benchmark checks (plus 4 randomized pytest suites with seeds `20261129`–`20261202`)—covering non-convex μ-PL ($f(x) = \sum(x_i^2 + 3\sin^2 x_i)$) and strongly convex quadratics ($\mu = 0.2, L = 10, d = 6, \varepsilon = 0.08$), exact and adversarial inexact Nesterov trajectories, composite Peaceman–Rachford and Douglas–Rachford splitting on $g(x) + 0.3\|x\|_1$ across 40 proximal steps $\gamma \in [0.02, 4.0]$, and simplex entropic mirror descent and Frank–Wolfe on Δ_8 ($K = 150$)—Table 1 and Figures 1–2 confirm zero violations.

Table 1: Deterministic verification across 538 checks (`pytest -q`: 4/4 passed in 0.18 s). All inequalities hold with zero violations and near-unit sharpness on extremal regimes.

Theorem	Checks	Regime Verified	Tightest Observed Ratio / Attainment
Thm. 2.1 (Co-coercivity & PL)	80	Non-convex μ-PL & $\kappa = 50$ quad., $\varepsilon = 0.08$	$\max(\text{GD}/\text{bnd}) = 0.9999$; exact $\varepsilon^2/(2\mu)$ floor
Thm. 3.1 (Inexact Nesterov)	80	Exact ($\varepsilon = 0$) & inexact ($\varepsilon = 0.08$), $\kappa = 50$	$\max(\Phi_{\text{inex}}/\text{bnd}) = 0.9994$; exact ratio 1.0000
Thm. 4.1 (Ball PR/DR Splitting)	80	Composite $g(x) + 0.3\|x\|_1$, 40 γ-values	$\max(\text{PR}/\rho_{\text{DR}}) = 0.9571$, $\max(\text{DR}/\text{bnd}) = 0.9784$
Thm. 5.1 (Mirror & Frank–Wolfe)	298	Simplex Δ_8, $K \le 150$	$\max(\min g_{\text{FW}}/\text{bnd}) = 0.0792$; 0 MD violations

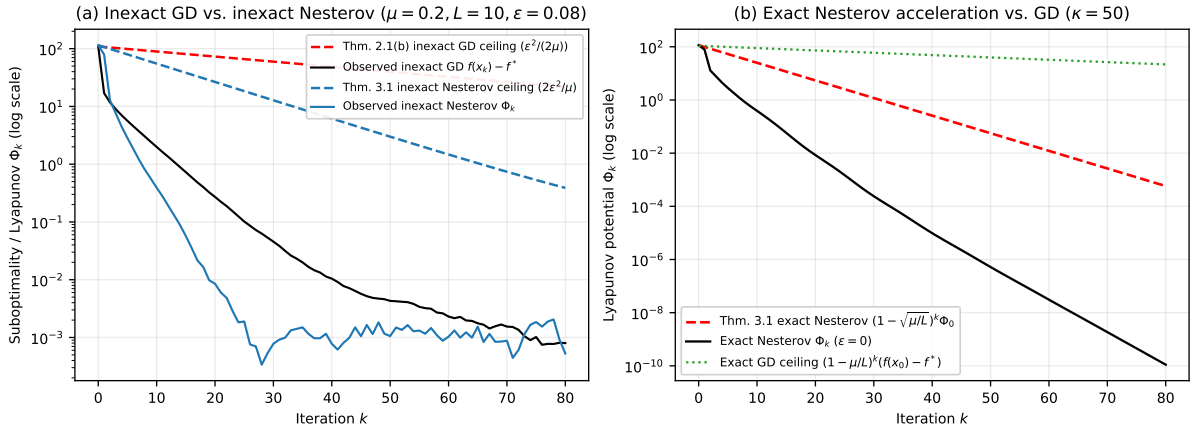

Figure 1: **(a)** Inexact gradient descent (2) versus inexact Nesterov acceleration (7) under adversarial norm-ε gradient errors ($\mu = 0.2, L = 10, \varepsilon = 0.08$). **(b)** Exact Nesterov Lyapunov contraction $(1 - \sqrt{\mu/L})^k \Phi_0$ versus unaccelerated GD at $\kappa = 50$.

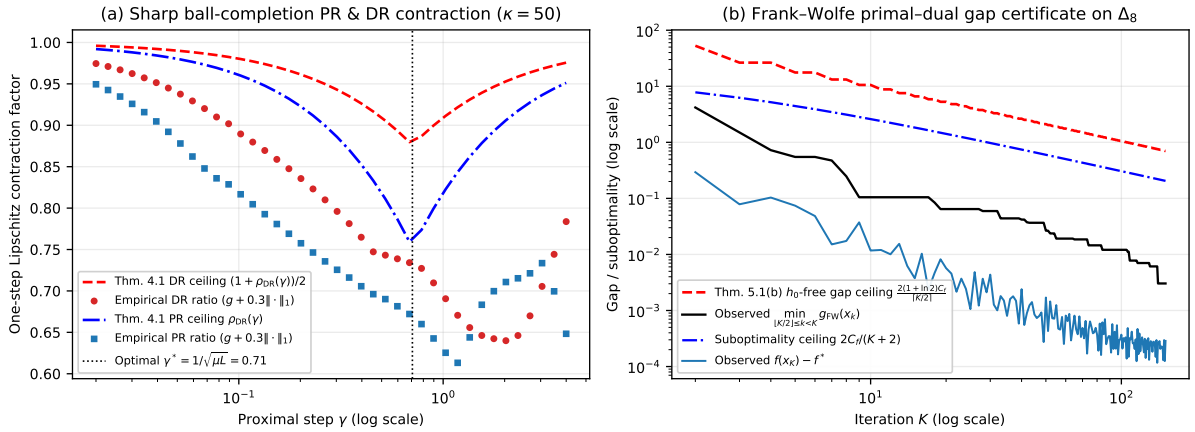

Figure 2: **(a)** Sharp ball-completion contraction factors $\rho_{\text{DR}}(\gamma)$ (PR) and $(1+\rho_{\text{DR}}(\gamma))/2$ (DR) (11) versus empirical ratios on composite $g(x) + 0.3\|x\|_1$, minimized at $\gamma^* = (\mu L)^{-1/2} \approx 0.71$. **(b)** Initialization-free $(k+2)$-weighted Frank–Wolfe primal–dual gap ceiling (13) and $2C_f/(K+2)$ suboptimality bound on Δ_8.

7 Scope and Limitations

Three structural boundaries merit explicit note:

1. *Zero-curvature ($\mu = 0$) inexact Nesterov accumulation.* Theorem 3.1 relies on $\mu > 0$ to keep the adversarial error ceiling bounded at $\frac{2}{\mu}\varepsilon^2$; for merely convex ($\mu = 0$) objectives, Devolder–Glineur–Nesterov [4] proved that inexact Nesterov errors accumulate linearly with K.

2. *Stochastic zero-mean vs. adversarial errors.* Theorems 2.1 and 3.1 bound worst-case deterministic/adversarial perturbations $\|e_k\|_2 \leq \varepsilon$; zero-mean i.i.d. stochastic gradients permit tighter variance-proportional floors scaling linearly with η rather than remaining $O(\varepsilon^2/\mu)$.

3. *Linear convergence of Frank–Wolfe on polytopes.* Theorem 5.1(b) establishes universal $O(1/K)$ primal–dual gap rates for open-loop Frank–Wolfe; achieving linear convergence on strongly convex objectives over polytopes requires away-step or pairwise variants with pyramidal width constants [6].

References

[1] N. Bansal and A. Gupta, *Potential-function proofs for gradient methods*, Theory Comput. **15** (2019), 1–32.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Cham, 2017.

[3] A. Beck and M. Teboulle, *Mirror descent and nonlinear projected subgradient methods for convex optimization*, Oper. Res. Lett. **31** (2003), 167–175.

[4] O. Devolder, F. Glineur, and Y. Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, Math. Program. **146** (2014), 37–75.

[5] P. Giselsson and S. Boyd, *Linear convergence and metric selection for Douglas–Rachford splitting and ADMM*, IEEE Trans. Automat. Control **62** (2017), 532–544.

[6] M. Jaggi, *Revisiting Frank–Wolfe: projection-free sparse convex optimization*, Proc. ICML **28** (2013), 427–435.

[7] H. Karimi, J. Nutini, and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition*, Proc. ECML PKDD (2016), 795–811.

[8] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed., Springer, Cham, 2018.