R-Linear Convergence of Barzilai–Borwein Methods on Strictly Convex Quadratics via Closed-Cone Homogeneous Dynamics

Shu-tai Yang · Ya-xiang Yuan

Abstract We give a new proof of R-linear convergence for BB1, BB2, and fixed positive spectral-weight variants on finite-dimensional strictly convex quadratics. The delayed recurrence is written as a first-order system on the closed cone of compatible consecutive-gradient pairs. Its transition is continuous and positively homogeneous of degree one, including at finite termination. A compactness theorem shows that pointwise convergence of every cone orbit implies a uniform finite-step contraction and hence a global R-linear estimate. By a realization lemma, every compatible state initiates a BB1 trajectory, so Raydan's theorem yields pointwise stability on the entire cone. Finally, the transformation $\boldsymbol{g} \mapsto \mathbf{W}^{1/2}\boldsymbol{g}$ conjugates every fixed positive spectral-weight rule to BB1. Consequently, BB1, BB2, and all such weighted rules have the same homogeneous growth radius.

Keywords Barzilai–Borwein method · spectral gradient method · positively homogeneous dynamics · closed cone · R-linear convergence

Mathematics Subject Classification (2020) 90C25 · 65K05 · 37C75

Acknowledgements. Shu-tai Yang thanks Shixiang Chen for supervising his undergraduate thesis at the University of Science and Technology of China, and Xiaowei Xu for supervising his project under the National College Students Innovation and Entrepreneurship Training Program at the same university. The present article grew out of those two undergraduate projects.

Funding. The work of Shu-tai Yang was supported by the National College Students Innovation and Entrepreneurship Training Program, administered by the University of Science and Technology of China (Project No. 202510358091). The work of Ya-xiang Yuan was supported by the National Natural Science Foundation of China (Grant No. 12288201).

Shu-tai Yang (Corresponding author)
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China
School of Mathematical Sciences, University of Science and Technology of China, Hefei, China
E-mail: yangshutai26@mails.ucas.ac.cn

Ya-xiang Yuan
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
E-mail: yyx@lsec.cc.ac.cn

1 Introduction

Consider the strictly convex quadratic problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{H}\boldsymbol{x} - \boldsymbol{b}^\top \boldsymbol{x}, \qquad \mathbf{H} = \mathbf{H}^\top \succ 0, \tag{1.1}$$

with minimizer $\boldsymbol{x}_* = \mathbf{H}^{-1}\boldsymbol{b}$ and gradients $\boldsymbol{g}_k = \mathbf{H}(\boldsymbol{x}_k - \boldsymbol{x}_*)$. A gradient step satisfies

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{g}_k, \qquad \boldsymbol{g}_{k+1} = (\mathbf{I} - \alpha_k \mathbf{H})\boldsymbol{g}_k. \tag{1.2}$$

The two classical Barzilai–Borwein (BB) rules choose, on a quadratic,

$$\alpha_k^{\mathrm{BB1}} = \frac{\boldsymbol{g}_{k-1}^\top \boldsymbol{g}_{k-1}}{\boldsymbol{g}_{k-1}^\top \mathbf{H}\boldsymbol{g}_{k-1}}, \qquad \alpha_k^{\mathrm{BB2}} = \frac{\boldsymbol{g}_{k-1}^\top \mathbf{H}\boldsymbol{g}_{k-1}}{\boldsymbol{g}_{k-1}^\top \mathbf{H}^2\boldsymbol{g}_{k-1}}. \tag{1.3}$$

Their defining feature is the one-step delay: the stepsize applied to $\boldsymbol{g}_k$ is determined by $\boldsymbol{g}_{k-1}$.

Barzilai and Borwein introduced these rules and analyzed their two-dimensional behavior [1]. Raydan proved global convergence of both choices on finite-dimensional strictly convex quadratics. Dai and Liao established R-linear convergence in arbitrary dimension for BB1 and recorded the analogous BB2 conclusion [3, 5]. The latter proof strengthens a spectral-component induction into a uniform finite-window estimate. Related coordinatewise mechanisms underlie the bounded-retard framework of Friedlander, Martínez, Molina, and Raydan and the Property A analysis of Dai [2, 4].

In this paper, we give a new convergence analysis based on the geometry of an augmented state. The state is a pair of consecutive gradients. Its realizable values form the cone

$$(\boldsymbol{u}, \boldsymbol{v}), \qquad \boldsymbol{v} = (\mathbf{I} - \alpha\mathbf{H})\boldsymbol{u}, \qquad \alpha \in [\lambda_{\max}(\mathbf{H})^{-1}, \lambda_{\min}(\mathbf{H})^{-1}].$$

The cone is closed and forward invariant, and its only state with first component zero is the origin. On this cone the delayed BB update is continuous, including at finite termination, and positively homogeneous of degree one.

The abstract step is a finite-dimensional compactness principle: pointwise convergence of every orbit of a continuous homogeneous self-map on a closed cone is equivalent to a uniform R-linear estimate. This equivalence follows from Shen and Hu [6, Theorem 5.1] in the single-map, degree-one case, and a related growth-rate characterization is given by Tuna [7]. We give a proof of the version needed here. A realization lemma makes Raydan's theorem applicable from every compatible state, providing the pointwise input for BB1. For a fixed positive spectral weight $\mathbf{W} = \omega(\mathbf{H})$, the map $(\boldsymbol{u}, \boldsymbol{v}) \mapsto (\mathbf{W}^{1/2}\boldsymbol{u}, \mathbf{W}^{1/2}\boldsymbol{v})$ conjugates the weighted pair dynamics to BB1. This transfers one decay factor to BB2 and to every fixed weighted rule and identifies their homogeneous growth radii.

Section 2 presents the compactness argument and records its growth radius. Section 3 constructs the compatible cone, establishes the spectral conjugacy, and derives estimates for gradient pairs, iterates, and objective gaps. A complementary sharp-factor analysis appears in [8].

2 Homogeneous dynamics on closed cones

Let X be a finite-dimensional normed space. A nonempty set $\Omega \subset X$ is called a *closed cone* if it is closed, $0 \in \Omega$, and

$$\boldsymbol{x} \in \Omega, \quad t \geq 0 \quad \Longrightarrow \quad t\boldsymbol{x} \in \Omega.$$

A map $T : \Omega \to \Omega$ is positively homogeneous of degree one if

$$T(t\boldsymbol{x}) = tT(\boldsymbol{x}), \qquad t \geq 0, \quad \boldsymbol{x} \in \Omega. \tag{2.1}$$

Positive homogeneity implies $T(0) = 0$. Throughout, iteration indices are nonnegative integers, and T^0 denotes the identity map on Ω, so that $T^0(\boldsymbol{x}) = \boldsymbol{x}$.

Definition 2.1. The origin is *globally R-linearly stable* for T on Ω if there exist $C \geq 1$ and $\rho \in (0, 1)$ such that

$$\|T^k(\boldsymbol{x})\| \leq C\rho^k \|\boldsymbol{x}\|, \qquad \boldsymbol{x} \in \Omega, \quad k \geq 0. \tag{2.2}$$

Theorem 2.2 ([6, Theorem 5.1], single-map case)**.** *Let Ω be a closed cone in X, and let $T : \Omega \to \Omega$ be continuous and positively homogeneous of degree one. The following statements are equivalent.*

(i) *For every $\boldsymbol{x} \in \Omega$, one has $T^k(\boldsymbol{x}) \to 0$.*
(ii) *There exist $C \geq 1$ and $\rho \in (0, 1)$ for which (2.2) holds.*
(iii) *There exist an integer $N \geq 1$ and $\theta \in (0, 1)$ such that*

$$\|T^N(\boldsymbol{x})\| \leq \theta \|\boldsymbol{x}\|, \qquad \boldsymbol{x} \in \Omega. \tag{2.3}$$

Proof. If $\Omega = \{0\}$, all three statements hold trivially. For the rest of the proof, assume $\Omega \neq \{0\}$ and set

$$\Sigma_\Omega := \{\boldsymbol{x} \in \Omega : \|\boldsymbol{x}\| = 1\}.$$

Then Σ_Ω is nonempty and compact. We prove the cycle (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) $\Rightarrow$ (iii).

(iii)$\Rightarrow$(ii). Continuity and positive homogeneity give

$$L_N := \max_{0 \leq j < N} \max_{\boldsymbol{y} \in \Sigma_\Omega} \|T^j(\boldsymbol{y})\| < \infty.$$

Write $k = qN + r$, with $0 \leq r < N$. Iterating (2.3) yields

$$\|T^k(\boldsymbol{x})\| \leq L_N \theta^q \|\boldsymbol{x}\| \leq \frac{L_N}{\theta} \left(\theta^{1/N}\right)^k \|\boldsymbol{x}\|.$$

Thus (ii) holds.

(ii)$\Rightarrow$(i). This follows immediately from $0 < \rho < 1$.

(i)$\Rightarrow$(iii). Fix $\eta \in (0, 1)$. For each $\boldsymbol{x} \in \Sigma_\Omega$, pointwise convergence gives an integer $n_{\boldsymbol{x}} \geq 1$ such that

$$\|T^{n_{\boldsymbol{x}}}(\boldsymbol{x})\| < \eta.$$

By continuity of $T^{n_{\boldsymbol{x}}}$, there is a relative neighborhood $U_{\boldsymbol{x}}$ of $\boldsymbol{x}$ in Σ_Ω on which

$$\|T^{n_{\boldsymbol{x}}}(\boldsymbol{y})\| \le \eta.$$

Choose a finite subcover $U_{\boldsymbol{x}_1}, \dots, U_{\boldsymbol{x}_s}$ and set

$$N_0 := \max_{1\le i\le s} n_{\boldsymbol{x}_i}, \qquad L := \max_{0\le j\le N_0} \max_{\boldsymbol{y}\in\Sigma_\Omega} \|T^j(\boldsymbol{y})\|.$$

For every $\boldsymbol{y} \in \Sigma_\Omega$, choose one $m(\boldsymbol{y}) \in \{1, \dots, N_0\}$ such that

$$\|T^{m(\boldsymbol{y})}(\boldsymbol{y})\| \le \eta. \tag{2.4}$$

Fix $\boldsymbol{x} \in \Omega \setminus \{0\}$. Set $\boldsymbol{z}_0 = \boldsymbol{x}$ and $s_0 = 0$. As long as $\boldsymbol{z}_q \neq 0$, define

$$m_q := m\left(\frac{\boldsymbol{z}_q}{\|\boldsymbol{z}_q\|}\right), \qquad s_{q+1} := s_q + m_q, \qquad \boldsymbol{z}_{q+1} := T^{m_q}(\boldsymbol{z}_q).$$

Positive homogeneity and (2.4) give

$$\|\boldsymbol{z}_{q+1}\| \le \eta\|\boldsymbol{z}_q\|, \qquad \|\boldsymbol{z}_q\| \le \eta^q\|\boldsymbol{x}\|. \tag{2.5}$$

If some $\boldsymbol{z}_q$ vanishes, all later iterates vanish. Otherwise $s_q \ge q$ and $s_q \to \infty$. For every iterate that has not already vanished, choose the unique q such that $s_q \le k < s_{q+1}$. Since $k - s_q \le N_0$,

$$\|T^k(\boldsymbol{x})\| = \|T^{k-s_q}(\boldsymbol{z}_q)\| \le L\|\boldsymbol{z}_q\| \le L\eta^q\|\boldsymbol{x}\|.$$

Moreover,

$$k < s_{q+1} \le (q+1)N_0, \qquad \text{so} \qquad q > \frac{k}{N_0} - 1.$$

For every $k \ge 1$, consequently,

$$\|T^k(\boldsymbol{x})\| \le \frac{L}{\eta}\left(\eta^{1/N_0}\right)^k\|\boldsymbol{x}\|. \tag{2.6}$$

The estimate is trivial after finite termination and at $\boldsymbol{x} = 0$. Since L and N_0 are independent of $\boldsymbol{x}$, choose an integer $N \ge 1$ large enough that

$$\theta := \frac{L}{\eta}\left(\eta^{1/N_0}\right)^N < 1.$$

Applying (2.6) at $k = N$ gives (2.3) for every $\boldsymbol{x} \in \Omega$. This proves (iii) and completes the equivalence. □

2.1 The homogeneous growth radius

For a nontrivial closed cone and a continuous positively homogeneous self-map, set

$$\Sigma_\Omega := \{\boldsymbol{x} \in \Omega : \|\boldsymbol{x}\| = 1\}$$

and define

$$a_k(T) := \max_{\boldsymbol{x}\in\Sigma_\Omega} \|T^k(\boldsymbol{x})\|, \qquad k \geq 0. \tag{2.7}$$

The maximum exists by compactness, and positive homogeneity gives

$$a_{k+\ell}(T) \leq a_k(T)a_\ell(T). \tag{2.8}$$

Indeed, for $\boldsymbol{x} \in \Sigma_\Omega$ with $T^\ell(\boldsymbol{x}) \neq 0$, normalize $T^\ell(\boldsymbol{x})$ and use homogeneity to obtain

$$\|T^{k+\ell}(\boldsymbol{x})\| \leq a_k(T)\|T^\ell(\boldsymbol{x})\| \leq a_k(T)a_\ell(T);$$

the case $T^\ell(\boldsymbol{x}) = 0$ is immediate. Hence the following limit exists:

$$r_\Omega(T) := \lim_{k\to\infty} a_k(T)^{1/k} = \inf_{k\geq 1} a_k(T)^{1/k}. \tag{2.9}$$

Corollary 2.3. *Under the assumptions of Theorem 2.2, with $\Omega \neq \{0\}$, pointwise convergence is equivalent to $r_\Omega(T) < 1$. Whenever these equivalent stability conditions hold, $r_\Omega(T)$ is the infimum of all $\rho \in (0, 1)$ for which an estimate* (2.2) *holds with some finite prefactor C.*

Proof. Pointwise convergence implies (2.2) by Theorem 2.2, hence $a_k(T) \leq C\rho^k$ and $r_\Omega(T) \leq \rho < 1$. Conversely, if $r_\Omega(T) < 1$, choose $\rho \in (r_\Omega(T), 1)$. By (2.9), there is N for which $a_N(T) < \rho^N$. Writing $k = qN + r$ and using (2.8) gives

$$a_k(T) \leq a_N(T)^q \max_{0\leq j<N} a_j(T) \leq C_\rho \rho^k$$

for a finite C_ρ. Positive homogeneity gives (2.2). The same argument proves that every $\rho > r_\Omega(T)$ is admissible, whereas any admissible rate bounds the limit in (2.9). □

Remark 2.4. The compact normalized section is essential. On ℓ^2, let

$$T\boldsymbol{e}_j = \left(1 - \frac{1}{j}\right)\boldsymbol{e}_j.$$

Then T is linear and $T^k\boldsymbol{x} \to 0$ for every $\boldsymbol{x} \in \ell^2$, but $\|T^k\| = 1$ for every k; no uniform R-linear estimate holds. Thus Theorem 2.2 does not extend to infinite-dimensional spaces without an additional compactness or uniformity hypothesis.

3 BB dynamics on compatible gradient pairs

Translate the minimizer in (1.1) to the origin and write

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \mathbf{H}\boldsymbol{x}, \qquad \boldsymbol{g} = \mathbf{H}\boldsymbol{x}, \qquad 0 < a := \lambda_{\min}(\mathbf{H}) \le \lambda_{\max}(\mathbf{H}) =: b. \tag{3.1}$$

A *positive spectral weight* is

$$\mathbf{W} = \omega(\mathbf{H}), \qquad \omega(\lambda) > 0 \quad \text{for every } \lambda \in \sigma(\mathbf{H}).$$

Thus $\mathbf{W} \succ 0$ and $\mathbf{WH} = \mathbf{HW}$. Write

$$\kappa_2(\mathbf{W}) := \|\mathbf{W}\|_2 \|\mathbf{W}^{-1}\|_2.$$

For $\boldsymbol{u} \neq 0$, define the delayed weighted Rayleigh step

$$\alpha_{\mathbf{W}}(\boldsymbol{u}) := \frac{\boldsymbol{u}^\top \mathbf{W}\boldsymbol{u}}{\boldsymbol{u}^\top \mathbf{WH}\boldsymbol{u}}. \tag{3.2}$$

The generalized Rayleigh quotient gives

$$\frac{1}{b} \le \alpha_{\mathbf{W}}(\boldsymbol{u}) \le \frac{1}{a}. \tag{3.3}$$

The choices $\mathbf{W} = \mathbf{I}$ and $\mathbf{W} = \mathbf{H}$ are BB1 and BB2, respectively. Starting from an arbitrary positive first step α_0, the weighted delayed recurrence is

$$\boldsymbol{g}_{k+1} = (\mathbf{I} - \alpha_k \mathbf{H})\boldsymbol{g}_k, \qquad \alpha_k = \alpha_{\mathbf{W}}(\boldsymbol{g}_{k-1}), \qquad k \ge 1, \tag{3.4}$$

with the usual zero continuation after termination.

3.1 The closed invariant cone

Let

$$\Omega_{\mathrm{BB}}(\mathbf{H}) := \left\{ (\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^n \times \mathbb{R}^n : \begin{array}{l} \boldsymbol{v} = (\mathbf{I} - \alpha\mathbf{H})\boldsymbol{u} \\ \text{for some } \alpha \in [b^{-1}, a^{-1}] \end{array} \right\} \tag{3.5}$$

and equip the product space with the Euclidean norm

$$\|(\boldsymbol{u}, \boldsymbol{v})\|_2 := \left(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2\right)^{1/2},$$

where vector norms without a subscript are Euclidean. Define

$$T_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v}) := \begin{cases} \big(\boldsymbol{v}, (\mathbf{I} - \alpha_{\mathbf{W}}(\boldsymbol{u})\mathbf{H})\boldsymbol{v}\big), & \boldsymbol{u} \neq 0, \\ (0, 0), & (\boldsymbol{u}, \boldsymbol{v}) = (0, 0). \end{cases} \tag{3.6}$$

There is no other state with $\boldsymbol{u} = 0$ in (3.5).

Proposition 3.1. *The set $\Omega_{\mathrm{BB}}(\mathbf{H})$ is a closed cone. For every fixed positive spectral weight $\mathbf{W}$, the transition*

$$T_{\mathbf{W}} : \Omega_{\mathrm{BB}}(\mathbf{H}) \longrightarrow \Omega_{\mathrm{BB}}(\mathbf{H})$$

is a continuous positively homogeneous self-map.

Proof. Closure under nonnegative scaling follows immediately from (3.5). Let $(\boldsymbol{u}^\ell, \boldsymbol{v}^\ell) \to (\boldsymbol{u}, \boldsymbol{v})$ with $(\boldsymbol{u}^\ell, \boldsymbol{v}^\ell) \in \Omega_{\mathrm{BB}}(\mathbf{H})$. Choose $\alpha_\ell \in [b^{-1}, a^{-1}]$ such that

$$\boldsymbol{v}^\ell = (\mathbf{I} - \alpha_\ell \mathbf{H})\boldsymbol{u}^\ell.$$

The interval is compact, so a subsequence satisfies $\alpha_\ell \to \alpha \in [b^{-1}, a^{-1}]$. Passing to the limit gives $\boldsymbol{v} = (\mathbf{I} - \alpha\mathbf{H})\boldsymbol{u}$, and the cone is closed.

If $(\boldsymbol{u}, \boldsymbol{v}) \in \Omega_{\mathrm{BB}}(\mathbf{H})$ and $\boldsymbol{u} \neq 0$, then (3.3) shows that

$$(\mathbf{I} - \alpha_{\mathbf{W}}(\boldsymbol{u})\mathbf{H})\boldsymbol{v}$$

is a legal one-step update of $\boldsymbol{v}$. Hence $T_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v})$ belongs to the same cone. The zero state maps to itself.

For $t > 0$, (3.2) gives $\alpha_{\mathbf{W}}(t\boldsymbol{u}) = \alpha_{\mathbf{W}}(\boldsymbol{u})$. Thus

$$T_{\mathbf{W}}(t\boldsymbol{u}, t\boldsymbol{v}) = tT_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v}).$$

The identity also holds at $t = 0$. Continuity away from the origin is immediate. At the origin, set

$$M_{\mathbf{H}} := \max_{\alpha \in [b^{-1}, a^{-1}]} \|\mathbf{I} - \alpha\mathbf{H}\|_2 < \infty.$$

For every nonzero compatible state,

$$\|(\mathbf{I} - \alpha_{\mathbf{W}}(\boldsymbol{u})\mathbf{H})\boldsymbol{v}\| \leq M_{\mathbf{H}}\|\boldsymbol{v}\|.$$

Therefore $T_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v}) \to (0, 0)$ whenever $(\boldsymbol{u}, \boldsymbol{v}) \to (0, 0)$ within the cone. This proves continuity. □

3.2 Spectral conjugacy

Set $\mathbf{S} = \mathbf{W}^{1/2}$ and define

$$\mathcal{S}_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v}) := (\mathbf{S}\boldsymbol{u}, \mathbf{S}\boldsymbol{v}). \tag{3.7}$$

Since $\mathbf{S}$ commutes with $\mathbf{H}$, $\mathcal{S}_{\mathbf{W}}$ maps $\Omega_{\mathrm{BB}}(\mathbf{H})$ onto itself.

Proposition 3.2. *Let $T_{\mathbf{I}}$ denote the BB1 transition. Then*

$$\mathcal{S}_{\mathbf{W}} T_{\mathbf{W}} = T_{\mathbf{I}} \mathcal{S}_{\mathbf{W}}, \qquad T_{\mathbf{W}} = \mathcal{S}_{\mathbf{W}}^{-1} T_{\mathbf{I}} \mathcal{S}_{\mathbf{W}}. \tag{3.8}$$

In particular,

$$r_{\Omega_{\mathrm{BB}}(\mathbf{H})}(T_{\mathbf{W}}) = r_{\Omega_{\mathrm{BB}}(\mathbf{H})}(T_{\mathbf{I}}). \tag{3.9}$$

Proof. For $\boldsymbol{u} \neq 0$,

$$\alpha_{\mathbf{I}}(\mathbf{S}\boldsymbol{u}) = \frac{(\mathbf{S}\boldsymbol{u})^\top(\mathbf{S}\boldsymbol{u})}{(\mathbf{S}\boldsymbol{u})^\top\mathbf{H}(\mathbf{S}\boldsymbol{u})} = \frac{\boldsymbol{u}^\top\mathbf{W}\boldsymbol{u}}{\boldsymbol{u}^\top\mathbf{W}\mathbf{H}\boldsymbol{u}} = \alpha_{\mathbf{W}}(\boldsymbol{u}).$$

Multiplying both components of (3.6) by $\mathbf{S}$ proves (3.8). Iteration gives $T_{\mathbf{W}}^k = \mathcal{S}_{\mathbf{W}}^{-1}T_{\mathbf{I}}^k\mathcal{S}_{\mathbf{W}}$. If $a_k(T)$ denotes (2.7) on $\Omega_{\mathrm{BB}}(\mathbf{H})$ and $\kappa(\mathcal{S}_{\mathbf{W}}) = \|\mathbf{S}\|_2\|\mathbf{S}^{-1}\|_2$, then

$$\kappa(\mathcal{S}_{\mathbf{W}})^{-1}a_k(T_{\mathbf{I}}) \leq a_k(T_{\mathbf{W}}) \leq \kappa(\mathcal{S}_{\mathbf{W}})a_k(T_{\mathbf{I}}). \tag{3.10}$$

Indeed, the upper bound follows from the conjugacy and homogeneity after normalizing $\mathcal{S}_{\mathbf{W}}\boldsymbol{z}$; the lower bound follows by interchanging $T_{\mathbf{W}}$ and $T_{\mathbf{I}}$. Taking kth roots in (3.10) proves (3.9). □

3.3 State-uniform R-linear convergence

Lemma 3.3. *Every state $(\boldsymbol{u}, \boldsymbol{v}) \in \Omega_{\mathrm{BB}}(\mathbf{H})$ is the consecutive-gradient state obtained after one admissible warm-up step. Starting from this state, its subsequent delayed weighted Rayleigh trajectory is exactly the orbit of $T_{\mathbf{W}}$.*

Proof. If $\boldsymbol{u} = 0$, compatibility gives $\boldsymbol{v} = 0$, and the assertion follows from the prescribed zero continuation. Hence assume $\boldsymbol{u} \neq 0$. Choose $\alpha \in [b^{-1}, a^{-1}]$ such that $\boldsymbol{v} = (\mathbf{I} - \alpha\mathbf{H})\boldsymbol{u}$. With $\boldsymbol{x}_0 = \mathbf{H}^{-1}\boldsymbol{u}$ and $\boldsymbol{x}_1 = \mathbf{H}^{-1}\boldsymbol{v}$, one has

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - \alpha\boldsymbol{u}.$$

Thus $\boldsymbol{u} \mapsto \boldsymbol{v}$ is a legal warm-up gradient step. In the BB1 case, if $\boldsymbol{s}_0 = \boldsymbol{x}_1 - \boldsymbol{x}_0$ and $\boldsymbol{y}_0 = \boldsymbol{v} - \boldsymbol{u}$, then

$$\boldsymbol{s}_0 = -\alpha\boldsymbol{u}, \qquad \boldsymbol{y}_0 = -\alpha\mathbf{H}\boldsymbol{u}, \qquad \frac{\boldsymbol{s}_0^\top\boldsymbol{s}_0}{\boldsymbol{s}_0^\top\boldsymbol{y}_0} = \alpha_{\mathbf{I}}(\boldsymbol{u}),$$

so the orbit is a genuine BB1 trajectory. For a general fixed weight, the next stepsize is $\alpha_{\mathbf{W}}(\boldsymbol{u})$ by definition, and hence

$$(\boldsymbol{u}, \boldsymbol{v}) \mapsto \big(\boldsymbol{v}, (\mathbf{I} - \alpha_{\mathbf{W}}(\boldsymbol{u})\mathbf{H})\boldsymbol{v}\big) = T_{\mathbf{W}}(\boldsymbol{u}, \boldsymbol{v}).$$

The same identity repeats until termination, after which both the algorithm and the map are continued by zeros. □

Theorem 3.4. *Fix $\mathbf{H} \succ 0$. There exist constants $C_{\mathbf{H}} \geq 1$ and $\rho_{\mathbf{H}} \in (0, 1)$, depending only on $\mathbf{H}$, such that every positive spectral weight $\mathbf{W} = \omega(\mathbf{H})$ satisfies*

$$\|T_{\mathbf{W}}^k(\boldsymbol{u}, \boldsymbol{v})\|_2 \leq C_{\mathbf{H}}\sqrt{\kappa_2(\mathbf{W})}\,\rho_{\mathbf{H}}^k\|(\boldsymbol{u}, \boldsymbol{v})\|_2, \qquad (\boldsymbol{u}, \boldsymbol{v}) \in \Omega_{\mathrm{BB}}(\mathbf{H}), \quad k \geq 0. \tag{3.11}$$

This includes BB1, BB2, and every fixed power weight $\mathbf{W} = \mathbf{H}^\tau$, $\tau \in \mathbb{R}$.

Proof. Raydan's convergence theorem shows that every BB1 trajectory on a strictly convex quadratic either terminates or has gradients tending to zero [5]. By Lemma 3.3, every $T_{\mathbf{I}}$-orbit on $\Omega_{\mathrm{BB}}(\mathbf{H})$ is such a trajectory. Hence $T_{\mathbf{I}}^k \boldsymbol{z} \to 0$ for every $\boldsymbol{z} \in \Omega_{\mathrm{BB}}(\mathbf{H})$, and Proposition 3.1 and Theorem 2.2 give constants $C_{\mathbf{H}} \geq 1$ and $\rho_{\mathbf{H}} \in (0,1)$, depending only on $\mathbf{H}$, for which (3.11) holds when $\mathbf{W} = \mathbf{I}$.

For general $\mathbf{W}$, Proposition 3.2 gives $T_{\mathbf{W}}^k = \mathcal{S}_{\mathbf{W}}^{-1} T_{\mathbf{I}}^k \mathcal{S}_{\mathbf{W}}$. The operator norms of $\mathcal{S}_{\mathbf{W}}$ and $\mathcal{S}_{\mathbf{W}}^{-1}$ for the product norm are $\|\mathbf{S}\|_2$ and $\|\mathbf{S}^{-1}\|_2$, respectively. Hence

$$\|T_{\mathbf{W}}^k \boldsymbol{z}\|_2 \leq \|\mathbf{S}^{-1}\|_2 \, C_{\mathbf{H}} \rho_{\mathbf{H}}^k \|\mathcal{S}_{\mathbf{W}} \boldsymbol{z}\|_2 \leq C_{\mathbf{H}} \|\mathbf{S}^{-1}\|_2 \|\mathbf{S}\|_2 \rho_{\mathbf{H}}^k \|\boldsymbol{z}\|_2.$$

Since $\mathbf{S} = \mathbf{W}^{1/2}$,

$$\|\mathbf{S}^{-1}\|_2 \|\mathbf{S}\|_2 = \sqrt{\kappa_2(\mathbf{W})}.$$

For $\mathbf{W} = \mathbf{H}^\tau$ the matrix is positive definite for every real τ. □

Corollary 3.5. *Consider BB1, BB2, or a fixed weighted rule* (3.4)*, with an arbitrary positive finite first step and zero continuation after termination. Then*

$$\|(\boldsymbol{g}_k, \boldsymbol{g}_{k+1})\|_2 \leq C_{\mathbf{H}} \sqrt{\kappa_2(\mathbf{W})} \, \rho_{\mathbf{H}}^{k-1} \|(\boldsymbol{g}_1, \boldsymbol{g}_2)\|_2, \qquad k \geq 1. \tag{3.12}$$

Consequently,

$$\|\boldsymbol{x}_k - \boldsymbol{x}_*\| \leq a^{-1} C_{\mathbf{H}} \sqrt{\kappa_2(\mathbf{W})} \, \rho_{\mathbf{H}}^{k-1} \|(\boldsymbol{g}_1, \boldsymbol{g}_2)\|_2, \tag{3.13}$$

$$0 \leq f(\boldsymbol{x}_k) - f(\boldsymbol{x}_*) \leq \frac{C_{\mathbf{H}}^2 \kappa_2(\mathbf{W})}{2a} \, \rho_{\mathbf{H}}^{2(k-1)} \|(\boldsymbol{g}_1, \boldsymbol{g}_2)\|_2^2. \tag{3.14}$$

Proof. If $\boldsymbol{g}_0 = 0$ or $\boldsymbol{g}_1 = 0$, the conclusion is immediate under zero continuation. Otherwise the first computed weighted Rayleigh step is $\alpha_1 = \alpha_{\mathbf{W}}(\boldsymbol{g}_0) \in [b^{-1}, a^{-1}]$. Hence $(\boldsymbol{g}_1, \boldsymbol{g}_2) \in \Omega_{\mathrm{BB}}(\mathbf{H})$ even when the arbitrary first step that produced $\boldsymbol{g}_1$ does not. Moreover,

$$(\boldsymbol{g}_k, \boldsymbol{g}_{k+1}) = T_{\mathbf{W}}^{k-1}(\boldsymbol{g}_1, \boldsymbol{g}_2), \qquad k \geq 1.$$

Applying Theorem 3.4 yields (3.12). The remaining estimates follow from

$$\|\boldsymbol{x} - \boldsymbol{x}_*\| \leq a^{-1} \|\boldsymbol{g}(\boldsymbol{x})\|$$

and

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_*) = \frac{1}{2} \boldsymbol{g}(\boldsymbol{x})^\top \mathbf{H}^{-1} \boldsymbol{g}(\boldsymbol{x}) \leq \frac{1}{2a} \|\boldsymbol{g}(\boldsymbol{x})\|^2.$$

□

4 Conclusion

The closed-cone construction gives a new convergence analysis by separating the algorithm-specific input, pointwise convergence from every compatible state, from a geometric rate-upgrade step. For BB1, Raydan's convergence theorem supplies the input; compactness and homogeneity yield the uniform R-linear rate. Spectral conjugacy then transfers the rate to BB2 and every fixed positive spectral weight while preserving the homogeneous growth radius.

The same proof architecture applies to a finite-memory spectral recurrence whenever its realizable states form a finite-dimensional closed invariant cone, the induced transition is continuous and degree-one positively homogeneous, and pointwise stability holds from every cone state. The central task is therefore the construction of the compatible state space. For the classical BB rules, a sharper evaluation of the common growth radius is developed in [8].

Statements and Declarations

Funding. The work of Shu-tai Yang was supported by the National College Students Innovation and Entrepreneurship Training Program, administered by the University of Science and Technology of China (Project No. 202510358091). The work of Ya-xiang Yuan was supported by the National Natural Science Foundation of China (Grant No. 12288201).

Competing interests. Ya-xiang Yuan is the Editor-in-Chief of the Journal of the Operations Research Society of China. The authors request independent editorial handling of this manuscript. The authors declare no other relevant financial or non-financial interests.

Author contributions. Shu-tai Yang conceived the study, developed the mathematical results, carried out the analysis, and wrote the original draft. Ya-xiang Yuan provided guidance on the research direction and reviewed the manuscript. Both authors read and approved the final manuscript.

Declaration on the use of generative AI. During the preparation of this manuscript, Shu-tai Yang used GPT-5.5 Thinking (OpenAI) for language and LaTeX editing, the presentation of mathematical notation, and assistance in checking mathematical arguments and references. He directed the use of the tool, independently checked and revised all model-assisted material, and made all scholarly judgments and final decisions. Both authors reviewed and approved the final manuscript and take full responsibility for its content.

Data availability. No datasets were generated or analyzed during the current study.

Code availability. No computational code was generated or used in this entirely theoretical study.

References

1. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA Journal of Numerical Analysis **8**(1), 141–148 (1988). https://doi.org/10.1093/imanum/8.1.141
2. Dai, Y.-H.: Alternate step gradient method. Optimization **52**(4–5), 395–415 (2003). https://doi.org/10.1080/02331930310001611547
3. Dai, Y.-H., Liao, L.-Z.: R-linear convergence of the Barzilai and Borwein gradient method. IMA Journal of Numerical Analysis **22**(1), 1–10 (2002). https://doi.org/10.1093/imanum/22.1.1
4. Friedlander, A., Martínez, J.M., Molina, B., Raydan, M.: Gradient method with retards and generalizations. SIAM Journal on Numerical Analysis **36**(1), 275–289 (1999). https://doi.org/10.1137/S003614299427315X
5. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA Journal of Numerical Analysis **13**(3), 321–326 (1993). https://doi.org/10.1093/imanum/13.3.321
6. Shen, J., Hu, J.: Stability of discrete-time switched homogeneous systems on cones and conewise homogeneous inclusions. SIAM Journal on Control and Optimization **50**(4), 2216–2253 (2012). https://doi.org/10.1137/110845215
7. Tuna, S.E.: Growth rate of switched homogeneous systems. Automatica **44**(11), 2857–2862 (2008). https://doi.org/10.1016/j.automatica.2008.03.017
8. Yang, S., Yuan, Y.-X.: The sharp worst-case asymptotic rate of the Barzilai–Borwein method in $\mathbb{R}^d$ and Hilbert spaces. Preprint, arXiv:2608.07839 (2026). https://doi.org/10.48550/arXiv.2608.07839