

A Sharper Theory of Ball-Proximal Optimization: Convergence and Radius Selection

Peter Richtárik Hanmin Li

King Abdullah University of Science and Technology
Thuwal 23955-6900, Saudi Arabia

28 September 2026

Abstract

We study the exact Euclidean ball-proximal point method for proper, closed, convex functions, where each iteration minimizes the objective over a ball centered at the current point. Retaining the objective gap in the decrease of squared distance yields sharper bounds on objective values, stationarity, and the symmetric Bregman distance to a minimizer. For constant radius $t > 0$ and initial distance $D_0 > 0$ to the solution set, the objective gap after K iterations is at most its initial value multiplied by $\exp(-2K^2t^2/D_0^2)$.

We characterize convergence for arbitrary positive radius sequences. If their sum diverges, the method reaches a minimizer in finitely many iterations whenever one exists; otherwise, the objective values converge to the infimum and the iterates escape every bounded set. If the radii are summable, the iterates converge to a possibly nonoptimal point. Self-contraction gives finite trajectory length for bounded iterates and a constant-radius termination bound $O_d(1 + D_0/t)$, whose implicit constant depends only on the dimension. A polyhedral family shows that the dimension-independent quadratic bound remains asymptotically sharp.

We also identify a minimum successful geometric decay factor and analyze adaptive radius rules based on subgradients or objective gaps, epigraph reformulation, and relaxed updates. Together, these results strengthen the foundations and convergence guarantees of the method without assuming smoothness or strong convexity.

Keywords: ball-proximal point method; convex optimization; finite termination; self-contracted sequences; radius selection; oracle complexity.

Mathematics Subject Classification (2020): Primary 90C25; Secondary 90C60, 65K05.

1 Introduction

We study the convex optimization problem

$$\min_{x \in \mathbb{R}^d} f(x), \tag{1}$$

where $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, closed, and convex. Unless stated otherwise, we assume neither that the infimum is attained nor that f is smooth. We focus on the exact ball-oracle iteration considered by Carmon et al. (2020, Appendix A) and subsequently developed as the ball-proximal (broximal) point framework by Gruntkowska et al. (2025). Starting from $x^0 \in \text{dom } f$, the ball-proximal point method (BPM) minimizes f over a Euclidean ball centered at the current point:

$$x^{k+1} \in \text{brox}_f^{t_k}(x^k) := \arg \min_{u \in \mathbb{B}(x^k, t_k)} f(u), \quad t_k > 0, \tag{BPM}$$

where $\mathbb{B}(x, t) = \{u : \|u - x\| \leq t\}$. A *call* to the exact ball oracle takes a center $x \in \text{dom } f$ and a radius $t > 0$ and returns a minimizer $u \in \text{brox}_f^t(x)$ of f over $\mathbb{B}(x, t)$; each BPM step consists of one such call. A step is *terminal* if its output is a global minimizer of f , and *nonterminal* otherwise. Equivalently, a step is terminal exactly when its ball contains a global minimizer. If the infimum is not attained, every step is nonterminal. Throughout the paper, we measure complexity by the number of oracle calls; each BPM iteration uses one such call.

Our analysis builds on two geometric properties of an exact step. First, retaining the objective gap in the decrease of squared distance yields sharper multi-step bounds on objective suboptimality. These bounds also give guarantees on the minimum subgradient norm and a symmetric Bregman distance to a minimizer, see [Sections 3, 3.3 and 3.4](#). Second, the geometry of an exact step relative to future iterates implies that bounded trajectories are self-contracted and have finite length in fixed dimension. This leads to the convergence characterization in [Section 4](#): nonsummable radii give finite termination when a minimizer exists and convergence of the objective values to the infimum otherwise. The same trajectory geometry yields a linear constant-radius step bound in fixed dimension, while a polyhedral construction shows that the dimension-independent quadratic bound remains asymptotically sharp.

We give a self-contained treatment intended to serve as a comprehensive reference for exact Euclidean convex BPM, rebuilding the inherited foundations and distinguishing them from the stronger guarantees established here. We then study how reformulation and radius choice affect the method, including epigraphical lifting, adaptive radius rules, and relaxed updates. We begin with the assumptions and notation, followed by seven equivalent formulations of an exact step and a comparison with prior work.

Standing assumptions. The assumptions below are invoked only when needed. In particular, [Assumption 2](#) (Attainment) is not required for results that allow the infimum to be unattained.

Assumption 1 (Convex objective). $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, closed, and convex, with $d < \infty$.

Assumption 2 (Attainment). The solution set $X^* = \arg \min f$ is nonempty, and we write $f_* = \min f$.

Assumption 3 (Exact BPM iteration). The starting point satisfies $x^0 \in \text{dom } f$, every radius is positive, and each update is an exact minimizer in BPM. Once a minimizer is reached, the sequence is held fixed. If $x^0 \in X^*$, no oracle call is made.

We say that BPM converges finitely if $x^N \in X^*$ for some finite N ; under the stopping convention above, this is finite termination. This property does not require that a point-only oracle provide a test for detecting optimality.

When [Assumption 2](#) (Attainment) holds, we use $D_k = \text{dist}(x^k, X^*)$ for the distance to the solution set and $\Delta_k = f(x^k) - f_*$ for the objective gap. For a constant-radius run, N_t counts iterations up to and including the first terminal step, with $N_t = 0$ at an optimal start and $N_t = \infty$ if no minimizer is ever reached. [Table 1](#) summarizes the main guarantees and points to their precise statements.

1.1 One optimality system, seven formulations

Before turning to convergence, we give seven equivalent descriptions of a single exact BPM step. The first is the ball-minimization formulation (BPM). The remaining six express the same step as a normalized implicit subgradient update, an explicit update for the ball envelope, a projection onto a sublevel set, a Fenchel-dual problem, a scalar dual problem, and an epigraph-constrained problem. Their equivalence is governed by a common first-order optimality condition.

To state these formulations for a generic exact step, suppose [Assumption 1](#) (Convex objective) holds, fix $x \in \text{dom } f$ and $t > 0$, and let $u \in \text{brox}_f^t(x)$.

Table 1: Convergence guarantees and radius policies at a glance. Each row separates the setting, the mathematical guarantee, and its location.

Quantity	Setting	Bound or conclusion	See
A. Accuracy after K iterations			
Objective gap	Fixed radius; sequence of nonterminal steps	$\Delta_K \leq \Delta_0 e^{-2K^2 t^2 / D_0^2}$	§3
Stationarity	Same setting; $K \geq 1$	$m_K \leq \frac{\Delta_0}{t} e^{-2(K-1)^2 t^2 / D_0^2}$	§3.3
Symmetric Bregman distance	Same setting; $K \geq 1$	$\mathcal{G}_K \leq \frac{D_0 \Delta_0}{t} e^{-2(K-1)^2 t^2 / D_0^2}$	§3.4
B. Convergence and termination			
Nonsummable radii	$\sum_{k=0}^{\infty} t_k = \infty$	$f(x^k) \downarrow \inf f$; finite termination if $X^* \neq \emptyset$, and $\ x^k\ \rightarrow \infty$ otherwise	§4
Summable radii	$\sum_{k=0}^{\infty} t_k < \infty$	$x^k \rightarrow \bar{x}$, $f(x^k) \downarrow f(\bar{x})$; finite length, possibly nonoptimal limit	§4
Iteration count	Fixed radius and dimension	$N_t = O_d\left(1 + \frac{D_0}{t}\right)$; no dimension-free linear bound	§4
Total length	Fixed radius; nearest-minimizer selection at termination	$S_t \leq \Psi_t(D_0) \leq \frac{D_0^2}{t} + \frac{t}{4}$	§4
Geometric radii	$t_k = t\rho^k$, $0 < t < D_0$	Minimum successful ρ_* ; $D_k = \Theta(\rho_*^k)$ there, without finite termination	App. D
C. Radius rules and reformulations			
Subgradient rule	$t_k = \tau \ h^k\ $, $h^k \in \partial f(x^k)$	$\Delta_K \leq \frac{D_0^2}{4\tau K}$	§6
Objective-gap rule	$t_k = \tau \Delta_k^\alpha$, known f_*	$\Delta_K \leq \left(\Delta_0^{-\alpha} + \frac{\alpha\tau K}{D_0}\right)^{-1/\alpha}$	§6
Epigraph lift	Start on the graph of f	Projected steps are BPM with induced radii; error and length conversions are explicit	§5
Radius ordering	Same objective and start	$s < t$ can give $N_s = 2 < N_t = 3$	App. E
Relaxation	Fixed $t > 0$, $0 < \lambda < 2$; centers in $\text{dom } f$	$x^k \rightarrow x^* \in X^*$	§7

$D_k = \text{dist}(x^k, X^*)$, $\Delta_k = f(x^k) - f_*$, and $m_K = \text{dist}(0, \partial f(x^K))$. The symmetric Bregman distance $\mathcal{G}_K$ is defined in §3.4; N_t includes the terminal step, and S_t includes its displacement.

Rows using X^* , D_k , or Δ_k assume attainment, except the two radius-summability rows, which explicitly distinguish the cases. Stronger implicit bounds and bounds over iterations $j, \dots, K-1$ appear in §3. Exact hypotheses are stated at the indicated locations.

We use the ball envelope introduced by Gruntkowska et al. (2025, Definition 5.1) and the convex conjugate:

$$F_t(y) = \inf_{\|v-y\| \leq t} f(v), \quad f^*(g) = \sup_z \{\langle g, z \rangle - f(z)\}. \quad (2)$$

The envelope F_t is itself proper, closed, and convex, but need not be smooth. For example, for $f(z) = |z|$ on $\mathbb{R}$,

$$F_t(x) \stackrel{(2)}{=} \max\{|x| - t, 0\},$$

which is nonsmooth at $x = \pm t$.

At iteration k of BPM, the generic notation above corresponds to $x = x^k$, $t = t_k$, and $u = x^{k+1}$. The proposition below first gives an optimality system that applies to both terminal and nonterminal iterations. The seven formulations are then stated for a nonterminal step, for which the normalized subgradient expressions are well defined. The terminal case is discussed immediately after the proposition.

Proposition 1.1 (Seven equivalent formulations of BPM). *Fix $x \in \text{dom } f$ and $t > 0$. The relation $u \in \text{brox}_f^t(x)$, where the operator is defined in BPM, is equivalent to the existence of g satisfying*

$$\boxed{g \in \partial f(u), \quad x - u \in t \partial \|\cdot\| (g)}. \quad (3)$$

The optimality system (3) applies to both terminal and nonterminal iterations. On a nonterminal step, the following seven formulations are equivalent, the output u is unique, and $\|x - u\| = t$.

(i) Exact ball minimization (the original formulation). *The relation $u \in \text{brox}_f^t(x)$ means*

$$u \in \arg \min_{v \in \mathbb{B}(x, t)} f(v). \quad (4)$$

This is the defining BPM step in BPM.

(ii) Normalized implicit subgradient step on f . *There exists a nonzero subgradient $g \in \partial f(u)$ satisfying*

$$u = x - t \frac{g}{\|g\|} = \text{prox}_{\lambda f}(x), \quad \lambda = \frac{t}{\|g\|}. \quad (5)$$

We call g a radial subgradient because it belongs to $\partial f(u)$ and is a positive multiple of $x - u$. Here the proximal map (Rockafellar, 1976) $\text{prox}_{\lambda f}(x)$ minimizes $f(v) + \|v - x\|^2 / (2\lambda)$. The subgradient is evaluated at the new point. An arbitrary subgradient of f there need not be radial.

(iii) Normalized explicit subgradient step on the ball envelope. *The same subgradient belongs to $\partial F_t(x)$, and*

$$u = x - t \frac{g}{\|g\|}, \quad g \in \partial F_t(x). \quad (6)$$

At such a center every envelope subgradient is nonzero and gives this same normalized direction. More precisely, for every $u \in \text{brox}_f^t(x)$, including an optimal point,

$$\partial F_t(x) = \{g \in \partial f(u) : \langle g, x - u \rangle = t \|g\|\}. \quad (7)$$

(iv) Projection onto the returned sublevel set. *With $C_u = \{z : f(z) \leq f(u)\}$,*

$$u = \text{Proj}_{C_u}(x), \quad \|x - u\| = t. \quad (8)$$

The full-radius condition is essential: projection onto a higher sublevel set may produce a point strictly inside the ball. The level $f(u)$ is determined by the ball-minimization problem itself and is not an independent parameter of the projection formulation.

(v) Fenchel-dual formulation. The subgradient solves

$$g \in \arg \min_h \{f^*(h) + t \|h\| - \langle x, h \rangle\}, \quad u \in \partial f^*(g), \quad x - u \in t \partial \|\cdot\| (g). \quad (9)$$

Equivalently, the primal and dual optimal values satisfy

$$F_t(x) = \max_g \{\langle x, g \rangle - f^*(g) - t \|g\|\}. \quad (10)$$

This dual problem has a solution even if $\partial f(x)$ is empty. When a previous subgradient $p \in \partial f(x)$ is available, it can also be written as

$$g \in \arg \min_h \{t \|h\| + D_{f^*}^x(h, p)\}, \quad D_{f^*}^x(h, p) = f^*(h) - f^*(p) - \langle x, h - p \rangle. \quad (11)$$

The term $D_{f^*}^x(h, p)$ is the generalized Bregman distance associated with f^* and the chosen subgradient $x \in \partial f^*(p)$ (Bregman, 1967; Burger, 2015). Thus the dual problem minimizes the norm penalty plus this generalized Bregman distance; neither smoothness nor strict convexity of f^* is required.

(vi) Lagrangian dual formulation. Dualizing the ball constraint gives

$$q_t(\gamma) = \inf_z \left\{ f(z) + \frac{\gamma}{2} (\|z - x\|^2 - t^2) \right\}, \quad F_t(x) = \max_{\gamma \geq 0} q_t(\gamma). \quad (12)$$

Here $q_t(0) = \inf f$. On a nonterminal step, an optimal $\gamma_\star > 0$ exists and

$$u = u_{\gamma_\star}, \quad u_\gamma = \text{prox}_{f/\gamma}(x), \quad \|u_{\gamma_\star} - x\| = t. \quad (13)$$

The function q_t is concave, is differentiable for $\gamma > 0$, and satisfies

$$q'_t(\gamma) = \frac{1}{2} (\|u_\gamma - x\|^2 - t^2). \quad (14)$$

The displacement $\|u_\gamma - x\|$ is continuous and nonincreasing in $\gamma > 0$. Thus an optimal multiplier $\gamma_\star$ selects a point on the proximal path whose distance from x is exactly the prescribed radius t . The optimal multiplier need not be unique.

(vii) Epigraph formulation with a cylindrical constraint. The ball-minimization problem can be lifted to the epigraph of f as

$$(u, f(u)) \in \arg \min_{v \in \mathbb{R}^d, s \in \mathbb{R}} \{s : f(v) \leq s, \|v - x\| \leq t\}. \quad (15)$$

The feasible set is $\text{epi } f \cap (\mathbb{B}(x, t) \times \mathbb{R})$, where $\text{epi } f = \{(v, s) : f(v) \leq s\}$. For any feasible v , minimizing over s gives $s = f(v)$, so the lifted problem is equivalent to minimizing $f(v)$ over $\mathbb{B}(x, t)$. Thus every minimizer v of the ball problem corresponds to the epigraph minimizer $(v, f(v))$, and conversely. This equivalence also holds on terminal iterations.

Proofs of these equivalences are given in Section C.1.

How the views are connected. Views (i) and (iv) describe the same point by exchanging the roles of objective value and distance. On a nonterminal step, let $\alpha = f(u) = F_t(x)$. Then

$$\min_{\|v-x\|\leq t} f(v) \stackrel{(4)}{=} \alpha, \quad \min_{f(v)\leq\alpha} \|v-x\| \stackrel{(8)}{=} t. \quad (16)$$

Thus view (i) fixes the radius t and determines the lowest attainable objective value α , whereas view (iv) fixes this attained level and recovers the same point as the nearest point in the corresponding sublevel set. In particular, α is determined by the ball-minimization problem rather than specified in advance.

Views (ii) and (iii) use the same subgradient at two different points. In the implicit formulation, $g \in \partial f(u)$ is evaluated at the new point u , while in the explicit formulation, the same subgradient satisfies $g \in \partial F_t(x)$ at the current center x . Both formulations therefore produce the same normalized direction from x to u .

View (v) is the Fenchel-dual formulation of the ball problem in view (i). The connection is provided by the ball envelope:

$$F_t \stackrel{(2)}{=} f \square \delta_{\mathbb{B}(0,t)}, \quad F_t^* = f^* + t \|\cdot\|. \quad (17)$$

Here $\square$ denotes infimal convolution, and δ_C is the indicator function of C , equal to zero on C and $+\infty$ outside C . The dual optimizers are precisely the elements of $\partial F_t(x)$. On a nonterminal step, these subgradients may have different magnitudes but determine the same normalized direction $x-u$. The Bregman representation in (11) additionally requires a subgradient $p \in \partial f(x)$ at the current center. Along a BPM trajectory, such a subgradient may be inherited from the preceding nonterminal step when it is retained by the oracle.

View (vi) gives a scalar alternative to the vector dual formulation. The multiplier γ_* selects a point on the proximal path in view (ii) whose displacement from x equals the prescribed radius t . The scalar and vector subgradients are related by

$$g \stackrel{(19)}{=} \gamma_*(x-u), \quad \gamma_* = \frac{\|g\|}{t}.$$

Since $\|u_\gamma - x\|$ is continuous and nonincreasing in γ , an optimal multiplier can be found by solving the one-dimensional equation $\|u_\gamma - x\| = t$, for example by a bracketing method when proximal evaluations are available. This is the Lagrange-multiplier form of the proximal and trust-region connections described in Gruntkowska et al. (2025, Sections 3 and 7).

Finally, view (vii) lifts the ball problem to the epigraph of f . The spatial constraint $\|v-x\| \leq t$ is unchanged, while the objective value is represented by the additional variable s . Since minimizing s subject to $f(v) \leq s$ forces $s = f(v)$ at optimality, the lifted problem has exactly the same minimizing v as view (i). This is the standard epigraph reformulation of an optimization problem (Boyd and Vandenberghe, 2004, Section 4.1.3).

Terminal iterations. On a terminal step, the ball $\mathbb{B}(x, t)$ contains a global minimizer, so every exact BPM output is itself a global minimizer. The optimality system (3) remains valid with $g = 0$: since $\partial \|\cdot\| (0)$ is the unit ball, the inclusion $x-u \in t\partial \|\cdot\| (0)$ is equivalent to $\|x-u\| \leq t$. The normalized subgradient formulations do not apply with $g = 0$, since $g/\|g\|$ is undefined.

In the scalar dual formulation, $\gamma = 0$ is optimal because $q_t(0) = \inf f = F_t(x)$, although this does not determine which minimizer in the ball is returned. Likewise, in the projection formulation, $C_u = X^*$, so $\text{Proj}_{C_u}(x)$ selects a nearest minimizer.

The normalized implicit and explicit subgradient descriptions, together with their radial geometry, build on Gruntkowska et al. (2025, Table 1 and Sections 3 and 5). The connection with sublevel projection also appears in proximal self-contraction theory (Danilidis et al., 2015).

A common example for the seven formulations. The following example illustrates how the seven formulations describe the same exact BPM step. Consider

$$f(z) = \frac{1}{2}(z_1^2 + 4z_2^2), \quad x = \left(2, \frac{5}{4}\right), \quad t = \sqrt{2}. \quad (18)$$

The ball minimization problem returns $u = (1, 1/4)$, for which

$$g = \nabla f(u) \stackrel{(18)}{=} (1, 1) = x - u, \quad f(u) = \frac{5}{8}, \quad \lambda \stackrel{(20)}{=} \frac{t}{\|g\|} = 1.$$

Thus the common optimality system (3) holds, and the same step can be read through views (i)–(vi) in Figures 1 to 4. For view (vii), the corresponding epigraph optimizer is $(u, f(u)) = ((1, 1/4), 5/8)$. Although this example is smooth for ease of visualization, the equivalences in Theorem 1.1 do not require smoothness.

1.2 Sharpening and extending the BPM theory

The original analysis of Grunkowska et al. (2025) establishes finite termination within $\lceil D_0^2/t^2 \rceil$ constant-radius iterations and the objective suboptimality bound $\Delta_K \leq \Delta_0(1 + t/D_0)^{-K}$. The results developed here sharpen and extend this theory in several directions. In Section 3, retaining the objective decrease in the bound on squared-distance decrease yields stronger multi-step bounds on objective suboptimality, which in turn give guarantees on the minimum subgradient norm and a symmetric Bregman distance to a minimizer. The original analysis already allows arbitrary positive radii and guarantees finite termination when their squared sum reaches the initial squared distance to the solution set (Grunkowska et al., 2025, Theorem 8.1 and Corollary 8.2). In Section 4, we strengthen this theory by characterizing convergence through the sum of the radii themselves and by treating problems whose infimum is not attained. When a minimizer exists, the same trajectory analysis improves the constant-radius step bound to linear dependence on D_0/t in fixed dimension, while a polyhedral family shows why the dimension-independent quadratic bound remains asymptotically sharp.

The later sections develop the effects of reformulation, adaptive radius selection, and relaxation beyond the convergence guarantees in the original analysis. Section 5 determines which reformulations preserve exact ball calls and shows how epigraphical lifting changes the induced radius sequence and associated guarantees. Section 6 studies the effect of radius choice, including geometric schedules, nonmonotonicity of the fixed-radius termination count, and adaptive rules based on subgradient or objective-gap information. Finally, Section 7 extends the analysis to relaxed updates of the ball-proximal step.

1.3 Related work

The exact recurrence BPM appears in Carmon et al. (2020, Appendix A, Algorithm 6), together with the segment contraction used in Theorem 3.1. The later work of Grunkowska et al. (2025) develops the broximal framework, its geometric and proximal connections, and the finite-termination analysis on which we build.

Ball optimization oracles also serve as primitives for accelerated convex optimization. Carmon et al. (2020) show that additional state and different query rules can reduce the number of oracle calls compared with the BPM recurrence. Carmon et al. (2021) further refine ball-oracle acceleration and develop an implementation for maximal-loss minimization through a smoothed objective. These works therefore address acceleration and the cost of implementing ball oracles.

Several subsequent works develop related broximal constructions in other settings. Non-Euclidean BPM replaces the Euclidean ball by more general norm geometry (Grunkowska and Richtárik, 2025); stabilized proximal methods combine proximal updates with trust-region-type constraints (Li et al., 2026); and broximal alignment extends related geometric ideas beyond the convex setting (Grunkowska et al., 2026). Local linear minimization methods also admit a

Figure 1. Ball minimization and sublevel projection



Figure 2. Implicit and envelope updates

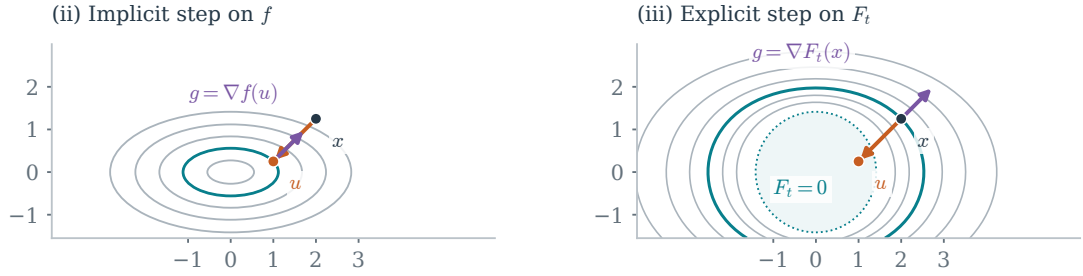


Figure 3. Vector duality

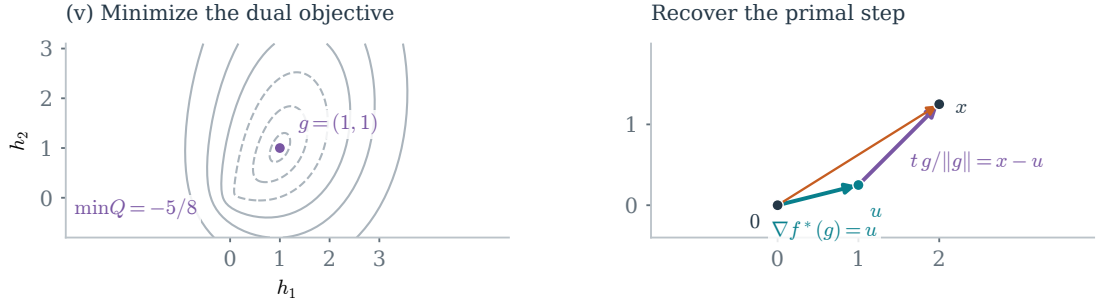
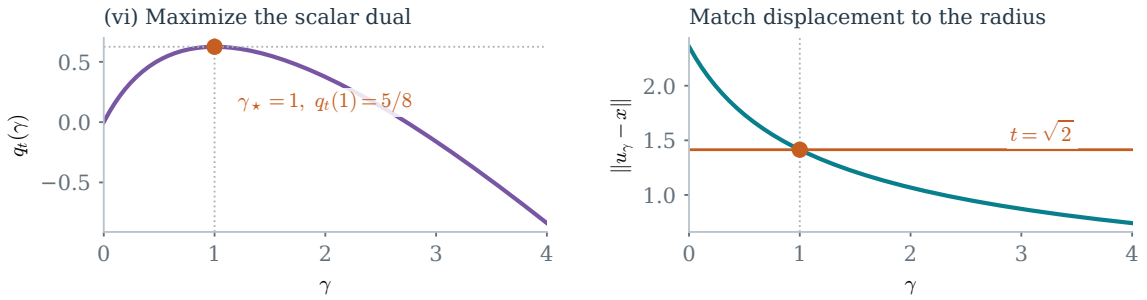


Figure 4. Scalar duality



Illustrations of formulations (i)–(vi) in [Theorem 1.1](#) for a single BPM step with $f(z) = \frac{1}{2}(z_1^2 + 4z_2^2)$, $x = (2, 5/4)$, and $t = \sqrt{2}$, yielding $u = (1, 1/4)$ and $g = (1, 1)$. [Figure 1](#) illustrates parts (i) and (iv): ball minimization (4) and projection onto the achieved sublevel $f(u) = 5/8$ in (8). [Figure 2](#) illustrates parts (ii) and (iii): the implicit step on f in (5) and the explicit step on F_t in (6), using the same subgradient g . [Figure 3](#) illustrates part (v), (9): the vector-dual minimum $Q(g) = -5/8$ and the reconstruction $x = \nabla f^*(g) + tg/\|g\|$. [Figure 4](#) illustrates part (vi): the scalar-dual maximum $q_t(1) = 5/8$ in (12) and the matching displacement $\|u_1 - x\| = t$ in (13).

forward–backward interpretation in which a proximal backward step is replaced by a broximal one (Richtárik et al., 2026).

The trajectory-length argument uses the classical theory of self-contracted curves and sequences (Daniilidis et al., 2015; Böhm and Daniilidis, 2022). The analysis of relaxed centers uses the cutter framework of Cegielski and Censor (2012) to obtain a Fejér-type decrease. The subgradient-based radius rule uses classical proximal-point analysis (Rockafellar, 1976) and the sharp value bound of Taylor et al. (2017); the objective-gap-based rules follow directly from the BPM segment contraction.

2 One-step geometry and progress

Throughout this section, we invoke [Assumption 1](#) (Convex objective) and [Assumption 3](#) (Exact BPM iteration) and impose [Assumption 2](#) (Attainment) only when a minimizer is required. Under [Assumption 2](#) (Attainment), let $p^k = \text{Proj}_{X^*}(x^k)$. The terminal and nonterminal cases introduced in [Section 1.1](#) are then characterized by $D_k \leq t_k$ and $D_k > t_k$, respectively.

2.1 Geometry of an exact step

The next lemma collects the geometric properties of an exact BPM step used below. A nonterminal step has a unique output on the boundary of the ball, together with a radial subgradient and an equivalent proximal representation. Under attainment, the lemma also characterizes the possible terminal outputs. These facts are established in [Gruntkowska et al. \(2025, Appendices D–E\)](#), we include the proof in [Section C.2](#).

Lemma 2.1 (Characterization of an exact BPM step). *Under [Assumption 1](#) (Convex objective), let $x \in \text{dom } f$ and $t > 0$. The set $\text{brox}_f^t(x)$ is nonempty and compact. If the step is nonterminal, the output u is unique and satisfies*

$$\|x - u\| = t, \quad g = c(x - u) \in \partial f(u), \quad c > 0. \quad (19)$$

Under [Assumption 2](#) (Attainment), if $D(x) := \text{dist}(x, X^) \leq t$, then $\text{brox}_f^t(x) = X^* \cap \mathbb{B}(x, t)$. At $D(x) = t$ this set is the singleton $\{\text{Proj}_{X^*}(x)\}$. On a nonterminal step,*

$$u = \text{prox}_{\lambda f}(x), \quad \lambda = \frac{1}{c} = \frac{t}{\|g\|}, \quad (20)$$

where $\text{prox}_{\lambda f}(x) = \arg \min_z \{f(z) + \|z - x\|^2 / (2\lambda)\}$.

The subgradient g in (19) need not be returned by the ball oracle. Its existence is sufficient for the analysis. When $D(x) < t$, a terminal step may have several possible outputs. For trajectory arguments that require a definite terminal point, we choose the minimizer nearest to the current center.

Definition 2.2 (Nearest-minimizer selection at termination). Under [Assumption 1](#) (Convex objective) and [Assumption 2](#) (Attainment), for $x \in \text{dom } f$ define

$$T_t(x) = \begin{cases} u, & D(x) > t, \quad \text{where } \text{brox}_f^t(x) = \{u\}, \\ \text{Proj}_{X^*}(x), & D(x) \leq t. \end{cases}$$

Then

$$\|x - T_t(x)\| = \min\{t, D(x)\}.$$

Before the first terminal step, this selection coincides with BPM, so it preserves the termination count. Selecting the nearest terminal minimizer requires additional information: a point-only exact ball oracle need not return that particular minimizer.

2.2 Progress from an exact step

We use the generalized Bregman distance associated with a specified subgradient (Bregman, 1967; Burger, 2015). For $g \in \partial f(u)$ and $z \in \text{dom } f$, this distance from z to u is defined by

$$D_f^g(z, u) = f(z) - f(u) - \langle g, z - u \rangle. \quad (21)$$

By the subgradient inequality, $D_f^g(z, u) \geq 0$. This quantity depends on the chosen subgradient and is not generally a metric. The following theorem gives the basic one step progress relations.

Theorem 2.3 (Exact progress identity). *Under [Assumption 1](#) (Convex objective) and [Assumption 3](#) (Exact BPM iteration), consider a nonterminal step with output $u = x^{k+1}$ and a subgradient $g^k = c_k(x^k - u)$ from [Theorem 2.1](#). For every $z \in \text{dom } f$,*

$$\|x^k - z\|^2 - \|u - z\|^2 = t_k^2 + \frac{2}{c_k} [f(u) - f(z) + D_f^{g^k}(z, u)]. \quad (22)$$

The objective decrease satisfies

$$f(x^k) - f(u) = t_k \|g^k\| + D_f^{g^k}(x^k, u). \quad (23)$$

In particular, whenever $f(z) \leq f(u)$,

$$\|u - z\|^2 \leq \|x^k - z\|^2 - t_k^2. \quad (24)$$

The proof is given in [Section C.28](#).

Equation (22) retains the objective-gap and Bregman terms in the distance decrease, while (23) relates objective decrease to the radial subgradient. The lower-level estimate (24) follows by discarding these additional nonnegative terms.

Example 2.4 (Slack in the squared-radius estimate). Take $f(x) = |x|$, $x^0 = 3$, and $t = 1$. The first ball is $[2, 4]$, so $u = 2$, $g = 1$, and $c = 1$. For the minimizer $z = 0$, the Bregman distance vanishes, and (22) gives

$$\underbrace{3^2 - 2^2}_{5} \stackrel{(22)}{=} \underbrace{1^2}_{\text{radius squared}} + \underbrace{2 \cdot 2}_{\text{objective-gap contribution}}.$$

The squared-radius estimate (24) retains only the first term, whereas the exact identity (22) accounts for the remaining four units through the objective gap. By (23), the objective decrease is $3 - 2 = t|g| = 1$. With the same fixed radius, the iterates are 3, 2, 1, 0, so the method reaches the minimizer in three iterations. In contrast, telescoping only the squared-radius decrease in (24) gives the generic bound $D_0^2/t^2 = 9$. This example shows that the squared-radius bound can be loose on a particular instance.

A nonterminal step also determines a halfspace containing the lower sublevel set. With $n_k = (x^k - u)/t_k$, the subgradient inequality gives

$$H_k := \{z : \langle n_k, z - u \rangle \leq 0\} \stackrel{(19)}{\supseteq} \{z : f(z) \leq f(u)\}, \quad u = \text{Proj}_{H_k}(x^k). \quad (25)$$

Thus the lower sublevel set lies in a halfspace through u , and u is the projection of x^k onto that halfspace.

Dropping the additional nonnegative terms in (22) recovers the classical squared-radius decrease.

Corollary 2.5 (Squared-radius bound). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), every nonterminal step satisfies $D_{k+1}^2 \leq D_k^2 - t_k^2$. Consequently, one of the first K iterations is terminal whenever*

$$\sum_{k=0}^{K-1} t_k^2 \geq D_0^2.$$

In particular, for a constant radius t ,

$$N_t \leq \left\lceil \frac{D_0^2}{t^2} \right\rceil. \quad (26)$$

For the stopped sequence,

$$D_K^2 \leq \max \left\{ D_0^2 - \sum_{k=0}^{K-1} t_k^2, 0 \right\},$$

where the radius sequence may be extended arbitrarily after stopping.

The proof is given in [Section C.29](#).

2.3 Subgradients associated with exact BPM steps

For a common radius, the corresponding comparison of radial subgradient norms appears in [Grunkowska et al. \(2025, Corollary E.13 and Remark E.14\)](#), including in the nonsmooth setting. The following argument also allows variable radii and arbitrary radial subgradients.

Proposition 2.6 (Nonincreasing norms of radial subgradients). *Under [Assumption 1](#) (Convex objective) and [Assumption 3](#) (Exact BPM iteration), suppose two consecutive iterations are nonterminal. Then any corresponding radial subgradients satisfy*

$$\left\langle g^k, \frac{g^{k+1}}{\|g^{k+1}\|} \right\rangle \geq \|g^{k+1}\|. \quad (27)$$

Consequently, $\|g^{k+1}\| \leq \|g^k\|$, and consecutive radial subgradient directions have positive inner product. Under [Assumption 2](#) (Attainment), after $K \geq 1$ iterations, taking the subgradient to be zero at and after a terminal output gives

$$\text{dist}(0, \partial f(x^K)) \leq \|g^{K-1}\| \leq \frac{f(x^0) - f(x^K)}{\sum_{k=0}^{K-1} t_k} \leq \frac{\Delta_0}{\sum_{k=0}^{K-1} t_k}. \quad (28)$$

The proof is given in [Section C.3](#). The subgradient bound will be used below to derive stationarity estimates and adaptive radius rules.

Corollary 2.7 (Monotonicity of the minimum subgradient norm). *Under [Assumption 1](#) (Convex objective), for $x \in \text{dom } \partial f$ and any exact output u , $m(u) \leq m(x)$, where $m(x) = \text{dist}(0, \partial f(x))$.*

The proof is given in [Section C.30](#).

3 Sharper objective-gap bounds for BPM

Throughout this section [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration) hold. We first study the objective gap

$$\Delta_k = f(x^k) - f_\star.$$

The analysis proceeds in two stages. We begin with a one-step contraction obtained by considering a feasible point on the segment from x^k to its projection $p^k = \text{Proj}_{X^\star}(x^k)$ onto the solution set. We then retain the actual objective decrease in each step and relate it to the decrease in squared distance, which leads to stronger multi-step bounds. In the second part of the section, we use these objective-gap estimates to control stationarity and a symmetric Bregman distance to a minimizer.

3.1 Objective-gap contraction and squared-distance decrease

The following result uses the segment argument of [Carmon et al. \(2020, Appendix A, Theorem 26\)](#), stated here under our assumptions and for variable radii.

Theorem 3.1 (Segment contraction). *If $D_k > 0$, then*

$$\Delta_{k+1} \leq \left(1 - \min\left\{\frac{t_k}{D_k}, 1\right\}\right) \Delta_k. \quad (29)$$

If $D_k = 0$, both gaps are zero. For $0 < t < D_0$ and constant radius, the stopped sequence satisfies

$$\Delta_K \leq \left(1 - \frac{t}{D_0}\right)^K \Delta_0. \quad (30)$$

The proof is given in [Section C.4](#). For comparison, [Grunkowska et al. \(2025, Corollary 8.2\)](#) gives the fixed-radius factor $(1 + t/D_0)^{-K}$. Since $1 - a \leq (1 + a)^{-1}$ for $0 < a < 1$, (30) improves this original geometric factor. We therefore use $(1 - t/D_0)^K$ as the reference geometric bound below.

Multiplying (29) along a sequence of nonterminal steps gives

$$\Delta_K \stackrel{(29)}{\leq} \Delta_0 \prod_{k < K} \left(1 - \frac{t_k}{D_k}\right). \quad (31)$$

For constant radius, [Theorem 2.5](#) further gives

$$\Delta_K \stackrel{(31),(24)}{\leq} \Delta_0 \prod_{k < K} \left(1 - \frac{t}{\sqrt{D_0^2 - kt^2}}\right).$$

If the first K iterations are nonterminal, then $D_k > t$ for every $k < K$, so all factors are positive. If a terminal step occurs earlier, then $\Delta_K = 0$.

The same one step geometry also yields a direct relation between objective suboptimality and distance to the solution set. The proof is given in [Section C.5](#).

Corollary 3.2 (Objective gap–distance comparison). *If $D_0 > 0$, then for arbitrary positive radii,*

$$\Delta_K \leq \frac{D_K}{D_0} \Delta_0. \quad (32)$$

The segment bound in (29) can be exact: for $f(x) = |x|$ with $x > t > 0$, it is attained with equality. It also applies to variable radii and does not require access to radial subgradients. However, it tracks objective decrease and distance decrease only through the current distance D_k . The next result retains the stronger coupling between these two quantities.

Theorem 3.3 (Refined distance decrease from objective progress). *At any nonterminal step,*

$$D_{k+1}^2 \leq D_k^2 - t_k^2 - \frac{2\Delta_{k+1}}{c_k}, \quad (33)$$

$$D_{k+1}^2 \leq D_k^2 - t_k^2 \frac{\Delta_k + \Delta_{k+1}}{\Delta_k - \Delta_{k+1}}. \quad (34)$$

In particular, the denominator is positive and the multiplier of t_k^2 in (34) exceeds one.

The proof is given in Section C.6. Equation (33) is the distance refinement in Gruntkowska et al. (2025, equation (23)). Eliminating the multiplier with their objective-decrease estimate gives (34). To make this relation explicit, define

$$q_k := \frac{\Delta_{k+1}}{\Delta_k}, \quad \beta_k := \frac{D_k^2 - D_{k+1}^2}{t_k^2}. \quad (35)$$

Here, q_k is the fraction of the objective gap remaining after the k -th step, while β_k measures the corresponding decrease in squared distance to the solution set, normalized by t_k^2 . The refined inequality (34) then gives

$$\beta_k \stackrel{(34),(35)}{\geq} \frac{1 + q_k}{1 - q_k}. \quad (36)$$

The above inequality shows that a small relative decrease in the objective gap must be accompanied by a large decrease in squared distance to the solution set. Indeed, if q_k is close to 1, then $(1 + q_k)/(1 - q_k)$ is large, hence β_k must also be large. Since the total decrease in squared distance is limited, such behavior cannot persist over many iterations with the same radius. We make this observation quantitative in the next subsection.

On the other hand, β_k cannot be arbitrarily large. At a nonterminal step, $\|x^{k+1} - x^k\| = t_k$, and the triangle inequality gives

$$D_{k+1} = \text{dist}(x^{k+1}, X^*) \geq D_k - \|x^{k+1} - x^k\| \stackrel{(19)}{=} D_k - t_k. \quad (37)$$

Therefore,

$$\beta_k \stackrel{(35)}{=} \frac{D_k^2 - D_{k+1}^2}{t_k^2} \stackrel{(37)}{\leq} \frac{D_k^2 - (D_k - t_k)^2}{t_k^2} = \frac{2D_k}{t_k} - 1. \quad (38)$$

Combining the lower and upper bounds on β_k yields

$$\frac{1 + q_k}{1 - q_k} \stackrel{(36),(38)}{\leq} \frac{2D_k}{t_k} - 1, \quad (39)$$

which is equivalent to

$$q_k \stackrel{(39)}{\leq} 1 - \frac{t_k}{D_k}.$$

Thus the refined distance inequality recovers the segment contraction, while retaining additional information about how objective-gap decrease is coupled to decrease in squared distance.

3.2 Refined multi-step objective-gap bounds

We now aggregate the relation between objective-gap decrease and squared-distance decrease across multiple iterations. Suppose the radius is constant, $t_k = t$, and define $S := D_0^2/t^2$. For the first K iterations, all assumed nonterminal,

$$\sum_{k=0}^{K-1} \beta_k \stackrel{(35)}{=} \frac{1}{t^2} \sum_{k=0}^{K-1} (D_k^2 - D_{k+1}^2) = \frac{D_0^2 - D_K^2}{t^2} < S. \quad (40)$$

Thus the quantities β_k share a finite total budget. Moreover, the inequality $\beta_k \geq (1 + q_k)/(1 - q_k)$ is equivalent to $q_k \leq (\beta_k - 1)/(\beta_k + 1)$. Since

$$\frac{\Delta_K}{\Delta_0} \stackrel{(35)}{=} \prod_{k < K} q_k, \quad (41)$$

we obtain

$$\frac{\Delta_K}{\Delta_0} \stackrel{(41),(36)}{\leq} \prod_{k < K} \frac{\beta_k - 1}{\beta_k + 1}.$$

The remaining question is therefore how large this product can be when the β_k satisfy the total-budget constraint above. To apply Jensen's inequality, take the negative logarithm of each factor and define $\phi(b) = -\log((b-1)/(b+1))$ for $b > 1$. This function is strictly convex, since $\phi''(b) = 4b/(b^2 - 1)^2 > 0$. Writing $\bar{\beta} = K^{-1} \sum_{k=0}^{K-1} \beta_k$, Jensen's inequality with equal weights $1/K$ gives

$$\frac{1}{K} \sum_{k=0}^{K-1} \phi(\beta_k) \stackrel{\text{Jensen}}{\geq} \phi(\bar{\beta}).$$

Multiplying by $-K$ reverses the inequality; exponentiating then gives

$$\prod_{k=0}^{K-1} \frac{\beta_k - 1}{\beta_k + 1} = \exp\left(-\sum_{k=0}^{K-1} \phi(\beta_k)\right) \stackrel{\text{Jensen}}{\leq} \exp(-K\phi(\bar{\beta})) = \left(\frac{\bar{\beta} - 1}{\bar{\beta} + 1}\right)^K.$$

Equality holds when all β_k are equal, so equal allocation gives the largest product for a fixed total. Finally, the fraction $(b-1)/(b+1)$ increases with $b > 1$, and the budget in (40) gives $\bar{\beta} < S/K$. Replacing $\bar{\beta}$ by S/K therefore yields the upper bound $((S-K)/(S+K))^K$. This leads to the next theorem.

Theorem 3.4 (Jensen objective-gap bound). *Let $t_k = t > 0$, $D_0 > 0$, and $S = D_0^2/t^2$. If the first $K \geq 1$ iterations are nonterminal, then $K < S$ and*

$$\Delta_K \leq \Delta_0 \left(\frac{S-K}{S+K}\right)^K \leq \Delta_0 \exp\left(-\frac{2K^2}{S}\right). \quad (42)$$

For $0 < \varepsilon < \Delta_0$, the first index with $\Delta_K \leq \varepsilon$ obeys

$$N_\varepsilon \leq \min \left\{ \lceil S \rceil, \left\lceil \frac{D_0}{t} \sqrt{\frac{1}{2} \log \frac{\Delta_0}{\varepsilon}} \right\rceil \right\}. \quad (43)$$

The proof is given in [Section C.7](#). For the first step in [Theorem 2.4](#), the segment bound is exact, giving $\Delta_1 \leq 2$, whereas the Jensen bound gives $\Delta_1 \leq 2.4$. Thus the Jensen bound is not necessarily sharper at short horizons, its advantage comes from aggregating information across multiple iterations.

The above Jensen bound and the geometric bound based on the initial distance capture different parts of the preceding analysis, and neither is uniformly stronger. For $a = t/D_0 \in (0, 1)$, a direct comparison gives

$$\left(\frac{S-K}{S+K}\right)^K \leq (1-a)^K \iff K \geq \frac{1}{a(2-a)}.$$

Thus the geometric bound based on the initial distance can be sharper over short horizons, as in [Theorem 2.4](#), whereas the Jensen bound becomes sharper once sufficiently many iterations are aggregated. The factor $((S-K)/(S+K))^K$ cannot be improved using only the constraints $\beta_k > 1$, $\sum_{k=0}^{K-1} \beta_k < S$, and $q_k \leq (\beta_k - 1)/(\beta_k + 1)$. To see this, take all β_k equal to a number b with $1 < b < S/K$, and set $q_k = (b - 1)/(b + 1)$. As b increases toward S/K , the product of the q_k approaches the stated factor arbitrarily closely. Equality is excluded by the strict budget constraint, since it would require $\sum_{k=0}^{K-1} \beta_k = S$. An actual trajectory satisfies additional constraints, however, and these give a strict improvement for fixed S and K .

Remark 3.5 (Using the final distance). The Jensen bound in [Theorem 3.4](#) uses only $\sum_{k < K} \beta_k < S$ and ignores the remaining distance D_K . Let $r = \Delta_K/\Delta_0 \in (0, 1)$. By [\(32\)](#), $D_K \geq rD_0$, so

$$\sum_{k=0}^{K-1} \beta_k \stackrel{(40)}{=} \frac{D_0^2 - D_K^2}{t^2} \stackrel{(32)}{\leq} S(1 - r^2). \quad (44)$$

Applying the same Jensen argument with this smaller budget gives

$$r \stackrel{(44)}{\leq} \left(\frac{S(1 - r^2) - K}{S(1 - r^2) + K}\right)^K, \quad r^2 + \frac{K}{S} \frac{1 + r^{1/K}}{1 - r^{1/K}} \leq 1. \quad (45)$$

Since each $\beta_k > 1$, we have $S(1 - r^2) > K$. The left-hand side of the second inequality is strictly increasing on $[0, 1)$, equals $K/S < 1$ at $r = 0$, and tends to infinity as $r \rightarrow 1$. Hence there is a unique $r_\star \in (0, 1)$ at which it equals 1, and [\(45\)](#) implies $r \leq r_\star$. Moreover, $r_\star < B = ((S-K)/(S+K))^K$, since substituting B gives $1 + B^2 > 1$. Thus [\(42\)](#) is not the exact worst-case factor for BPM at fixed S and K , although its explicit form remains useful. For $K = 1$, the root equation simplifies to $(1 - r_\star)^2 = 1/S$, so $r_\star = 1 - t/D_0$ recovers the segment bound. Sharpness of the implicit bound for general K remains open.

3.3 From objective decrease to stationarity

Each nonterminal step provides a stationarity certificate at its returned point. Indeed, the implicit formulation [\(5\)](#) gives $g^k \in \partial f(x^{k+1})$, and hence $m(x^{k+1}) \leq \|g^k\|$. Moreover, the objective decrease identity [\(23\)](#) in [Theorem 2.3](#) gives

$$\Delta_k - \Delta_{k+1} \stackrel{(23)}{=} t_k \|g^k\| + D_f^{g^k}(x^k, x^{k+1}) \stackrel{(21)}{\geq} t_k \|g^k\|.$$

Thus a bound on the preceding objective gap can be transferred directly to a bound on stationarity at the returned point.

Corollary 3.6 (Stationarity from objective decrease). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), for every $K \geq 1$ and arbitrary positive radii,*

$$m(x^K) \leq \frac{\Delta_{K-1} - \Delta_K}{t_{K-1}} \leq \frac{\Delta_{K-1}}{t_{K-1}}. \quad (46)$$

For a constant radius $0 < t < D_0$, combining this with (30) gives

$$m(x^K) \leq \frac{\Delta_0}{t} \left(1 - \frac{t}{D_0}\right)^{K-1}, \quad K \geq 1. \quad (47)$$

At and after the first optimal output, $m(x^K) = 0$. For differentiable f , $m(x^K) = \|\nabla f(x^K)\|$, so (47) also gives a gradient-norm bound without assuming a Lipschitz gradient.

The proof is given in [Section C.31](#).

This does not imply that $m(x^k)$ contracts by a uniform factor strictly below one at every iteration; the monotonicity in [Theorem 2.7](#) still applies.

Refining the stationarity bound over several iterations. [Theorem 3.6](#) uses only the last step. A stronger estimate follows by summing objective decrease over iterations $j, \dots, K-1$. By the monotonicity of the radial subgradient norms in [Theorem 2.6](#), for $j \leq i < K$,

$$\|g^i\| \stackrel{(27)}{\geq} \|g^{K-1}\| \geq m(x^K).$$

Hence the decrease accumulated from iteration j to K controls the final stationarity measure through the total radius used over those iterations. Combining this bound over iterations $j, \dots, K-1$ with the objective-gap bounds above yields the refined stationarity bounds below.

Theorem 3.7 (Refined stationarity bounds). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), for arbitrary positive radii and every $K \geq 1$,*

$$m(x^K) \leq \min_{0 \leq j < K} \frac{\Delta_j - \Delta_K}{\sum_{i=j}^{K-1} t_i} \leq \min_{0 \leq j < K} \frac{\Delta_j}{\sum_{i=j}^{K-1} t_i}. \quad (48)$$

For a constant radius $0 < t < D_0$, set $S = D_0^2/t^2$ and suppose the first K iterations are nonterminal. Then $K < S$ and

$$m(x^K) \leq \frac{\Delta_0}{t} \left(\frac{S - K + 1}{S + K - 1} \right)^{K-1} \leq \frac{\Delta_0}{t} \exp \left(-\frac{2(K-1)^2}{S} \right). \quad (49)$$

For integers $0 \leq j < S$, define

$$E_j := \min \left\{ \left(1 - \frac{t}{D_0}\right)^j, \left(\frac{S-j}{S+j}\right)^j \right\}, \quad (50)$$

Then

$$m(x^K) \leq \frac{\Delta_0}{t} \min_{0 \leq j < K} \frac{E_j}{K-j}. \quad (51)$$

When f is differentiable, the same bounds apply to $\|\nabla f(x^K)\|$.

The proof is given in [Section C.8](#). The bound over iterations $j, \dots, K-1$ (48) uses the actual gaps Δ_j and Δ_K , whereas (51) replaces them by the objective-gap bounds derived above. In (51), the index j determines the first iteration included in the bound and is used only in the analysis, it does not affect the BPM iteration. Choosing $j = 0$ uses the full radius sum and gives $\Delta_0/(Kt)$, while choosing $j = K-1$ reduces to the single step estimate with the better of

the geometric and Jensen bounds for Δ_{K-1} . Intermediate values of j trade a smaller objective gap later in the trajectory against a shorter remaining radius sum.

Shrinking radii. For variable radii, convergence of the objective gap alone need not imply convergence to stationarity. Indeed, in (46), the denominator may shrink at the same rate as the objective decrease. The second schedule in Theorem 4.10, with $f(x) = |x|$, $x^0 = 1$, and $t_k = 2^{-k-1}$, satisfies

$$D_k = \Delta_k = 2^{-k}, \quad m(x^k) = 1 \quad \text{for every finite } k.$$

Thus both the distance to the solution set and the objective gap converge geometrically to zero, while the minimum subgradient norm remains constant.

3.4 Symmetric Bregman distance to a minimizer

The exact progress identity (22), evaluated at a minimizer $x^* \in X^*$, contains the term

$$f(u) - f_* + D_f^g(x^*, u).$$

We interpret this term as the symmetric Bregman distance to a minimizer and bound it using the radial subgradient estimates developed above. In the dual formulation of Theorem 1.1, the radial subgradient g is also a dual solution associated with the exact ball subproblem. Since $0 \in \partial f(x^*)$, this is the symmetric Bregman distance associated with the subgradients g at u and 0 at x^* (Burger, 2015, Definition 2.2). It also admits the following representation in terms of the convex conjugate.

Proposition 3.8 (Symmetric Bregman distance and its dual representation). *Under Assumption 1 (Convex objective) and Assumption 2 (Attainment), fix $x^* \in X^*$, let $u \in \text{dom } \partial f$, and choose $g \in \partial f(u)$. The two Bregman distances for the conjugate satisfy*

$$D_{f^*}^u(0, g) := f^*(0) - f^*(g) + \langle u, g \rangle = f(u) - f_*, \quad (52)$$

$$D_{f^*}^{x^*}(g, 0) := f^*(g) - f^*(0) - \langle x^*, g \rangle = D_f^g(x^*, u) \geq 0. \quad (53)$$

Define their sum, the symmetric Bregman distance to x^* , by

$$\mathcal{G}_{x^*}(u, g) := D_{f^*}^u(0, g) + D_{f^*}^{x^*}(g, 0) = \langle g, u - x^* \rangle. \quad (54)$$

Equivalently, $\mathcal{G}_{x^*}(u, g) = D_f^g(x^*, u) + D_f^0(u, x^*)$. Then

$$0 \leq f(u) - f_* \leq \mathcal{G}_{x^*}(u, g) \leq \|g\| \|u - x^*\|. \quad (55)$$

For the exact stopped iteration of Assumption 3 (Exact BPM iteration), set $x^* = \text{Proj}_{X^*}(x^0)$ and $\mathcal{G}_K = \mathcal{G}_{x^*}(x^K, g^{K-1})$, where $g^{K-1} \in \partial f(x^K)$ is the radial subgradient associated with the step from x^{K-1} to x^K . At and after termination, we use the convention $g^{K-1} = 0$. For $K \geq 1$ and arbitrary positive radii,

$$\mathcal{G}_K \leq D_0 \min_{0 \leq j < K} \frac{\Delta_j - \Delta_K}{\sum_{i=j}^{K-1} t_i}. \quad (56)$$

For a constant radius $0 < t < D_0$ and the first K iterations, all assumed nonterminal,

$$\mathcal{G}_K \leq \frac{D_0 \Delta_0}{t} \min_{0 \leq j < K} \frac{E_j}{K - j}, \quad (57)$$

where E_j is defined in (50). At and after termination, $\mathcal{G}_K = 0$. Since the two Bregman distances for the conjugate are nonnegative, each is also bounded by the right-hand sides of (56) and (57).

The proof is given in [Section C.9](#). The quantity $\mathcal{G}_{x^*}(u, g)$ is exactly $f(u) - f_* + D_f^g(x^*, u)$ from the exact progress identity (22). For the iterate x^K , (55) gives $\mathcal{G}_K \leq D_0 \|g^{K-1}\|$. The proof of [Theorem 3.7](#) bounds $\|g^{K-1}\|$ by the same quantity for iterations $j, \dots, K-1$ appearing in (48), which gives (56). The fixed-radius bound (57) follows in the same way from (51).

Interpretation of the symmetric Bregman distance. The quantity $\mathcal{G}_{x^*}(u, g)$ depends on both a reference minimizer x^* and a subgradient g , so it is primarily an analytical measure rather than a stopping criterion based only on the returned point. It is also different from the Fenchel–Young gap $f(u) + f^*(g) - \langle u, g \rangle$, which vanishes whenever $g \in \partial f(u)$ and therefore does not measure progress toward optimality here.

A vanishing $\mathcal{G}_{x^*}(u, g)$ also does not imply that the subgradient g converges to zero. In the shrinking-radius example of [Theorem 4.10](#), with $f(x) = |x|$ and $x^k = 2^{-k}$, the radial subgradients satisfy $g^{k-1} = 1$, while $\mathcal{G}_0(x^k, 1) = 2^{-k} \rightarrow 0$. Thus convergence of the symmetric Bregman distance should be distinguished from convergence to stationarity.

4 Trajectory convergence and finite termination under arbitrary radii

The behavior of an exact BPM trajectory depends strongly on whether the total requested radius is finite or infinite. The following theorem gives the complete alternative for positive radii without assuming that a minimizer exists. The rest of the section establishes the geometric argument behind finite termination and then examines what can happen when the radii are summable.

Theorem 4.1 (General convergence under arbitrary radii). *Let $f : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfy [Assumption 1](#) (Convex objective), with $d \geq 1$, and use the exact stopped iteration of [Assumption 3](#) (Exact BPM iteration) from $x^0 \in \text{dom } f$. For any positive radius sequence (t_k) , exactly one of the following cases applies.*

(i) *If $\sum_k t_k < \infty$, then the trajectory has finite total length and*

$$x^k \rightarrow \bar{x} \in \text{dom } f, \quad f(x^k) \downarrow f(\bar{x}).$$

The limit $\bar{x}$ need not be a minimizer.

(ii) *If $\sum_k t_k = \infty$ and a minimizer exists, then the iteration attains a minimizer after finitely many iterations.*

(iii) *If $\sum_k t_k = \infty$ and no minimizer exists, then*

$$f(x^k) \downarrow \inf f, \quad \|x^k\| \rightarrow \infty,$$

including when $\inf f = -\infty$.

In particular,

$$\sum_{k=0}^{\infty} t_k = \infty \implies f(x^k) \downarrow \inf f. \quad (58)$$

Moreover, for a prescribed positive radius sequence in any fixed finite dimension $d \geq 1$, the condition $\sum_k t_k = \infty$ is necessary for $f(x^k) \downarrow \inf f$ to hold for every proper closed convex objective and every starting point in its domain.

The proof is given in [Section C.10](#). The theorem gives a sharp distinction between summable

and nonsummable radii. Nonsummable radii guarantee convergence of the objective values to $\inf f$. When a minimizer exists, they further force finite termination. A summable radius sequence cannot guarantee value convergence uniformly over all convex objectives and starting points, although it may still lead to asymptotic or finite convergence to a minimizer on a particular problem.

4.1 Self-contraction and finite termination

We now turn to the finite termination statement in [Theorem 4.1](#) and sharpen it quantitatively in fixed dimension. For any nonterminal step j and any later iterate x^m , $m > j$, the monotonicity of the objective implies that x^m belongs to the sublevel set used in the projection formulation of the j -th step. Applying the corresponding projection inequality gives (59), and hence the trajectory is self-contracted. In finite dimension, bounded self-contracted sequences have finite total length. Since every nonterminal BPM step satisfies $\|x^{k+1} - x^k\| = t_k$, this length bound directly controls the total radius spent before termination.

Self-contraction and finite length are classical for proximal sequences. See [Daniilidis et al. \(2015, Proposition 4.16 and Theorem 4.17\)](#) and [Böhm and Daniilidis \(2022, Corollary 3.8 and Section 3.4\)](#). Each nonterminal brox step is also a proximal step with a positive parameter, so those results give an alternative route to the same conclusion. We give a direct argument in the radius notation used here.

Definition 4.2 (Self-contracted sequence). A finite or infinite sequence (z^k) is self-contracted if $\|z^j - z^m\| \leq \|z^i - z^m\|$ whenever $i \leq j \leq m$.

Lemma 4.3 (Distance decrease toward later iterates). *Under [Assumption 1](#) (Convex objective) and [Assumption 3](#) (Exact BPM iteration), every finite prefix consisting of nonterminal outputs is self-contracted. More precisely, for $j < m$ in that prefix,*

$$\|x^{j+1} - x^m\|^2 \leq \|x^j - x^m\|^2 - t_j^2. \quad (59)$$

The proof is given in [Section C.11](#).

Lemma 4.4 (Finite length of bounded self-contracted sequences). *There is a finite constant C_d , depending only on dimension, such that every bounded self-contracted sequence in $\mathbb{R}^d$ has total length at most C_d times its diameter. One may take*

$$C_1 = 2, \quad C_d = 4\sqrt{d}(1 + 4\sqrt{d})^d \quad (d \geq 2).$$

This is an instance of the Euclidean self-contraction theory of [Daniilidis et al. \(2015\)](#). For completeness, [Section C.12](#) gives a discrete proof using a finite spherical net. Related finite-net arguments appear in [Daniilidis et al. \(2019, Theorem 4.5 and Remark 4.6, accepted manuscript\)](#). The explicit constant is sufficient here and is not claimed to be sharp.

Together, these results bound the total radius that can be spent along any sequence of nonterminal steps.

Theorem 4.5 (Bound on the sum of radii before termination). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), every prefix of K nonterminal iterations satisfies*

$$\sum_{k=0}^{K-1} t_k \leq 2C_d D_0. \quad (60)$$

For a constant radius and $D_0 > 0$,

$$\left\lceil \frac{D_0}{t} \right\rceil \leq N_t \leq \min \left\{ \left\lceil \frac{D_0^2}{t^2} \right\rceil, 1 + \left\lfloor \frac{2C_d D_0}{t} \right\rfloor \right\}. \quad (61)$$

Consequently, for fixed dimension, the worst-case complexity is $\Theta_d(D_0/t)$ as $D_0/t \rightarrow \infty$.

The proof is left to [Section C.13](#). Equation (60) makes the finite-termination statement in [Theorem 4.1\(ii\)](#) quantitative: if the run remained nonterminal while $\sum_k t_k = \infty$, its partial radius sums would eventually exceed $2C_d D_0$. For a constant radius, the same bound gives a linear dependence on D_0/t in fixed dimension, improving the dimension-free quadratic upper bound.

The next result adds the converse and characterizes exactly which prescribed radius sequences force finite termination on every problem with a minimizer.

Theorem 4.6 (Universal criterion for finite termination). *Fix a positive radius sequence (t_k) and a finite dimension $d \geq 1$. The following are equivalent:*

- (i) $\sum_{k=0}^{\infty} t_k = \infty$;
- (ii) for every f satisfying [Assumption 1](#) (Convex objective) and [Assumption 2](#) (Attainment) and every $x^0 \in \text{dom } f$, the stopped exact iteration attains a minimizer after finitely many iterations.

The proof is given in [Section C.14](#). Sufficiency follows from (60): on any particular problem, termination must occur before the partial radius sum exceeds $2C_d D_0$. The converse shows that no summable positive schedule has this guarantee uniformly over all convex objectives and starting points. For example, the harmonic schedule $t_k = 1/(k+1)$ forces finite termination even though $\sum_k t_k^2 < \infty$.

The bound (60) controls only the nonterminal part of the trajectory. Under a constant radius and nearest-minimizer selection at termination, the terminal displacement can also be bounded, yielding a bound on the full trajectory length.

Proposition 4.7 (Trajectory length including the terminal step). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), suppose that the nearest-minimizer selection at termination of [Theorem 2.2](#) is used, the radius is constant, $t_k \equiv t > 0$, and $D_0 > 0$. Let*

$$J = \left\lceil \frac{D_0^2}{t^2} \right\rceil - 1, \quad \Psi_t(D_0) = Jt + \sqrt{D_0^2 - Jt^2}.$$

Then the full trajectory length S_t , including the terminal displacement, satisfies

$$S_t \leq \Psi_t(D_0) \leq \frac{D_0^2}{t} + \frac{t}{4}.$$

If D_0^2/t^2 is an integer, then $Jt + \sqrt{D_0^2 - Jt^2} = D_0^2/t$. If $t \geq D_0$, the first step is terminal and nearest-minimizer selection at termination gives $S_t = D_0$.

The proof is given in [Section C.15](#). The nearest-minimizer selection at termination is needed to control the final displacement.

4.2 Summable radii and trajectory convergence

Theorem 4.6 identifies nonsummability as the exact condition for finite termination uniformly over all problems with a minimizer. When the radii are summable, finite termination is no longer guaranteed. Nevertheless, **Theorem 4.1** shows that the trajectory still has finite length and converges. We now quantify this convergence by relating the remaining radius sum to the distance and objective error relative to the limiting point. The proof of the following theorem is given in [Section C.16](#).

Theorem 4.8 (Finite length and tail bounds). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), every stopped exact BPM trajectory satisfies*

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| < \infty, \quad x^k \longrightarrow \bar{x} \in \text{dom } f, \quad f(x^k) \downarrow f(\bar{x}). \quad (62)$$

If no step is terminal, then necessarily $\sum_{k=0}^{\infty} t_k < \infty$. Moreover, for any radial subgradient $g^0 \in \partial f(x^1)$ from the first step,

$$\|x^k - \bar{x}\| \leq \sum_{j=k}^{\infty} t_j \quad (k \geq 0), \quad 0 \leq f(x^k) - f(\bar{x}) \leq \|g^0\| \sum_{j=k}^{\infty} t_j \quad (k \geq 1). \quad (63)$$

If $\sum_{k=0}^{\infty} t_k < \infty$ is assumed, then [Assumption 2](#) (Attainment) is unnecessary: (62) still holds, and on an infinite nonterminal run so does (63).

The preceding bounds use the remaining radius sum. The next corollary instead controls the actual length of the remaining trajectory in terms of the current distance to the solution set.

Corollary 4.9 (Remaining trajectory length). *Under [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration), use the nearest-minimizer selection at termination of [Theorem 2.2](#). Let $\bar{x} = \lim_k x^k$ and $L_k = \sum_{i=k}^{\infty} \|x^{i+1} - x^i\|$. Then, for every $k \geq 0$,*

$$\|x^k - \bar{x}\| \leq L_k \leq 2C_d D_k. \quad (64)$$

In particular, if $D_k \leq A\rho^k$ for some $A > 0$ and $0 < \rho < 1$, then $\bar{x} \in X^$, and both L_k and $\|x^k - \bar{x}\|$ decay at least geometrically, with upper bound $2C_d A\rho^k$.*

The proof is given in [Section C.17](#). Although summable radii give point and objective value convergence, the limit still depends on the radius schedule. Even for the same objective, the trajectory may terminate at a minimizer, converge to one asymptotically, or converge to a nonoptimal point.

Example 4.10 (Three behaviors under summable radii). Let $f(x) = |x|$ and $x^0 = D > 0$. Consider the following three summable radius schedules.

- (i) If $t_k = D 2^{-k}$, then the first radius equals D , so the first step reaches the minimizer 0.
- (ii) If $t_k = D 2^{-k-1}$, then $x^k = D 2^{-k}$ for every $k \geq 0$. Hence no step is terminal, but $x^k \rightarrow 0$.
- (iii) If $t_k = D 2^{-k-2}$, then $x^k = D/2 + D 2^{-k-1}$ for every $k \geq 0$. Hence the trajectory is again infinite, but now $x^k \rightarrow D/2$, which is not optimal.

Thus, even for the same objective and starting point, summable radii can lead to finite

termination, asymptotic convergence to a minimizer, or convergence to a nonoptimal point. These formulas follow from $x^{k+1} = \max\{x^k - t_k, 0\}$. In the two infinite trajectories, both bounds in (63) hold with equality, with $\|g^0\| = 1$.

4.3 Dimension dependence of the linear bound

The linear step bound in (61) is a fixed dimensional result because its constant depends on d . This dependence is unavoidable. The linear bound cannot hold with a constant independent of dimension. Indeed, the construction below shows that, as the dimension increases, the number of iterations can scale like D_0^2/t^2 rather than D_0/t .

Theorem 4.11 (Polyhedral lower bound for the step count). *For every integer $n \geq 2$, every $t > 0$, and every $\epsilon \in (0, 1/2)$, there is a finite convex piecewise-affine function on $\mathbb{R}^n$ with a unique minimizer such that BPM performs exactly n radius- t iterations from $x^0 = 0$ and*

$$D_0^2 = t^2[n + 2(n-1)\epsilon]. \quad (65)$$

If $\epsilon = o(1/n)$, then $D_0^2/t^2 = n + o(1)$ and

$$\frac{N_t}{\lceil D_0^2/t^2 \rceil} \longrightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Consequently, no constant independent of dimension can bound N_t by CD_0/t for all convex objectives.

The proof is given in [Section C.18](#). The lower bound is asymptotic over increasing dimensions, while for each fixed dimension the linear bound in [Theorem 4.5](#) still applies. In this family, the preterminal trajectory has length $(n-1)t$. Consequently, any constant L_d satisfying $\sum_{k < N_t-1} t \leq L_d D_0$ uniformly over all BPM trajectories in dimension d must satisfy $L_d = \Omega(\sqrt{d})$.

5 Reformulations and preservation of BPM geometry

Equivalent formulations of the same optimization problem need not induce the same BPM trajectory, because BPM depends on the geometry of Euclidean balls. We first identify transformations that preserve exact ball calls. We then study the epigraph reformulation, which preserves the underlying optimization problem but changes distances, and hence the radius geometry. The epigraph reformulation itself is classical ([Boyd and Vandenberghe, 2004](#), Section 4.1.3).

Proposition 5.1 (Transformations preserving exact ball calls). *Under [Assumption 1](#) (Convex objective), the following transformations preserve exact BPM iterations.*

- (i) *Let $x = a + sQy$, where $s > 0$ and Q is orthogonal. Exact BPM iterations in y -space with radius t correspond to exact BPM iterations in x -space with radius st . In particular, translations and orthogonal changes of coordinates preserve the radii.*
- (ii) *Let $\tilde{f} = \chi \circ f$ on $\text{dom } f$, with $\tilde{f} = +\infty$ elsewhere, where χ is strictly increasing on the finite range of f and $\tilde{f}$ is proper, closed and convex. Then f and $\tilde{f}$ have the same exact ball-minimizer sets at every center and radius, and the same solution set. Hence, under the same selections, they generate the same fixed-radius trajectories and termination counts.*

The proof is given in [Section C.19](#). Positive affine changes of the objective are a special case

of part (ii). Although the exact ball calls are preserved, function gaps, subgradient norms, and hence radius rules based on them need not be. A general nonsingular linear change of variables falls outside this invariance: it maps Euclidean balls to ellipsoids and therefore does not preserve the same BPM radius geometry.

5.1 Epigraphical lifting and induced radii

For f satisfying [Assumption 1](#) (Convex objective), define on $\mathbb{R}^{d+1}$

$$F(x, s) := s + \delta_{\text{epi } f}(x, s), \quad \text{epi } f = \{(x, s) : f(x) \leq s\}, \quad (66)$$

where δ_C denotes the indicator function of a set C . Then F is proper, closed, and convex, with $\inf F = \inf f$ and, under attainment, $\arg \min F = X^\star \times \{f_\star\}$. We apply BPM to F using Euclidean balls in $\mathbb{R}^{d+1}$, initialized at the graph point $z^0 = (x^0, f(x^0))$ with $x^0 \in \text{dom } f$. This initialization therefore requires the value $f(x^0)$.

Theorem 5.2 (Epigraphical step correspondence). *Every BPM iterate on F from the above initialization remains on the graph: let $z^k = (x^k, f(x^k))$. At a nonterminal radius- t_k step, define*

$$\delta_k = f(x^k) - f(x^{k+1}), \quad r_k = \|x^{k+1} - x^k\|.$$

Then $\delta_k > 0$, $r_k > 0$, and

$$t_k^2 = r_k^2 + \delta_k^2, \quad x^{k+1} \in \text{brox}_f^{r_k}(x^k). \quad (67)$$

Conversely, any nonterminal radius r BPM step $x \mapsto u$ for f lifts to an exact step for F from $(x, f(x))$ with radius $\sqrt{r^2 + (f(x) - f(u))^2}$. Moreover, if $G^k = c_k(z^k - z^{k+1}) \in \partial F(z^{k+1})$, $c_k > 0$, is a radial subgradient for the lifted step, then

$$\lambda_k = 1 - c_k \delta_k \in (0, 1), \quad g^k = \frac{c_k}{\lambda_k}(x^k - x^{k+1}) \in \partial f(x^{k+1}), \quad G^k = (\lambda_k g^k, 1 - \lambda_k). \quad (68)$$

At a terminal lifted step, $x^{k+1} \in X^\star$, and the projected output is an exact BPM output for f with radius $\sqrt{t_k^2 - \Delta_k^2} \geq D_k$. Conversely, a terminal radius r step for f lifts to a terminal step for F with radius $\sqrt{r^2 + \Delta_k^2}$. These terminal correspondences hold for any optimal output selected by the terminal step.

The proof is given in [Section C.20](#) and allows extended-valued f . The theorem shows that epigraphical lifting preserves projected BPM trajectories after an induced change of radii, but does not preserve a prescribed radius schedule. This differs from the cylindrical epigraph formulation in (15). There, only the horizontal displacement is constrained, $\|v - x\| \leq t$, whereas a Euclidean ball in the lifted space imposes $\|v - x\|^2 + (s - f(x))^2 \leq t^2$. Thus part of the lifted radius is spent in the vertical direction, and the projected nonterminal radius is generally strictly smaller than t . Initialization on the graph is essential. For example, if $f \equiv 0$ and $z^0 = (0, 2)$, the lifted method may take purely vertical nonterminal steps even though the projected point is already optimal.

For the same center and radius t , the projected lifted output is feasible for the BPM subproblem on f . Hence BPM attains an objective value no larger than the projected lifted step. This comparison applies only to the current step. The two methods generally produce different next centers, so it does not imply an ordering of their subsequent trajectories.

5.2 Convergence guarantees under epigraphical lifting

Under attainment, define the lifted distance to the solution set $R_k := \text{dist}(z^k, \arg \min F)$. Since $z^k = (x^k, f(x^k))$ and $\arg \min F = X^\star \times \{f_\star\}$,

$$F(z^k) - \inf F \stackrel{(66)}{=} \Delta_k, \quad R_k^2 = D_k^2 + \Delta_k^2. \quad (69)$$

Hence the general BPM bounds apply to F with dimension $d + 1$, initial distance R_0 , and the radii used in the lifted space. In particular, for a constant lifted radius t , every sequence of nonterminal steps satisfies

$$\Delta_K \stackrel{(42)}{\leq} \Delta_0 \exp\left(-\frac{2K^2 t^2}{R_0^2}\right), \quad N_t^{\text{epi}} \stackrel{(26)}{\leq} \left\lceil \frac{R_0^2}{t^2} \right\rceil. \quad (70)$$

Here N_t^{epi} denotes the termination count for lifted BPM. In particular, $R_0 = \sqrt{D_0^2 + \Delta_0^2}$, so these bounds depend on the lifted distance rather than on D_0 alone.

Proposition 5.3 (Stationarity and Bregman-distance relations). *For every $x \in \text{dom } f$,*

$$\text{dist}(0, \partial F(x, f(x))) = \frac{m(x)}{\sqrt{1 + m(x)^2}}, \quad (71)$$

where the right-hand side is interpreted as one if $\partial f(x) = \emptyset$. Consequently, a lifted stationarity bound gives

$$\text{dist}(0, \partial F(x, f(x))) \leq \varepsilon < 1 \implies m(x) \leq \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}}.$$

By contrast, the bound $\text{dist}(0, \partial F(x, f(x))) \leq 1$ alone gives no finite upper bound on $m(x)$. Under attainment, fix $x^\star \in X^\star$. At a nonterminal paired output $(u, f(u))$, let $G = (\lambda g, 1 - \lambda)$ be the paired subgradient from (68). Then

$$\langle G, (u, f(u)) - (x^\star, f_\star) \rangle = \lambda \mathcal{G}_{x^\star}(u, g) + (1 - \lambda)(f(u) - f_\star). \quad (72)$$

The primal gap is unchanged,

$$F(u, f(u)) - \inf F = f(u) - f_\star,$$

and the complementary Bregman distance satisfies

$$D_F^G((x^\star, f_\star), (u, f(u))) = \lambda D_f^g(x^\star, u).$$

The proof is given in [Section C.21](#). Applying the earlier BPM bounds to F gives bounds in terms of the lifted distance R_k , related to the original quantities by (69). Stationarity and primal–dual quantities are instead related by (71) and (72). The epigraphical lift also does not preserve additional regularity, F is extended-valued and affine along feasible vertical rays, so smoothness or strong convexity of f does not carry over to F .

Theorem 5.4 (Summability and length bounds). *For an infinite nonterminal lifted iteration initialized at $z^0 = (x^0, f(x^0))$, the lifted radii t_k and the corresponding projected radii r_k from (67) satisfy*

$$\sum_{k=0}^{\infty} t_k < \infty \iff \sum_{k=0}^{\infty} r_k < \infty.$$

Consequently, for any prescribed positive lifted radius sequence (t_k) , the convergence alterna-

tives of [Theorem 4.1](#) also hold for the projected iterates: if $\sum_k t_k = \infty$, finite termination occurs when f attains its minimum, while otherwise $f(x^k) \downarrow \inf f$ and $\|x^k\| \rightarrow \infty$. If $\sum_k t_k < \infty$, then $x^k \rightarrow \bar{x} \in \text{dom } f$, and $f(x^k) \downarrow f(\bar{x})$, where $\bar{x}$ need not be optimal. Under attainment and a nonoptimal start, every sequence of nonterminal steps satisfies

$$\sum_{k=0}^{K-1} t_k \leq 2C_d D_0 + \Delta_0 - \Delta_K < 2C_d D_0 + \Delta_0. \quad (73)$$

Hence, for a constant lifted radius t ,

$$N_t^{\text{epi}} \leq \min \left\{ \left\lceil \frac{D_0^2 + \Delta_0^2}{t^2} \right\rceil, \left\lceil \frac{2C_d D_0 + \Delta_0}{t} \right\rceil \right\}. \quad (74)$$

If the nearest-minimizer selection at termination of [Theorem 2.2](#) is used, the total lifted trajectory length is at most $2C_d D_0 + \Delta_0$.

The proof is given in [Section C.22](#). The bound (73) combines the projected trajectory-length bound (60) with the total decrease in objective value, which is why the term $2C_d D_0$ depends on the original dimension d rather than the lifted dimension $d + 1$.

Example 5.5 (Fixed-radius counts under objective scaling). Let $f(x) = a|x|$, $a > 0$, and $x^0 = D_0 > 0$. Ordinary BPM has $N_t = \lceil D_0/t \rceil$. The lifted trajectory initialized at $(D_0, f(D_0))$ follows the line segment $s = ax$ from (D_0, aD_0) to $(0, 0)$, so

$$N_t^{\text{epi}} = \left\lceil \frac{\sqrt{1+a^2} D_0}{t} \right\rceil.$$

Hence N_t^{epi}/N_t can grow arbitrarily large as $a \rightarrow \infty$.

Positive objective scaling leaves fixed-radius BPM unchanged by [Theorem 5.1](#), but changes the geometry of the epigraph lift. Thus it can change the lifted fixed-radius termination count by an arbitrarily large factor. The verification is given in [Section C.23](#).

6 Adaptive radius selection

The choice of radius can substantially affect the behavior of BPM. A larger radius can produce a better first objective value yet require more iterations to reach a minimizer: [Theorem E.1](#) gives $0 < s < t$ with $N_s = 2$ but $N_t = 3$. For geometric schedules $t_k = t\rho^k$ with $0 < t < D_0$, [Theorem D.1](#) identifies the smallest successful factor ρ_* . At this factor the trajectory converges to a minimizer without reaching one in finitely many iterations, whereas any smaller factor leaves a nonoptimal limit. These results show why radius selection must account for the trajectory that the radii generate.

We therefore study adaptive rules that choose the radius using information available at the current iterate, specifically subgradient or objective gap information. The guarantees below require no smoothness, strong convexity, or growth condition. Throughout, [Assumption 1](#) (Convex objective), [Assumption 2](#) (Attainment), and [Assumption 3](#) (Exact BPM iteration) hold.

6.1 Subgradient-based radius selection

If $t_k \geq \theta D_k$ for some fixed $0 < \theta < 1$, [Theorems 2.5](#) and [3.1](#) give

$$D_{k+1} \stackrel{(24)}{\leq} \sqrt{1-\theta^2} D_k, \quad \Delta_{k+1} \stackrel{(29)}{\leq} (1-\theta) \Delta_k.$$

Thus a radius comparable to the current distance to the solution set gives geometric contraction in both distance and objective gap. Since D_k is generally unavailable, we next consider an adaptive rule based on an available subgradient. Its convergence rate follows by relating the resulting BPM steps to the proximal-point method (Rockafellar, 1976; Taylor et al., 2017).

Theorem 6.1 (Subgradient-based radius selection). *Suppose $x^0 \in \text{dom } \partial f$. At iteration k , choose an available $h^k \in \partial f(x^k)$. If $h^k = 0$, stop; otherwise set $t_k = \tau \|h^k\|$ for some $\tau > 0$. For the first $K \geq 1$ iterations, all assumed nonterminal, let g^k be a radial subgradient and let $\lambda_{k+1} = t_k / \|g^k\|$ be the corresponding proximal parameter from (20). Then, for $k < K$,*

$$\lambda_{k+1} \geq \tau, \quad \Delta_K \leq \frac{D_0^2}{4 \sum_{k=0}^{K-1} \lambda_{k+1}} \leq \frac{D_0^2}{4\tau K}.$$

Under the stopped convention of Assumption 3 (Exact BPM iteration), the final bound remains valid after termination.

The proof is given in Section C.24. The rule is implementable whenever a subgradient is available at the current iterate. In particular, the radial subgradient from Theorem 2.1 can be reused at the next iteration if the oracle returns it. A zero subgradient provides a stopping certificate, whereas a nonzero subgradient may still be returned at a nonsmooth minimizer. The constant $1/4$ is inherited from the proximal point estimate of Taylor et al. (2017, Theorem 4.1) and is also sharp for this radius rule: the one-dimensional absolute-value example in Section C.24 attains equality in both bounds of Theorem 6.1.

6.2 Objective gap based radius selection

If f_* is known, the current objective gap can be used directly to choose the radius. For fixed $\tau > 0$ and $\alpha > 0$, set $t_k = \tau \Delta_k^\alpha$ whenever $\Delta_k > 0$.

Theorem 6.2 (Objective gap based radius selection). *Suppose $\Delta_0 > 0$ and, at each nonoptimal iterate, choose $t_k = \tau \Delta_k^\alpha$. Then the resulting BPM iterates satisfy*

$$\Delta_K \leq \left(\Delta_0^{-\alpha} + \frac{\alpha\tau K}{D_0} \right)^{-1/\alpha}. \quad (75)$$

Consequently, for $0 < \varepsilon < \Delta_0$, it suffices to take

$$K \geq \left\lceil \frac{D_0}{\alpha\tau} (\varepsilon^{-\alpha} - \Delta_0^{-\alpha}) \right\rceil$$

to obtain $\Delta_K \leq \varepsilon$.

The proof is given in Section C.25. In particular, $\alpha = 1$ gives an $O(K^{-1})$ rate, while $\alpha = 1/2$ gives $O(K^{-2})$. These rates arise from different radius schedules, rather than from accelerating a fixed ball-oracle iteration. For a fixed initial radius t_0 , the rule can be written as $t_k = t_0(\Delta_k/\Delta_0)^\alpha$, so a smaller α decreases the radius more slowly as the objective gap closes.

Proposition 6.3 (Radius selection with a lower bound). *Suppose the assumptions of Theorem 6.2 hold, but only a finite lower bound $\ell \leq f_*$ is known. If the radius is chosen as*

$t_k = \tau(f(x^k) - \ell)^\alpha$, then the objective-gap bound (75) remains valid. Moreover, if $\ell < f_\star$ and $D_0 > 0$, a terminal step occurs within

$$\left\lceil \frac{D_0^2}{\tau^2(f_\star - \ell)^{2\alpha}} \right\rceil$$

iterations.

The proof is given in [Section C.26](#). A valid lower bound keeps the radius positive at every nonoptimal point, an upper estimate need not. For $f(x) = x^2/2$ and $u > 0$, the rule $t = \tau \max\{f(x) - u, 0\}^\alpha$ gives $t = 0$ whenever $f(x) \leq u$, even if x is not optimal.

7 Relaxed BPM iterations

Besides choosing the radius, one can also modify how far the next center moves toward the projected BPM output. Let T_t denote the projected selection from [Theorem 2.2](#). We consider the relaxed update $R_{t,\lambda}(x) := x + \lambda(T_t(x) - x)$, where $t > 0$ and $0 < \lambda < 2$. The convergence analysis relies on the cutter property of T_t ([Cegielski and Censor, 2012](#)). To continue the iteration, the relaxed center must remain in $\text{dom } f$. This is automatic for $0 < \lambda \leq 1$, and for all $0 < \lambda < 2$ when f is finite-valued. For extended-valued objectives and $\lambda > 1$, this must be checked separately.

Theorem 7.1 (Convergence of relaxed BPM). *Assume [Assumption 1](#) (Convex objective) and [Assumption 2](#) (Attainment). Fix $t > 0$ and $0 < \lambda < 2$, and let $x^{k+1} = R_{t,\lambda}(x^k)$ from some $x^0 \in \text{dom } f$, assuming that every iterate remains in $\text{dom } f$. Then x^k converges to a minimizer. Moreover, for every $x \in \text{dom } f$ and $z \in X^\star$,*

$$\|R_{t,\lambda}(x) - z\|^2 \leq \|x - z\|^2 - \lambda(2 - \lambda) \min\{t, D(x)\}^2. \quad (76)$$

Once $D(x^k) \leq t$, the distance to the solution set contracts as $D(x^{k+1}) \leq |1 - \lambda|D(x^k)$.

The proof is given in [Section C.27](#). When $\lambda \neq 1$, an optimal ball output does not necessarily make the next relaxed iterate optimal, so the relaxed sequence may continue after the first such output. For $\lambda > 1$, the objective values need not be monotone.

8 Discussion

Taken together, the results in this paper reveal two distinct roles of the radius sequence. In any fixed finite dimension, nonsummable radii force finite termination whenever a minimizer exists, whereas summable radii may lead to finite termination, asymptotic convergence to a minimizer, or convergence to a nonoptimal point. For a constant radius, the worst-case step count is linear in D_0/t when the dimension is fixed, but this dimension dependence cannot be removed: as the dimension grows, the quadratic dependence on D_0/t can be asymptotically attained.

Two sharpness questions remain open. The first concerns the dependence on dimension in the trajectory-length bound. [Theorem 4.11](#) shows that the corresponding constant must grow at least as $\sqrt{d}$, while the upper bound in [Theorem 4.4](#) leaves a gap in its dependence on d . The second concerns the optimal worst-case objective-gap bound for BPM. The Jensen estimate (42) is sharp for its scalar relaxation, but [Theorem 3.5](#) shows that retaining the gap–distance relation gives a strictly smaller factor for fixed S and K . Determining the optimal bound for convex BPM trajectories, and the asymptotic sharpness of its dependence on S and K , remains open.

The complexity bounds in this paper are stated in terms of exact ball calls. Converting them

into computational work would additionally require a model for solving the ball subproblems and obtaining the required subgradients.

Acknowledgments and declarations

Funding. This work was supported by funding from King Abdullah University of Science and Technology (KAUST): i) KAUST Baseline Research Scheme, ii) Center of Excellence for Generative AI (award no. 5940).

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Use of AI tools. The authors used OpenAI’s ChatGPT and Codex to assist with manuscript drafting and revision, mathematical exploration, and checks of proofs, calculations, and references. The authors reviewed and verified the resulting material and take full responsibility for the final manuscript.

References

- Axel Böhm and Aris Daniilidis. Ubiquitous algorithms in convex optimization generate self-contracted sequences. *Journal of Convex Analysis*, 29(1):119–128, 2022. URL <https://journalofconvexanalysis.com/articles/jca29006/>.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. URL https://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.
- L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967. URL [https://doi.org/10.1016/0041-5553\(67\)90040-7](https://doi.org/10.1016/0041-5553(67)90040-7).
- Martin Burger. Bregman distances in inverse problems and partial differential equations. *arXiv preprint arXiv:1505.05191*, 2015. URL <https://arxiv.org/abs/1505.05191>.
- Y. Carmon, A. Jambulapati, Y. Jin, and A. Sidford. Thinking inside the ball: Near-optimal minimization of the maximal loss. In *Proceedings of COLT*, volume 134 of *PMLR*, pages 866–882, 2021. URL <https://proceedings.mlr.press/v134/carmon21a.html>.
- Yair Carmon, Arun Jambulapati, Qijia Jiang, Yujia Jin, Yin Tat Lee, Aaron Sidford, and Kevin Tian. Acceleration with a ball optimization oracle. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL <https://arxiv.org/abs/2003.08078>.
- Andrzej Cegielski and Yair Censor. Extrapolation and local acceleration of an iterative process for common fixed point problems. *Journal of Mathematical Analysis and Applications*, 394(2): 809–818, 2012. doi: 10.1016/j.jmaa.2012.04.072. URL <https://doi.org/10.1016/j.jmaa.2012.04.072>.
- Aris Daniilidis, Guy David, Estibalitz Durand-Cartagena, and Antoine Lemenant. Rectifiability of self-contracted curves in the Euclidean space and applications. *Journal of Geometric Analysis*, 25(2):1211–1239, 2015. URL <https://arxiv.org/abs/1211.3227>.

- Aris Daniilidis, Robert Deville, and Estibalitz Durand-Cartagena. Metric and geometric relaxations of self-contracted curves. *Journal of Optimization Theory and Applications*, 182(1):81–109, 2019. doi: 10.1007/s10957-018-1408-0. URL <https://doi.org/10.1007/s10957-018-1408-0>. Theorem numbering follows the [accepted manuscript](#).
- Kaja Grunkowska and Peter Richtárik. Non-Euclidean broximal point method: A blueprint for geometry-aware optimization, 2025. URL <https://arxiv.org/abs/2510.00823>. Preprint, arXiv:2510.00823.
- Kaja Grunkowska, Hanmin Li, Aadi Rane, and Peter Richtárik. The ball-proximal (= broximal) point method: a new algorithm, convergence theory, and applications, 2025. URL <https://arxiv.org/abs/2502.02002>. Preprint, arXiv:2502.02002.
- Kaja Grunkowska, Hanmin Li, Xun Qian, and Peter Richtárik. Broximal alignment for global non-convex optimization, 2026. URL <https://arxiv.org/abs/2604.13483>. Preprint, arXiv:2604.13483.
- Hanmin Li, Kaja Grunkowska, and Peter Richtárik. Stabilized proximal point method via trust region control, 2026. URL <https://arxiv.org/abs/2604.02943>. Preprint, arXiv:2604.02943.
- Peter Richtárik, Kaja Grunkowska, and Hanmin Li. Local LMO: Constrained gradient optimization via a local linear minimization oracle, 2026. URL <https://arxiv.org/abs/2605.08850>. Preprint, arXiv:2605.08850.
- R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976. URL <https://doi.org/10.1137/0314056>.
- Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3):1283–1313, 2017. URL <https://doi.org/10.1137/16M108104X>. Also available at <https://arxiv.org/abs/1512.07516>.

Appendix

Contents

1	Introduction	1
1.1	One optimality system, seven formulations	2
1.2	Sharpening and extending the BPM theory	7
1.3	Related work	7
2	One-step geometry and progress	9
2.1	Geometry of an exact step	9
2.2	Progress from an exact step	10
2.3	Subgradients associated with exact BPM steps	11
3	Sharper objective-gap bounds for BPM	12
3.1	Objective-gap contraction and squared-distance decrease	12
3.2	Refined multi-step objective-gap bounds	14
3.3	From objective decrease to stationarity	15
3.4	Symmetric Bregman distance to a minimizer	17
4	Trajectory convergence and finite termination under arbitrary radii	18
4.1	Self-contraction and finite termination	19
4.2	Summable radii and trajectory convergence	21
4.3	Dimension dependence of the linear bound	22
5	Reformulations and preservation of BPM geometry	22
5.1	Epigraphical lifting and induced radii	23
5.2	Convergence guarantees under epigraphical lifting	24
6	Adaptive radius selection	25
6.1	Subgradient-based radius selection	25
6.2	Objective gap based radius selection	26
7	Relaxed BPM iterations	27
8	Discussion	27
A	Notation	32
B	Auxiliary results	34
B.1	Auxiliary geometry for the Euclidean length bound	34
C	Proofs	34
C.1	Proof of Theorem 1.1	34
C.2	Proof of Theorem 2.1	36
C.3	Proof of Theorem 2.6	37
C.4	Proof of Theorem 3.1	37
C.5	Proof of Theorem 3.2	38
C.6	Proof of Theorem 3.3	38
C.7	Proof of Theorem 3.4	38
C.8	Proof of Theorem 3.7	39
C.9	Proof of Theorem 3.8	39
C.10	Proof of Theorem 4.1	40

C.11 Proof of Theorem 4.3	41
C.12 Proof of Theorem 4.4	41
C.13 Proof of Theorem 4.5	41
C.14 Proof of Theorem 4.6	42
C.15 Proof of Theorem 4.7	42
C.16 Proof of Theorem 4.8	43
C.17 Proof of Theorem 4.9	43
C.18 Proof of Theorem 4.11	44
C.19 Proof of Theorem 5.1	45
C.20 Proof of Theorem 5.2	45
C.21 Proof of Theorem 5.3	46
C.22 Proof of Theorem 5.4	46
C.23 Verification of Theorem 5.5	46
C.24 Proof of Theorem 6.1	47
C.25 Proof of Theorem 6.2	47
C.26 Proof of Theorem 6.3	48
C.27 Proof of Theorem 7.1	48
C.28 Proof of Theorem 2.3	48
C.29 Proof of Theorem 2.5	48
C.30 Proof of Theorem 2.7	49
C.31 Proof of Theorem 3.6	49
D Geometric radius schedules and the minimum successful factor	49
D.1 Proof of Theorem D.1	51
E Nonmonotonicity of the fixed-radius termination count	52
F Key concepts	55

A Notation

The symbols below are grouped by their role in the analysis. Quantities involving the solution set or the optimal value assume attainment unless stated otherwise. The last column points to the defining result or discussion.

Table 2: Notation for exact Euclidean BPM.

Symbol	Meaning or definition	Location
1. Problem and Euclidean geometry		
f, d	Proper, closed, convex objective on $\mathbb{R}^d$; d is finite.	Assump. 1
$\text{dom } f, \partial f(x)$	Effective domain $\{x : f(x) < +\infty\}$ and convex subdifferential at x .	§1
X^*, f_*	Solution set $\arg \min f$ and attained minimum $\min f$.	Assump. 2
$\langle u, v \rangle, \ u\ $	Euclidean inner product and its induced norm.	§1
$\mathbb{B}(x, t)$	Closed ball $\{u : \ u - x\ \leq t\}$, with $t > 0$.	§1
$\mathbb{S}^{d-1}$	Unit sphere $\{u \in \mathbb{R}^d : \ u\ = 1\}$.	Lem. B.1
$\text{dist}(x, C), \text{Proj}_C(x)$	Distance to a set and Euclidean projection onto a nonempty closed convex set.	§2
2. Iterates, BPM steps, and subgradients		
x^k, t_k, t	Iterate, radius of the step from x^k , and a constant radius; K denotes a number of iterations.	Assump. 3
$\text{brox}_f^t(x)$	Set of exact minimizers of f on $\mathbb{B}(x, t)$. A nonterminal step has a unique output.	(BPM)
u, g^k, c_k	Returned point $u = x^{k+1}$ and radial subgradient $g^k = c_k(x^k - u) \in \partial f(u)$, with $c_k > 0$ on a nonterminal step.	(19)
$\text{prox}_{\lambda f}(x)$	Minimizer of $f(u) + \ u - x\ ^2 / (2\lambda)$; the corresponding proximal parameter is $\lambda = 1/c = t / \ g\ $.	(20)
$T_t(x)$	Unique nonterminal output, or the nearest minimizer $\text{Proj}_{X^*}(x)$ when the step is terminal.	Def. 2.2
h^k, τ, α	Available subgradient at x^k ; positive radius-rule parameters in $t_k = \tau \ h^k\ $ or $t_k = \tau \Delta_k^\alpha$.	§6
$R_{t,\lambda}(x)$	Relaxed update $x + \lambda(T_t(x) - x)$, with relaxation factor $0 < \lambda < 2$.	§7
3. Error measures and termination counts		
$D_k, D(x)$	Distances to the solution set $\text{dist}(x^k, X^*)$ and $\text{dist}(x, X^*)$; $p^k = \text{Proj}_{X^*}(x^k)$.	§2
Δ_k	Objective gap $f(x^k) - f_*$.	§3
$m(x), m_K$	Minimum subgradient norm $\text{dist}(0, \partial f(x))$ and $m(x^K)$; the distance to the empty set is $+\infty$.	Cor. 2.7
$D_f^g(z, u)$	Generalized Bregman distance $f(z) - f(u) - \langle g, z - u \rangle$, with $g \in \partial f(u)$.	(21)
$\mathcal{G}_{x^*}(u, g)$	Symmetric Bregman distance to a minimizer $\langle g, u - x^* \rangle$, for $x^* \in X^*$ and $g \in \partial f(u)$.	(54)
$\mathcal{G}_K$	$\mathcal{G}_{x^*}(x^K, g^{K-1})$, with $x^* = \text{Proj}_{X^*}(x^0)$; set to zero at and after termination.	§3.4
N_t, N_ε	Iterations to termination at constant radius t , including the terminal step; first index with $\Delta_K \leq \varepsilon$, respectively.	(26), (43)

Continued on the next page

Table 2. Notation for exact Euclidean BPM (continued)

Symbol	Meaning or definition	Location
4. Squared-distance decrease and trajectory length		
S, q_k, β_k	$S = D_0^2/t^2$, $q_k = \Delta_{k+1}/\Delta_k$, and $\beta_k = (D_k^2 - D_{k+1}^2)/t_k^2$, on nonterminal iterations.	§3
$r, r_\star$	In the final-distance refinement, $r = \Delta_K/\Delta_0$ and $r_\star$ is the unique root defining its implicit upper bound.	Rem. 3.5
E_j	Minimum of the geometric and Jensen bounds on the relative objective gap at integer horizon j .	(50)
C_d	Dimension-dependent constant bounding the length of a bounded self-contracted sequence by its diameter.	Lem. 4.4
$S_t, \Psi_t(D_0)$	Full fixed-radius trajectory length, including the projected terminal step, and its explicit upper bound.	Prop. 4.7
$\bar{x}, L_k$	Limit of the stopped trajectory and remaining length $\sum_{i=k}^{\infty} \ x^{i+1} - x^i\ $.	Cor. 4.9
$\rho, \rho_\star$	Decay factor in $t_k = t\rho^k$ and the minimum successful factor when $0 < t < D_0$.	Thm. D.1
$\mathcal{T}_k(\rho)$	Remaining geometric radius sum $t\rho^k/(1 - \rho)$.	App. D
5. Envelope, duality, and epigraph lifting		
F_t, f^*	Ball envelope $F_t(x) = \inf_{\ u-x\ \leq t} f(u)$ and convex conjugate $f^*(g) = \sup_x \{\langle g, x \rangle - f(x)\}$.	(2)
$q_t(\gamma), u_\gamma$	Scalar dual value and proximal point $\text{prox}_{f/\gamma}(x)$, for $\gamma > 0$.	(12)–(13)
$\delta_C, \text{epi } f$	Indicator of a set (0 on C , $+\infty$ outside) and epigraph $\{(x, s) : f(x) \leq s\}$.	(66)
F, z^k, R_k	Lifted objective $F(x, s) = s + \delta_{\text{epi } f}(x, s)$, lifted iterate $z^k = (x^k, f(x^k))$, and lifted distance to the solution set R_k .	§5
r_k, δ_k	Projected displacement $\ x^{k+1} - x^k\ $ and value decrease $f(x^k) - f(x^{k+1})$; $t_k^2 = r_k^2 + \delta_k^2$.	(67)
G^k, λ_k	Radial subgradient of the lifted objective $(\lambda_k g^k, 1 - \lambda_k)$ and its coefficient $\lambda_k = 1 - c_k \delta_k \in (0, 1)$.	(68)
N_t^{epi}	Fixed-radius termination count for lifted BPM initialized on the graph.	(74)

Context distinguishes three uses of λ : a positive proximal parameter, a relaxation factor in $(0, 2)$, and a coefficient in the epigraph subgradient representation in $(0, 1)$. Likewise, F_t is the ball envelope, whereas F is the lifted objective; S is a squared-distance budget, whereas S_t is a trajectory length.

B Auxiliary results

B.1 Auxiliary geometry for the Euclidean length bound

Lemma B.1 (Directions with pairwise nonnegative inner products). *Let $\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : \|u\| = 1\}$. If a nonempty set $U \subset \mathbb{S}^{d-1}$ satisfies $\langle u, v \rangle \geq 0$ for every $u, v \in U$, then there exists $w \in \mathbb{S}^{d-1}$ such that $\langle w, u \rangle \geq 1/\sqrt{d}$ for every $u \in U$.*

Proof. Replace U by its closure and let p be the point of minimum norm in its convex hull. By Carathéodory's theorem, $p = \sum_{i=1}^m a_i u_i$ with $m \leq d+1$, $a_i \geq 0$, and $\sum_i a_i = 1$. Pairwise nonnegativity gives $\|p\|^2 \geq \sum_i a_i^2 \geq 1/(d+1)$, so $p \neq 0$. Since p is the projection of the origin onto $\text{conv } U$, $\langle p, u \rangle \geq \|p\|^2$ for every $u \in U$. It remains to sharpen the representation to $m \leq d$. If $\text{conv } U$ has affine dimension at most $d-1$, this follows directly from Carathéodory's theorem. Otherwise, the supporting hyperplane $\{u : \langle p, u \rangle = \|p\|^2\}$ contains p . In any convex representation of p , every point with positive weight must lie on this hyperplane, so Carathéodory's theorem there again gives a representation with at most d points. Hence $\|p\|^2 \geq 1/d$. Setting $w = p/\|p\|$ gives $\langle w, u \rangle \geq 1/\sqrt{d}$ for every $u \in U$. $\square$

For $\delta \in (0, 1]$, the unit sphere $\mathbb{S}^{d-1}$ has a finite δ -net $V \subset \mathbb{S}^{d-1}$ with $|V| \leq (1 + 2/\delta)^d$. Indeed, take a maximal δ -separated subset of $\mathbb{S}^{d-1}$. The disjoint open balls of radius $\delta/2$ centered at its points lie inside the ball of radius $1 + \delta/2$. Volume comparison gives the cardinality bound, while maximality gives the net property.

C Proofs

C.1 Proof of Theorem 1.1

Proof. Optimality system. By Theorem 2.1, every step from $x \in \text{dom } f$ attains a finite value. If the output u is nonterminal, it is unique, $\|x - u\| = t$, and there exists $g = c(x - u) \in \partial f(u)$ with $c > 0$. Hence

$$g \in \partial f(u), \quad x - u \stackrel{(19)}{=} t \partial \|\cdot\| (g).$$

If u is terminal, then $0 \in \partial f(u)$ and $\|x - u\| \leq t$, so the same system holds with $g = 0$. Conversely, suppose

$$g \in \partial f(u), \quad x - u \in t \partial \|\cdot\| (g).$$

If $g = 0$, then u is a feasible global minimizer. If $g \neq 0$, then $x - u = tg/\|g\|$, and for every $z \in \mathbb{B}(x, t)$,

$$f(z) \geq f(u) + \langle g, z - u \rangle \stackrel{(19)}{=} f(u) + \langle g, z - x \rangle + t \|g\| \geq f(u).$$

Thus u is an exact ball minimizer. This proves (3).

(i): *Exact ball minimization.* This is the defining relation $u \in \text{brox}_f^t(x)$ in (4).

(ii): *Normalized implicit subgradient formulation.* On a nonterminal step, the optimality system gives $g = c(x - u) \neq 0$ and $\|x - u\| = t$, hence

$$x - u \stackrel{(19)}{=} t \frac{g}{\|g\|}.$$

Moreover, $0 \in \partial f(u) + c(u - x)$, so

$$u \stackrel{(20)}{=} \text{prox}_{\lambda f}(x), \quad \lambda = \frac{1}{c} = \frac{t}{\|g\|}.$$

This proves (5). Conversely, any $g \in \partial f(u) \setminus \{0\}$ satisfying $x - u = tg / \|g\|$ satisfies the optimality system and therefore gives an exact ball minimizer.

(iii): *Normalized explicit envelope formulation.* Write $F_t = f \square \delta_{\mathbb{B}(0,t)}$. Convexity follows directly from convexity of f and the ball, and F_t is proper because $F_t(y) \leq f(y) < \infty$ for every $y \in \text{dom } f$. To see that F_t is closed, let $y_n \rightarrow y$ with $\liminf_n F_t(y_n) < \infty$. Passing to a subsequence, assume the limit exists and choose minimizers v_n with $\|v_n - y_n\| \leq t$. Then (v_n) is bounded, so along a further subsequence $v_n \rightarrow v$ with $\|v - y\| \leq t$. Closedness of f gives

$$F_t(y) \stackrel{(2)}{\leq} f(v) \leq \liminf_n f(v_n) = \liminf_n F_t(y_n).$$

The same argument rules out a limit of $-\infty$ by properness of f . Thus F_t is proper, closed, and convex. Writing $y = v + w$ with $\|w\| \leq t$ gives

$$F_t^*(g) = f^*(g) + t \|g\|. \quad (77)$$

Fix an exact output u at x , so $F_t(x) = f(u)$. For any g ,

$$f(u) + f^*(g) - \langle g, u \rangle \geq 0, \quad t \|g\| - \langle g, x - u \rangle \geq 0. \quad (78)$$

Their sum is $F_t(x) + F_t^*(g) - \langle g, x \rangle$. Hence $g \in \partial F_t(x)$ exactly when both terms vanish, namely when

$$g \in \partial f(u), \quad \langle g, x - u \rangle \stackrel{(78)}{=} t \|g\|.$$

This proves (7). On a nonterminal step, $0 \notin \partial f(u)$, and equality in Cauchy–Schwarz gives

$$x - u \stackrel{(7)}{=} t \frac{g}{\|g\|}$$

for every $g \in \partial F_t(x)$. This proves (6).

(iv): *Projection formulation.* Let u be nonterminal and $C_u = \{z : f(z) \leq f(u)\}$. For every $z \in C_u$,

$$\langle g, z - u \rangle \leq f(z) - f(u) \leq 0.$$

Since $g = c(x - u)$ with $c > 0$, this is the projection optimality condition for $u = \text{Proj}_{C_u}(x)$. Together with $\|x - u\| = t$, this proves (8). Conversely, suppose $u = \text{Proj}_{C_u}(x)$ and $\|x - u\| = t$. If some $z \in \mathbb{B}(x, t)$ satisfied $f(z) < f(u)$, then $v = (z + u)/2$ would satisfy $f(v) < f(u)$ and $\|v - x\| < t$, contradicting $\text{dist}(x, C_u) = t$. Hence u is an exact ball minimizer. Taking $\alpha = f(u)$ gives (16).

(v): *Vector dual formulation.* The identity $F_t^* = f^* + t \|\cdot\|$ gives

$$F_t(x) \stackrel{(77)}{=} \max_g \{ \langle x, g \rangle - f^*(g) - t \|g\| \},$$

and the maximizing set is exactly $\partial F_t(x)$. Hence the maximum is attained. For any dual optimizer g and exact primal output u , the two nonnegative terms above have zero sum and therefore vanish separately. Thus

$$u \in \partial f^*(g), \quad x - u \in t \partial \|\cdot\|(g),$$

which proves (9). Conversely, these inclusions give the optimality system and therefore primal optimality and equality of primal and dual values. On a nonterminal step, $g \neq 0$, so

$$u \stackrel{(3)}{=} x - t \frac{g}{\|g\|}.$$

If $p \in \partial f(x)$ is available, conjugacy gives $x \in \partial f^*(p)$ and $f^*(p) < \infty$. Moreover,

$$t \|h\| + D_{f*}^x(h, p) \stackrel{(11)}{=} f^*(h) + t \|h\| - \langle x, h \rangle + \langle x, p \rangle - f^*(p),$$

so (11) differs from the vector dual objective only by a constant. Thus the two minimization problems are equivalent.

(vi): *Scalar multiplier formulation.* For $\gamma \geq 0$ and every feasible z ,

$$f(z) + \frac{\gamma}{2}(\|z - x\|^2 - t^2) \leq f(z),$$

so $q_t(\gamma) \leq F_t(x)$. For a nonterminal output, let $g = c(x - u)$ be a radial subgradient. Since $0 \in \partial f(u) + c(u - x)$, the point u minimizes $f(z) + c\|z - x\|^2/2$. Together with $\|u - x\| = t$, this gives $q_t(c) = f(u) = F_t(x)$. Hence the scalar maximum is attained at a positive multiplier. For $\gamma > 0$, let

$$u_\gamma = \text{prox}_{f/\gamma}(x).$$

The minimizer is unique and depends continuously on $\gamma > 0$. Indeed, on a compact positive parameter interval, comparison with $z = x$ and an affine lower bound on f keep the minimizers bounded. Every cluster point at a limiting parameter solves the corresponding strongly convex problem and is therefore its unique minimizer. Set

$$a(\gamma) = \frac{\|u_\gamma - x\|^2 - t^2}{2}.$$

For $\eta > \gamma > 0$, evaluating each parameter's objective at the other parameter's minimizer gives

$$a(\eta) \stackrel{(12)}{\leq} \frac{q_t(\eta) - q_t(\gamma)}{\eta - \gamma} \stackrel{(12)}{\leq} a(\gamma). \quad (79)$$

Continuity therefore yields

$$q'_t(\gamma) \stackrel{(79)}{=} \frac{1}{2}(\|u_\gamma - x\|^2 - t^2),$$

which proves (14). Since q_t is an infimum of affine functions of γ , it is concave, and the same inequalities show that $\|u_\gamma - x\|$ is nonincreasing. Every positive maximizing multiplier satisfies $\|u_\gamma - x\| = t$. Its proximal optimality condition gives $g = \gamma(x - u_\gamma)$, so u_γ is the exact ball output. Conversely, any positive γ satisfying $\|u_\gamma - x\| = t$ gives the same radial subgradient and attains the scalar maximum. The maximizing multiplier need not be unique.

(vii): *Epigraph formulation.* If u minimizes f over $\mathbb{B}(x, t)$, then $(u, f(u))$ is feasible for (15), and every feasible (v, s) satisfies $s \geq f(v) \geq f(u)$. Hence $(u, f(u))$ is optimal. Conversely, if (u, s) is optimal, then $s = f(u)$; otherwise lowering s to $f(u)$ preserves feasibility and decreases the objective. For every $v \in \mathbb{B}(x, t) \cap \text{dom } f$, the feasible pair $(v, f(v))$ then gives $f(u) \leq f(v)$. Thus u is an exact ball minimizer, and

$$\arg \min_{\substack{f(v) \leq s \\ \|v - x\| \leq t}} s \stackrel{(15)}{=} \{(v, f(v)) : v \in \text{brox}_f^t(x)\}.$$

This proves (vii). □

C.2 Proof of Theorem 2.1

Proof. Since $x \in \text{dom } f$, the compact ball $\mathbb{B}(x, t)$ contains a point where f is finite. Since f is proper and closed, it attains a finite minimum on $\mathbb{B}(x, t)$. Hence $\text{brox}_f^t(x)$ is nonempty and compact. Suppose the step is nonterminal and let $u \in \text{brox}_f^t(x)$. Then u must lie on the boundary of the ball: otherwise u would be a local minimizer of the convex function f , hence a global minimizer. The output is unique as well, since two distinct minimizers would have an

interior midpoint of no larger value, which would again be a global minimizer. Thus $\|x - u\| = t$. Under [Assumption 2](#) (Attainment), if $D(x) \leq t$, the ball contains a global minimizer, and hence $\text{brox}_f^t(x) = X^* \cap \mathbb{B}(x, t)$. If $D(x) = t$, every point in this intersection is a projection of x onto the closed convex set X^* , so uniqueness of the projection gives $\text{brox}_f^t(x) = \{\text{Proj}_{X^*}(x)\}$. It remains to prove the subgradient for a nonterminal step. Choose $z \in \text{ri}(\text{dom } f)$. Points on the segment from x toward z , sufficiently close to x , lie in $\text{ri}(\text{dom } f) \cap \mathbb{B}^\circ(x, t)$, so the convex subdifferential sum rule applies. Fermat's rule gives $0 \in \partial f(u) + N_{\mathbb{B}(x, t)}(u)$. Since $\|x - u\| = t$, we have $N_{\mathbb{B}(x, t)}(u) = \{c(u - x) : c \geq 0\}$, and therefore, for some $g \in \partial f(u)$ and $c \geq 0$,

$$\|x - u\| = t, \quad g = c(x - u) \in \partial f(u).$$

If $c = 0$, then $0 \in \partial f(u)$ and u is globally optimal, contradicting nonterminality. Hence $c > 0$. Finally, setting $\lambda = 1/c$ gives $0 \in \partial f(u) + (u - x)/\lambda$, which is the optimality condition for the proximal subproblem. Therefore

$$u = \text{prox}_{\lambda f}(x), \quad \lambda = \frac{1}{c} = \frac{t}{\|g\|}.$$

□

C.3 Proof of [Theorem 2.6](#)

Proof. For two consecutive nonterminal iterations, monotonicity of ∂f at x^{k+1} and x^{k+2} gives

$$\langle g^k - g^{k+1}, x^{k+1} - x^{k+2} \rangle \geq 0.$$

Using $x^{k+1} - x^{k+2} = t_{k+1}g^{k+1}/\|g^{k+1}\|$ and dividing by $t_{k+1} > 0$ yields [\(27\)](#). Cauchy–Schwarz then gives $\|g^{k+1}\| \leq \|g^k\|$, and [\(27\)](#) also shows that the two radial subgradient directions have positive inner product. Now suppose the first K iterations are nonterminal. Summing [\(23\)](#) and using the monotonicity just proved gives

$$f(x^0) - f(x^K) \stackrel{(23)}{\geq} \sum_{k=0}^{K-1} t_k \|g^k\| \stackrel{(27)}{\geq} \left(\sum_{k=0}^{K-1} t_k \right) \|g^{K-1}\|.$$

Since $g^{K-1} \in \partial f(x^K)$, $\text{dist}(0, \partial f(x^K)) \leq \|g^{K-1}\|$, and under [Assumption 2](#) (Attainment), $f(x^0) - f(x^K) \leq \Delta_0$. This proves [\(28\)](#). If a terminal output has already occurred, then by convention $g^{K-1} = 0$ and $0 \in \partial f(x^K)$, so the same inequalities hold immediately. □

C.4 Proof of [Theorem 3.1](#)

Proof. Suppose first that $D_k > 0$, and let $p^k = \text{Proj}_{X^*}(x^k)$ and $a_k = \min\{t_k/D_k, 1\}$. The point $v = (1 - a_k)x^k + a_k p^k$ satisfies $\|v - x^k\| = a_k D_k \leq t_k$ and is therefore feasible for the k th step. Exact minimization and convexity give

$$f(x^{k+1}) \stackrel{\text{(BPM)}}{\leq} f(v) \leq (1 - a_k)f(x^k) + a_k f_\star.$$

Subtracting $f_\star$ yields [\(29\)](#). If $D_k = 0$, then $x^k \in X^*$, and both gaps are zero by the stopping convention. Now let $t_k = t$ with $0 < t < D_0$. Before termination, $D_k \leq D_0$, so $1 - t/D_k \leq 1 - t/D_0$ on every nonterminal step. Iterating [\(29\)](#) therefore gives [\(30\)](#). At and after a terminal step the objective gap is zero, so the same bound holds for the stopped sequence. □

C.5 Proof of Theorem 3.2

Proof. Consider a nonterminal step and let $p^{k+1} = \text{Proj}_{X^*}(x^{k+1})$. Since $g^k \in \partial f(x^{k+1})$, the subgradient inequality gives $\Delta_{k+1} \leq \|g^k\| D_{k+1}$, while (23) gives $\Delta_k - \Delta_{k+1} \geq t_k \|g^k\|$. Hence

$$\frac{\Delta_{k+1}}{\Delta_k} \stackrel{(23)}{\leq} \frac{D_{k+1}}{D_{k+1} + t_k} \leq \frac{D_{k+1}}{D_k},$$

where the second inequality uses $D_k \leq D_{k+1} + \|x^{k+1} - x^k\| = D_{k+1} + t_k$. Multiplying these inequalities over a sequence of nonterminal steps gives $\Delta_K/\Delta_0 \leq D_K/D_0$, proving (32). If a terminal step occurs earlier, then the stopped sequence has $\Delta_K = D_K = 0$, so the same bound holds. $\square$

C.6 Proof of Theorem 3.3

Proof. Let $p^k = \text{Proj}_{X^*}(x^k)$. Applying (22) with $z = p^k$ and discarding the nonnegative Bregman term gives

$$D_k^2 - \|x^{k+1} - p^k\|^2 \stackrel{(22)}{\geq} t_k^2 + \frac{2\Delta_{k+1}}{c_k}.$$

Since $D_{k+1} \leq \|x^{k+1} - p^k\|$, this yields (33). By (23), $\Delta_k - \Delta_{k+1} \geq t_k \|g^k\| = c_k t_k^2 > 0$. Hence

$$\frac{2\Delta_{k+1}}{c_k} \stackrel{(23),(19)}{\geq} \frac{2t_k^2 \Delta_{k+1}}{\Delta_k - \Delta_{k+1}}.$$

Substituting this bound into (33) and combining the two t_k^2 terms gives (34). Since the step is nonterminal, $\Delta_{k+1} > 0$, so the multiplier in (34) is strictly larger than one. $\square$

C.7 Proof of Theorem 3.4

Proof. For the first K nonterminal iterations, set $q_k = \Delta_{k+1}/\Delta_k \in (0, 1)$ and $\beta_k = (D_k^2 - D_{k+1}^2)/t^2 > 1$. By Theorem 3.3,

$$q_k \stackrel{(34)}{\leq} \frac{\beta_k - 1}{\beta_k + 1}, \quad \sum_{k=0}^{K-1} \beta_k = \frac{D_0^2 - D_K^2}{t^2} < S, \quad (80)$$

where the strict inequality follows from $D_K > 0$. Since every $\beta_k > 1$, this also implies $K < S$. Let $h(b) = \log((b-1)/(b+1))$. On $(1, \infty)$, $h'(b) = 2/(b^2 - 1) > 0$ and $h''(b) = -4b/(b^2 - 1)^2 < 0$, so h is increasing and strictly concave. Jensen's inequality therefore gives

$$\begin{aligned} \log \frac{\Delta_K}{\Delta_0} &\stackrel{(35)}{=} \sum_{k=0}^{K-1} \log q_k \\ &\stackrel{(80)}{\leq} K h \left(\frac{1}{K} \sum_{k=0}^{K-1} \beta_k \right) \\ &\stackrel{(80)}{\leq} K \log \frac{S - K}{S + K}. \end{aligned}$$

Exponentiating proves the first inequality in (42). With $v = K/S \in (0, 1)$, the inequality $\log((1-v)/(1+v)) \leq -2v$ yields $\Delta_K \leq \Delta_0 \exp(-2K^2/S)$. For (43), Theorem 2.5 gives $N_\varepsilon \leq \lceil S \rceil$. Now set

$$\widehat{K} = \left\lceil \frac{D_0}{t} \sqrt{\frac{1}{2} \log \frac{\Delta_0}{\varepsilon}} \right\rceil.$$

If $\widehat{K} \geq \lceil S \rceil$, the first bound already gives $N_\varepsilon \leq \widehat{K}$. Otherwise $\widehat{K} < S$. If a terminal step occurs before $\widehat{K}$, then $\Delta_{\widehat{K}} = 0 \leq \varepsilon$; if the first $\widehat{K}$ iterations are nonterminal, the exponential bound gives $\Delta_{\widehat{K}} \leq \varepsilon$. Thus $N_\varepsilon \leq \widehat{K}$, proving (43). $\square$

C.8 Proof of Theorem 3.7

Proof. Suppose first that the first K iterations are nonterminal, and choose radial subgradients $g^i \in \partial f(x^{i+1})$ for $0 \leq i < K$. For any $0 \leq j < K$, summing (23) from $i = j$ to $K - 1$ and discarding the nonnegative Bregman distances gives

$$\Delta_j - \Delta_K \stackrel{(23)}{\geq} \sum_{i=j}^{K-1} t_i \|g^i\|. \quad (81)$$

By Theorem 2.6, $\|g^i\| \geq \|g^{K-1}\| \geq m(x^K)$ for $j \leq i < K$. Hence

$$\Delta_j - \Delta_K \stackrel{(81),(27)}{\geq} \left(\sum_{i=j}^{K-1} t_i \right) m(x^K).$$

Dividing by the positive radius sum and minimizing over j gives the first inequality in (48). The second follows from $\Delta_K \geq 0$. If a terminal output occurs before or at K , then $m(x^K) = 0$, while $\Delta_j \geq \Delta_K$ for every $j < K$, so (48) remains valid. Now let $t_k = t$ with $0 < t < D_0$, and suppose the first K iterations are nonterminal. By Theorem 3.4, $K < S$. Taking $j = K - 1$ in (48) gives $m(x^K) \leq \Delta_{K-1}/t$. For $K \geq 2$, applying Theorem 3.4 at horizon $K - 1$ yields

$$\Delta_{K-1} \stackrel{(42)}{\leq} \Delta_0 \left(\frac{S - K + 1}{S + K - 1} \right)^{K-1} \stackrel{(42)}{\leq} \Delta_0 \exp \left(-\frac{2(K-1)^2}{S} \right),$$

which proves (49). For $K = 1$, the same bound reduces to $m(x^1) \leq \Delta_0/t$. Finally, for every $0 \leq j < K$, (30) and (42) give $\Delta_j \leq \Delta_0 E_j$. Since $\sum_{i=j}^{K-1} t_i = (K - j)t$, the second inequality in (48) gives

$$m(x^K) \stackrel{(48),(50)}{\leq} \frac{\Delta_0}{t} \frac{E_j}{K - j}.$$

Minimizing over $0 \leq j < K$ proves (51). If f is differentiable, then $m(x^K) = \|\nabla f(x^K)\|$, giving the stated specialization. $\square$

C.9 Proof of Theorem 3.8

Proof. Since $x^\star \in X^\star$, we have $0 \in \partial f(x^\star)$, hence $x^\star \in \partial f^\star(0)$ and $f^\star(0) = -f_\star$. Similarly, $g \in \partial f(u)$ implies $u \in \partial f^\star(g)$ and $f^\star(g) = \langle u, g \rangle - f(u)$. Therefore

$$D_{f^\star}^u(0, g) \stackrel{(11)}{=} f^\star(0) - f^\star(g) + \langle u, g \rangle = f(u) - f_\star, \quad (82)$$

which proves (52). Moreover,

$$D_{f^\star}^{x^\star}(g, 0) \stackrel{(11)}{=} f^\star(g) - f^\star(0) - \langle x^\star, g \rangle = f_\star - f(u) - \langle g, x^\star - u \rangle \stackrel{(21)}{=} D_f^g(x^\star, u), \quad (83)$$

proving (53). Both quantities are nonnegative by convexity. Adding them gives

$$\mathcal{G}_{x^\star}(u, g) \stackrel{(82),(83)}{=} \langle g, u - x^\star \rangle,$$

which is (54). Hence

$$0 \leq f(u) - f_\star \stackrel{(82),(83)}{\leq} \mathcal{G}_{x^\star}(u, g) \stackrel{(54)}{\leq} \|g\| \|u - x^\star\|,$$

proving (55). Now consider the stopped iteration and set $x^\star = \text{Proj}_{X^\star}(x^0)$. Before termination, (24) gives $\|x^K - x^\star\| \leq D_0$, and therefore $\mathcal{G}_K \leq D_0 \|g^{K-1}\|$. For any $0 \leq j < K$, summing (23) gives

$$\Delta_j - \Delta_K \stackrel{(23)}{\geq} \sum_{i=j}^{K-1} t_i \|g^i\|. \quad (84)$$

By Theorem 2.6, $\|g^i\| \geq \|g^{K-1}\|$ for $j \leq i < K$, so

$$\left(\sum_{i=j}^{K-1} t_i \right) \|g^{K-1}\| \stackrel{(84),(27)}{\leq} \Delta_j - \Delta_K.$$

Combining the last two inequalities and minimizing over j proves (56). At and after termination, the convention $g^{K-1} = 0$ gives $\mathcal{G}_K = 0$, so the same bound holds. Finally, for a constant-radius sequence of nonterminal steps, $\Delta_j \leq \Delta_0 E_j$ and $\sum_{i=j}^{K-1} t_i = (K-j)t$. Substituting these into (56) gives

$$\mathcal{G}_K \stackrel{(56),(50)}{\leq} \frac{D_0 \Delta_0}{t} \frac{E_j}{K-j}.$$

Minimizing over $0 \leq j < K$ proves (57). Since the two Bregman distances are nonnegative and sum to $\mathcal{G}_K$, each satisfies the same upper bounds. $\square$

C.10 Proof of Theorem 4.1

Proof. If $\sum_k t_k < \infty$, the final assertion of Theorem 4.8 applies without assuming attainment and gives finite trajectory length together with

$$x^k \stackrel{(62)}{\rightarrow} \bar{x} \in \text{dom } f, \quad f(x^k) \stackrel{(62)}{\downarrow} f(\bar{x}).$$

This proves (i). Suppose now that $\sum_k t_k = \infty$. If a minimizer exists, Theorem 4.6 gives finite termination, proving (ii). It remains to consider $\arg \min f = \emptyset$. Every step is then nonterminal, so $\|x^{k+1} - x^k\| = t_k$ for every k . Since x^k is feasible for the k th ball problem, the values $f(x^k)$ are nonincreasing. Let $\ell = \lim_k f(x^k)$. Suppose $\ell > \inf f$. Choose $z \in \text{dom } f$ with $f(z) < \ell$. Then $f(z) < f(x^{k+1})$ for every k , so (24) gives

$$\|x^{k+1} - z\|^2 \stackrel{(24)}{\leq} \|x^k - z\|^2 - t_k^2.$$

Hence the trajectory is bounded. By Theorem 4.3, it is self-contracted, and Theorem 4.4 therefore gives finite total length. But every step has length t_k , so

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| \stackrel{(19)}{=} \sum_{k=0}^{\infty} t_k = \infty,$$

contradicting the finite-length conclusion. Therefore $f(x^k) \downarrow \inf f$. Suppose next that $\|x^k\|$ does not tend to infinity. Then some bounded subsequence has a cluster point $\bar{x}$. By closedness of f ,

$$f(\bar{x}) \leq \liminf_j f(x^{k_j}) = \inf f.$$

If $\inf f$ is finite, then $\bar{x}$ attains the infimum, contradicting $\arg \min f = \emptyset$. If $\inf f = -\infty$, then $f(\bar{x}) = -\infty$, contradicting properness. Thus $\|x^k\| \rightarrow \infty$, proving (iii). The two nonsummable cases also give (58). For universal necessity, use the construction in the proof of [Theorem 4.6](#). For any summable schedule with $T = \sum_k t_k$, the objective $f(x) = \|x\|$ and starting point $x^0 = (T+1)e_1$ produce $x^k \rightarrow e_1$ and $f(x^k) \rightarrow 1 > 0 = \inf f$. Thus no summable positive schedule guarantees value convergence to the infimum on every problem. $\square$

C.11 Proof of [Theorem 4.3](#)

Proof. Since x^k is feasible for the k -th ball problem, the objective values are nonincreasing. Thus, for $j < m$ in a sequence of nonterminal steps, $f(x^m) \leq f(x^{j+1})$. Applying (24) to the j th step with $z = x^m$ gives $\|x^{j+1} - x^m\|^2 \leq \|x^j - x^m\|^2 - t_j^2$, which is (59). In particular, $\|x^{j+1} - x^m\| \leq \|x^j - x^m\|$. For $i \leq j < m$, applying this comparison successively from i to $j-1$ gives $\|x^j - x^m\| \leq \|x^i - x^m\|$. The case $j = m$ is immediate, so the prefix is self-contracted. $\square$

C.12 Proof of [Theorem 4.4](#)

It suffices to consider a finite sequence $z^0, \dots, z^m$. Let D be its diameter and $\ell_i = \|z^{i+1} - z^i\|$. Discard zero-length steps and, for each $i < m$, define

$$U_i := \left\{ \frac{z^j - z^i}{\|z^j - z^i\|} : j > i \right\}.$$

No future point equals z^i . If $i < j < k$, set $a = z^j - z^i$ and $b = z^k - z^i$. Self-contraction gives $\|b - a\| \leq \|b\|$, hence $\langle a, b \rangle \geq \|a\|^2/2 \geq 0$. Thus the directions in U_i have pairwise nonnegative inner products, and [Theorem B.1](#) gives $v_i \in \mathbb{S}^{d-1}$ with $\langle v_i, u \rangle \geq 1/\sqrt{d}$ for every $u \in U_i$. Suppose first that $d \geq 2$. Take a $\delta = 1/(2\sqrt{d})$ net V and choose $w_i \in V$ with $\|w_i - v_i\| \leq \delta$. Then $\langle w_i, u \rangle \geq 1/(2\sqrt{d})$ for every $u \in U_i$. For $j > i$, self-contraction gives $\|z^{i+1} - z^j\| \leq \|z^i - z^j\|$, and therefore $\ell_i \leq 2\|z^i - z^j\|$. It follows that $\langle w_i, z^j - z^i \rangle \geq \ell_i/(4\sqrt{d})$. For $w \in V$, define $b_i(w) = \min_{j \geq i} \langle w, z^j \rangle$. These quantities are nondecreasing in i . If $w_i = w$, then $\langle w, z^j \rangle \geq \langle w, z^i \rangle + \ell_i/(4\sqrt{d})$ for every $j > i$, so $b_{i+1}(w) - b_i(w) \geq \ell_i/(4\sqrt{d})$. Summing over the indices with $w_i = w$ gives

$$\sum_{\substack{i=0 \\ w_i=w}}^{m-1} \ell_i \leq 4\sqrt{d}(b_m(w) - b_0(w)) \leq 4\sqrt{d}D.$$

Summing over $w \in V$ and using $|V| \leq (1+4\sqrt{d})^d$ gives the stated constant. For $d = 1$, all future points lie on the same side of the current point. Let $I_i = [\min_{j \geq i} z^j, \max_{j \geq i} z^j]$. Then z^i is an endpoint of I_i . The estimate $\ell_i \leq 2\|z^i - z^j\|$ for every $j > i$ shows that removing z^i moves this endpoint inward by at least $\ell_i/2$. Hence

$$\ell_i \leq 2(|I_i| - |I_{i+1}|),$$

and telescoping gives total length at most $2D$. Applying the finite bound to every prefix proves the result for infinite bounded self-contracted sequences.

C.13 Proof of [Theorem 4.5](#)

Proof. Let $p^0 = \text{Proj}_{X^*}(x^0)$. For every iterate before terminality, (24) with $z = p^0$ gives $\|x^k - p^0\| \leq D_0$. Hence every sequence of nonterminal steps has diameter at most $2D_0$. By [Theorems 4.3](#) and [4.4](#), its total length is at most $2C_d D_0$. Since every nonterminal displacement has length t_k , this proves $\sum_{k < K} t_k \leq 2C_d D_0$, which is (60). Now suppose the radius is constant. The

first $N_t - 1$ iterations are nonterminal, so $(N_t - 1)t \leq 2C_d D_0$, and therefore $N_t \leq 1 + \lfloor 2C_d D_0/t \rfloor$. The other upper bound in (61) follows from Theorem 2.5. For the lower bound, let $x^{N_t} \in X^*$ be the terminal output. The total trajectory length satisfies

$$D_0 \leq \|x^{N_t} - x^0\| \leq \sum_{k=0}^{N_t-1} \|x^{k+1} - x^k\| \stackrel{\text{(BPM)}}{\leq} N_t t.$$

Thus $N_t \geq \lceil D_0/t \rceil$. This lower bound is attained in one dimension by $f(x) = |x|$, and together with the fixed-dimensional upper bound yields the stated $\Theta_d(D_0/t)$ complexity. $\square$

C.14 Proof of Theorem 4.6

Proof. Assume first that $\sum_k t_k = \infty$. If the iteration remained nonterminal indefinitely, then (60) would give $\sum_{k < K} t_k \leq 2C_d D_0$ for every K , contradicting $\sum_k t_k = \infty$. Hence a terminal step must occur after finitely many steps, and its output is a minimizer. Conversely, suppose $T = \sum_k t_k < \infty$. Take $f(x) = \|x\|$ and $x^0 = (T + 1)e_1$. We claim that

$$x^k = \left(T + 1 - \sum_{j=0}^{k-1} t_j \right) e_1.$$

Indeed, $\|x^k\| = 1 + \sum_{j \geq k} t_j > t_k$, so $\mathbb{B}(x^k, t_k)$ does not contain the minimizer 0. Since x^k lies on the positive e_1 -axis, the point in this ball nearest to the origin is $x^k - t_k e_1$. Thus $x^{k+1} = x^k - t_k e_1$, which gives the claimed formula at index $k + 1$. Every step is therefore nonterminal, so the iteration never attains the minimizer and (ii) fails. $\square$

C.15 Proof of Theorem 4.7

Proof. There are $n = N_t - 1$ nonterminal displacements of length t and a projected terminal displacement $r = D_n \leq t$. Let $p = \text{Proj}_{X^*}(x^0)$. Applying (24) with $z = p$ to the first n iterations gives

$$\|x^n - p\|^2 \stackrel{(24)}{\leq} D_0^2 - nt^2. \quad (85)$$

Since the terminal step is $x^{n+1} = \text{Proj}_{X^*}(x^n)$, the projection inequality gives $\|x^{n+1} - p\|^2 \leq \|x^n - p\|^2 - r^2$. Hence $nt^2 + r^2 \leq D_0^2$, so

$$S_t = nt + r \stackrel{(85)}{\leq} nt + \sqrt{D_0^2 - nt^2}.$$

By Theorem 2.5, $N_t \leq \lceil D_0^2/t^2 \rceil$, and therefore $n \leq J$. The function $n \mapsto n + \sqrt{D_0^2/t^2 - n}$ is nondecreasing on the integers $0, \dots, J$: for $n < J$, its increment is $1 + \sqrt{a-1} - \sqrt{a} \geq 0$, where $a = D_0^2/t^2 - n > 1$. Thus $S_t \leq Jt + \sqrt{D_0^2 - Jt^2}$. Set $q = D_0^2/t^2 - J \in (0, 1]$. Then $J = D_0^2/t^2 - q$, and hence

$$Jt + \sqrt{D_0^2 - Jt^2} = \frac{D_0^2}{t} + t(\sqrt{q} - q) \leq \frac{D_0^2}{t} + \frac{t}{4},$$

since $0 \leq \sqrt{q} - q \leq 1/4$ for $q \in [0, 1]$. If D_0^2/t^2 is an integer, then $q = 1$ and the first bound equals D_0^2/t . If $t \geq D_0$, the initial ball contains a minimizer, so the first step is terminal and nearest-minimizer selection at termination gives $S_t = D_0$. $\square$

C.16 Proof of Theorem 4.8

Proof. If the initial point is optimal or a terminal step occurs, the stopped sequence is eventually a fixed minimizer. It therefore has finite total length and satisfies (62). The terminal displacement is finite, although no radius-independent bound on it is asserted. Suppose henceforth that every step is nonterminal. Under attainment, Theorem 4.5 bounds every partial sum of radii by $2C_d D_0$, so $\sum_k t_k < \infty$. If summability is assumed instead, this conclusion requires no attainment assumption. In either case, full-radius motion gives $\sum_k \|x^{k+1} - x^k\| = \sum_k t_k < \infty$. Hence (x^k) is Cauchy in $\mathbb{R}^d$ and converges to some $\bar{x}$. For $m > k$, the triangle inequality gives $\|x^m - x^k\| \leq \sum_{j=k}^{m-1} t_j$. Letting $m \rightarrow \infty$ proves the first estimate in (63). Choose a radial subgradient $g^k \in \partial f(x^{k+1})$ at each step. By Theorem 2.6, $\|g^k\| \leq M := \|g^0\|$ for all $k \geq 0$. The objective values are nonincreasing and also bounded below. Indeed, the subgradient inequality at x^1 gives

$$f(x^k) \geq f(x^1) + \langle g^0, x^k - x^1 \rangle \quad (k \geq 1),$$

and the convergent sequence (x^k) is bounded. Hence $f(x^k) \downarrow \ell$ for some $\ell \in \mathbb{R}$. Lower semicontinuity gives $f(\bar{x}) \leq \ell < \infty$, so $\bar{x} \in \text{dom } f$. For every $k \geq 1$, applying the subgradient inequality at x^k with subgradient g^{k-1} to $\bar{x}$ gives

$$f(\bar{x}) \geq f(x^k) + \langle g^{k-1}, \bar{x} - x^k \rangle \stackrel{(27)}{\geq} f(x^k) - M \|\bar{x} - x^k\|. \quad (86)$$

Passing to the limit yields $f(\bar{x}) \geq \ell$. Thus $f(\bar{x}) = \ell$, proving value convergence. The same inequality, together with $f(x^k) \geq \ell$ and the distance tail estimate, gives

$$0 \leq f(x^k) - f(\bar{x}) \stackrel{(86)}{\leq} M \sum_{j=k}^{\infty} t_j,$$

which is the second estimate in (63). No assumption on $\partial f(x^0)$ is needed, since the first nonterminal output supplies the initial subgradient. $\square$

C.17 Proof of Theorem 4.9

Proof. The limit exists by Theorem 4.8. Fix k and let $p = \text{Proj}_{X^*}(x^k)$. For every nonterminal step after k , (24) with $z = p$ shows that the distance to p cannot increase. If a terminal step occurs, nearest-minimizer selection at termination and the projection inequality give the same conclusion. Hence $\|x^i - p\| \leq \|x^k - p\| = D_k$ for every $i \geq k$, so the tail has diameter at most $2D_k$. The tail is also self-contracted. Indeed, if the update from x^i is nonterminal and $j > i$, monotonicity gives $f(x^j) \leq f(x^{i+1})$, so (24) yields $\|x^{i+1} - x^j\| \leq \|x^i - x^j\|$. If the update from x^i is terminal, every future point equals x^{i+1} by the stopping convention. Iterating these comparisons proves self-contraction. Applying Theorem 4.4 to the tail gives $L_k \leq 2C_d D_k$, while the triangle inequality on finite tails and passage to the limit give $\|x^k - \bar{x}\| \leq L_k$. If $D_k \leq A\rho^k$, substitution gives the stated geometric bounds. Moreover, $D_k \rightarrow 0$, and since X^* is closed, $x^k \rightarrow \bar{x}$ implies $\bar{x} \in X^*$. Finally, nearest-minimizer selection at termination is essential for a radius-independent bound on the terminal displacement. Take $f(x, y) = x^2/2$ and $x^0 = (1, 0)$. For any $t > 1$, an exact ball oracle may return the minimizer $(0, \sqrt{t^2 - 1})$, whose displacement from x^0 is t , although $D_0 = 1$. Nearest-minimizer selection at termination instead returns $(0, 0)$. On an infinite nonterminal run, BPM with arbitrary terminal selection and BPM with nearest-minimizer selection coincide, so no additional terminal-selection assumption is needed there. $\square$

C.18 Proof of Theorem 4.11

Proof. Let G be the $n \times n$ tridiagonal matrix with diagonal entries one and adjacent entries $\epsilon \in (0, 1/2)$. For every $v \in \mathbb{R}^n$,

$$v^\top G v = \sum_{j=0}^{n-1} v_j^2 + 2\epsilon \sum_{j=0}^{n-2} v_j v_{j+1} \geq (1 - 2\epsilon) \|v\|^2,$$

so G is positive definite. Choose unit vectors $n_0, \dots, n_{n-1}$ with Gram matrix G , and define

$$x^0 = 0, \quad x^{j+1} = x^j - t n_j, \quad x^* = x^n, \quad \rho = \frac{\epsilon}{1 + \epsilon}.$$

For $j = 0, \dots, n-2$, set $u_j = x^{j+1}$, $a_j = \rho^j$, $g_j = a_j n_j$, and $F_j = a_j t \epsilon$. Let $\eta = a_{n-2} \epsilon / [2(1 + \epsilon)]$ and define

$$f(z) = \max \left\{ 0, \max_{0 \leq j \leq n-2} [F_j + \langle g_j, z - u_j \rangle], \max_{0 \leq j < n} \eta |\langle n_j, z - x^* \rangle| \right\}. \quad (87)$$

After expanding the absolute values, f is the maximum of finitely many affine functions and is therefore finite, closed, convex, and piecewise affine. We first verify that the affine piece with gradient g_i is active at u_i . For $i > j$, $\langle g_j, u_i - u_j \rangle = -a_j t \epsilon$, so $F_j + \langle g_j, u_i - u_j \rangle = 0 \leq F_i$. If $i = j - 1$, then $F_j + \langle g_j, u_i - u_j \rangle = F_j + a_j t = F_{j-1}$. If $i \leq j - 2$, then

$$F_j + \langle g_j, u_i - u_j \rangle = F_j + a_j t(1 + \epsilon) \leq F_i,$$

because $F_i - F_j \geq a_j t((1 + \epsilon)^2 / \epsilon - \epsilon) \geq a_j t(1 + \epsilon)$. Finally, when $i = j$ the value is exactly F_i . Thus none of the other affine pieces exceeds F_i at u_i . At x^* , each of these affine pieces vanishes. Indeed, $x^* - u_j = -t \sum_{r=j+1}^{n-1} n_r$, and only the $r = j + 1$ term has nonzero inner product with n_j , giving $F_j + \langle g_j, x^* - u_j \rangle = F_j - a_j t \epsilon = 0$. It remains to control the symmetric closing pieces. Since $u_i - x^* = t \sum_{r=i+1}^{n-1} n_r$, for all relevant i and j ,

$$0 \leq \langle n_j, u_i - x^* \rangle \leq t(1 + 2\epsilon). \quad (88)$$

Hence

$$\eta |\langle n_j, u_i - x^* \rangle| \stackrel{(88)}{\leq} \frac{a_{n-2} t \epsilon (1 + 2\epsilon)}{2(1 + \epsilon)} \leq a_i t \epsilon = F_i.$$

Therefore $f(u_i) = F_i$, and the active affine piece gives $g_i \in \partial f(u_i)$. We now verify that u_i is the exact radius- t response at x^i . Since $g_i = (a_i/t)(x^i - u_i)$ and $\|x^i - u_i\| = t$, every $z \in \mathbb{B}(x^i, t)$ satisfies

$$\langle g_i, z - u_i \rangle = \frac{a_i}{t} (\langle x^i - u_i, z - x^i \rangle + t^2) \geq 0.$$

The subgradient inequality then gives $f(z) \geq f(u_i)$, so u_i is an exact ball minimizer. Moreover, $f(x^*) = 0$. If $f(z) = 0$, the symmetric pieces in (87) force $\langle n_j, z - x^* \rangle = 0$ for every j . Since G is positive definite, the vectors $n_0, \dots, n_{n-1}$ form a basis of $\mathbb{R}^n$, and hence $z = x^*$. Thus x^* is the unique minimizer. For $j \leq n-2$,

$$\|x^j - x^*\|^2 = t^2 [(n - j) + 2(n - j - 1)\epsilon] > t^2. \quad (89)$$

Thus the first $n - 1$ iterations are nonterminal, and their exact outputs are uniquely the prescribed points x^{j+1} . At x^{n-1} , the distance to x^* is exactly t , so the n th step is terminal. Therefore $N_t = n$. Taking $j = 0$ in the same distance formula gives

$$D_0^2 \stackrel{(89)}{=} t^2 [n + 2(n - 1)\epsilon],$$

which is (65). If $\epsilon = o(1/n)$, then $D_0^2/t^2 = n + o(1)$, and since the excess over n is positive and tends to zero, $\lceil D_0^2/t^2 \rceil = n + 1$ for all sufficiently large n . Hence $N_t / \lceil D_0^2/t^2 \rceil = n/(n + 1) \rightarrow 1$. Finally, a dimension-independent bound $N_t \leq C D_0/t$ would imply $n \leq C \sqrt{n + o(1)}$, which is impossible as $n \rightarrow \infty$. $\square$

C.19 Proof of Theorem 5.1

Proof. For part (i), under the change of variables $x = a + sQy$, orthogonality gives $\|x - x_0\| = s\|y - y_0\|$. Thus the radius- t ball centered at y_0 corresponds exactly to the radius- st ball centered at $x_0 = a + sQy_0$. Substitution preserves objective values and therefore identifies the corresponding exact ball-minimizer sets. For part (ii), strict monotonicity of χ gives $f(x) \leq f(y)$ if and only if $\tilde{f}(x) \leq \tilde{f}(y)$ for all $x, y \in \text{dom } f$. Since the center of every ball call lies in the domain and has finite objective value, points outside the domain, where both objectives equal $+\infty$, cannot be minimizers. Hence f and $\tilde{f}$ have the same exact ball-minimizer sets at every center and radius. The same order equivalence shows that their global minimizer sets coincide. Therefore, under the same selections, induction gives identical fixed-radius trajectories and hence identical termination counts. $\square$

C.20 Proof of Theorem 5.2

We first record the subgradients of F at a graph point. If $(v, \beta) \in \partial F(u, f(u))$, testing points vertically above $(u, f(u))$ gives $\lambda := 1 - \beta \geq 0$. Testing $(y, f(y))$ for $y \in \text{dom } f$ then gives

$$\langle v, y - u \rangle \leq \lambda(f(y) - f(u)). \quad (90)$$

Conversely, these conditions imply the subgradient inequality at every point of the epigraph. If $\lambda > 0$, (90) is equivalent to $v/\lambda \in \partial f(u)$. If $\lambda = 0$, it means $v \in N_{\text{dom } f}(u)$, where $\langle v, y - u \rangle \leq 0$ for every $y \in \text{dom } f$. No closedness of $\text{dom } f$ is needed. At a point (u, s) strictly above the graph, testing both $(u, f(u))$ and points vertically above (u, s) forces $\beta = 1$, and then $v \in N_{\text{dom } f}(u)$.

Proof. The epigraph is nonempty, closed, and convex, so F is proper, closed, and convex. Its infimum and minimizers follow by minimizing first over $s \geq f(x)$. Consider a nonterminal output (u, s) from $z = (x, f(x))$. By Theorem 2.1 applied to F , there is $c > 0$ such that

$$(v, \beta) \stackrel{(19)}{=} c(x - u, f(x) - s) \in \partial F(u, s), \quad \|z - (u, s)\| \stackrel{(19)}{=} t.$$

Since the step is nonterminal, its objective value is strictly smaller than that of the center, so $s < f(x)$. Suppose $s > f(u)$. The preceding subgradient characterization gives $v \in N_{\text{dom } f}(u)$. Testing $x \in \text{dom } f$ then gives $c\|x - u\|^2 \leq 0$, so $u = x$, contradicting $f(u) \leq s < f(x)$. Hence every nonterminal output lies on the graph. Terminal outputs also lie on the graph because $\arg \min F = X^* \times \{f_\star\}$. Induction therefore shows that every iterate remains on the graph. Set $\delta = f(x) - f(u) > 0$. At the graph output, the subgradient characterization gives $\lambda := 1 - c\delta \geq 0$. If $\lambda = 0$, then $c(x - u) \in N_{\text{dom } f}(u)$, and testing x again forces $u = x$, contradicting $\delta > 0$. Thus $0 < \lambda < 1$, and

$$g = \frac{c}{\lambda}(x - u) \stackrel{(90)}{\in} \partial f(u), \quad (v, \beta) = (\lambda g, 1 - \lambda).$$

In particular, $r = \|x - u\| > 0$, while the full-radius identity gives $t^2 = r^2 + \delta^2$. If $\|y - x\| \leq r$, then $\langle x - u, y - u \rangle \geq r^2 - r\|y - x\| \geq 0$, and the subgradient inequality yields $f(y) \geq f(u)$. Hence $u \in \text{brox}_f^r(x)$, proving (67) and (68). Conversely, let u be a nonterminal radius- r BPM output for f from x . By Theorem 2.1, there is $a > 0$ such that $g = a(x - u) \in \partial f(u)$, with $\|x - u\| = r$ and $\delta = f(x) - f(u) > 0$. Set $\lambda = (1 + a\delta)^{-1}$ and $c = a(1 + a\delta)^{-1}$. Then $(\lambda g, 1 - \lambda) = c(x - u, \delta)$ belongs to $\partial F(u, f(u))$ by (90). Since this subgradient is radial from $(x, f(x))$ to $(u, f(u))$, it certifies an exact lifted step of radius $\sqrt{r^2 + \delta^2}$. Finally, consider a terminal lifted step of radius t . Its output has the form $(u, f_\star)$ with $u \in X^*$, and feasibility gives $\|x - u\|^2 + \Delta^2 \leq t^2$. Thus u is an exact terminal output for f at radius $\sqrt{t^2 - \Delta^2} \geq D$. Conversely, any terminal radius- r output $u \in X^*$ for f satisfies $\|x - u\| \leq r$, so $(u, f_\star)$ is feasible for the lifted ball of radius $\sqrt{r^2 + \Delta^2}$ and is therefore a terminal lifted output. This proves the terminal correspondences and completes the proof. $\square$

C.21 Proof of Theorem 5.3

Proof. Since $\arg \min F = X^\star \times \{f_\star\}$ and $F(x, f(x)) = f(x)$, the gap and distance identities in (69) follow immediately. Applying Theorems 2.5 and 3.4 to F with initial distance R_0 gives (70). Suppose first that $\partial f(x) \neq \emptyset$. By (90), for fixed $\lambda > 0$ the minimum squared norm of a lifted subgradient $(\lambda g, 1 - \lambda)$ is $\lambda^2 m(x)^2 + (1 - \lambda)^2$. Minimizing over $\lambda \geq 0$ gives

$$\lambda = \frac{1}{1 + m(x)^2}, \quad \min_{\lambda \geq 0} [\lambda^2 m(x)^2 + (1 - \lambda)^2] = \frac{m(x)^2}{1 + m(x)^2}.$$

The branch $\lambda = 0$ has squared norm $\|v\|^2 + 1 \geq 1$, while $(0, 1) \in \partial F(x, f(x))$. If $\partial f(x) = \emptyset$, no branch with $\lambda > 0$ is available, so the minimum subgradient norm of F is one. This proves (71). Solving the resulting inequality for $m(x)$ gives the stated inverse estimate. For the paired subgradients, $G = (\lambda g, 1 - \lambda)$ and $\mathcal{G}_{x^\star}(u, g) = \langle g, u - x^\star \rangle$, so

$$\langle G, (u, f(u)) - (x^\star, f_\star) \rangle \stackrel{(68), (54)}{=} \lambda \mathcal{G}_{x^\star}(u, g) + (1 - \lambda)(f(u) - f_\star), \quad (91)$$

which is (72). Since the primal gap is $f(u) - f_\star$, subtracting it from the lifted symmetric Bregman distance gives

$$\begin{aligned} D_F^G((x^\star, f_\star), (u, f(u))) &\stackrel{(91), (69)}{=} \lambda (\mathcal{G}_{x^\star}(u, g) - (f(u) - f_\star)) \\ &\stackrel{(53), (52)}{=} \lambda D_f^g(x^\star, u). \end{aligned}$$

□

C.22 Proof of Theorem 5.4

Proof. On an infinite nonterminal corresponding run, $r_k \leq t_k = \sqrt{r_k^2 + \delta_k^2} \leq r_k + \delta_k$. Thus $\sum_k t_k < \infty$ immediately implies $\sum_k r_k < \infty$. Conversely, suppose $\sum_k r_k < \infty$. Then (x^k) converges and is contained in a compact set. Since f is proper and closed, it is bounded below on this set. As the values are nonincreasing, they converge to a finite limit, and hence $\sum_k \delta_k = f(x^0) - \lim_k f(x^k) < \infty$. The preceding inequality then gives $\sum_k t_k < \infty$. Apply Theorem 4.1 to the projected BPM run with radii r_k . If $\sum_k t_k = \infty$, an infinite nonterminal run has $\sum_k r_k = \infty$. Under attainment this is impossible, so finite termination occurs; without attainment, $f(x^k) \downarrow \inf f$ and $\|x^k\| \rightarrow \infty$. If $\sum_k t_k < \infty$, then $\sum_k r_k < \infty$, and an infinite projected run satisfies $x^k \rightarrow \bar{x} \in \text{dom } f$ and $f(x^k) \downarrow f(\bar{x})$. Finite termination gives the corresponding conclusions immediately. Under attainment, Theorem 4.5 gives $\sum_{k < K} r_k \leq 2C_d D_0$ for every sequence of nonterminal steps. Summing $t_k \leq r_k + \delta_k$ and using $\sum_{k < K} \delta_k = \Delta_0 - \Delta_K$ gives (73); the strict inequality follows from $\Delta_K > 0$. Let $B = 2C_d D_0 + \Delta_0$. For a constant lifted radius t , if the first $M = \lceil B/t \rceil$ iterations were all nonterminal, then (73) would give $Mt < B$, contradicting $Mt \geq B$. This gives the second bound in (74), while the first follows from (70). Finally, under nearest-minimizer selection at termination for F , projection onto $\arg \min F = X^\star \times \{f_\star\}$ selects the nearest original minimizer in the horizontal coordinate. By Theorem 4.9, the total projected length is at most $2C_d D_0$. The total vertical variation is at most Δ_0 , so the triangle inequality bounds the total lifted trajectory length by $2C_d D_0 + \Delta_0$. □

C.23 Verification of Theorem 5.5

Proof. For $x > 0$, a radius- t BPM step gives $(x - t)_+$, so $N_t = \lceil D_0/t \rceil$. For the lifted method, Theorem 5.2 shows that every nonterminal output projects to $u = x - r > 0$ for some $r > 0$. Since $f(x) - f(u) = ar$, (67) gives $t^2 = (1 + a^2)r^2$, hence $r = t/\sqrt{1 + a^2}$. A lifted step is terminal exactly when its ball contains $(0, 0)$, that is, when $t \geq \sqrt{1 + a^2}x$. Thus each nonterminal

step advances a lifted distance t along the segment from (D_0, aD_0) to $(0, 0)$, whose length is $\sqrt{1+a^2} D_0$, and therefore

$$N_t^{\text{epi}} = \left\lceil \frac{\sqrt{1+a^2} D_0}{t} \right\rceil.$$

Since N_t is independent of a , letting $a \rightarrow \infty$ gives the asserted unbounded separation. $\square$

C.24 Proof of Theorem 6.1

Proof. By monotonicity of ∂f , $\langle h^k - g^k, x^k - x^{k+1} \rangle \geq 0$. Since $x^k - x^{k+1} = \lambda_{k+1} g^k$ with $\lambda_{k+1} > 0$, this gives $\langle h^k - g^k, g^k \rangle \geq 0$, and hence $\|g^k\| \leq \|h^k\|$. Therefore

$$\lambda_{k+1} \stackrel{(20)}{=} \frac{t_k}{\|g^k\|} = \tau \frac{\|h^k\|}{\|g^k\|} \geq \tau. \quad (92)$$

By (20), every nonterminal update satisfies $x^{k+1} = \text{prox}_{\lambda_{k+1}f}(x^k)$. The tight proximal-point estimate of Taylor et al. (2017, Theorem 4.1) therefore gives

$$\Delta_K \leq \frac{D_0^2}{4 \sum_{k=0}^{K-1} \lambda_{k+1}} \stackrel{(92)}{\leq} \frac{D_0^2}{4\tau K}.$$

After termination, $\Delta_K = 0$ by Assumption 3 (Exact BPM iteration), so the final bound remains valid for the stopped sequence.

Sharpness. Fix $D > 0$, $\tau > 0$, and an integer $K \geq 1$, and take $f(x) = a|x|$ on $\mathbb{R}$, with $x^0 = D$ and $a = D/(2\tau K)$. The minimizer is 0, so $D_0 = D$. For $0 \leq k < K$, the unique subgradient at the positive center is $h^k = a$, and the prescribed radius is $t_k = \tau a = D/(2K)$. Induction gives $x^k = D - kD/(2K)$ for $0 \leq k \leq K$, so all K iterations are nonterminal and $x^K = D/2$. Their radial subgradients are $g^k = a$, hence

$$\lambda_{k+1} \stackrel{(20)}{=} \frac{t_k}{|g^k|} = \tau, \quad \Delta_K = \frac{aD}{2} = \frac{D_0^2}{4\tau K} = \frac{D_0^2}{4 \sum_{k=0}^{K-1} \lambda_{k+1}}.$$

Thus both upper bounds are attained for every prescribed horizon K . $\square$

C.25 Proof of Theorem 6.2

Proof. If termination occurs, the claim is immediate. Before termination, $t_k < D_k \leq D_0$, so Theorem 3.1 gives

$$\Delta_{k+1} \stackrel{(29)}{\leq} \Delta_k \left(1 - \frac{t_k}{D_k}\right) \leq \Delta_k(1 - a\Delta_k^\alpha), \quad a = \tau/D_0, \quad (93)$$

with $0 < a\Delta_k^\alpha < 1$. Since $(1 - z)^{-\alpha} \geq 1 + \alpha z$ for $0 \leq z < 1$, we obtain

$$\Delta_{k+1}^{-\alpha} \stackrel{(93)}{\geq} \Delta_k^{-\alpha} (1 - a\Delta_k^\alpha)^{-\alpha} \geq \Delta_k^{-\alpha} + \alpha a.$$

Summing over $k = 0, \dots, K-1$ gives $\Delta_K^{-\alpha} \geq \Delta_0^{-\alpha} + \alpha\tau K/D_0$, and inversion yields (75). Solving this bound for $\Delta_K \leq \varepsilon$ gives the stated iteration count. $\square$

C.26 Proof of Theorem 6.3

Proof. Since $\ell \leq f_\star$, we have $f(x^k) - \ell \geq \Delta_k$, and hence $t_k \geq \tau \Delta_k^\alpha$. If a step is terminal, the objective-gap bound is immediate. Otherwise, Theorem 3.1 gives the same upper recurrence as in the proof of Theorem 6.2, so (75) follows without comparing the two trajectories. If $\ell < f_\star$, then every nonoptimal iterate satisfies $t_k \geq \tau(f_\star - \ell)^\alpha > 0$. Thus, if the first K iterations were all nonterminal, Theorem 2.5 would give $K\tau^2(f_\star - \ell)^{2\alpha} < D_0^2$. Therefore, among the first

$$\left\lceil \frac{D_0^2}{\tau^2(f_\star - \ell)^{2\alpha}} \right\rceil$$

iterations, at least one must be terminal. $\square$

C.27 Proof of Theorem 7.1

Proof. Let $p = T_t(x)$ and $d = x - p$. For every $z \in X^\star$, the radial subgradient in the nonterminal case and the projection property in the terminal case give

$$\langle d, p - z \rangle \geq 0. \quad (94)$$

Since $R_{t,\lambda}(x) = x - \lambda d$ and $x - z = d + (p - z)$,

$$\|R_{t,\lambda}(x) - z\|^2 \stackrel{(94)}{\leq} \|x - z\|^2 - \lambda(2 - \lambda) \|d\|^2.$$

Because $\|d\| = \min\{t, D(x)\}$, this is (76). Summing (76) along the iterates gives $\min\{t, D(x^k)\} \rightarrow 0$, and hence $D(x^k) \rightarrow 0$. The sequence is Fejér monotone with respect to $X^\star$ and therefore bounded. Any cluster point belongs to $X^\star$ because $D(x^k) \rightarrow 0$ and $X^\star$ is closed. Let $\bar{x} \in X^\star$ be such a cluster point. Then $\|x^k - \bar{x}\|$ is nonincreasing by (76), while a subsequence converges to $\bar{x}$. Thus $\|x^k - \bar{x}\| \rightarrow 0$, so the whole sequence converges to $\bar{x}$. Finally, if $D(x) \leq t$, then $p = \text{Proj}_{X^\star}(x)$ and

$$D(R_{t,\lambda}(x)) \leq \|R_{t,\lambda}(x) - p\| = |1 - \lambda|D(x),$$

which proves the final estimate. $\square$

C.28 Proof of Theorem 2.3

Proof. Expanding $x^k - z = (x^k - u) + (u - z)$ gives the cross term

$$\begin{aligned} \langle x^k - u, u - z \rangle &\stackrel{(19)}{=} c_k^{-1} \langle g^k, u - z \rangle \\ &\stackrel{(21)}{=} c_k^{-1} [f(u) - f(z) + D_f^{g^k}(z, u)]. \end{aligned}$$

which yields (22). Evaluating the definition of the Bregman distance at $z = x^k$ and using $\langle g^k, x^k - u \rangle = c_k t_k^2 = t_k \|g^k\|$ gives (23). Finally, if $f(z) \leq f(u)$, then $f(u) - f(z) \geq 0$ and $D_f^{g^k}(z, u) \geq 0$, so (24) follows directly from (22). $\square$

C.29 Proof of Theorem 2.5

Proof. For a nonterminal step, take $z = \text{Proj}_{X^\star}(x^k)$ in (24). Since $z \in X^\star$, minimizing the resulting distance over $X^\star$ gives $D_{k+1}^2 \leq D_k^2 - t_k^2$. If the first K iterations were all nonterminal, telescoping would yield $0 < D_K^2 \leq D_0^2 - \sum_{k < K} t_k^2$. Hence $\sum_{k < K} t_k^2 \geq D_0^2$ forces a terminal step among the first K iterations. For $t_k = t$, taking $K = \lceil D_0^2/t^2 \rceil$ gives (26). Finally, before termination the same telescoping argument applies, while at and after termination $D_K = 0$, which gives the stated bound for the stopped sequence. $\square$

C.30 Proof of Theorem 2.7

Proof. Suppose first that u is a nonterminal output, and let $g = c(x - u) \in \partial f(u)$ be its radial subgradient. Since $\|x - u\| = t$, we have $(x - u)/t = g/\|g\|$. For any $p \in \partial f(x)$, monotonicity of ∂f gives $\langle p - g, x - u \rangle \geq 0$, and hence $\langle p, (x - u)/t \rangle \geq \|g\| \geq m(u)$. By Cauchy–Schwarz, $\|p\| \geq m(u)$. Taking the infimum over $p \in \partial f(x)$ yields $m(x) \geq m(u)$. If u is terminal, then u is a global minimizer, so $0 \in \partial f(u)$ and therefore $m(u) = 0$. Thus $m(u) \leq m(x)$ in all cases. $\square$

C.31 Proof of Theorem 3.6

Proof. Suppose first that the K -th output is nonterminal, and let $g^{K-1} \in \partial f(x^K)$ be its radial subgradient. By (23),

$$t_{K-1} \|g^{K-1}\| \stackrel{(23)}{\leq} \Delta_{K-1} - \Delta_K. \quad (95)$$

Since $m(x^K) \leq \|g^{K-1}\|$ and $\Delta_K \geq 0$, division by $t_{K-1} > 0$ gives

$$m(x^K) \stackrel{(95)}{\leq} \frac{\Delta_{K-1} - \Delta_K}{t_{K-1}} \leq \frac{\Delta_{K-1}}{t_{K-1}},$$

which is (46). If a terminal output has already occurred, then $0 \in \partial f(x^K)$ and hence $m(x^K) = 0$, so the same inequalities hold. Now let $t_k = t$ with $0 < t < D_0$. By (30),

$$\Delta_{K-1} \stackrel{(30)}{\leq} \left(1 - \frac{t}{D_0}\right)^{K-1} \Delta_0.$$

Substituting this into (46) gives (47). Finally, if f is differentiable, $\partial f(x^K) = \{\nabla f(x^K)\}$, so $m(x^K) = \|\nabla f(x^K)\|$. $\square$

D Geometric radius schedules and the minimum successful factor

Theorems 4.1 and 4.6 leave the behavior of summable radius schedules problem dependent. As Theorem 4.10 shows, even for the same objective and starting point, one summable schedule may attain a minimizer, another may converge to one only asymptotically, and another may converge to a nonoptimal point. We now examine how these possibilities arise within the one-parameter family

$$t_k = t\rho^k, \quad 0 < \rho < 1.$$

For fixed f , x^0 , and t , only the decay factor ρ varies. The question is whether there is a smallest value of ρ for which the trajectory still converges to a minimizer. Write

$$\mathcal{T}_k(\rho) = \sum_{j=k}^{\infty} t\rho^j = \frac{t\rho^k}{1-\rho}$$

for the total radius remaining after iteration k . Call $\rho \in (0, 1)$ successful if the corresponding trajectory converges to a point in X^* . The theorem below shows that the successful factors have a minimum ρ_* . At this minimum successful factor, the trajectory converges to a minimizer without finite termination, and ρ_* can be characterized through finite prefixes.

Theorem D.1 (Minimum successful geometric factor). *Under Assumption 1 (Convex objective), Assumption 2 (Attainment), and Assumption 3 (Exact BPM iteration), fix $x^0 \in \text{dom } f$ and $t > 0$, and let $D_k(\rho) = \text{dist}(x^k(\rho), X^*)$. If $t \geq D_0$, every $\rho \in (0, 1)$ attains*

a minimizer in at most one step, so there is no smallest positive successful factor. Suppose $0 < t < D_0$. The first output x^1 is independent of ρ . Set $d_1 = \text{dist}(x^1, X^\star) > 0$ and

$$a = \frac{d_1}{t + d_1}, \quad b = \frac{d_1}{\sqrt{t^2 + d_1^2}}.$$

Then:

(i) The successful factors have a minimum $\rho_\star$, with

$$0 < a \leq \rho_\star < b < 1. \quad (96)$$

Every $\rho \in [b, 1)$ attains a minimizer finitely, whereas every $\rho < \rho_\star$ converges to a nonoptimal limit.

(ii) The trajectory at $\rho_\star$ never attains a minimizer finitely, but converges to some $\bar{x} \in X^\star$. For every $k \geq 0$,

$$\frac{t\rho_\star^k}{2C_d(1 - \rho_\star)} \leq D_k(\rho_\star) \leq \|x^k(\rho_\star) - \bar{x}\| \leq \frac{t\rho_\star^k}{1 - \rho_\star}. \quad (97)$$

Consequently,

$$D_k(\rho_\star)^{1/k} \rightarrow \rho_\star, \quad \|x^k(\rho_\star) - \bar{x}\|^{1/k} \rightarrow \rho_\star.$$

If $M = \|g^0\|$ for any radial subgradient from the first step, then $\Delta_k \leq Mt\rho_\star^k/(1 - \rho_\star)$ for every $k \geq 1$.

(iii) A factor ρ is successful if and only if

$$D_k(\rho) \leq \mathcal{T}_k(\rho) \quad \text{for every } k \geq 1. \quad (98)$$

For $K \geq 1$, define

$$A_K = \{\rho \in [a, b] : D_j(\rho) \leq \mathcal{T}_j(\rho), \ 1 \leq j \leq K\}.$$

Each A_K is nonempty and compact, and

$$\rho_\star = \lim_{K \rightarrow \infty} \min A_K, \quad \min A_K \text{ is nondecreasing in } K. \quad (99)$$

The proof is given in [Section D.1](#). The finite-prefix characterization is exact but not intended as an efficient computational procedure, since evaluating A_K requires distances to $X^\star$ and a global scalar search over ρ . In particular, the theorem does not assert that the successful factors form an interval. If ρ fails, then some finite prefix satisfies $D_k(\rho) > \mathcal{T}_k(\rho)$, certifying that the remaining radius budget is insufficient to reach any minimizer. At $\rho_\star$, every step is nonterminal, so the remaining trajectory length equals $\mathcal{T}_k(\rho_\star)$ and (97) gives the asymptotic convergence rate. For $f(x) = |x|$ with $x^0 = D_0 > t$, direct summation gives $\rho_\star = 1 - t/D_0 = a$. In this example, any decrease below $\rho_\star$ destroys convergence to the minimizer.

The dependence on trajectory geometry can be seen by comparing two objectives with the same D_0/t . For the quadratic example in [Section 1.1](#), take $f(z) = \frac{1}{2}(z_1^2 + 4z_2^2)$, $x^0 = (2, 5/4)$, and $t = \sqrt{2}$. The first output is $x^1 = (1, 1/4)$, so $D_0 = \sqrt{89}/4$ and $d_1 = \sqrt{17}/4 > D_0 - t$. Hence (96) gives

$$\rho_\star \stackrel{(96)}{\geq} \frac{d_1}{t + d_1} > 1 - \frac{t}{D_0}.$$

For the objective $f(z) = \|z\|$ at the same starting point and radius, the trajectory is radial and $\rho_\star = 1 - t/D_0$. Thus the two problems have the same initial distance and radius but different minimum successful factors.

D.1 Proof of Theorem D.1

Proof. If $t \geq D_0$, the initial ball meets $X^\star$, so the first step is terminal for every $\rho \in (0, 1)$. Hence assume $0 < t < D_0$. We may use the nearest-minimizer selection at termination, since it leaves all nonterminal iterates and the first terminal step unchanged. Write $x^k(\rho)$ for the resulting stopped trajectory and $z(\rho) = \lim_k x^k(\rho)$, which exists by Theorem 4.8. Since each displacement is at most its radius,

$$\|z(\rho) - x^k(\rho)\| \leq \mathcal{T}_k(\rho). \quad (100)$$

Continuity in ρ . We first show continuity of a single selected step. Let $y_n \rightarrow y$, $r_n \rightarrow r > 0$, and $f(y_n) \leq f(x^0)$. Suppose first that $D(y) \geq r$. The limiting ball problem has a unique selected output u : the step is nonterminal if $D(y) > r$, while at $D(y) = r$ uniqueness follows from Theorem 2.1. Let u_n denote the selected outputs for (y_n, r_n) . Since $\|u_n - y_n\| \leq r_n$, the sequence (u_n) is bounded. For fixed $\epsilon \in (0, 1)$, set $v_n = (1 - \epsilon)u + \epsilon y_n$. Since $\|u - y\| = r$, we have $\|v_n - y_n\| \rightarrow (1 - \epsilon)r < r$, so v_n is feasible for all sufficiently large n . Hence

$$f(u_n) \leq f(v_n) \leq (1 - \epsilon)f(u) + \epsilon f(x^0).$$

If a subsequence u_{n_j} converges to $\bar{u}$, then $\bar{u}$ is feasible for the limiting ball and lower semicontinuity gives $f(\bar{u}) \leq (1 - \epsilon)f(u) + \epsilon f(x^0)$. Letting $\epsilon \downarrow 0$ yields $f(\bar{u}) \leq f(u)$. Since u minimizes the limiting ball problem, $\bar{u}$ is also a minimizer, and uniqueness gives $\bar{u} = u$. Thus $u_n \rightarrow u$. If $D(y) < r$, then $D(y_n) < r_n$ for all sufficiently large n , so the nearby steps are terminal and their selected outputs are $\text{Proj}_{X^\star}(y_n) \rightarrow \text{Proj}_{X^\star}(y)$. Thus the selected step is continuous in its center and radius. Induction now shows that $x^k(\rho)$ is continuous in ρ for every fixed k . On any compact interval $[c, d] \subset (0, 1)$, (100) gives

$$\|z(\rho) - x^k(\rho)\| \stackrel{(100)}{\leq} \frac{td^k}{1 - d},$$

uniformly in $\rho \in [c, d]$. Hence z is continuous as a locally uniform limit of the continuous maps x^k . Since $X^\star$ is closed, the successful set $\{\rho \in (0, 1) : z(\rho) \in X^\star\}$ is relatively closed in $(0, 1)$.

The set of factors yielding finite termination is open. Suppose $x^n(\rho) \in X^\star$. Since $D(x^n(\rho)) = 0 < t\rho^n$, continuity of $x^n(\rho)$, of the distance to $X^\star$, and of $t\rho^n$ implies that, for every sufficiently nearby ρ' , $D(x^n(\rho')) < t(\rho')^n$. The next step is therefore terminal. Hence the set of factors that attain a minimizer finitely is open.

Bounds on the minimum successful factor. If ρ is successful, then (100) at $k = 1$ gives

$$d_1 \stackrel{(100)}{\leq} \frac{t\rho}{1 - \rho},$$

and therefore $\rho \geq a$. For $\rho \geq b$,

$$\sum_{k=1}^{\infty} t^2 \rho^{2k} = \frac{t^2 \rho^2}{1 - \rho^2} \geq d_1^2.$$

Suppose, to the contrary, that every step after x^1 is nonterminal. By (33), the step from x^1 satisfies

$$D_2^2 \stackrel{(33)}{\leq} d_1^2 - t^2 \rho^2 - \frac{2\Delta_2}{c_1}, \quad \Delta_2 > 0. \quad (101)$$

Each later nonterminal step decreases squared distance by at least its squared radius. Hence for every $N \geq 2$,

$$D_N^2 \stackrel{(101),(24)}{\leq} d_1^2 - \sum_{k=1}^{N-1} t^2 \rho^{2k} - \frac{2\Delta_2}{c_1}.$$

Since the infinite sum is at least d_1^2 , the right-hand side is negative for all sufficiently large N , a contradiction. Thus every $\rho \in [b, 1)$ attains a minimizer finitely. The successful set is therefore nonempty, and its intersection with $[a, b]$ is compact. Let $\rho_\star$ be its minimum. Since the set of factors yielding finite termination is open and the iteration terminates finitely at b , $\rho_\star < b$. The same openness shows that $\rho_\star$ itself cannot attain finitely, since otherwise some smaller nearby factor would also be successful. Thus every step at $\rho_\star$ is nonterminal, while $z(\rho_\star) \in X^\star$. Every $\rho < \rho_\star$ is unsuccessful and therefore has a nonoptimal limit.

Rate at the minimum successful factor. Let $\bar{x} = z(\rho_\star)$. Since every step at $\rho_\star$ is nonterminal, each displacement has length $t\rho_\star^j$, so the remaining trajectory length from $x^k(\rho_\star)$ is exactly $\mathcal{T}_k(\rho_\star)$. By Theorem 4.9,

$$\mathcal{T}_k(\rho_\star) \stackrel{(64)}{\leq} 2C_d D_k(\rho_\star).$$

Since $\bar{x} \in X^\star$, (100) also gives

$$D_k(\rho_\star) \leq \|x^k(\rho_\star) - \bar{x}\| \stackrel{(100)}{\leq} \mathcal{T}_k(\rho_\star).$$

Together these inequalities give (97), and taking k th roots yields the two root-rate limits. The objective-gap estimate follows from (63) with the optimal limit $\bar{x}$.

Finite-prefix characterization. If ρ is successful, then $z(\rho) \in X^\star$, so (100) gives $D_k(\rho) \leq \mathcal{T}_k(\rho)$ for every $k \geq 1$. Conversely, if these inequalities hold for every k , then $\mathcal{T}_k(\rho) \rightarrow 0$ implies $D_k(\rho) \rightarrow 0$. Since $x^k(\rho) \rightarrow z(\rho)$ and $X^\star$ is closed, $z(\rho) \in X^\star$, so ρ is successful. For each K , continuity of the finite iterates makes A_K compact, and $b \in A_K$, so A_K is nonempty. Moreover, $A_{K+1} \subseteq A_K$. Hence $m_K := \min A_K$ is nondecreasing and converges to some $m_\infty \leq \rho_\star$. For every fixed j , $m_K \in A_j$ whenever $K \geq j$, so closedness of A_j gives $m_\infty \in A_j$. Thus m_∞ belongs to every A_j , and the preceding characterization shows that m_∞ is successful. By minimality of $\rho_\star$, $m_\infty \geq \rho_\star$. Therefore $m_\infty = \rho_\star$, proving (99). $\square$

E Nonmonotonicity of the fixed-radius termination count

The upper bounds on N_t improve as the fixed radius t increases, but this does not imply that the actual termination count is nonincreasing in t . Changing the radius also changes the trajectory. The following counterexample shows that a larger radius can produce a lower objective value after the first step while placing the next iterate farther from the solution set, resulting in an additional step.

Proposition E.1 (The termination count need not decrease with the radius). *There exist a finite, continuous, coercive convex function on $\mathbb{R}^2$ with a unique minimizer, a starting point x^0 , and radii $0 < s < t$ such that*

$$N_s = 2, \quad N_t = 3.$$

Moreover, the first radius- t step attains a lower objective value than the first radius- s step.

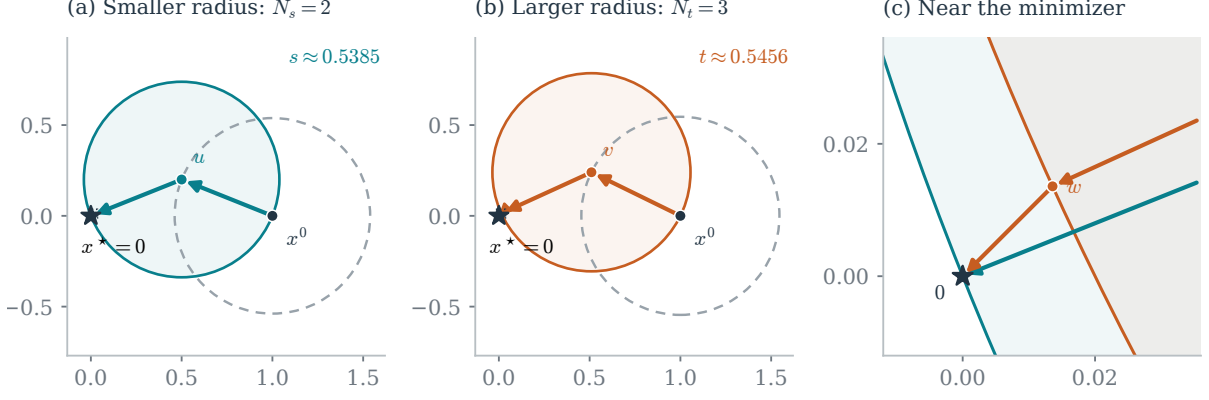


Figure 5: Two fixed-radius BPM trajectories from the same starting point. For radius s , the path is $x^0 \rightarrow u \rightarrow 0$; for radius t , it is $x^0 \rightarrow v \rightarrow w \rightarrow 0$. Dashed circles show the first balls and colored circles the second balls. Panels (a)–(b) use identical scales, while panel (c) magnifies the neighborhood of the minimizer: zero lies on the smaller second ball but outside the larger second ball. The third ball centered at w is omitted.

Proof. Take $x^0 = (1, 0)$ and

$$f(x, y) = \max \left\{ \frac{|x|}{100}, \frac{|y|}{100}, x - \frac{2}{5}y - \frac{37}{100}, \frac{49}{100}x - \frac{6}{25}y - \frac{1473}{10000} \right\}. \quad (102)$$

Since $f(x, y) \geq \max\{|x|, |y|\}/100$ and $f(0, 0) = 0$, the function is finite, continuous, coercive, and convex, with unique minimizer $(0, 0)$. Set

$$s = \frac{\sqrt{29}}{10}, \quad t = \frac{\sqrt{2977}}{100},$$

so that $0 < s < t$.

For radius s , direct calculation gives the first output $u = (1/2, 1/5)$. Indeed, the third affine term in (102) is uniquely active at u , with $f(u) = 1/20$ and $\nabla f(u) = (1, -2/5) = 2(x^0 - u)$, while $\|x^0 - u\| = \|u\| = s$. Thus Theorem 2.1 verifies the first step, and the second radius- s ball contains the minimizer. Hence $N_s = 2$.

For radius t , the first output is similarly $v = (51/100, 6/25)$. The fourth affine term is uniquely active at v , with $f(v) = 9/200$ and $\nabla f(v) = (49/100, -6/25) = x^0 - v$, and $\|x^0 - v\| = t$. Hence Theorem 2.1 verifies the first radius- t step. However, $\|v\|^2 = 3177/10000 > t^2$, so the second step is nonterminal. If w is its output, Theorem 2.5 gives $\|w\|^2 \leq \|v\|^2 - t^2 = 1/50 < t^2$. Therefore the third ball contains the minimizer, and $N_t = 3$.

Finally, $f(v) = 9/200 < 1/20 = f(u)$. Thus the larger radius gives a better first objective value but requires an additional step to attain the minimizer. $\square$

For completeness, the second output of the radius- t trajectory shown in Figure 5 is $w = \alpha(1, 1)$ with $\alpha = (15 - \sqrt{209})/40$.

Table 3: Comparison of the first BPM steps in Figure 5. The larger radius gives a lower objective value but leaves its first output farther than one radius from the minimizer.

Quantity	Smaller radius s	Larger radius t
Radius	$\sqrt{29}/10 \approx 0.5385$	$\sqrt{2977}/100 \approx 0.5456$
First output	$u = (1/2, 1/5)$	$v = (51/100, 6/25)$
Objective value	$f(u) = 1/20 = 0.050$	$f(v) = 9/200 = 0.045$
Distance to the minimizer	$\ u\ = s$	$\ v\ = \sqrt{3177}/100 \approx 0.5636 > t$
Second ball contains the minimizer	Yes	No
Iterations to termination	$N_s = 2$	$N_t = 3$

F Key concepts

Table 4 collects the named concepts used in the paper, grouped by their role. The last column links to the first mention (including the abstract and overview table) and, where different, to the definition or precise formulation. The notation index is in [Section A](#).

Table 4: Key concepts: definitions and a guide to their introduction.

Concept	Definition or meaning in this paper	First appearance; formulation
1. Exact steps and equivalent formulations		
Ball-proximal point method (BPM)	Each step minimizes f over the closed Euclidean ball centered at the current iterate: $x^{k+1} \in \text{brox}_f^{t_k}(x^k)$. The brox operator is the set of these minimizers.	Abstract , p. 1; update (BPM).
Exact ball oracle	Given $x \in \text{dom } f$ and $t > 0$, returns a minimizer of f on $\mathbb{B}(x, t)$. Each BPM iteration uses one oracle call.	§1 , p. 1; definition after the update .
Terminal and nonterminal steps	A step is terminal if its output belongs to X^* , equivalently if its ball meets X^* ; otherwise it is nonterminal. A nonterminal output is unique and lies on the ball boundary.	§1 , p. 2; Lem. 2.1 .
Radial subgradient	A subgradient $g \in \partial f(u)$ of the form $g = c(x - u)$, $c > 0$, for a nonterminal step from x to u . Its norm need not be the minimum subgradient norm.	Prop. 1.1(ii) , p. 4; (19).
Proximal map	$\text{prox}_{\lambda f}(x)$ is the unique minimizer of $f(v) + \ v - x\ ^2 / (2\lambda)$, $\lambda > 0$. A nonterminal BPM step is proximal with $\lambda = t / \ g\ $.	Prop. 1.1(ii) , p. 4; (5).
Ball envelope	The value function $F_t(x) = \inf_{\ u-x\ \leq t} f(u)$. It is convex but need not be differentiable. A BPM step is a normalized explicit subgradient step on this envelope.	§1.1 , p. 2; (2).
Sublevel-set projection	A nonterminal output satisfies $u = \text{Proj}_{\{v: f(v) \leq f(u)\}}(x)$ and $\ x - u\ = t$. The level is determined by the ball problem.	§1.1 , p. 2; Prop. 1.1(iv) .
Fenchel-dual formulation	The ball value equals $\max_g \{\langle x, g \rangle - f^*(g) - t \ g\ \}$, where $f^*(g) = \sup_v \{\langle g, v \rangle - f(v)\}$ is the convex conjugate.	§1.1 , p. 2; Prop. 1.1(v) .
Lagrangian dual formulation	Maximize $q_t(\gamma) = \inf_v \{f(v) + \frac{\gamma}{2}(\ v - x\ ^2 - t^2)\}$ over $\gamma \geq 0$. A positive optimal multiplier yields the nonterminal proximal output.	§1.1 , p. 2 (scalar dual); Prop. 1.1(vi) .

Continued on the next page

Table 4. Key concepts (continued)

Concept	Definition or meaning in this paper	First appearance; formulation
2. Error measures and convergence bounds		
Objective gap	$\Delta_k = f(x^k) - f_*$ measures function-value error when the minimum is attained.	Abstract, p. 1; notation in §1.
Distance to the solution set	$D_k = \text{dist}(x^k, X^*)$ is the distance to a nearest minimizer, rather than necessarily to a fixed reference minimizer.	Abstract, p. 1; notation in §1.
Minimum subgradient norm; stationarity	$m(x) = \text{dist}(0, \partial f(x))$, with $m(x) = +\infty$ if $\partial f(x) = \emptyset$. For convex f , $m(x) = 0$ characterizes a minimizer.	Abstract, p. 1; Cor. 2.7.
Generalized Bregman distance	$D_f^g(z, u) = f(z) - f(u) - \langle g, z - u \rangle$ for a specified $g \in \partial f(u)$. It is nonnegative, depends on the chosen subgradient, and is not generally a metric.	Prop. 1.1(v), p. 4 (for f^*); (21).
Symmetric Bregman distance to a minimizer	$\mathcal{G}_{x^*}(u, g) = D_f^g(x^*, u) + D_f^0(u, x^*) = \langle g, u - x^* \rangle$, using $g \in \partial f(u)$ and $0 \in \partial f(x^*)$.	Abstract, p. 1; Prop. 3.8.
Fenchel–Young gap	$f(u) + f^*(g) - \langle u, g \rangle$. It is zero whenever $g \in \partial f(u)$, so for radial subgradients it does not measure progress toward a minimizer.	§3.4, p. 18 (interpretation).
Finite convergence; finite termination	Some finite iterate lies in X^* . Under the stopped-sequence convention, the method then terminates. This property does not supply an oracle test for detecting optimality.	Abstract, p. 1; definition in §1.
Segment contraction	Comparing with a feasible point on the segment to a nearest minimizer gives $\Delta_{k+1} \leq (1 - t_k/D_k)_+ \Delta_k$ when $D_k > 0$.	§1.3, p. 7; Thm. 3.1.
Jensen objective-gap bound	For K nonterminal constant-radius steps and $S = D_0^2/t^2$, $\Delta_K/\Delta_0 \leq ((S - K)/(S + K))^K$. Jensen’s inequality combines the decreases in squared distance across steps.	§3.2, p. 14 (derivation); Thm. 3.4.

Continued on the next page

Table 4. Key concepts (continued)

Concept	Definition or meaning in this paper	First appearance; formulation
3. Trajectory geometry, radius rules, and reformulations		
Self-contracted sequence	$\ z^j - z^m\ \leq \ z^i - z^m\ $ for $i \leq j \leq m$: the distance to any later point cannot increase along earlier iterates.	Abstract, p. 1; Def. 4.2.
Finite trajectory length	The total displacement $\sum_{k=0}^{\infty} \ x^{k+1} - x^k\ $ is finite. Bounded self-contracted sequences have this property in finite dimension.	Abstract, p. 1; Lem. 4.4.
Summable and nonsummable radii	Positive radii are summable if $\sum_{k=0}^{\infty} t_k < \infty$, and nonsummable otherwise. The latter condition guarantees finite termination whenever a minimizer exists.	Abstract, p. 1; Thm. 4.1.
Nearest-minimizer selection at termination	Choose $\text{Proj}_{X^*}(x)$ when the ball meets X^* ; otherwise use the unique nonterminal output. This defines the single-valued selection T_t .	Table 1, p. 3; Def. 2.2.
Epigraph reformulation	Minimize $F(x, s) = s + \delta_{\text{epi } f}(x, s)$, where $\text{epi } f = \{(x, s) : f(x) \leq s\}$. Balls in the lifted space induce different radii in the original space.	Abstract, p. 1; (66).
Subgradient-based radius rule	Set $t_k = \tau \ h^k\ $ with $h^k \in \partial f(x^k)$ and $\tau > 0$; stop if $h^k = 0$. An available radial subgradient from the preceding step may be reused.	Abstract, p. 1; Thm. 6.1.
Objective-gap-based radius rule	Set $t_k = \tau \Delta_k^\alpha$ with $\tau, \alpha > 0$ and known f_* . The paper also analyzes replacement of f_* by a valid lower bound.	Abstract, p. 1; Thm. 6.2.
Minimum successful geometric factor	For $t_k = t\rho^k$, a factor is successful if the trajectory converges to a minimizer. When $0 < t < D_0$, the successful factors have a minimum; they are not asserted to form an interval.	Abstract, p. 1; Thm. D.1.
Relaxed BPM; cutter property	The update is $R_{t,\lambda}(x) = x + \lambda(T_t(x) - x)$, $0 < \lambda < 2$. The cutter property is $\langle x - T_t(x), z - T_t(x) \rangle \leq 0$ for $z \in X^*$.	Relaxation: Abstract, p. 1. Cutter: §1.3; §7.
Fejér monotonicity	$\ x^{k+1} - z\ \leq \ x^k - z\ $ for every $z \in X^*$. Unlike self-contraction, the reference points belong to a fixed set.	§1.3, p. 7 (Fejér-type decrease); (76).