A harmonic framework for stepsize selection in gradient methods

Giulia Ferrandi[1], Michiel E. Hochstenbach[1] and Nataša Krejić[2]

[1]Department of Mathematics and Computer Science, TU Eindhoven, PO Box 513, Eindhoven, 5600 MB, The Netherlands.
[2]Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Trg D.Obradovića 4, Novi Sad, 21000, Serbia.

Contributing authors: g.ferrandi@tue.nl; m.e.hochstenbach@tue.nl; natasak@uns.ac.rs;

Abstract

We study the use of inverse harmonic Rayleigh quotients with target for the stepsize selection in gradient methods for nonlinear unconstrained optimization problems. This provides not only an elegant and flexible framework to parametrize and reinterpret existing stepsize schemes, but also gives inspiration for new flexible and tunable families of steplengths. In particular, we analyze and extend the adaptive Barzilai–Borwein method to a new family of stepsizes. While this family exploits negative values for the target, we also consider positive targets. We present a convergence analysis for quadratic problems extending results by Dai and Liao (2002), and carry out experiments outlining the potential of the approaches.

Keywords: Unconstrained optimization, harmonic Rayleigh quotient, gradient methods, framework for steplength selection, ABB method, Hessian spectral properties

AMS Classification: 65K05 , 90C20 , 90C30 , 65F15 , 65F10

1 Introduction

We study the unconstrained optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$

for convex quadratic and general nonlinear continuously differentiable functions $f:\mathbb{R}^n\to\mathbb{R}$. We consider the popular gradient method

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \beta_k\,\mathbf{g}_k \;=\; \mathbf{x}_k - \alpha_k^{-1}\,\mathbf{g}_k,$$

where $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$ and $\beta_k > 0$ is the steplength. It is convenient to introduce a separate notation α_k for the inverse of the stepsize β_k, since both play important roles; α_k corresponds to (harmonic) Rayleigh quotients, which are scalars providing second-order information (on the Hessian).

As usual, write $\mathbf{s}_{k-1} = \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}$. Two popular stepsizes are the Barzilai–Borwein (BB) steplengths [1]

$$\beta_k^{\text{BB1}} = \frac{\mathbf{s}_{k-1}^T\mathbf{s}_{k-1}}{\mathbf{y}_{k-1}^T\mathbf{s}_{k-1}}, \qquad \beta_k^{\text{BB2}} = \frac{\mathbf{y}_{k-1}^T\mathbf{s}_{k-1}}{\mathbf{y}_{k-1}^T\mathbf{y}_{k-1}}.$$

We denote their inverses by α_k^{BB1} and α_k^{BB2}, respectively. In case of convex quadratic problems

$$\min_{\mathbf{x}\in\mathbb{R}^n} \tfrac{1}{2}\,\mathbf{x}^TA\mathbf{x} - \mathbf{b}^T\mathbf{x}, \tag{1}$$

where A is $n\times n$ symmetric positive definite (SPD), and $\mathbf{b}\in\mathbb{R}^n$, the BB steps are the inverses of the Rayleigh quotient and harmonic Rayleigh quotient,

$$\beta_k^{\text{BB1}} = \frac{\mathbf{s}_{k-1}^T\mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T A\,\mathbf{s}_{k-1}}, \qquad \beta_k^{\text{BB2}} = \frac{\mathbf{s}_{k-1}^T A\,\mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T A^2\,\mathbf{s}_{k-1}}.$$

We refer to [2] for a nice recent review on various steplength options.

In this paper, we will consider a general framework for these and other stepsizes by introducing a harmonic Rayleigh quotient including a *target* τ. We recall the harmonic Rayleigh–Ritz extraction for matrix eigenvalue problems in Section 2. The general form of this extraction features a target $\tau\in\mathbb{R}\cup\{\pm\infty\}$. This target is analyzed and exploited in Section 3, to develop a new general framework for all possible stepsizes. We will see that the BB stepsizes correspond to $\tau = 0$ or $\tau = \pm\infty$. This may not only add towards a new understanding and interpretation of known strategies, but also suggests new competitive schemes. Section 4 closer studies the Adaptive Barzilai–Borwein method (ABB) [3], and provides a new theoretical justification for it. We also showcase the potential of the framework by introducing new families generalizing the ABB method. As is common (see, e.g., [2]) we first consider the convex quadratic problem. Convergence results for this case, extending those of [4], are presented in Section 5. The extension of the harmonic steplength to general nonlinear problems is treated in Section 6. Finally, we carry out numerical experiments and summarize some conclusions in Sections 7 and 8.

2 Harmonic extraction and harmonic Rayleigh quotients

The harmonic Rayleigh–Ritz extraction has been introduced in the context of eigenvalue problems (see, e.g., [5, 6], [7, Sec. 4.4], [8]) to extract promising approximate (interior) eigenpairs from a subspace. Consider the eigenproblem $A\mathbf{x} = \lambda\mathbf{x}$ for a given square A. Although A does not necessarily need to be symmetric or real for the harmonic extraction method, in our optimization context we are interested in real symmetric matrices SPD matrices A.

Suppose that we wish to extract promising approximate eigenpairs from a low-dimensional search space $\mathcal{U}$ for which the columns of $U \in \mathbb{R}^{n\times d}$ form an orthogonal basis, where usually $d \ll n$. We are interested in finding approximate eigenpairs $(\theta, \mathbf{u}) \approx (\lambda, \mathbf{x})$, where $\mathbf{u}$ is of the form $\mathbf{u} = U\mathbf{c} \approx \mathbf{x}$, with $\mathbf{c} \in \mathbb{R}^d$ of unit 2-norm. The standard Rayleigh–Ritz extraction imposes the Galerkin condition

$$AU\mathbf{c} - \theta\, U\mathbf{c} \perp \mathcal{U}.$$

This leads to d approximate eigenpairs $(\theta_j, U\mathbf{c}_j)$, for $j = 1, \ldots, d$, obtained from the eigenpairs $(\theta_j, \mathbf{c}_j)$ of U^TAU.

Denote the eigenvalues of A by $0 < \lambda_1 \le \cdots \le \lambda_n$. The standard Rayleigh–Ritz extraction enjoys a good reputation for exterior eigenvalues (see, e.g., [9]), which means the largest or smallest few eigenvalues, in our case of symmetric A. However, for interior eigenvalues near a target $\tau \in (\lambda_1, \lambda_n)$, the harmonic Rayleigh–Ritz extraction tends to produce approximate eigenvectors of better quality. This approach works as follows (see, e.g., [7, Sec. 4.4] for more details).

Let τ be not equal to an eigenvalue; in the context of eigenvalue problems, τ is typically chosen close to the eigenvalues of interest. Eigenvalues near τ are exterior eigenvalues of $(A - \tau I)^{-1}$, which is a favorable situation to impose a Galerkin condition. Therefore, the idea is to impose such a condition involving this shifted and inverted matrix. To avoid having to work with an inverse of a (potentially large) matrix, a modified Galerkin condition

$$(A - \tau I)^{-1}\, U\,\widetilde{\mathbf{c}} - (\widetilde{\theta} - \tau)^{-1}\, U\,\widetilde{\mathbf{c}} \perp (A - \tau I)^2\, \mathcal{U}$$

is considered. We note that this is equivalent to the Galerkin condition $(A - \tau I)^{-1}\,\mathbf{u} - (\theta - \tau)^{-1}\,\mathbf{u} \perp (A - \tau I)\,\mathcal{U}$ for $\mathbf{u} \in (A - \tau I)\,\mathcal{U}$, which considers this extraction from a different viewpoint.

This implies that the quantities of interest are $(\widetilde{\theta}_j, \widetilde{\mathbf{c}}_j)$, for $j = 1, \ldots, d$, the eigenpairs of the pencil $(U^T(A - \tau I)\,AU,\ U^T(A - \tau I)\,U)$; and the associated vectors $\widetilde{\mathbf{u}}_j = U\widetilde{\mathbf{c}}_j$. This means that the relation between a harmonic Ritz vector $\widetilde{\mathbf{u}} = U\widetilde{\mathbf{c}}$ and the corresponding harmonic Ritz value is

$$\widetilde{\theta} = \frac{\widetilde{\mathbf{u}}^T (A-\tau I)^2\, \widetilde{\mathbf{u}}}{\widetilde{\mathbf{u}}^T (A-\tau I)\, \widetilde{\mathbf{u}}} + \tau \ = \ \frac{\widetilde{\mathbf{u}}^T (A-\tau I)\, A\, \widetilde{\mathbf{u}}}{\widetilde{\mathbf{u}}^T (A-\tau I)\, \widetilde{\mathbf{u}}}. \tag{2}$$

We will exploit this quantity in the next section to introduce a general harmonic framework for the choice of steplengths.

3 A harmonic framework for stepsize selection

Inspired by (2), we now propose and study the use of harmonic Rayleigh quotients of the form

$$\alpha_k(\tau_k) = \frac{\mathbf{s}_{k-1}^T (A-\tau_k I)\, A\, \mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T (A-\tau_k I)\, \mathbf{s}_{k-1}} \tag{3}$$

in the context of gradient methods, where the τ_k are targets that may be varied throughout the process. In contrast to the use of the target for eigenvalue problems, where the τ_k are typically selected inside or very close to the interval $[\lambda_1, \lambda_n]$ (as discussed in Section 2), we investigate strategies with τ-values outside this interval, as well as schemes where these targets may sometimes be inside.

The stepsize we consider is given by the inverse harmonic Rayleigh quotient

$$\beta_k(\tau_k) = \frac{\mathbf{s}_{k-1}^T (A-\tau_k I)\, \mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T (A-\tau_k I)\, A\, \mathbf{s}_{k-1}}. \tag{4}$$

We will refer to these steps as “TBB steps”: Barzilai–Borwein type of steps using a harmonic Rayleigh quotient with target τ_k. In the rest of this section we will consider various aspects of gradient methods with TBB steps as in (4). In particular, we will discuss strategies for picking τ in Section 3.7.

3.1 Properties of the TBB stepsize

First, recall that $\alpha^{\mathrm{BB1}} \le \alpha^{\mathrm{BB2}}$ and therefore $\beta^{\mathrm{BB2}} \le \beta^{\mathrm{BB1}}$; in fact (see, e.g., [2])

$$\alpha^{\mathrm{BB1}}/\alpha^{\mathrm{BB2}} = \beta^{\mathrm{BB2}}/\beta^{\mathrm{BB1}} = \cos^2(\mathbf{s}_{k-1}, A\mathbf{s}_{k-1}). \tag{5}$$

The following proposition summarizes several basic but essential properties of stepsize (4).

Proposition 1 *Let $\mathbf{s} \in \mathbb{R}^n$ be not equal to a multiple of an eigenvector. The function $\beta(\tau) = \frac{\mathbf{s}^T (A-\tau I)\, \mathbf{s}}{\mathbf{s}^T (A-\tau I)\, A\, \mathbf{s}}$ enjoys the following properties.*

(i) $\beta(\tau)$ is defined for all $\tau \in \mathbb{R}$ with exception of $\alpha^{\mathrm{BB2}} = \frac{\mathbf{s}^T A^2\, \mathbf{s}}{\mathbf{s}^T A\, \mathbf{s}}$, and is a strictly monotonically decreasing function on $(-\infty, \alpha^{\mathrm{BB2}})$ and $(\alpha^{\mathrm{BB2}}, \infty)$.

(ii) *Alternative expressions are* $\beta(\tau) = \beta^{\rm BB1}\,\frac{\tau-\alpha^{\rm BB1}}{\tau-\alpha^{\rm BB2}} = \beta^{\rm BB2}\,\frac{\beta^{\rm BB1}\tau-1}{\beta^{\rm BB2}\tau-1}$.

(iii) $\beta(0) = \beta^{\rm BB2}$ *and* $\lim\limits_{\tau\to\pm\infty} \beta(\tau) = \beta^{\rm BB1}$.

(iv) *For* $-\infty < \tau < 0$, *it holds* $\beta^{\rm BB2} < \beta(\tau) < \beta^{\rm BB1}$.

(v) *For* $0 < \tau < \lambda_1$, *we have* $\frac{1-\beta^{\rm BB1}\lambda_1}{1-\beta^{\rm BB2}\lambda_1} \cdot \beta^{\rm BB2} < \beta(\tau) < \beta^{\rm BB2}$.

(vi) *For* $\tau > \lambda_n$, *it holds that* $\beta^{\rm BB1} < \beta(\tau) < \frac{\lambda_n-\alpha^{\rm BB1}}{\lambda_n-\alpha^{\rm BB2}} \cdot \beta^{\rm BB1}$.

(vii) β *is a bijection from* $\mathbb{R}\backslash\{\alpha^{\rm BB2}\}$ *to* $\mathbb{R}\backslash\{\beta^{\rm BB1}\}$, *and from* $\mathbb{R}\cup\{\pm\infty\}-\{\alpha^{\rm BB2}\}$ *to* $\mathbb{R}$.

Proof The derivative of β with respect to τ is given by

$$\beta'(\tau) = \frac{(\mathbf{s}^TA\,\mathbf{s})^2 - (\mathbf{s}^TA^2\,\mathbf{s})\,(\mathbf{s}^T\mathbf{s})}{(\mathbf{s}^T(A-\tau I)\,A\,\mathbf{s})^2}.$$

The numerator is equal to $\|A\mathbf{s}\|^2\,\|\mathbf{s}\|^2\cos^2(A\mathbf{s},\mathbf{s}) - \|A\mathbf{s}\|^2\,\|\mathbf{s}\|^2 < 0$, since $\mathbf{s}$ is assumed to be not equal to an eigenvector; part (i) follows from this. Item (ii) is obtained by factoring out $\mathbf{s}^T\mathbf{s}$ in the numerator, and $\mathbf{s}^TA\mathbf{s}$ in the denominator. Part (iii) follows directly from (ii). Since β is defined everywhere and strictly decreasing on the interval $(-\infty, 0)$ we get item (iv). Part (v) is derived from (ii) by the fact that β on the interval $(0, \lambda_1)$ is defined everywhere and strictly decreasing. The factor $\frac{1-\beta^{\rm BB1}\lambda_1}{1-\beta^{\rm BB2}\lambda_1}$ is less than one, since $\beta^{\rm BB1}$, $\beta^{\rm BB2} < \lambda_1^{-1}$ (again by the fact that $\mathbf{s}$ is not a multiple of an eigenvector) and $\beta^{\rm BB2} < \beta^{\rm BB1}$. Item (vi) follows from the fact that β is defined everywhere and strictly decreasing on the interval (λ_n, ∞). The factor $\frac{\lambda_n-\alpha^{\rm BB1}}{\lambda_n-\alpha^{\rm BB2}}$ is greater than one in view of $\alpha^{\rm BB1}$, $\alpha^{\rm BB2} < \lambda_n$ and $\alpha^{\rm BB1} < \alpha^{\rm BB2}$.

□

It is particularly item (vii) that implies that *the harmonic Rayleigh quotient forms a framework or parametrization for all possible steplengths*: together with target $\tau = \pm\infty$, we have a one-to-one relation between targets in $\mathbb{R} \cup \{\pm\infty\}$ and any real stepsize (positive or negative). We stress that, because of the pole of β in $\tau = \alpha^{\rm BB2}$, the stepsize might be unbounded for $\tau \in (\lambda_1, \lambda_n)$, which evidently is unwanted. Note that in the (unlikely) case that $\mathbf{s}$ is equal to an eigenvector corresponding to eigenvalue λ, $\beta(\tau)$ is equal to the constant function $\beta(\tau) \equiv \lambda^{-1}$ (with exception of the "hole" at $\tau = \lambda$). Figure 1 gives an impression of the properties in Proposition 1 for a typical situation.

For completeness, we also list some characteristics of the inverse stepsize $\alpha(\tau)$ (see (3)), the harmonic Rayleigh quotient.

Proposition 2 *(i) The function* $\alpha(\tau) = \frac{\mathbf{s}^T(A-\tau I)\,A\,\mathbf{s}}{\mathbf{s}^T(A-\tau I)\,\mathbf{s}}$ *is defined for all* $\tau\backslash\{\alpha^{\rm BB1}\}$, *and is a strictly monotonically increasing function on the intervals* $(-\infty, \alpha^{\rm BB1})$ *and* $(\alpha^{\rm BB1}, \infty)$.

(ii) *Alternative expression are* $\alpha(\tau) = \alpha^{\rm BB1}\,\frac{\tau-\alpha^{\rm BB2}}{\tau-\alpha^{\rm BB1}} = \alpha^{\rm BB2}\,\frac{\beta^{\rm BB2}\,\tau-1}{\beta^{\rm BB1}\,\tau-1}$.

(iii) $\lim\limits_{\tau\to\pm\infty} \alpha(\tau) = \alpha^{\rm BB1}$ *and* $\alpha(0) = \alpha^{\rm BB2}$.

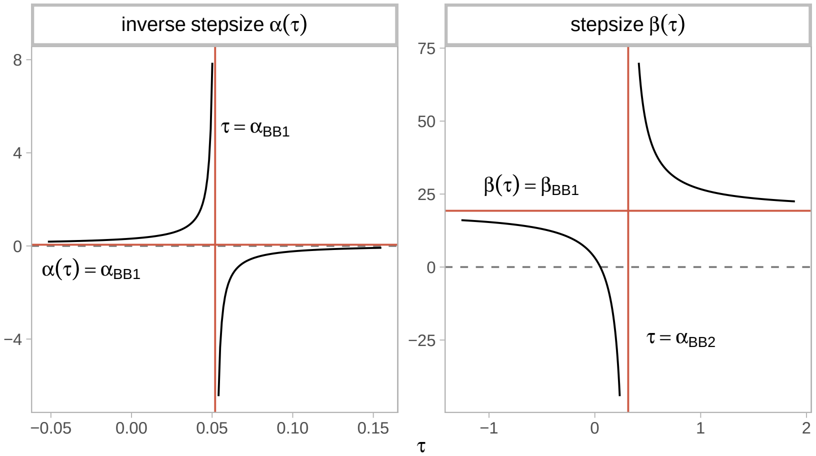

Fig. 1 Harmonic Rayleigh quotient (left) and its inverse, the stepsize (right), as a function of τ for the convex quadratic case $A = \text{diag}(\frac{1}{100}, \frac{1}{99}, \ldots, \frac{1}{2}, 1)$, where $\mathbf{s} = (1, \ldots, 1)^T$.

Proof The derivative of α with respect to τ satisfies

$$\alpha'(\tau) = \frac{(\mathbf{s}^T A^2\, \mathbf{s})\,(\mathbf{s}^T \mathbf{s}) - (\mathbf{s}^T A\, \mathbf{s})^2}{(\mathbf{s}^T (A - \tau I)\, \mathbf{s})^2}.$$

The result now follows from a reasoning similar to Proposition 1. Part (ii) can be derived by factoring out a factor of $\mathbf{s}^T\mathbf{s}$, $\mathbf{s}^T A\mathbf{s}$, or $\mathbf{s}^T A^2\mathbf{s}$ from the numerator or denominator. Item (iii) is straightforward. □

3.2 Sensitivity of the stepsize with respect to the target

We now study the sensitivity of the steplength $\beta(\tau)$ as function of τ, in particular around $\tau = 0$ and $\tau = -\infty$, which correspond to β^{BB1} and β^{BB2}, respectively. We first consider the situation of small τ; recall that $\beta(0) = \beta^{\text{BB2}}$.

Proposition 3 *For $\tau \to 0$, we have up to higher-order terms in τ*

$$\frac{\beta(\tau) - \beta^{\text{BB2}}}{\beta^{\text{BB2}}} = -\tau\,(\beta^{\text{BB1}} - \beta^{\text{BB2}}).$$

Proof For $\tau \to 0$ it holds that (cf. Proposition 1(ii))

$$\beta(\tau) = \beta^{\text{BB2}} \cdot \frac{1 - \tau\,\beta^{\text{BB1}}}{1 - \tau\,\beta^{\text{BB2}}} = \beta^{\text{BB2}} \cdot (1 - \tau\,(\beta^{\text{BB1}} - \beta^{\text{BB2}})) + \mathcal{O}(\tau^2).$$

□

In agreement with Figure 1 and Proposition 1, an appreciable interpretation of this result is that for small negative τ, the stepsize $\beta(\tau)$ increases from β^{BB2} (for $\tau = 0$) towards the larger stepsize β^{BB1} (corresponding to $\tau = -\infty$). Moreover, the rate of change for $\tau \to 0$ is asymptotically proportional to the difference between β^{BB1} and β^{BB2}. As a side note, from (5) we have that $\beta^{\mathrm{BB1}} - \beta^{\mathrm{BB2}} = \beta^{\mathrm{BB2}} \tan^2(\mathbf{s}, \mathbf{y})$.

Next, let us investigate the asymptotic situation $\tau \to \pm\infty$. To this end, we exploit the transformed variable $\zeta = \tau^{-1}$ and consider the expression $\widehat{\beta}(\zeta) := \beta(\zeta^{-1}) = \beta(\tau) = \frac{\zeta\, \mathbf{s}^T A\, \mathbf{s} - \mathbf{s}^T \mathbf{s}}{\zeta\, \mathbf{s}^T A^2\, \mathbf{s} - \mathbf{s}^T A\, \mathbf{s}}$ for $\zeta \to 0$.

Proposition 4 *For $\tau \to \pm\infty$, we have up to higher-order terms in τ^{-1}*

$$\frac{\beta(\tau) - \beta^{\mathrm{BB1}}}{\beta^{\mathrm{BB1}}} = -\tau^{-1}\,(\alpha^{\mathrm{BB1}} - \alpha^{\mathrm{BB2}}).$$

Proof For $\zeta \to 0$ it holds that (cf. Proposition 1(ii))

$$\widehat{\beta}(\zeta) = \beta^{\mathrm{BB1}} \cdot \frac{1 - \zeta\, \alpha^{\mathrm{BB1}}}{1 - \zeta\, \alpha^{\mathrm{BB2}}} = \beta^{\mathrm{BB1}} \cdot (1 - \zeta\,(\alpha^{\mathrm{BB1}} - \alpha^{\mathrm{BB2}})) + \mathcal{O}(\zeta^2).$$

□

Again, this result has a nice meaning: for small negative τ^{-1} (i.e., large negative τ), the stepsize $\beta(\tau^{-1})$ decreases from β^{BB1} (for $\tau^{-1} = 0$) towards the smaller stepsize β^{BB2} (associated with $\tau = 0$). Moreover, the more β^{BB1} differs from β^{BB2}, the faster $\beta_k(\tau^{-1})$ decreases as function of τ^{-1}. For small positive τ^{-1} (which means large positive τ), the steplength increases, and thus gets larger than β^{BB1}; cf. Figure 1. In Sections 3.7, 4.3 and 7 we will discuss and experiment with strategies involving both negative and positive values of τ.

3.3 Pseudocode for gradient method with TBB steps

In Algorithm 1 we give a pseudocode for a gradient method based on TBB steps. We exploit a relative stopping criterion in line 4, which may be replaced by any other reasonable stopping rule.

Algorithm 1 A TBB method for strictly convex quadratic functions

Input: function $f(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^T A \mathbf{x} - \mathbf{b}^T \mathbf{x}$ with A SPD, initial guess $\mathbf{x}_0$, initial stepsize $\beta_0 > 0$, tolerance tol

Output: approximation to minimizer $\mathrm{argmin}_{\mathbf{x}}\, f(\mathbf{x})$

1: Set $\mathbf{g}_0 = A\mathbf{x}_0 - \mathbf{b}$

for $k = 0, 1, \ldots$

2: Set $\mathbf{s}_k = -\beta_k\, \mathbf{g}_k$ and update $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$

3: Compute the gradient $\mathbf{g}_{k+1} = A\mathbf{x}_{k+1} - \mathbf{b}$

4: **if** $\|\mathbf{g}_{k+1}\| \le \mathsf{tol} \cdot \|\mathbf{g}_0\|$, **return**, **end**

5: $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$

6: Choose τ_{k+1}, compute TBB step $\beta_{k+1}(\tau_{k+1})$ according to (4)

Clearly, the choice of targets τ_k in Line 6 is a crucial aspect of the method. We discuss some options for this particularly in Section 3.7 and 4.3. In Section 6 we also consider practically important details such as the choice of β_0.

3.4 Secant conditions

In this subsection we consider an equivalent formulation of the TBB stepsize

$$\beta_k(\tau) = \frac{\mathbf{s}_{k-1}^T\,(\mathbf{y}_{k-1} - \tau_k\ \mathbf{s}_{k-1})}{\mathbf{y}_{k-1}^T\,(\mathbf{y}_{k-1} - \tau_k\ \mathbf{s}_{k-1})}, \tag{6}$$

where $\mathbf{y}_{k-1} = A\mathbf{s}_{k-1}$; this will be useful in Section 6 for generic problems where the Hessian changes over the iterations.

We recall from [1] that a justification of the BB steps is the fact that they approximate the Hessian matrix by a scalar multiple of the identity, as follows. It is reasonable that an approximation B_k to the Hessian approximately satisfies the secant equation $\mathbf{y}_{k-1} = B_k\mathbf{s}_{k-1}$. The BB steps solve the secant equation in a least-squares sense [1]:

$$\alpha_k^{\text{BB1}} = \underset{\alpha}{\operatorname{argmin}}\,\|\mathbf{y}_{k-1} - \alpha\,\mathbf{s}_{k-1}\|, \qquad \alpha_k^{\text{BB2}} = \underset{\alpha}{\operatorname{argmin}}\,\|\mathbf{s}_{k-1} - \alpha^{-1}\,\mathbf{y}_{k-1}\|, \tag{7}$$

which results in an approximation of the form $B_k = \alpha I$ to the Hessian, where α is α_k^{BB1} or α_k^{BB2}, respectively.

As the TBB step (4) involves the shifted matrix $A - \tau I$ (where τ may vary over the iterations), this suggests us to consider a shifted secant equation

$$\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1} = (B_k - \tau I)\,\mathbf{s}_{k-1}. \tag{8}$$

By replacing $\mathbf{y}_{k-1}$ by $\mathbf{y}_k - \tau\,\mathbf{s}_{k-1}$ and α by $\alpha - \tau$ in the second secant condition in (7), we obtain that the TBB step satisfies a modified secant condition, which is equivalent to the second equation of (7) for $\tau = 0$, but not equivalent to the first or second one for any other target value.

Proposition 5 *Let $\tau \neq \alpha_k^{\text{BB1}} = \frac{\mathbf{y}_{k-1}^T\mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T\mathbf{s}_{k-1}}$. Then the inverse TBB step $\alpha_k = \beta_k^{-1}$ satisfies*

$$\alpha_k(\tau) = \frac{\mathbf{y}_{k-1}^T\,(\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1})}{\mathbf{s}_{k-1}^T\,(\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1})} = \underset{\alpha}{\operatorname{argmin}}\,\|\mathbf{s}_{k-1} - (\alpha - \tau)^{-1}(\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1})\|. \tag{9}$$

Proof The result follows by setting to zero the derivative of the square of the objective function in (9) with respect to α, which gives

$$(\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1})^T(\mathbf{s}_{k-1} - (\alpha - \tau)^{-1}(\mathbf{y}_{k-1} - \tau\,\mathbf{s}_{k-1})) = 0.$$

□

In addition to this interpretation as modified secant condition, when the target is located outside $[\lambda_1, \lambda_n]$, we can also think of the TBB step as the scalar least squares solution to the following problem involving a certain weighted norm, as follows. Define the standard weighted norm associated with a given SPD matrix W by $\|\mathbf{x}\|_W^2 := \mathbf{x}^T W \mathbf{x}$.

Proposition 6 *The least squares solution to the weighted secant equation satisfies*

$$\underset{\alpha}{\operatorname{argmin}} \|\mathbf{y}_{k-1} - \alpha\, \mathbf{s}_{k-1}\|_W = \frac{\mathbf{y}_{k-1}^T\, W\, \mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T\, W\, \mathbf{s}_{k-1}}. \tag{10}$$

Proof The result follows by setting $-\mathbf{y}_{k-1}^T W \mathbf{s}_{k-1} + \alpha\, \mathbf{s}_{k-1}^T W \mathbf{s}_{k-1}$, the derivative of $\frac{1}{2}\,\|\mathbf{y}_{k-1} - \alpha\, \mathbf{s}_{k-1}\|_W^2$ with respect to α, to zero. □

The BB1 and BB2 steps can be obtained from this proposition by taking $W = I$ and $W = A$, respectively; cf. (7). The TBB step can be derived by choosing $W = A - \tau I$ for $\tau < \lambda_1$ or $W = \tau I - A$ for $\tau > \lambda_n$, which gives an SPD weight matrix in both cases.

In conclusion, the BB1, BB2, and TBB steps approximate the Hessian by a positive scalar multiple of the identity of the form $B_k = \alpha_k I \approx A$. The scalars satisfy one or both (weighted) secant conditions.

3.5 Regularization

Another viewpoint on harmonic steps (4) with a target τ outside $[\lambda_1, \lambda_n]$ is as *regularization of the Hessian.* First consider taking a shift $\tau < 0$. As we replace A by $A - \tau I$ for $\tau < 0$ this yields a "more positive definite" shifted Hessian. As one indicator, the condition number

$$\kappa(A) = \lambda_n \,/\, \lambda_1 \tag{11}$$

of A is modified to $\frac{\lambda_n - \tau}{\lambda_1 - \tau}$ by this shift. Consider the function $\varphi : (-\infty, 0] \to [1, \infty)$ given by $\varphi(t) = \frac{\lambda_n - t}{\lambda_1 - t}$. Since this function is strictly monotonically decreasing on the domain $(-\infty, 0]$, we conclude that $\kappa(A - \tau I) < \kappa(A)$. More precisely, we have the following first-order estimate.

Proposition 7 *For $\tau \to 0$ we have*

$$\kappa(A - \tau I) = \kappa(A) + \frac{\lambda_n - \lambda_1}{\lambda_1^2} \cdot \tau + \mathcal{O}(\tau^2).$$

Proof Straightforward using the linear approximation $\varphi(t) \approx \varphi(0) + \varphi'(0)\, t$. □

In fact, for $\tau < 0$, the shifted condition number $\kappa(A - \tau I)$ may be considerably smaller than $\kappa(A)$, especially if the Hessian is nearly singular. In conclusion, also in view of (8) and Proposition 5, *harmonic stepsizes with* $\tau < 0$ *may be viewed as satisfying a secant condition on a regularized Hessian.* Note that $\lim_{\tau \to -\infty} \kappa(A - \tau I) = 1$.

Moreover, for $\tau > \lambda_n$, we have a similar situation. It is not difficult to show that $\kappa(A - \tau I) = \kappa(A)$ when $\tau = \lambda_n + \lambda_1$ (which is usually close to λ_n). For $\tau > \lambda_n + \lambda_1$, the condition number of the shifted matrix $A - \tau I$ decreases monotonically, with the analogous property $\lim_{\tau \to \infty} \kappa(A - \tau I) = 1$.

3.6 Connections with other stepsizes

We would like to point out that quotients of the form

$$\frac{\mathbf{s}^T\, p(A)\, A\, \mathbf{s}}{\mathbf{s}^T\, p(A)\, \mathbf{s}}, \tag{12}$$

for certain polynomials p, have also been considered in different contexts in [10, (2.6)], [8], and [11]. The harmonic Rayleigh quotient (3) is a special case of (12), but a very practical instance for several reasons. First, it gives a clear connection with the harmonic Rayleigh–Ritz extraction for eigenvalue problems as seen in Section 2. Second, as we have seen in Section 3.1, by taking first-order polynomials p in τ, we have a one-to-one correspondence between the target τ and the stepsize β.

The introduction of an adjustable parameter in the stepsize has been first proposed in [12], where the authors present a convex combination of BB1 and BB2 steps,

$$\beta_k^{\mathrm{CON}}(\zeta_k) = \zeta_k\, \beta_k^{\mathrm{BB1}} + (1 - \zeta_k)\, \beta_k^{\mathrm{BB2}},$$

where $\zeta_k \in [0, 1]$. Note that $\beta_k^{\mathrm{CON}}(0) = \beta_k^{\mathrm{BB2}}$ and $\beta_k^{\mathrm{CON}}(1) = \beta_k^{\mathrm{BB1}}$. Its inverse minimizes a linear combination of secant conditions, i.e.,

$$\operatorname*{argmin}_{\alpha} \|\zeta\, (\alpha\, \mathbf{s}_{k-1} - \mathbf{y}_{k-1}) + (1 - \zeta)\, (\mathbf{s}_{k-1} - \alpha^{-1}\, \mathbf{y}_{k-1})\|.$$

In [12], several strategies are considered to choose ζ_k: fixed, randomly from the uniform distribution over $[0, 1]$, or imitating the behavior of the cyclic gradient methods (cf. [12, pp. 56–57] and references therein). In the next section, we show the link between this convex combination steplength and the TBB step. This not only suggests relevant strategies to select ζ_k, or rather τ_k for our TBB methods, but also gives a far wider range of options.

3.7 Strategies to select targets

There is a one-to-one correspondence between the parameter ζ_k in Section 3.6 and the target τ_k in the TBB stepsize: with the choice

$$\tau_k = -\frac{\zeta_k}{1-\zeta_k}\, \alpha_k^{\mathrm{BB2}}, \tag{13}$$

the corresponding TBB step will coincide with the stepsize in [12]: $\beta_k(\tau_k) = \beta_k^{\mathrm{CON}}(\zeta_k)$. In addition, since $\zeta_k \in [0,1]$, the corresponding target values lie in $\tau_k \in [-\infty, 0]$. From Proposition 1 we conclude that $\beta_k^{\mathrm{BB2}} \le \beta_k^{\mathrm{CON}}(\zeta_k) \le \beta_k^{\mathrm{BB1}}$.

Given the relation between ζ_k and τ_k, all strategies mentioned in [12] correspond to negative targets τ_k for the TBB steplengths (4). In the next section, we analyze new schemes for the choice of negative targets; here we focus on positive targets τ_k. As this yields steplengths $\beta_k(\tau_k) > \beta_k^{\mathrm{BB1}}$, this is not equivalent to any of the stepsizes determined by ζ_k in [12]. Inspired by the expression in (13), where the target is a negative factor times the inverse BB2 stepsize α^{BB2}, we consider positive targets of the form

$$\tau_k = \rho\, \alpha_k^{\mathrm{BB2}}, \quad \rho > 1. \tag{14}$$

This gives us an *affine* (rather than convex) combination of BB1 and BB2:

$$\beta_k(\tau_k) = \tfrac{\rho}{\rho-1}\, \beta_k^{\mathrm{BB1}} - \tfrac{1}{\rho-1}\, \beta_k^{\mathrm{BB2}}.$$

This new stepsize is located in the right branch of the hyperbola (right plot of Figure 1). We will make use of the following bounds for the inverse stepsize:

$$\alpha_k(\tau_k) = \tfrac{\rho-1}{\rho-\cos^2(\mathbf{s}_{k-1},\mathbf{y}_{k-1})}\, \alpha_k^{\mathrm{BB1}} \;\in\; [\tfrac{\rho-1}{\rho}\, \alpha_k^{\mathrm{BB1}},\, \alpha_k^{\mathrm{BB1}}] \;\subseteq\; [\tfrac{\rho-1}{\rho}\, \lambda_1,\; \lambda_n].$$

As in (13), we may let ρ vary through the iterations. In the numerical experiments, we will consider the strategy $\tau_1 = 0$ and

$$\tau_k = k\, \alpha_k^{\mathrm{BB2}}, \qquad \text{for } k = 2, 3, \ldots. \tag{15}$$

Since $\tau_k \ge k\, \lambda_1$ for all k, the sequence $\{\tau_k\}$ converges to infinity; therefore, the corresponding inverse stepsizes $\beta_k(\tau_k)$ behave asymptotically as α_k^{BB1}. In the long run, there exists a $\rho > 1$ such that $\alpha_k(\tau_k) \in [\frac{\rho-1}{\rho}\lambda_1,\, \lambda_n]$ for all $k > \rho - 1$.

The inverse stepsizes obtained from (14) and (15) are bounded from below by a multiple of α_k^{BB1}, which will play a key role in the global convergence of the resulting gradient method in Section 5.

4 An analysis and extension of the ABB scheme

We now present an analysis and "continuous extension" of the ABB method [3]. We propose a new harmonic family of stepsizes, using various options for the target τ_k throughout the process. These strategies aim to retain the advantages of the ABB method, while being more flexible and tunable than the original approach.

The motivations for this adaptation are the following. First, it is a popular method. Second, the method contains a threshold parameter η, the choice of which may be seen as a bit arbitrary. Third, it is a good showcase of the possibilities that the harmonic framework offers.

We first briefly recall the ABB approach. The stepsize is selected as

$$\beta_k^{\text{ABB}} = \begin{cases} \beta_k^{\text{BB2}}, & \text{if } \beta_k^{\text{BB2}} < \eta\, \beta_k^{\text{BB1}}, \\ \beta_k^{\text{BB1}}, & \text{otherwise.} \end{cases}$$

Here, η is a user-selected parameter with a common choice $\eta = 0.8$; see, e.g., [2]. This strategy adaptively picks BB1 and BB2 steps based on the value of $\beta_k^{\text{BB2}}/\beta_k^{\text{BB1}} = \cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$.

The idea of the ABB stepsize is to take a larger step when $\cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \approx 1$, and a smaller step when this is not the case. If $\cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$ is close to 1, this means that $\mathbf{g}_{k-1}$ (or, equivalently, $\mathbf{s}_{k-1}$) is close to an eigenvector of A corresponding to an eigenvalue $\lambda > 0$. Thus, as we will see in Section 5, by the step β_k^{BB1} we are particularly reducing the gradient component corresponding to λ significantly. When we are far from an eigenvalue, we prefer to take shorter steps, such as β_k^{BB2}, since we aim to reduce several gradient components; this fosters the gradient method to take a new longer step in the next iterations. There exist several variants of the ABB method; the interested reader may refer to [3, 13] for such ideas.

4.1 A theoretical foundation for the ABB method

As discussed, a key statement for the ABB method is: "if $\cos(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \approx 1$, then $\mathbf{s}_{k-1}$ is close to an eigenvector" (see, e.g., [2, pp. 179–180]). The following new result quantifies this statement for the quadratic case. We need the assumption that λ is a simple eigenvalue, since otherwise an eigenvector is not uniquely defined. We consider the situation $\cos^2(\mathbf{s}, A\mathbf{s}) \approx 1$, so that $\sin(\mathbf{s}, A\mathbf{s})$ is small.

Proposition 8 *Let $(\lambda, \mathbf{x})$ be an eigenpair of A, where λ is a simple eigenvalue. Let $\mathbf{s} \approx \mathbf{x}$ be an approximate eigenvector. Then, up to higher-order terms in $\angle(\mathbf{s}, \mathbf{x})$,*

$$\frac{\lambda}{\max_{\lambda_i \neq \lambda} |\lambda_i - \lambda|} \cdot \sin(\mathbf{s}, A\mathbf{s}) \lesssim \tan(\mathbf{s}, \mathbf{x}) \lesssim \frac{\lambda}{\min_{\lambda_i \neq \lambda} |\lambda_i - \lambda|} \cdot \sin(\mathbf{s}, A\mathbf{s}).$$

Proof Without loss of generality we may assume that $A = \text{diag}(\lambda, \Lambda)$, where Λ is an $(n-1) \times (n-1)$ diagonal matrix containing all eigenvalues different from λ, and that $\mathbf{s}$ is of the form $[1,\ \mathbf{z}]^T$, a perturbation of $\mathbf{x}$, the first canonical basis vector. This means $\tan(\mathbf{s},\mathbf{x}) = \|\mathbf{z}\|$; our goal is to connect this quantity to $\angle(\mathbf{s}, A\mathbf{s})$ via $\sin(\mathbf{s}, A\mathbf{s})$. We have $As = [\lambda,\ \Lambda\,\mathbf{z}]^T$, and

$$\cos^2(\mathbf{s}, A\mathbf{s}) = \frac{(\mathbf{s}^T A\mathbf{s})^2}{\|\mathbf{s}\|^2\,\|A\mathbf{s}\|^2} = \frac{(1+\lambda^{-1}\,\mathbf{z}^T\Lambda\,\mathbf{z})^2}{(1+\|\mathbf{z}\|^2)\,(1+\lambda^{-2}\,\|\Lambda\,\mathbf{z}\|^2)}.$$

We now twice use the Taylor expansion $(1-t)^{-1} = 1+t+\mathcal{O}(t^2)$ for small t, so that

$$\begin{aligned}(1+\|\mathbf{z}\|^2)^{-1} &= 1-\|\mathbf{z}\|^2+\mathcal{O}(\|\mathbf{z}\|^4),\\ (1+\lambda^{-2}\,\|\Lambda\,\mathbf{z}\|^2)^{-1} &= 1-\lambda^{-2}\,\|\Lambda\,\mathbf{z}\|^2+\mathcal{O}(\|\mathbf{z}\|^4).\end{aligned}$$

This yields that up to $\mathcal{O}(\|\mathbf{z}\|^4)$ terms we have

$$\sin^2(\mathbf{s}, A\mathbf{s}) = \|\mathbf{z}\|^2 + \lambda^{-2}\,\|\Lambda\,\mathbf{z}\|^2 - 2\lambda^{-1}\,\mathbf{z}^T\Lambda\,\mathbf{z} = \|\mathbf{z}-\lambda^{-1}\,\Lambda\,\mathbf{z}\|^2.$$

Multiplication by $\lambda > 0$ gives $\|\Lambda\,\mathbf{z}-\lambda\mathbf{z}\| = \lambda\,\sin(\mathbf{s}, A\mathbf{s})$. The result now follows from

$$\min_{\lambda_i\neq\lambda} |\lambda_i-\lambda|\ \|\mathbf{z}\| \le \|\Lambda\,\mathbf{z}-\lambda\mathbf{z}\| \le \max_{\lambda_i\neq\lambda} |\lambda_i-\lambda|\ \|\mathbf{z}\|.$$

□

Next, we investigate the sensitivity of the BB steps for quadratic problems where the direction is close to an eigenvector.

Proposition 9 *Let* $\mathbf{s}$ *be an approximation of an eigenvector* $\mathbf{x}$ *corresponding to a simple eigenvalue* λ*. Up to higher-order terms in* $\angle(\mathbf{s},\mathbf{x})$*, we have*

$$\frac{|\beta_k^{\mathrm{BB1}}-\lambda^{-1}|}{\lambda^{-1}} \lesssim \frac{\max_i |\lambda_i-\lambda|}{\lambda}\cdot\tan^2(\mathbf{s},\mathbf{x}),$$

$$\frac{|\beta_k^{\mathrm{BB2}}-\lambda^{-1}|}{\lambda^{-1}} \lesssim \frac{\lambda_n}{\lambda}\cdot\frac{\max_i |\lambda_i-\lambda|}{\lambda}\cdot\tan^2(\mathbf{s},\mathbf{x}).$$

Proof With the same notation as in the proof of Proposition 8, we have for the BB1 step

$$\frac{\mathbf{s}^T\mathbf{s}}{\mathbf{s}^T A\,\mathbf{s}} = \lambda^{-1}\,\frac{1+\mathbf{z}^T\mathbf{z}}{1+\lambda^{-1}\,\mathbf{z}^T\Lambda\,\mathbf{z}} = \lambda^{-1}\,(1+\mathbf{z}^T(1-\lambda^{-1}\Lambda)\,\mathbf{z})+\mathcal{O}(\|\mathbf{z}\|^4),$$

and for the BB2 step

$$\frac{\mathbf{s}^T A\,\mathbf{s}}{\mathbf{s}^T A^2\,\mathbf{s}} = \frac{\lambda+\mathbf{z}^T\Lambda\,\mathbf{z}}{\lambda^2+\mathbf{z}^T\Lambda^2\,\mathbf{z}} = \lambda^{-1}\,(1+\mathbf{z}^T\lambda^{-1}\Lambda\,(I-\lambda^{-1}\Lambda)\,\mathbf{z})+\mathcal{O}(\|\mathbf{z}\|^4).$$

□

One interpretation of this proposition is that, indeed, it may be a good idea to take BB1 steps rather than BB2 steps if s is close to an eigenvector, since BB2 steps are more sensitive with respect to perturbations in that direction, as seen as the extra factor λ_n/λ in the upper bound. This factor can be very large ($\approx \kappa(A)$) for small eigenvalues. Therefore, this result forms a clear mathematical motivation for the ABB scheme.

4.2 Sensitivity of $\beta(\tau)$ with respect to s

The following result is an extension of Proposition 9 to the harmonic step with target (4). We will see that it reduces to Proposition 9 in the case of $\tau = 0$ or $\tau \to \pm\infty$.

Proposition 10 *Let* $\mathbf{s}$ *be an approximation of an eigenvector* $\mathbf{x}$ *corresponding to a simple eigenvalue* λ*. Up to higher-order terms in* $\angle(\mathbf{s}, \mathbf{x})$*, we have*

$$\frac{|\beta(\tau) - \lambda^{-1}|}{\lambda^{-1}} \lesssim \frac{\max_i |\lambda_i - \tau|}{|\lambda - \tau|} \cdot \frac{\max_i |\lambda_i - \lambda|}{\lambda} \cdot \tan^2(\mathbf{s}, \mathbf{x}).$$

Proof With the notation as in Proposition 9, and using similar techniques we get

$$\begin{aligned} \beta(\tau) &= \frac{\lambda - \tau + \mathbf{z}^T(\Lambda - \tau I)\,\mathbf{z}}{\lambda\,(\lambda - \tau) + \mathbf{z}^T\Lambda\,(\Lambda - \tau I)\,\mathbf{z}} \\ &= \lambda^{-1}\,(1 + (\lambda - \tau)^{-1}\,\mathbf{z}^T\,(\Lambda - \tau I)\,(I - \lambda^{-1}\Lambda)\,\mathbf{z}) + \mathcal{O}(\|\mathbf{z}\|^4). \end{aligned}$$

□

Note that when $\tau < 0$, the first factor on the right-hand side is equal to $\frac{\lambda_n - \tau}{\lambda - \tau}$. Interestingly, this factor converges to 1 for $\tau \to \pm\infty$, which reduces to the first inequality in Proposition 9; this corresponds to the BB1 step, the inverse Rayleigh quotient. When $\tau \to 0$, this factor converges to that in the second inequality in Proposition 9; this corresponds to the BB2 step, the inverse harmonic Rayleigh quotient for zero target.

4.3 A new family of stepsizes

The ABB strategy may be viewed as "discrete", in the sense that just two types of stepsizes are possible: the BB1 or the BB2 step. We will now propose a new "continuous" variant of ABB parameterized by choosing appropriate τ_k. We design this strategy to have a similar behavior as ABB: when $\cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \approx 1$, the steps are close to the BB1 step, while the steps should be close to the BB2 step when $\cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \approx 0$. Therefore, we are interested in a function of $\cos(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$ such that when $\tau_k \to -\infty$ we recover BB1, while $\tau_k = 0$ yields the BB2 step. One choice to attain this is to use a cotangent function:

$$\tau_k = -\cot(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}).$$

Indeed, this choice has the two types of desired asymptotic behavior. To further tune the speed by which we approach the two BB steps when $\cos(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$ approaches 0 or 1, we will also introduce two extra parameters, and consider

$$\tau_k = -\frac{\cos^q(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})}{\sin^r(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})}, \tag{16}$$

for q, $r > 0$. For example, if we want our gradient method to have shorter steps more often than long ones, we may keep $r = 1$ but select a higher value of q, e.g., $q = 2$. This mimics the effect of setting η relatively close to 1, as it is done in [2]. The following result ensures that this "cotangent step" (16) indeed may be regarded as a continuous extension of the ABB step, having similar properties for $\angle(\mathbf{s}, \mathbf{y})$ close to 0 or $\pi/2$.

Proposition 11 *For $q, r > 0$, consider $\beta(\tau)$ as in (4), where τ is defined by (16). We have that $\beta(\tau) \to \beta^{\mathrm{BB1}}$ when $\angle(\mathbf{s}, \mathbf{y}) \to 0$ and $\beta(\tau) \to \beta^{\mathrm{BB2}}$ when $\angle(\mathbf{s}, \mathbf{y}) \to \pi/2$. Moreover, $\beta(\tau)$ is a decreasing function of $\angle(\mathbf{s}, \mathbf{y})$.*

Proof Since sin is strictly increasing and cos is strictly decreasing on $(0, \frac{\pi}{2})$, the function τ defined by (16) is a strictly increasing function of $\angle(\mathbf{s}, \mathbf{y})$, ranging from $-\infty$ for $\angle(\mathbf{s}, \mathbf{y}) \to 0$ to 0 for $\angle(\mathbf{s}, \mathbf{y}) = \frac{\pi}{2}$. Therefore, in combination with Section 3.1, we conclude that β decreases from β^{BB1} to β^{BB2}. □

We point out that in the quadratic case the targets (16) are negative, which implies that the corresponding stepsize has the same bounds as β_k^{CON} in Section 3.7: $\beta_k^{\mathrm{BB1}} \le \beta_k(\tau_k) \le \beta_k^{\mathrm{BB2}}$. We will test various choices for $q, r > 0$ in the experiments in Section 7.

5 Convergence analysis

In this section, we extend a few results on BB steps to the TBB step with the choices for the target τ_k described in Sections 3.7 and 4.3. Global convergence of the gradient method with BB steps has been proven by Raydan [14] for strictly convex quadratic functions. Dai and Liao [4] and Dai [15] show R-linear convergence of the method for a class of BB stepsizes. We follow [15, Thm. 4.1] and [4, Thm. 2.5] to extend the results to the TBB steps.

Before stating the assumptions that the TBB stepsize must satisfy, we introduce some expressions that will be useful for the following results. We first decompose the gradient along an orthonormal basis of eigenvectors of A. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the (orthonormal) eigenvectors associated with the eigenvalues $\lambda_1, \dots, \lambda_n$. The gradient can be expressed as linear combination of these eigenvectors (cf., e.g., [4, (2.2)])

$$g_k = \sum_{i=1}^{n} \gamma_i^k \, \mathbf{v}_i.$$

Therefore, the TBB step expressed in the eigenvalues of A and the γ_i^{k-1} is

$$\alpha_k(\tau) = \frac{\sum_i (\gamma_i^{k-1})^2 \, \lambda_i \, (\lambda_i - \tau)}{\sum_i (\gamma_i^{k-1})^2 \, (\lambda_i - \tau)}, \tag{17}$$

provided the denominator is nonzero. In particular, for $\tau \to \pm\infty$, this stepsize converges to the BB1 stepsize, which is expressed as (cf., e.g., [4, Eq. (2.18)])

$$\alpha_k^{\mathrm{BB1}} = \frac{\sum_i (\gamma_i^{k-1})^2\ \lambda_i}{\sum_i (\gamma_i^{k-1})^2}. \tag{18}$$

For strictly convex quadratic functions, we have the following recursive formula for the gradients g_k and, as a consequence, for their coefficients (cf., e.g., [14, Eq. (8)])

$$\mathbf{g}_{k+1} = (I - \beta_k A)\,\mathbf{g}_k, \tag{19}$$
$$\gamma_i^{k+1} = (1 - \beta_k \lambda_i)\,\gamma_i^k. \tag{20}$$

Equations (19)–(20) can also be applied to the error $\mathbf{e}_k = \mathbf{x}_k - A^{-1}\mathbf{b}$ and its components e_i^k in the directions of the eigenvectors. Using these equations, the following results can be obtained for the error components as well. The only complication is that higher powers of the eigenvalues appear:

$$\alpha_k(\tau) = \frac{\sum_i (e_i^{k-1})^2\,\lambda_i^3\,(\lambda_i - \tau)}{\sum_i (e_i^{k-1})^2\,\lambda_i^2\,(\lambda_i - \tau)},$$

since $\beta_k s_k = A\mathbf{e}_k$ for all k. This expression extends [14, Eq. (12) and p. 325]. To prove the convergence of the gradient method for strictly convex quadratic functions, it is sufficient to show that $\|\mathbf{g}_k\| \to 0$. Since we chose an orthonormal basis, $\|g_k\|^2 = \sum_i (\gamma_i^k)^2$ and thus we aim to show $\gamma_i^k \to 0$ for $i = 1, \ldots, n$. We remark that working with the γ_i^k is equivalent to assume that our Hessian matrix A is diagonal.

5.1 Assumptions for the TBB stepsize

We state the assumptions on the TBB stepsize, needed to get R-linear convergence. We adapt [15, Property A] to the TBB steplengths proposed in Sections 3.7 and 4.3:

Assumption 1 The inverse stepsize α_k satisfies Assumption 1 if there exist positive constants $\xi_{\mathrm{low}} \in (\frac{1}{2}, 1]$, $\xi^{\mathrm{up}} \geq 1$ and M_2 such that, for any k,

(i) $\xi_{\mathrm{low}} \cdot \lambda_1 \leq \alpha_k \leq \xi^{\mathrm{up}} \cdot \lambda_n$;

(ii) for any $1 \leq \ell \leq n-1$ and $\varepsilon > 0$, if $\sum_{i=1}^{\ell} (\gamma_i^{k-1})^2 \leq \varepsilon$ and $(\gamma_{\ell+1}^{k-1})^2 \geq M_2\,\varepsilon$, then $\alpha_k \geq \frac{\xi_{\mathrm{low}}}{\xi_{\mathrm{low}}+1/2}\,\lambda_{\ell+1}$.

[15, Property A] also includes retards in the BB steps, but, as we are not interested to deal with retards in this paper, we will not include them to ease the notation in what follows. The interested reader is referred to [10] for the definition of BB steps with retards, and to [15] for the proof of R-linear convergence under this property. Secondly, we note that [15, Property A] requires $\xi_{\text{low}} = 1$, while we allow a looser lower bound for the inverse stepsize. In other words, admissible stepsizes are larger than the largest eigenvalue of A^{-1}. Finally, our upper bound in (i) has a more specific shape than the one set in [15, Property A], which is some $M_1 \geq \lambda_1$. This will enable us to express some bounds as a function of the condition number of A. Given all the differences between Assumption 1 and [15, Property A], it is worthwhile to analyse the new situation.

We show that targets (14), (15), and (16) all lead to a stepsize that satisfies Assumption 1. In addition, we show a useful bound for the $(\ell+1)$st gradient component. Analogue proofs, with the same line of thought, can be found in, e.g, [15, Corollary 4.2] or [14, Lemma 2]. First notice that, with these choices for the target, for k sufficiently large:

$$\xi_{\text{low}} \cdot \alpha_k^{\text{BB1}} \leq \alpha_k(\tau_k) \leq \xi^{\text{up}} \cdot \lambda_n \tag{21}$$

for certain $0 < \xi_{\text{low}} \leq 1$ and $\xi^{\text{up}} \geq 1$. Consequently,

$$(\xi^{\text{up}})^{-1} \cdot \lambda_n^{-1} \leq \beta_k(\tau_k) \leq \xi_{\text{low}}^{-1} \cdot \beta_k^{\text{BB1}}.$$

Lemma 12 *Let the inverse stepsize $\alpha_k(\tau_k)$ satisfy (21) with $\frac{1}{2} < \xi_{\text{low}} \leq 1$ and $\xi^{\text{up}} \geq 1$. Then such stepsize satisfies Assumption 1. In addition, given k, there exists a constant $c \in (0,1)$ such that*

$$|\gamma_{\ell+1}^{k+1}| \leq c\, |\gamma_{\ell+1}^{k}|. \tag{22}$$

Proof Part (i) of Assumption 1 immediately follows from the bounds on the BB1 step. Given the hypotheses in (ii) of Assumption 1, Equation (18) and $M_2 = (\xi_{\text{low}} - \frac{1}{2})^{-1}$, it follows that

$$\begin{aligned}\alpha_k(\tau_k) &\geq \xi_{\text{low}} \, \frac{\sum_{i=1}^n (\gamma_i^{k-1})^2 \lambda_i}{\sum_{i=1}^n (\gamma_i^{k-1})^2} \geq \xi_{\text{low}}\, \lambda_{\ell+1} \, \frac{\sum_{i=\ell+1}^n (\gamma_i^{k-1})^2}{\sum_{i=\ell+1}^n (\gamma_i^{k-1})^2 + \varepsilon} \\ &\geq \xi_{\text{low}}\, \lambda_{\ell+1} \, \frac{(\xi_{\text{low}} - \frac{1}{2})^{-1}\varepsilon}{(\xi_{\text{low}} - \frac{1}{2})^{-1}\varepsilon + \varepsilon} = \frac{\xi_{\text{low}}}{\xi_{\text{low}} + \frac{1}{2}}\, \lambda_{\ell+1}.\end{aligned}$$

Since $\alpha_k(\tau_k) \leq \xi^{\text{up}}\, \lambda_n$,

$$1 - \frac{\xi_{\text{low}} + \frac{1}{2}}{\xi_{\text{low}}} \leq 1 - \lambda_{\ell+1}\, \beta_k(\tau_k) \leq 1 - \frac{\lambda_{\ell+1}}{\xi^{\text{up}}\, \lambda_n}.$$

Given (20), this implies $|\gamma_{\ell+1}^{k+1}| \leq c|\gamma_{\ell+1}^{k}|$ for constant $c := \max\{(2\xi_{\text{low}})^{-1}, 1 - \lambda_{\ell+1}\, (\xi^{\text{up}}\, \lambda_n)^{-1}\} < 1$. □

Remark 13 If $\tau_k = \rho\, \alpha_k^{\text{BB2}}$ (cf. (14)), Lemma 12 does not hold for all the values of ρ: we must restrict ourselves to $\rho > 2$. Nevertheless, in the non-quadratic case, the gradient method is endowed with a line search, where bounds on the stepsize are provided by the user (see, e.g., [16] and Section 6). In this context, we may try also $\rho \leq 2$, which corresponds to $\xi_{\text{low}} < \frac{1}{2}$.

5.2 Bounds on gradient components

We establish two bounds on the gradient components. To do so, the bounds in Assumption 1 are used, with an appropriate restriction on ξ_{low}. We then show that the first gradient component converges to 0 under certain conditions. Let us start with the following lemma, which is an extension of [14, Lemma 1]: this holds for $\xi_{\text{low}} = \xi^{\text{up}} = 1$ and it is stated for the components of the error, that enjoy the same recursive formula as the components of the gradient.

Lemma 14 *Under Assumption 1, γ_1^k converges to zero Q-linearly, i.e., there exists a constant $c_1 \in (0,1)$ such that*

$$|\gamma_1^{k+1}| \leq c_1\, |\gamma_1^k|. \tag{23}$$

Proof From part (i) in Assumption 1, we have

$$1 - \xi_{\text{low}}^{-1} \leq 1 - \lambda_1\, \beta_k(\tau_k) \leq 1 - (\kappa(A)\, \xi^{\text{up}})^{-1},$$

with $\kappa(A)$ the condition number as in (11). Thus, when applying these bounds to (20), we see that (23) holds with $c_1 = \max\{\xi_{\text{low}}^{-1} - 1,\ 1 - (\kappa(A)\, \xi^{\text{up}})^{-1}\}$. The conditions on ξ_{low} and ξ^{up} guarantee that c_1 is indeed in the interval $(0,1)$. □

Unfortunately, it is not possible to prove the Q-linear convergence of the other gradient components, but the following inequality will play a role later. This result generalizes [4, Lemma 2.1].

Lemma 15 *Under Assumption 1, there exists a constant $c_2 > 0$ such that for $i = 2, \ldots, n$*

$$|\gamma_i^{k+1}| \leq c_2\, |\gamma_i^k|. \tag{24}$$

Proof From part (i) in Assumption 1, we have

$$1 - \xi_{\text{low}}^{-1}\, \kappa(A) \leq 1 - \lambda_i\, \beta_k(\tau_k) \leq 1 - (\kappa(A)\, \xi^{\text{up}})^{-1}.$$

Application of these bounds to (20) implies that (24) holds with positive constant $c_2 := \max\{\xi_{\text{low}}^{-1}\, \kappa(A) - 1,\ 1 - (\kappa(A)\, \xi^{\text{up}})^{-1}\}$. □

We note that usually c_2 will be $\xi_{\text{low}}^{-1}\, \kappa(A) - 1$, and that this quantity may be very large. However, this still provides us with a needed tool for proving the convergence of the gradient method for quadratics. In addition, we remark that Lemma 15 still holds in general for $\xi_{\text{low}} \in (0,1]$ and $\xi^{\text{up}} \geq 1$.

5.3 Proof of R-linear convergence

Finally, we are able to state the R-linear convergence result under Assumption 1, which closely follows the line of the proof of [15, Thm. 4.1]. The key parts of the proof are the bounds on the gradient components from Lemma 14 and Lemma 15, the result on the $(\ell+1)$st gradient component of Lemma 12. Our contribution is to adapt all these results that were already in [15] but derived from [15, Property A]. A slightly less general proof of R-linear convergence was previously presented in [4].

Theorem 16 *Let f be a strictly convex quadratic function and let $\mathbf{x}^* = A^{-1}\mathbf{b}$ be its unique minimizer. Let $\{\mathbf{x}_k\}$ be the sequence generated by the gradient method where the stepsize satisfies Assumption 1. Then, either $\mathbf{g}_k = \mathbf{0}$ for some finite k, or the sequence $\{\|\mathbf{g}_k\|\}$ converges to zero R-linearly.*

Proof Let $G(k,\ell) := \sum_{i=1}^{\ell} (\gamma_i^k)^2$, $\delta_1 = c_1^2$, $\delta_2 = c_2^2$ and $\delta = c^2$. In particular, notice that $G(k,n) = \|\mathbf{g}_k\|^2$. Assume also that $c_2 > 1$, otherwise we would immediately conclude the proof due to Lemma 15.

Part I. First we prove that, for an integer $1 \le \ell \le n-1$ and given $k \ge 1$, if there exists some $\varepsilon_\ell \in (0, M_2^{-1})$ and integer m_ℓ such that

$$G(k+j,\ell) \le \varepsilon_\ell \|\mathbf{g}_k\|^2 \quad \text{for} \quad j \ge m_\ell \tag{25}$$

then there exists $j_0 \in \{m_\ell, \dots, m_\ell + \Delta_\ell + 1\}$, with $\Delta_\ell = \Delta_\ell(M_2, \varepsilon_\ell, \delta_2, \delta, m_\ell)$, such that

$$(\gamma_{\ell+1}^{k+j_0})^2 \le M_2\, \varepsilon_\ell \|\mathbf{g}_k\|^2.$$

Assume that $(\gamma_{\ell+1}^{k+j})^2 > M_2\, \varepsilon_\ell \|\mathbf{g}_k\|^2$ for all $j \in \{m_\ell, \dots, m_\ell + \Delta_\ell\}$. We show that the thesis holds for $j_0 = m_\ell + \Delta_\ell + 1$. First, apply Lemma 12 $\Delta_\ell + 1$ times, and Lemma 15 m_ℓ times to obtain

$$(\gamma_{\ell+1}^{k+m_\ell+\Delta_\ell+1})^2 \le \delta^{\Delta_\ell+1} (\gamma_{\ell+1}^{k+m_\ell})^2 \le \delta^{\Delta_\ell+1}\, \delta_2^{m_\ell}\, (\gamma_{\ell+1}^{k})^2.$$

Then choose Δ_ℓ as the smallest integer that solves $\delta^{\Delta_\ell+1}\delta_2^{m_\ell} \le M_2\, \varepsilon_\ell$ (such Δ_ℓ exists due to the choice of ε_ℓ and the fact that $\delta < 1$) and complete the first proof:

$$(\gamma_{\ell+1}^{k+m_\ell+\Delta_\ell+1})^2 \le M_2\, \varepsilon_\ell \|\mathbf{g}_k\|^2.$$

Part II. Let $m_{\ell+1} = m_\ell + \Delta_\ell + 1$ and $\varepsilon_{\ell+1} = (1 + M_2\, \delta_2^2)\, \varepsilon_\ell$. If (25) holds, we show that

$$G(k+j,\ell+1) \le \varepsilon_{\ell+1} \|\mathbf{g}_k\|^2 \quad \text{for} \quad j \ge m_{\ell+1}.$$

Since $G(k+j,\ell+1) = G(k+j,\ell) + (\gamma_{\ell+1}^{k+j})^2$ and $m_{\ell+1} > m_\ell$, it is sufficient to prove that

$$(\gamma_{\ell+1}^{k+j})^2 \le M_2\, \delta_2^2\, \varepsilon_\ell \|\mathbf{g}_k\|^2 \quad \text{for} \quad j \ge m_{\ell+1}.$$

From the first result, there are infinitely many pairs of indices j_1, j_2 with $j_2 \ge j_1 + 2 > j_1 \ge j_0 \ge m_\ell$ such that

$$\begin{aligned}
(\gamma_{\ell+1}^{k+j})^2 &\le M_2\, \varepsilon_\ell \|\mathbf{g}_k\|^2 \quad \text{for} \quad j = j_1, j_2, \\
(\gamma_{\ell+1}^{k+j})^2 &> M_2\, \varepsilon_\ell \|\mathbf{g}_k\|^2 \quad \text{for} \quad j \in \{j_1+1, \dots, j_2-1\}.
\end{aligned}$$

From Lemma 12 it holds that $(\gamma_{\ell+1}^{k+j})^2 \le \delta\,(\gamma_{\ell+1}^{k+j-1})^2 < (\gamma_{\ell+1}^{k+j-1})^2$ for $j \in \{j_1+3,\dots,j_2+1\}$, since $\delta < 1$. This results in a chain of inequalities, halting at $j = j_1+2$, which corresponds to the rightmost term; to get any further, we apply Lemma 15:

$$(\gamma_{\ell+1}^{k+j})^2 \le (\gamma_{\ell+1}^{k+j_1+3})^2 \le (\gamma_{\ell+1}^{k+j_1+2})^2 \le \delta_2^2\,(\gamma_{\ell+1}^{k+j_1})^2.$$

Note that the last inequality also holds when $j = j_1,\ j_1+1$ (in place of $j = j_1+2$), since we assumed $\delta_2 > 1$. Thus our conclusion is

$$(\gamma_{\ell+1}^{k+j})^2 \le M_2\,\delta_2^2\,\varepsilon_\ell\,\|\mathbf{g}_k\|^2, \qquad j_1 \le j \le j_2+1.$$

Since j_1 and j_2 are chosen arbitrarily and $j_0 \le m_{\ell+1}$, the result automatically holds for any $j \ge m_{\ell+1}$.

Part III. Finally, we prove by induction that (25) holds for all $1 \le \ell \le n$ with

$$\varepsilon_\ell = \tfrac{1}{4}\,(1 + M_2\,\delta_2^2)^{\ell-n}.$$

For $\ell = 1$ and from Lemma 14, the first component of the gradient satisfies $G(k+j,1) \le \delta_1^j\,\|\mathbf{g}_k\|^2$. As in the first step, we ask $\delta_1^j \le \varepsilon_1$ and get $j \ge m_1$, with $m_1 = \lceil \frac{\log \varepsilon_1}{\log \delta_1} \rceil$. Once the thesis is true for some $1 \le \ell \le n-1$, the second step shows that it also holds for $\ell+1$, with $m_{\ell+1} = m_\ell + \Delta_\ell + 1$ and $\varepsilon_{\ell+1} = \frac{1}{4}\,(1+M_2\,\delta_2^2)\,(1+M_2\,\delta_2^2)^{\ell-n}$. We can conclude that the thesis holds for $\ell = n$, and thus

$$\|\mathbf{g}_{k+m_n}\|^2 \le \tfrac{1}{4}\,\|\mathbf{g}_k\|^2,$$

where m_n does not depend on k. Renumbering the indices as in [4] enables us to conclude that the $\mathbf{g}_k$ converge to zero R-linearly. □

We remark that, when the objective function is quadratic, Theorem 16 shows that no line search is required to guarantee the convergence of the gradient method with TBB stepsizes. For generic unconstrained optimization problems, we add a line search procedure with a condition of *sufficient decrease* in the next section.

6 Generic nonlinear functions

We now turn our attention to generic (non-quadratic) continuous differentiable functions f. As the expression (4) is not suitable since the Hessian is usually not available, we use the generalization (6). Just as for the quadratic case (see Proposition 1), the well-known BB1 and BB2 stepsizes are retrieved for $\tau_k \to \pm\infty$ and $\tau_k = 0$, respectively. In the quadratic case, the secant equation $\mathbf{y}_{k-1} = B_k \mathbf{s}_{k-1}$ holds for $B_k \equiv A$, and this allows for the interpretation of BB steps and the TBB step as Rayleigh quotients of A. When f is a generic function, the average Hessian $B_k = \int_0^1 \nabla^2 f(\mathbf{x}_k + t\mathbf{s}_k)\,dt$ satisfies the secant equation (cf., e.g., [17, Eq. (6.11)]), and thus we can still think at BB steps and the TBB step as Rayleigh quotients, which approximate the eigenvalues of this average B_k instead of the ones of $\nabla^2 f(\mathbf{x}_k)$. We note that, under the condition that B_k is SPD, all results of Sections 3 and 4 continue to hold for generic functions when replacing A by B_k.

Algorithm 2 shows a pseudocode for TBB-step methods for general nonlinear unconstrained optimization problems. As usual (cf., e.g., [2, Alg. 1]), unlike the quadratic case of Algorithm 1, safeguards are added for the steplength.

We also include the nonmonotone line search strategy from [18] (see Line 2). Line 3 features a well-known condition of sufficient decrease, where we take the common values of the line search parameters $c_{\mathrm{ls}} = 10^{-4}$, $\sigma_{\mathrm{ls}} = \frac{1}{2}$ (cf. [17, p. 33]), and $M = 10$ in the experiments. Global convergence of Algorithm 2 is proven in [16, Thm. 2.1]. In particular, it is possible to show R-linear convergence of the algorithm for uniformly convex functions (cf. [19, Thm. 3.1, Eq. (31)]). One of the key ingredients of [16, Thm. 2.1] is the uniform bounds on the stepsize, i.e., $\beta_k \in [\beta_{\min}, \beta_{\max}]$ for all k. Since the TBB stepsize (4) with safeguard lies in this interval, convergence of the algorithm is guaranteed; still, the initial starting steplength in the nonmonotone line search is different, which might lead to a smaller number of backtracking steps.

The last two features that may significantly affect the speed of the algorithm are the initial stepsize and the treatment of uphill directions, or, equivalently, negative steplengths. Popular choices for the initial stepsize are $\beta_0 = 1$ (cf., e.g., [2, 16]) or $\beta_0 = \|\mathbf{g}_0\|^{-1}$ (cf., e.g., [15]), where the norm is the Euclidean norm or the ∞-norm. Since the Hessian may be indefinite, we have to deal with possible uphill directions. We do this in line 7, where a negative steplength is replaced by a certain $\widehat{\beta} > 0$. A possible choice is $\widehat{\beta} = \beta_{\max}$ (see, e.g., [2]), but this stepsize is huge and could cause overflow problems. Raydan [16] proposes to set $\widehat{\beta} = \min(\max(\|\mathbf{g}_k\|_2^{-1},\, 10^{-5}),\, 1)$, which is an attempt to move away from the uphill direction, while keeping $\|\widehat{\beta}\,\mathbf{g}_k\|$ moderate. Others (e.g., [11, 20]) simply use $\widehat{\beta} = \|\mathbf{g}_k\|^{-1}$, as it is done for the first stepsize. We adopt the interesting alternative of [21] and reuse the previous steplength $\widehat{\beta} = \beta_k$; this strategy resembles the cyclic gradient method, where the same BB stepsize is reused for several iterations [20].

Algorithm 2 A TBB method for general nonlinear functions

Input: continuous differentiable function f, initial guess $\mathbf{x}_0$, initial stepsize $\beta_0 > 0$, tolerance tol; safeguarding parameters $\beta_{\max} > \beta_{\min} > 0$; line search parameters c_{ls}, $\sigma_{\mathrm{ls}} \in (0,1)$; memory integer $M > 0$; replacement for negative stepsizes $\widehat{\beta} > 0$

Output: approximation to minimizer $\mathrm{argmin}_{\mathbf{x}}\, f(\mathbf{x})$

1: Set $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$
  for $k = 0, 1, \dots$
2: $\nu_k = \beta_k, \;\; f_{\mathrm{ref}} = \max\{\, f(\mathbf{x}_{k-j}) \,:\, 0 \le j \le \min(k, M-1)\,\}$
3: **while** $f(\mathbf{x}_k - \nu_k\, \mathbf{g}_k) > f_{\mathrm{ref}} - c_{\mathrm{ls}}\, \nu_k\, \|\mathbf{g}_k\|^2$ **do** $\nu_k = \sigma_{\mathrm{ls}}\, \nu_k$ **end**
4: Set $\mathbf{s}_k = -\nu_k\, \mathbf{g}_k$ and update $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$
3: Compute the gradient $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$
4: **if** $\|\mathbf{g}_{k+1}\| \le \mathsf{tol} \cdot \|\mathbf{g}_0\|$, **return**, **end**
5: $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$
6: Choose τ_{k+1}, compute the TBB step $\beta_{k+1}(\tau_{k+1})$ according to (6)
7: **if** $\beta_{k+1} < 0$, set $\beta_{k+1} = \widehat{\beta}$
8: Set $\beta_{k+1} = \min(\max(\beta_{k+1},\, \beta_{\min}),\, \beta_{\max})$

Algorithms 1 and 2 can be combined with preconditioning or scaling. Scaling may be viewed as the simplest case of preconditioning, that is, by a diagonal SPD matrix. Scaling is a powerful and efficient technique; we refer to [22] for scaling techniques for unconstrained optimization problems. The combination of preconditioning and BB steps for quadratic problems has been discussed in [23]. The use of scaling or more general preconditioning is outside the scope of this paper.

7 Numerical experiments

We test different target strategies on strictly convex quadratics (Algorithm 1) and generic differentiable functions (Algorithm 2). The purpose of these experiments is to show numerically that the introduction of an adaptive target in the stepsizes (4) can sometimes lead to better convergence results, in terms of number of iterations and function evaluations.

7.1 Strictly convex quadratic functions

For the problems of the form (1), we experiment with examples from the Suite Sparse Matrix Collection [24]. The selected matrices A are 65 symmetric positive definite matrices with a number of rows in $[10^2, 10^4]$, and an estimated condition number $\leq 10^8$ (the condition number is estimated via the routine condest in the Matrix R package). The vector $\mathbf{b}$ is chosen so that the solution of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}^* = \mathbf{e}$, the vector of all ones. For all problems, the starting vector is $\mathbf{x}_0 = -10\,\mathbf{e}$, and the initial stepsize is $\beta_0 = 1$. The algorithm stops when $\|\mathbf{g}_k\| \leq \mathsf{tol}\,\|\mathbf{g}_0\|$ with $\mathsf{tol} = 10^{-6}$, or when $5 \cdot 10^4$ iterations are reached. The problem nos4 is scaled by the Euclidean norm of the first gradient.

Table 1 reports all the implemented target strategies, divided into three groups: known schemes from the literature, positive targets, and targets inspired by the cotangent function. With respect to the positive targets, Remark 13 suggests the use of $\rho > 2$ in (14) to ensure the convergence of the corresponding gradient method. The aim of setting $\rho = 2.01$ is to stay close to this lower bound and take the largest possible stepsizes (which are larger than β^{BB1}). An approach with $\rho = 100$ picks positive targets τ_k such that the corresponding stepsize β_k is close to, but still larger than β^{BB1}.

We compare the performances of different stepsizes by the required number of iterations to get convergence, by means of the performance profile [25]. The cost of solving each problem is normalized based on the minimum cost for that problem, to get the *performance ratio* [25]. The most efficient method solves the given problem with performance ratio 1, while all other methods solve it with a performance ratio at least 1. We plot the ratio of problems solved by a method within a certain factor of the smallest cost; this results in a cumulative distribution for each method. The algorithms are rated based on the maximum cost that one is willing to pay to get convergence. An infinite cost is assigned whenever a method is not able to solve a problem to the tolerance within the maximum number of iterations.

Table 1 Strategies for the stepsize.

Method	Target τ_k	Reference
BB1	$\pm\infty$	Cf. [1]
BB2	0	Cf. [1]
ABB	(NA)	Cf. [3], $\eta = 0.8$
IBB2 2.01	$2.01\,\alpha_k^{\mathrm{BB2}}$	Eq. (14)
IBB2 100	$100\,\alpha_k^{\mathrm{BB2}}$	Eq. (14)
ITER	$k\,\alpha_k^{\mathrm{BB2}}$	Eq. (15)
COT 11	$-\cot(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$	Eq. (16)
COT H1	$-\cos^{1/2}(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \;/\; \sin(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$	Eq. (16)
COT 1H	$-\cos(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \;/\; \sin^{1/2}(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$	Eq. (16)
COT 21	$-\cos^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \;/\; \sin(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$	Eq. (16)
COT 12	$-\cos(\mathbf{s}_{k-1}, \mathbf{y}_{k-1}) \;/\; \sin^2(\mathbf{s}_{k-1}, \mathbf{y}_{k-1})$	Eq. (16)

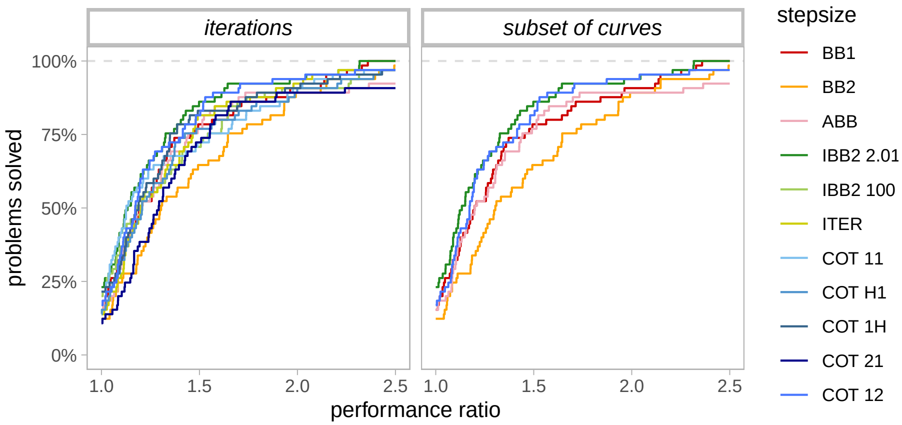

Fig. 2 Performance profile for strictly convex quadratic problems. On the right we plot COT 12 (blue) and IBB2 2.01 (green) against BB1 (red), BB2 (orange), ABB (pink).

Figure 2 reports the performance profile of the different stepsizes, based on the number of iterations. The performance ratio is considered in the interval [1, 2.5], as at 2.5 almost all methods reach the maximal number of solved problems. From the left plot of Figure 2, it seems that the new stepsize schemes (the cotangent family and positive multiples of α^{BB2}) are quite competitive. On the right plot of Figure 2, we take a subset of the curves to make this difference clearer.

We collect some information coming from the performance profile in Table 2. All methods are robust, since they all solve more than 95% of the given problems. Among them, BB1, ITER, IBB2 2.01, IBB2 100 solve all problems, and IBB2 2.01 does it with the lower average cost. IBB2 2.01 seems to be the winner on this set of quadratic problems: it solves the highest rate of problems with cost 1, with a low average cost and the most narrow cost range. COT 12 has the most narrow range of costs, with the lowest average, but it is less robust than IBB2 2.01. Two methods, BB2 and COT 21 have higher (i.e. worse) average performance ratio than all other methods. By looking at

Table 2 Gradient method with TBB steps. From left to right: proportion of solved problems, proportion of problems solved with unit cost, i.e., performance ratio (PR) equal to 1, average and standard deviation of the performance ratio, range of the performance ratio.

Stepsize	Solved	PR = 1	Avg	Sd	Range
IBB2 2.01	100 %	23 %	1.28	0.33	[1.00, 2.32]
IBB2 100	100 %	20 %	1.38	0.43	[1.00, 2.89]
BB1	100 %	15 %	1.36	0.38	[1.00, 2.36]
ITER	100 %	14 %	1.40	0.48	[1.00, 3.94]
COT H1	98 %	12 %	1.39	0.42	[1.00, 2.92]
BB2	98 %	12 %	1.47	0.40	[1.00, 2.49]
COT 1H	97 %	22 %	1.29	0.32	[1.00, 2.43]
COT 11	97 %	22 %	1.33	0.37	[1.00, 2.34]
COT 12	97 %	17 %	1.26	0.27	[1.00, 2.29]
ABB	95 %	17 %	1.33	0.38	[1.00, 2.76]
COT 21	95 %	11 %	1.41	0.53	[1.00, 4.42]

the performance profile, we add that those bad performances are related to a small number of problems, since the profile of COT 21 is close to the profile of other methods at 2.5.

7.2 Unconstrained optimization

We take some generic differentiable functions from the collection in [16, 26, 27] and the suggested starting points therein, as listed in Table 3.

Table 3 Unconstrained optimization test problems.

Name	Reference	Name	Reference
Brown almost linear	[27]	generalized White and Holst	[26]
Broyden tridiagonal	[27]	penalty 1	[27]
extended Rosenbrock	[27]	strictly convex 1	[16]
extended Powell	[27]	strictly convex 2	[16]
Hager	[26]	trigonometric	[27]
generalized Rosenbrock	[26]	variably dimensioned	[27]
generalized tridiagonal 1	[26]		

We remark that the problems that come from the collection [27] are nonlinear least squares problems

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2} \sum_{i=1}^{m} f_i^2(x),$$

where m is a function of n. For all the test functions, we pick $n \in \{10^2, 10^3, 10^4\}$. The generalized Rosenbrock, generalized White and Holst and extended Powell objective functions have been scaled by the Euclidean norm of the first gradient.

The gradient method for unconstrained optimization problems requires the tuning of more parameters than the gradient method for quadratic functions. We maintain the choices made in [2] and set $\beta_{\min} = 10^{-30}$, $\beta_{\max} = 10^{30}$,

$c_{\text{ls}} = 10^{-4}$, $\sigma_{\text{ls}} = \frac{1}{2}$, $M = 10$ and $\beta_0 = 1$. One may argue that the bounds on the stepsize are extremely large: the aim of this choice is to accept the BB stepsize as frequently as possible. Again the algorithm stops when $\|\mathbf{g}_k\| \leq \mathsf{tol}\,\|\mathbf{g}_0\|$, with $\mathsf{tol} = 10^{-8}$, or when $5 \cdot 10^4$ iterations are reached. All different steps in Table 1 are tested.

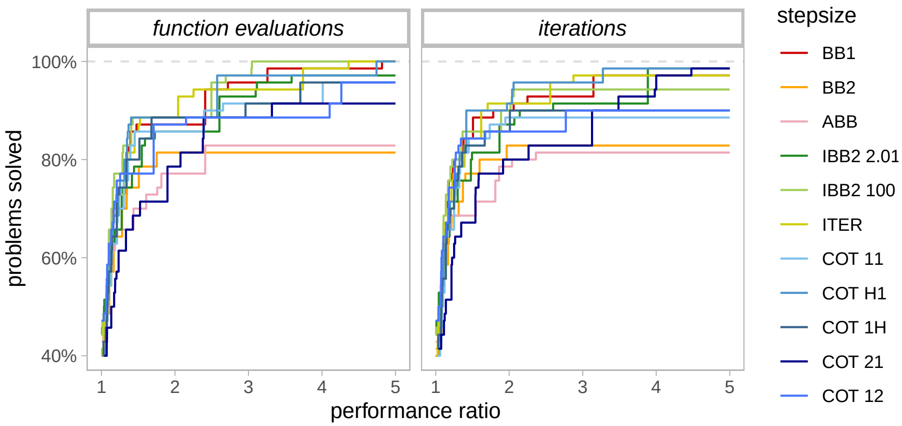

Fig. 3 Performance profiles for generic unconstrained optimization problems.

As the performance profile, we may consider two different costs: the number of function evaluations and the number of iterations. The latter corresponds to the number of gradient evaluations, since the line search in Algorithm 2 does not require the computation of the gradient at the new tentative iterate. The performance profiles are showed in Figure 3 in the range [1, 5]. In the computation of the performance ratio, only those problems that were solved by at least one method are reported. Most of the methods perform equally well; only BB2 and ABB solve a smaller amount of problem than other methods within cost 5. As in the quadratic case, COT 21 starts lower than the other curves, but joins them when the performance ratio grows.

As in the quadratic case, Table 4 reports some statistics on the performance ratios, based on the number of function evaluations. The steps ABB and BB2 solve the smallest number of problems, while again IBB2 2.01, IBB2 100, ITER and BB1 solve all problems. In addition, IBB2 100 has a narrow cost range, low average cost and standard deviation. COT 1H has similar features but is less robust.

At first glance, it is surprising that all methods solve more than 34% of problems with cost 1. This fact is explained by Table 5, that shows the average performance ratio along with its standard deviation for each problem, taking into account all methods. Some values are missing because none of the considered methods was able to solve the corresponding problems. The performance ratio is given by the number of function evaluations. Problems strictly convex 1, variably dimensioned, Brown almost linear and penalty 1 show the same exact number of function evaluations independently from the selected target. This

Table 4 Gradient method with nonmonotone line search. From left to right: proportion of solved problems, proportion of problems solved with unit cost, i.e., performance ratio (PR) equal to 1, average and standard deviation of the performance ratio, range of the performance ratio. The performance ratio is based on the number of function evaluations.

Stepsize	Solved	PR = 1	Avg	Sd	Range
IBB2 100	100 %	44 %	1.28	0.58	[1.00, 3.05]
IBB2 2.01	100 %	44 %	1.48	0.93	[1.00, 5.15]
ITER	100 %	40 %	1.33	0.74	[1.00, 4.36]
BB1	100 %	40 %	1.34	0.79	[1.00, 4.82]
COT H1	100 %	39 %	1.26	0.67	[1.00, 4.74]
COT 21	100 %	34 %	1.51	0.95	[1.00, 5.90]
COT 12	96 %	47 %	1.46	1.03	[1.00, 4.26]
COT 1H	96 %	37 %	1.39	0.80	[1.00, 3.71]
COT 11	96 %	37 %	1.43	0.88	[1.00, 4.01]
ABB	87 %	43 %	1.81	2.06	[1.00, 8.03]
BB2	87 %	41 %	1.75	2.02	[1.00, 7.91]

Table 5 Average and standard deviation of the performance ratio per problem and size, based on the number of function evaluations.

	$n = 100$		$n = 1000$		$n = 10000$	
	avg	sd	avg	sd	avg	sd
Brown almost linear	1.00	0.00	1.00	0.00	1.00	0.00
Broyden tridiagonal	1.20	0.16	1.27	0.19	1.30	0.60
extended Powell	3.29	1.31	2.28	1.10	1.68	0.52
extended Rosenbrock	3.26	2.64	3.26	2.64	3.26	2.64
Hager	1.16	0.08	1.15	0.17	1.25	0.22
generalized Rosenbrock	1.26	0.32	1.03	0.05		
generalized tridiagonal 1	1.09	0.06	1.08	0.04	1.08	0.05
generalized White and Holst	1.15	0.12				
penalty 1	1.00	0.00	1.00	0.00	1.00	0.00
strictly convex 1	1.00	0.00	1.00	0.00	1.00	0.00
strictly convex 2	1.27	0.20	1.65	0.69	2.19	1.47
trigonometric	1.39	0.68	1.15	0.27		
variably dimensioned	1.00	0.00	1.00	0.00	1.00	0.00

means that each inverse stepsize α_k is an eigenvalue for the average Hessian B_k, thus for any target choice, the stepsize does not change (cf. comment to Proposition 1). It is also interesting to notice how the performances of the methods does not change with the size of the problem for extended Rosenbrock and generalized tridiagonal 1.

In summary we saw that the TBB steps may improve the well-known BB steps and the ABB method. While TBB steps with positive target are robust and tend to solve the majority of the problems with a small cost, the "cotangent" family shows more variability in the performances.

8 Conclusions

We have developed a harmonic framework for stepsize selection in gradient methods for unconstrained nonlinear optimization. The harmonic steplength (4) depending on targets τ_k is inspired by the harmonic Rayleigh–Ritz extraction for matrix eigenvalue problems.

Because of the one-to-one relation between target and stepsize, this gives a general framework with new viewpoints and interpretations. Compared to the eigenproblem context, where the target is commonly chosen inside the spectrum, in our situation we have studied both strategies with the target outside the spectrum and schemes that sometimes pick the target inside. Targets on the negative real axis lead to stepsizes between BB2 and BB1. This yields connection with schemes such as [12]. We have analyzed and extended the popular ABB method. While the original ABB approach only allows a choice between two stepsizes based on a single parameter, we have introduced a new competitive family of stepsizes with tunable parameters, that enjoy the same key idea but are more flexible. Additionally, we have considered new families of positives targets, leading to steplengths larger than the BB1 steps. The use of harmonic stepsizes requires a very modest number of extra vector operations per iteration, with an equal number of function and gradient evaluations, and therefore this approach is only marginally more expensive per iteration compared to other strategies. On the other hand, the experiments show that the number of iterations may be considerably lower for some problems. The experiments suggest that both the cotangent family and the approaches with positive targets seem competitive.

For an analysis of the new schemes, we have extended convergence results from Dai and Liao [4] in Section 5. In view of the TBB steps, instead of $\lambda_1 \le \alpha_k \le \lambda_n$, we have studied the more general setting $\xi_{\text{low}}\, \lambda_1 \le \alpha_k \le \xi^{\text{up}}\, \lambda_n$, particularly for $\frac{1}{2} < \xi_{\text{low}} \le 1$ and $\xi^{\text{up}} \ge 1$.

An R implementation of the methods described in this paper can be obtained from `github.com/gferrandi/tbbr`.

Acknowledgment: This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 812912.

References

[1] Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**(1), 141–148 (1988)

[2] Di Serafino, D., Ruggiero, V., Toraldo, G., Zanni, L.: On the steplength selection in gradient methods for unconstrained optimization. Appl. Math. Comput. **318**, 176–195 (2018)

[3] Zhou, B., Gao, L., Dai, Y.H.: Gradient methods with adaptive step-sizes. Comput. Optim. Appl. **35**(1), 69–86 (2006)

[4] Dai, Y.H., Liao, L.Z.: R-linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**(1), 1–10 (2002)

[5] Morgan, R.B.: Computing interior eigenvalues of large matrices. Linear Algebra Appl. **154**, 289–309 (1991)

[6] Paige, C.C., Parlett, B.N., van der Vorst, H.A.: Approximate solutions and eigenvalue bounds from Krylov subspaces. Numer. Linear Algebra Appl. **2**(2), 115–133 (1995)

[7] Stewart, G.W.: Matrix Algorithms. Vol. II. SIAM, Philadelphia, PA (2001)

[8] Hochstenbach, M.E.: Generalizations of harmonic and refined Rayleigh–Ritz. Electron. Trans. Numer. Anal. **20**, 235–252 (2005)

[9] Parlett, B.N.: The Symmetric Eigenvalue Problem. SIAM, Philadelphia, PA (1998)

[10] Friedlander, A., Martínez, J.M., Molina, B., Raydan, M.: Gradient method with retards and generalizations. SIAM J. Numer. Anal. **36**(1), 275–289 (1998)

[11] Huang, Y., Dai, Y., Liu, X., Zhang, H.: On the acceleration of the Barzilai–Borwein method. Comput. Optim. Appl., 1–24 (2022)

[12] Dai, Y.H., Huang, Y., Liu, X.W.: A family of spectral gradient methods for optimization. Comput. Optim. Appl. **74**(1), 43–65 (2019)

[13] Frassoldati, G., Zanni, L., Zanghirati, G.: New adaptive stepsize selections in gradient methods. J. Ind. Manag. **4**(2), 299 (2008)

[14] Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. **13**(3), 321–326 (1993)

[15] Dai, Y.H.: Alternate step gradient method. Optimization **52**(4-5), 395–415 (2003)

[16] Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**(1), 26–33 (1997)

[17] Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York, NY, USA (2006)

[18] Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**(4), 707–716 (1986)

[19] Dai, Y.H.: On the nonmonotone line search. J. Optim. Theory Appl. **112**(2), 315–330 (2002)

[20] Dai, Y.H., Hager, W.W., Schittkowski, K., Zhang, H.: The cyclic Barzilai–Borwein method for unconstrained optimization. IMA J. Numer. Anal. **26**(3), 604–627 (2006)

[21] Park, Y., Dhar, S., Boyd, S., Shah, M.: Variable metric proximal gradient method with diagonal Barzilai–Borwein stepsize. In: ICASSP 2020, pp. 3597–3601 (2020). IEEE

[22] Luengo, F., Raydan, M., Glunt, W., Hayden, T.L.: Preconditioned spectral gradient method. Numer. Algorithms **30**(3), 241–258 (2002)

[23] Molina, B., Raydan, M.: Preconditioned Barzilai–Borwein method for the numerical solution of partial differential equations. Numer. Algorithms **13**(1), 45–60 (1996)

[24] Davis, T.A., Hu, Y.: The University of Florida sparse matrix collection. ACM Trans. Math. Softw. **38**(1), 1–25 (2011)

[25] Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)

[26] Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim **10**(1), 147–161 (2008)

[27] Moré, J.J., Garbow, B.S., Hillstrom, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**(1), 17–41 (1981)