JOTA manuscript No.
(will be inserted by the editor)

Analysis and Performance of the Barzilai-Borwein Step-Size Rules for Optimization Problems in Hilbert Spaces

Behzad Azmi · Karl Kunisch

Received: date / Accepted: date

Abstract The Barzilai and Borwein gradient method has received a significant amount of attention in different fields of optimization. This is due to its simplicity, computational cheapness, and efficiency in practice. In this research, based on spectral analysis techniques, root-linear global convergence for the Barzilai and Borwein method is proven for strictly convex quadratic problems posed in infinite-dimensional Hilbert spaces. The applicability of these results is demonstrated for two optimization problems governed by partial differential equations.

Keywords Barzilai-Borwein method · Hilbert spaces · R-linear rate of convergence · PDE-constrained optimization

Mathematics Subject Classification (2000) 65K05 · 49J20 · 49K20 · 93C20

1 Introduction

First-order (gradient) methods are progressively getting more attention since it has been realized that for a suitable choice of the step-length, using the negative gradient as the search direction may give rise to very efficient algorithmic behavior. As a pioneering work, we can refer to the method proposed by Barzilai and Borwein in [1] abbreviated as the BB-method. In this work, the authors demonstrated that choosing an appropriate step-length leads to a significant acceleration over the steepest descent method. The BB-method incorporates the quasi-Newton property, by approximating the Hessian matrix by a scalar times the identity which satisfies the secant condition in the sense of least squares. Despite the simplicity

Behzad Azmi, Corresponding author
Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Science, Altenberger Straße 69, A-4040 Linz, Austria
behzad.azmi@ricam.oeaw.ac.at

Karl Kunisch
Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010 Graz, Austria and Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenberger Straße 69, A-4040 Linz, Austria
karl.kunisch@uni-graz.at

and cheapness, this method has exhibited a surprisingly efficient numerical behaviour. This stimulated a significant amount of research. In the original work [1], the authors established R-superlinear (R stands for root) convergence for two-dimensional strictly convex quadratic problems. Later, Raydan [2] and Dai and Liao [3] proved, respectively, global convergence and R-linear convergence rate of the BB-method for any finite-dimensional strictly convex quadratic problem. One of the important feature of this method is the nonmonotonicity in the values of the objective function and of the gradient norm. A deep analysis of the asymptotic behaviour of the BB-method was given in [4,5]. In these works, the surprising computational efficiency of the algorithm in relation to its nonmonotonicity was discussed and several circumstances were presented under which the performance of the BB-method (without globalization) is competitive, or even, superior to conjugate gradient methods. This occurs, for instance, when a low accuracy for the solution of problem is required, or when significant round-off errors are present, and the objective functions is made up of a quadratic function plus a small non-quadratic term (near quadratic).

In this work we study the BB-method within the scope of PDE-constrained optimization. For optimization problems governed by partial different equations, every function evaluation is typically carried out through solving a partial differential equation (state equation). Hence function evaluations can be computationally very expensive and it is desirable to avoid them as far as possible. Moreover, due to numerical discretization, the presence of round-off and truncation errors is inevitable and, depending on the discretization procedure, the finite-dimensional approximation for the gradient of the original problem need not coincide with the gradient of the finite-dimensional approximation for the original problem (optimization and discretization do not commute). A wide range of models arising from industry and natural science can be formulated as optimization problems governed by linear and semilinear partial differential equations. For these problems, the corresponding reduced formulations lead to infinite-dimensional quadratic and near quadratic unconstrained optimization problems.

In view of the above discussion, we are motivated to study the BB-method for problems posed in infinite-dimensional Hilbert spaces.

As mentioned before, numerous results have been published on the BB-method, but, to the best of our knowledge, for optimization problems posed in infinite-dimensional spaces, there still does not exist a rigorous theory. Here we take a step in this direction and we analyse the convergence of the BB-method. Inspired by the result in [3] and based on the spectral theorem, we establish R-linear global convergence of the BB-method for strictly convex quadratic problems defined by bounded uniformly positive self-adjoint operators. Our theoretical framework is supported by PDE constrained optimal control problems.

The rest of paper is organized as follows: The optimization problem and the BB-method are specified in Section 2. In Section 3, we first recall some concepts from the spectral theory for bounded self-adjoint operators. We then deal with the global convergence analysis for strictly convex quadratic functions defined by bounded self-adjoint operators. In Section 4, an optimal control problem for the wave equation is presented. Finally, in Section 5 numerical results are given.

2 Problem Formulation and Algorithm

Here we are concerned with the following quadratic programming in an abstract Hilbert space $\mathcal{H}$

$$\min_{u\in\mathcal{H}} \mathcal{F}(u) := \frac{1}{2}(\mathcal{A}u, u) - (b, u), \tag{QP}$$

where $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is a bounded self-adjoint uniformly positive operator and $b \in \mathcal{H}$. $\mathcal{H}$ is endowed with an inner product $(\cdot, \cdot)$ from which is derived the norm $\|\cdot\|$. The Barzilai-Borwein iterations for solving (QP) are defined by

$$u_{k+1} = u_k - \frac{1}{\alpha_k}\mathcal{G}_k, \tag{1}$$

where $\mathcal{G}_k := \mathcal{G}(u_k)$ and $\mathcal{G} : \mathcal{H} \to \mathcal{H}$ stands for the gradient of $\mathcal{F}$. This gradient is defined by $\mathcal{G} := \mathcal{R} \circ \mathcal{F}'$, where $\mathcal{F}' : \mathcal{H} \to \mathcal{H}'$ is the first derivative of $\mathcal{F}$, and $\mathcal{R} : \mathcal{H}' \to \mathcal{H}$ is the Riesz isomorphism, with $\mathcal{H}'$ denoting the dual space of $\mathcal{H}$. Thus for every $\delta u \in \mathcal{H}$, we have $\mathcal{F}'(u)\delta u = (\mathcal{G}(u), \delta u)$, with $(\cdot, \cdot)$ denoting the inner product in $\mathcal{H}$. Furthermore, the step-size $\alpha_k > 0$ is chosen according to either

$$\alpha_k^{BB1} := \frac{(\mathcal{S}_{k-1}, \mathcal{Y}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{S}_{k-1})}, \quad \text{or} \quad \alpha_k^{BB2} := \frac{(\mathcal{Y}_{k-1}, \mathcal{Y}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{Y}_{k-1})}, \tag{2}$$

where $\mathcal{S}_{k-1} := u_k - u_{k-1}$ and $\mathcal{Y}_{k-1} := \mathcal{G}_k - \mathcal{G}_{k-1}$. With these specifications we are prepared to specify Algorithm 1 which will be investigated in this paper.

Algorithm 1 BB-gradient

Require: Let initial iterates $u_{-1}, u_0 \in \mathcal{H}$ with $u_{-1} \neq u_0$ have been given.
1: Set $k = 0$.
2: If $\|\mathcal{G}_k\| = 0$ stop.
3: Choose α_k equal to either α_k^{BB1} or α_k^{BB2}.
4: Set $u_{k+1} = u_k - \frac{1}{\alpha_k}\mathcal{G}_k$, $k = k + 1$, and go to Step 2.

3 Convergence Analysis

Here the convergence of Algorithm 1 applied to (QP) is investigated. The analysis is inspired by the works [3, 2] for finite-dimensional quadratic problems, but we deal with the more general class of self-adjoint operators on infinite-dimensional spaces whose spectrum may be a continuous set, or a countably infinite set with a finite number of accumulation points. This requires a new analysis based on the spectral theorem, see also Remark 3.1 below. Strictly convex quadratic problems are not only of importance in their own right, but also as a model for twice continuously Fréchet-differentiable functions in a neighbourhood of strong minima.
Due to the structure of (QP), we infer that $\mathcal{G}_k := \mathcal{G}(u_k) = \mathcal{A}u_k - b$ and it can easily be shown for every $k \geq 0$ that

$$\alpha_k^{BB1} = \frac{(\mathcal{S}_{k-1}, \mathcal{A}\mathcal{S}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{S}_{k-1})} \quad \text{and} \quad \alpha_k^{BB2} = \frac{(\mathcal{S}_{k-1}, \mathcal{A}^2\mathcal{S}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{A}\mathcal{S}_{k-1})}. \tag{3}$$

Moreover, using (3) and the fact that $\mathcal{S}_{k-1} = -\frac{1}{\alpha_{k-1}}\mathcal{G}_{k-1}$ for every $k \geq 1$, we can infer that

$$\alpha_k^{BB1} = \frac{(\mathcal{G}_{k-1}, \mathcal{A}\mathcal{G}_{k-1})}{(\mathcal{G}_{k-1}, \mathcal{G}_{k-1})} \quad \text{and} \quad \alpha_k^{BB2} = \frac{(\mathcal{G}_{k-1}, \mathcal{A}^2\mathcal{G}_{k-1})}{(\mathcal{G}_{k-1}, \mathcal{A}\mathcal{G}_{k-1})} \quad \text{for } k \geq 1. \tag{4}$$

We define the numerical range $\mathcal{W}(\mathcal{A}) \subset \mathbb{R}$ of $\mathcal{A}$ by

$$\mathcal{W}(\mathcal{A}) := \{(u, \mathcal{A}u) : u \in \mathcal{H}, \|u\| = 1\}.$$

This set is convex and contains all the eigenvalues of $\mathcal{A}$. Moreover using (3) and the fact that

$$\alpha_k^{BB2} = \frac{(\mathcal{S}_{k-1}, \mathcal{A}^2\mathcal{S}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{A}\mathcal{S}_{k-1})} = \frac{(\bar{\mathcal{S}}_{k-1}, \mathcal{A}\bar{\mathcal{S}}_{k-1})}{(\bar{\mathcal{S}}_{k-1}, \bar{\mathcal{S}}_{k-1})},$$

with $\bar{\mathcal{S}}_{k-1} := \mathcal{A}^{\frac{1}{2}}\mathcal{S}_{k-1}$, we infer that $\alpha_k^{BB1}, \alpha_k^{BB2} \in \mathcal{W}(\mathcal{A})$ for all $k \geq 0$. Therefore, if we define the strictly positive constants $\delta_{\inf}$ and $\delta_{\sup}$ by

$$\delta_{\inf} := \inf \mathcal{W}(\mathcal{A}), \quad \delta_{\sup} := \sup \mathcal{W}(\mathcal{A}),$$

we can write

$$\alpha_k^{BB1}, \alpha_k^{BB2} \in [\delta_{\inf}, \delta_{\sup}] \quad \text{for all } k \geq 0. \tag{5}$$

To exclude trivial cases we assume throughout that $\delta_{\inf} < \delta_{\sup}$. For the following analysis we recall some facts from spectral theory. The spectrum $\sigma(\mathcal{A})$ of $\mathcal{A}$ is a closed subset of the interval $[\delta_{\inf}, \delta_{\sup}]$ with $\delta_{\inf}, \delta_{\sup} \in \sigma(\mathcal{A})$ and since $\mathcal{A}$ is a normal operator, we have $\overline{\mathcal{W}(\mathcal{A})} = \mathbf{conv}(\sigma(\mathcal{A})) = [\delta_{\inf}, \delta_{\sup}]$, where $\mathbf{conv}(S)$ denotes the convex hull of the set S. Hence the interval $[\delta_{\inf}, \delta_{\sup}]$ is completely determined by the spectrum $\sigma(\mathcal{A})$.

Further, due to the spectral theorem [6,7], there exists a unique spectral measure E on $\mathbb{R}$ which is supported on $\sigma(\mathcal{A})$, and whose range is the set of orthogonal projections in $\mathcal{H}$, such that

$$\mathcal{A} = \int_{\sigma(\mathcal{A})} \lambda \, dE_\lambda.$$

For every bounded measurable function $f : \sigma(\mathcal{A}) \to \mathbb{R}$, the operator $f(\mathcal{A})$ is defined by

$$f(\mathcal{A}) = \int_{\sigma(\mathcal{A})} f(\lambda) \, dE_\lambda, \tag{6}$$

and for every $x, y \in \mathcal{H}$ we have

$$(f(\mathcal{A})x, y) = \int_{\sigma(\mathcal{A})} f(\lambda) d(E_\lambda x, y), \tag{7}$$

where $d(E_\lambda x, y)$ stands for the integration with respect to the Borel measure $A \mapsto (E_A x, y)$ where $A \subseteq \sigma(\mathcal{A})$ is an arbitrary Borel set.
From (1) we have

$$\mathcal{G}_{k+1} = \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k \quad \text{for all } k = 0, 1, \ldots. \tag{8}$$

For $\mathcal{G}_0 \in \mathcal{H}$ we find

$$\mathcal{G}_0 = \int_{\sigma(\mathcal{A})} dE_\lambda \, \mathcal{G}_0, \quad \text{and} \quad \|\mathcal{G}_0\|^2 = \int_{\sigma(\mathcal{A})} d(E_\lambda \mathcal{G}_0, \mathcal{G}_0).$$

Using (6) and (8), we have

$$\mathcal{G}_1 = \frac{1}{\alpha_0}(\alpha_0 - \mathcal{A})\mathcal{G}_0 = \int_{\sigma(\mathcal{A})} \frac{1}{\alpha_0}(\alpha_0 - \lambda)\, dE_\lambda \mathcal{G}_0,$$

and, in a similar manner, we obtain

$$\mathcal{G}_k = \int_{\sigma(\mathcal{A})} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)\right] dE_\lambda \mathcal{G}_0 \quad \text{for every } k = 0, 1, \ldots .$$

where $\prod_{p=0}^{-1} = 1$. Moreover, we can write for $k = 0, 1, \ldots$

$$\begin{aligned} \|\mathcal{G}_{k+1}\|^2 &= \left(\frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k, \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k\right) \\ &= \left(\frac{1}{\alpha_k^2}(\alpha_k \mathcal{I} - \mathcal{A})^2 \mathcal{G}_k, \mathcal{G}_k\right) = \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k}\right)^2 d(E_\lambda \mathcal{G}_k, \mathcal{G}_k). \end{aligned} \tag{9}$$

Similarly, we have

$$\begin{aligned} \|\mathcal{G}_k\|^2 &= \left(\left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_0, \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_0\right) \\ &= \left(\left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)^2\right] \mathcal{G}_0, \mathcal{G}_0\right) = \int_{\sigma(\mathcal{A})} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_0, \mathcal{G}_0). \end{aligned} \tag{10}$$

We define $\gamma_{\mathcal{A}} := \frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}$ and $\rho_{\mathcal{A}} := \frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\sup}}$ and, from now on, for simplicity we used the notation

$$d_{\lambda,k} := d(E_\lambda \mathcal{G}_k, \mathcal{G}_k) \quad \text{for } k \geq 0.$$

These quantities will be used frequently in the proofs. First we investigate the special case in which $\delta_{\sup} < 2\delta_{\inf}$. In this case, it can be shown that $\gamma_{\mathcal{A}} < 1$.

Theorem 3.1 *Let $\delta_{\sup} < 2\delta_{\inf}$. Then the sequence $\{u_k\}_k$ generated by Algorithm 1 converges Q-linearly to the solution u^* of* (QP) *with the rate $\gamma_{\mathcal{A}}$.*

Proof Recall that by (9), we have for $k \geq 0$ that

$$\|\mathcal{G}_{k+1}\|^2 = \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k}\right)^2 d_{\lambda,k}. \tag{11}$$

Since $\delta_{\sup} < 2\delta_{\inf}$, it follows for every $k \geq 0$ and $\lambda \in \sigma(\mathcal{A})$ that

$$\left|\frac{\alpha_k - \lambda}{\alpha_k}\right|^2 \leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}\right)^2 = \gamma_{\mathcal{A}}^2 < 1. \tag{12}$$

Using (11) and (12), we obtain

$$\|\mathcal{G}_{k+1}\|^2 \leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}\right)^2 \int_{\sigma(\mathcal{A})} d_{\lambda,k} = \gamma_{\mathcal{A}}^2 \|\mathcal{G}_k\|^2 \quad \text{for every } k \geq 0. \tag{13}$$

Therefore, we can conclude that $\|\mathcal{G}_k\|^2 \leq \gamma_{\mathcal{A}}^{2k} \|\mathcal{G}_0\|^2$ for $k \geq 0$, and this completes the proof. □

If we lift the condition $\delta_{\sup} < 2\delta_{\inf}$, we attain the following result.

Theorem 3.2 *Let $\{u_k\}_k$ be the sequence generated by Algorithm 1 for finding the global minimum u^* of* (QP)*. Then either $u_k = u^*$ for a finite k, or the sequence $\{u_k\}_k$ converges R-linearly to u^*.*

The proof requires several lemmas and will be given in the remainder of this section. First, we need to define some quantities that will be used throughout the results. For any given $\eta > 0$, we denote $a_i := \delta_{\inf} + (i-1)\eta$ for every i with $1 \le i \le n^u_\eta$, and

$$b_i := \begin{cases} \delta_{\inf} + i\eta & \text{for } 1 \le i \le n^u_\eta - 1, \\ \delta_{\sup} & \text{for } i = n^u_\eta, \end{cases}$$

where $n^u_\eta := \lfloor \frac{\delta_{\sup}-\delta_{\inf}}{\eta} \rfloor + 1$. Then, clearly, $b_{i-1} = a_i$ for every $i = 2, \dots, n^u_\eta$ and we can define the following family of pairwise disjoint intervals

$$I_i = \begin{cases} [a_i, b_i) & \text{for } 1 \le i \le n^u_\eta - 1, \\ [a_{n^u_\eta}, b_{n^u_\eta}] & \text{for } i = n^u_\eta. \end{cases} \tag{14}$$

By construction it is clear that $|I_i| \le \eta$ for every $i = 1, \dots, n^u_\eta$ where $|I_i|$ denotes the length of I_i , and

$$\sigma(\mathcal{A}) \subseteq [\delta_{\inf}, \delta_{\sup}] = \bigcup_{i=1}^{n^u_\eta} I_i. \tag{15}$$

For $i = 1, \cdots, n^u_\eta$, we define

$$(g^k_i)^2 := \int_{I_i} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0}, \tag{16}$$

and attain

$$\begin{aligned} \|\mathcal{G}_k\|^2 &= \int_{\sigma(\mathcal{A})} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} = \int_{\delta_{\inf}}^{\delta_{\sup}} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} \\ &= \sum_{i=1}^{n^u_\eta} \int_{I_i} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} = \sum_{i=1}^{n^u_\eta} (g^k_i)^2. \end{aligned} \tag{17}$$

Moreover, we define

$$G(k, \ell) := \sum_{i=1}^{\ell} (g^k_i)^2 \quad \text{for every } k \ge 0 \text{ and } 1 \le \ell \le n^u_\eta, \tag{18}$$

where, n^u_η defined as above, depends on $\eta > 0$, and g^k_i given in (16). Then it is clear that

$$G(k, n^u_\eta) = \sum_{i=1}^{n^u_\eta} (g^k_i)^2 = \|\mathcal{G}_k\|^2 \quad \text{for every } k \ge 0.$$

In the following lemma we show that there exists an index n^l_η such that the sequences $\{g^k_i\}_k$ with $1 \le i \le n^l_\eta$ converge to zero Q-linearly as k tends to infinity.

Lemma 3.1 *For every $\eta \in\,]0, \rho_{\mathcal{A}}\delta_{\inf}]$, there exists an integer n_η^l with $1 \le n_\eta^l \le n_\eta^u$ such that for every $1 \le i \le n_\eta^l$, the sequences $\{g_i^k\}_k$ converge to zero Q-linearly with the factor $\rho_{\mathcal{A}}$ as k tends to infinity.*

Proof Choose $n_\eta^l \in \{1, \cdots, n_\eta^u\}$ as the largest integer such that

$$\bigcup_{i=1}^{n_\eta^l} I_i \subseteq [\delta_{\inf}, (1+\rho_{\mathcal{A}})\delta_{\inf}].$$

Observe that this is well-defined since $\eta \le \rho_{\mathcal{A}}\delta_{\inf}$. Moreover, for every $\lambda \in I_i$ with $1 \le i \le n_\eta^l$ and every $p \ge 1$ we have the following two cases:

1. If $\alpha_p - \lambda \ge 0$, then we have

$$\left|\frac{\alpha_p - \lambda}{\alpha_p}\right| = \frac{\alpha_p - \lambda}{\alpha_p} \le \rho_{\mathcal{A}} < 1.$$

2. If $\alpha_p - \lambda < 0$, then clearly both of α_p and λ belong to $[\delta_{\inf}, (1+\rho_{\mathcal{A}})\delta_{\inf}]$ and we can write

$$\left|\frac{\alpha_p - \lambda}{\alpha_p}\right| = \frac{\lambda - \alpha_p}{\alpha_p} \le \frac{(1+\rho_{\mathcal{A}})\delta_{\inf} - \delta_{\inf}}{\delta_{\inf}} = \rho_{\mathcal{A}}.$$

Therefore, we obtain

$$\begin{aligned}(g_i^{k+1})^2 &= \int_{I_i} \left[\prod_{p=0}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \le \int_{I_i} \left(\frac{\alpha_k - \lambda}{\alpha_k}\right)^2 \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\ &\le \rho_{\mathcal{A}}^2 \int_{I_i} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} = \rho_{\mathcal{A}}^2 (g_i^k)^2.\end{aligned} \tag{19}$$

This concludes the proof. □

Next we prove the following useful lemmas, which will be used later.

Lemma 3.2 *For any interval length $\eta \in\,]0, \frac{\delta_{\inf}}{2}[$, every integer ℓ with $n_\eta^l \le \ell \le n_\eta^u$, and $k \ge 1$, the following property holds:*
If the following condition

$$G(k+j, \ell) \le \overline{\zeta}\|\mathcal{G}_k\|^2 \quad \textit{for all } j \ge \overline{r} \tag{20}$$

holds for some positive numbers $\overline{r} \in \mathbb{N}$ and $\overline{\zeta} \in \mathbb{R}_+$, then there exists an integer $\hat{j} \in \{\overline{r}, \cdots, \overline{r} + \Theta + 1\}$ such that

$$(g_{\ell+1}^{k+\hat{j}})^2 \le 2\overline{\zeta}\|\mathcal{G}_k\|^2,$$

where $\Theta = \Theta(\overline{\zeta}, \overline{r}) := \left\lceil \frac{\log(2\overline{\zeta}\gamma_{\mathcal{A}}{}^{-2(\overline{r}+1)})}{2\log c} \right\rceil$ with $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2} + \frac{\eta}{\delta_{\inf}}\}$.

Proof Supposing that

$$(g_{\ell+1}^{k+j})^2 > 2\bar{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j \in \{\bar{r}, \cdots, \bar{r}+\Theta\}, \tag{21}$$

we will show that

$$(g_{\ell+1}^{k+\bar{r}+\Theta+1})^2 \leq 2\bar{\zeta}\|\mathcal{G}_k\|^2.$$

Due to (16), we have for every $k \geq 0$ that

$$\begin{aligned}
(g_{\ell+1}^{k+\bar{r}+1})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=0}^{k+\bar{r}} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\
&= \int_{I_{\ell+1}} \left[\prod_{p=k}^{k+\bar{r}} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\
&\leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}\right)^{2(\bar{r}+1)} \int_{I_{\ell+1}} \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\
&= \gamma_{\mathcal{A}}^{2(\bar{r}+1)} (g_{\ell+1}^k)^2 \leq \gamma_{\mathcal{A}}^{2(\bar{r}+1)} \|\mathcal{G}_k\|^2.
\end{aligned} \tag{22}$$

Due to Algorithm 1, for every $j \in \{\bar{r}, \cdots, \bar{r}+\Theta\}$ we have one of the cases $\alpha_{k+j} = \alpha_{k+j}^{BB1}$ or $\alpha_{k+j} = \alpha_{k+j}^{BB2}$. Further, using (8), the fact that $\mathcal{A}$ is self-adjoint, and the spectral property (7), we have for every $k \geq 0$ and $q = 0, 1, 2$, that

$$\begin{aligned}
(\mathcal{G}_k, \mathcal{A}^q\mathcal{G}_k) &= \left(\left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_0, \mathcal{A}^q \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_0\right) \\
&= \left(\mathcal{A}^q \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)^2\right] \mathcal{G}_0, \mathcal{G}_0\right) = \int_{\sigma(\mathcal{A})} \lambda^q \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\
&= \int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda^q \left[\prod_{p=0}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}.
\end{aligned} \tag{23}$$

Now, by using (4) and (23), we can write for $j \in \{\bar{r}, \cdots, \bar{r}+\Theta\}$ that

$$\alpha_{k+j+1}^{BB1} = \frac{(\mathcal{G}_{k+j}, \mathcal{A}\mathcal{G}_{k+j})}{(\mathcal{G}_{k+j}, \mathcal{G}_{k+j})} = \frac{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda \left[\prod_{p=0}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}}{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \left[\prod_{p=0}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}}, \tag{24}$$

and

$$\alpha_{k+j+1}^{BB2} = \frac{(\mathcal{G}_{k+j}, \mathcal{A}^2\mathcal{G}_{k+j})}{\mathcal{G}_{k+j}, \mathcal{A}\mathcal{G}_{k+j})} = \frac{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda^2 \left[\prod_{p=0}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}}{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda \left[\prod_{p=0}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}}. \tag{25}$$

Moreover, due to (16) and (20), we have

$$\begin{aligned}\int_{\bigcup_{i=1}^{\ell} I_i}\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} &= \sum_{i=1}^{\ell}\int_{I_i}\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}\\ &= \sum_{i=1}^{\ell}(g_i^{k+j})^2 = G(k+j,\ell) \le \overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\end{aligned} \tag{26}$$

For every $\lambda\in\bigcup_{i=1}^{\ell} I_i$, we have $\lambda\le a_{\ell+1}$. Thus, by (26), we can write

$$\begin{aligned}\int_{\bigcup_{i=1}^{\ell} I_i}\lambda\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} &\le a_{\ell+1}\int_{\bigcup_{i=1}^{\ell} I_i}\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}\\ &= a_{\ell+1}G(k+j,\ell) \le a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\end{aligned} \tag{27}$$

From (24) and (26), we obtain

$$\frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}} \le \alpha_{k+j+1}^{BB1} \le \delta_{\sup} \quad \text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}, \tag{28}$$

where $\mathcal{Z} := \int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}$. From (25), (27), and the fact that $\lambda\ge a_{\ell+1}$ for every $\lambda\in\bigcup_{i=\ell+1}^{n_\eta^u} I_i$, it follows for all $j\in\{\overline{r},\cdots,\overline{r}+\Theta\}$ that

$$\begin{aligned}\frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}} &= \frac{a_{\ell+1}^2\mathcal{Z}}{a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2+\alpha_{\ell+1}\mathcal{Z}}\\ &\le \frac{a_{\ell+1}\int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\lambda\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}}{a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2+\int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\lambda\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}} \le \alpha_{k+j+1}^{BB2} \le \delta_{\sup}.\end{aligned} \tag{29}$$

Now, using the fact that

$$\mathcal{Z} \ge \int_{I_{\ell+1}}\left[\prod_{p=0}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} = (g_{\ell+1}^{k+j})^2,$$

and by (21), (28), and (29), we infer that for a chosen $\alpha_{k+j+1}=\alpha_{k+j+1}^{BB1}$ or $\alpha_{k+j+1}=\alpha_{k+j+1}^{BB2}$ and for all $j\in\{\overline{r},\cdots,\overline{r}+\Theta\}$ that

$$\frac{2a_{\ell+1}}{3} = \frac{a_{\ell+1}\mathcal{Z}}{\frac{1}{2}\mathcal{Z}+\mathcal{Z}} \le \frac{a_{\ell+1}\mathcal{Z}}{\frac{1}{2}(g_{\ell+1}^{k+j})^2+\mathcal{Z}} \le \frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}} \le \alpha_{k+j+1} \le \delta_{\sup}. \tag{30}$$

Now for $\lambda\in[a_{\ell+1},b_{\ell+1}]$ and for $j\in\{\overline{r},\cdots,\overline{r}+\Theta\}$ we have the following two cases:

1. If $\alpha_{k+j+1}-\lambda\ge 0$, then by (5) we have

$$\left|1-\frac{\lambda}{\alpha_{k+j+1}}\right| = \left(1-\frac{\lambda}{\alpha_{k+j+1}}\right) \le \rho_{\mathcal{A}} < 1.$$

2. If $\alpha_{k+j+1} - \lambda < 0$, then by (30) and using the fact that $\lambda \leq b_{\ell+1} \leq a_{\ell+1} + \eta$ for $\lambda \in I_{\ell+1}$, we obtain

$$\left|1 - \frac{\lambda}{\alpha_{k+j+1}}\right| = \left(\frac{\lambda}{\alpha_{k+j+1}} - 1\right) \leq \left(\frac{b_{\ell+1}}{\alpha_{k+j+1}} - 1\right) \leq \left(\frac{a_{\ell+1}+\eta}{\alpha_{k+j+1}} - 1\right)$$
$$\leq \frac{3}{2} + \frac{\eta}{\alpha_{k+j+1}} - 1 \leq \frac{1}{2} + \frac{\eta}{\delta_{\inf}} < 1,$$

where in the last inequality we have used that $\eta < \frac{\delta_{\inf}}{2}$.

Hence, by the fact that $c = \max\{\rho_{\mathcal{A}}, \frac{1}{2} + \frac{\eta}{\delta_{\inf}}\}$, we have for every $j \in \{\overline{r}, \cdots, \overline{r} + \Theta\}$ and $\lambda \in [a_{\ell+1}, b_{\ell+1}]$ that

$$\left|1 - \frac{\lambda}{\alpha_{k+j+1}}\right| \leq c < 1. \tag{31}$$

Finally, by using (16) and (31) we obtain for every $j \in \{\overline{r}, \cdots, \overline{r} + \Theta\}$ that

$$\begin{aligned}(g_{\ell+1}^{k+j+2})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=0}^{k+j+1}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\ &= \int_{I_{\ell+1}} \left|1 - \frac{\lambda}{\alpha_{k+j+1}}\right|^2 \left[\prod_{p=0}^{k+j}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\ &\leq c^2 \int_{I_{\ell+1}} \left[\prod_{p=0}^{k+j}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} = c^2 (g_{\ell+1}^{k+j+1})^2.\end{aligned} \tag{32}$$

Using (22), (32), and the definitions of Θ, we obtain

$$(g_{\ell+1}^{k+\overline{r}+\Theta+1})^2 \leq c^{2\Theta}(g_{\ell+1}^{k+\overline{r}+1})^2 \leq c^{2\Theta}\gamma_{\mathcal{A}}^{2(\overline{r}+1)}(g_{\ell+1}^{k})^2 \leq c^{2\Theta}\gamma_{\mathcal{A}}^{2(\overline{r}+1)}\|\mathcal{G}_k\|^2 \leq 2\overline{\zeta}\|\mathcal{G}_k\|^2,$$

and the proof is complete. □

Lemma 3.3 *Let $\delta_{\sup} \geq 2\delta_{\inf}$. Moreover, assume that for any $\eta \in \,]0, \frac{\delta_{\inf}}{2}[$, integer ℓ with $n_\eta^l \leq \ell \leq n_\eta^u$, and $k \geq 0$, there exist $r_\ell \in \mathbb{N}$ and $\zeta_\ell \in \mathbb{R}_+$ such that the condition*

$$G(k+j, \ell) \leq \zeta_\ell \|\mathcal{G}_k\|^2 \quad \text{for all } j \geq r_\ell \tag{33}$$

holds. Then we show that for the choice of

$$\zeta_{\ell+1} := (1 + 2\gamma_{\mathcal{A}}^4)\zeta_\ell, \quad \text{and} \quad r_{\ell+1} := r_\ell + \Theta_\ell + 1,$$

with $\Theta_\ell := \Theta(\zeta_\ell, r_\ell)$ defined as in Lemma 3.2, we have

$$G(k+j, \ell+1) \leq \zeta_{\ell+1}\|\mathcal{G}_k\|^2 \quad \text{for all } j \geq r_{\ell+1}.$$

Proof First, observe that

$$G(k+j,\ell+1) = G(k+j,\ell) + (g_{\ell+1}^{k+j})^2.$$

Therefore, using (33) we only need to show that for every $j \geq r_{\ell+1}$

$$(g_{\ell+1}^{k+j})^2 \leq 2\gamma_{\mathcal{A}}^4 \zeta_\ell \|\mathcal{G}_k\|^2. \tag{34}$$

Due to Lemma 3.2 for $\overline{\zeta} = \zeta_\ell$ and $\overline{r} = r_\ell$, there exists an integer $j_1 \in \{r_\ell, \cdots, r_\ell + \Theta_\ell + 1\}$ such that

$$(g_{\ell+1}^{k+j_1})^2 \leq 2\zeta_\ell \|\mathcal{G}_k\|^2.$$

Now let us introduce a shifting variable which we initialize by $j_s = j_1$. Assume that $j_2 \geq j_s = j_1$ is an index, for which we have

$$(g_{\ell+1}^{k+j})^2 \leq 2\zeta_\ell \|\mathcal{G}_k\|^2 \quad \text{for all } j_1 \leq j \leq j_2, \tag{35}$$

and

$$(g_{\ell+1}^{k+j_2+1})^2 > 2\zeta_\ell \|\mathcal{G}_k\|^2. \tag{36}$$

Note that if this case does not arise, clearly, (34) holds for all $j \geq j_s = j_1$ and since $\gamma_{\mathcal{A}} \geq 1$ the proof is finished. Further, we can write

$$(g_{\ell+1}^{k+j+1})^2 > 2\zeta_\ell \|\mathcal{G}_k\|^2 \quad \text{for all } j_2 \leq j \leq j_3 - 2, \tag{37}$$

where $j_3 \geq j_2 + 2$ is the first integer greater than j_2 for which we have

$$(g_{\ell+1}^{k+j_3})^2 \leq 2\zeta_\ell \|\mathcal{G}_k\|^2. \tag{38}$$

Existence of such an index is justified using Lemma 3.2 for $\overline{r} = j_2$ and $\overline{\zeta} = \zeta_\ell$. Now, by (37) and using the same argument as in the proof of Lemma 3.2, where we have shown that from (21) implies (30), we can infer that

$$\frac{2}{3} a_{\ell+1} \leq \alpha_{k+j+2} \leq \delta_{\sup} \quad \text{for every } j_2 \leq j \leq j_3 - 2.$$

Continuing the argument from the proof of Lemma 3.2 we infer that

$$(g_{\ell+1}^{k+j+3})^2 \leq c^2 (g_{\ell+1}^{k+j+2})^2 \quad \text{for every } j_2 \leq j \leq j_3 - 2, \tag{39}$$

where $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2} + \frac{\eta}{\delta_{\inf}}\} < 1$. Finally, using (16) and (35), we have for $r = 1, 2$

$$\begin{aligned}
(g_{\ell+1}^{k+j_2+r})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=0}^{k+j_2+r-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} \\
&= \int_{I_{\ell+1}} \left[\prod_{p=k+j_2}^{k+j_2+r-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] \left[\prod_{p=0}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} \\
&= \int_{I_{\ell+1}} \left[\prod_{p=1}^{r} \left(\frac{\alpha_{k+j_2+p-1} - \lambda}{\alpha_{k+j_2+p-1}} \right)^2 \right] \left[\prod_{p=0}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} \\
&\leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}} \right)^{2r} \int_{I_{\ell+1}} \left[\prod_{p=0}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d_{\lambda,0} = \gamma_{\mathcal{A}}^{2r} (g_{\ell+1}^{k+j_2})^2.
\end{aligned} \tag{40}$$

Now since $c < 1$ and $\gamma_{\mathcal{A}} \geq 1$ due to the fact that $\delta_{\sup} \geq 2\delta_{\inf}$, we obtain from (35), (39), and (40) that

$$(g_{\ell+1}^{k+j+3})^2 \leq \gamma_{\mathcal{A}}^4 (g_{\ell+1}^{k+j_2})^2 \leq 2\zeta_\ell \gamma_{\mathcal{A}}^4 \|\mathcal{G}_k\|^2 \quad \text{for every } j_2 - 2 \leq j \leq j_3 - 2, \tag{41}$$

and as a consequence, we obtain

$$(g_{\ell+1}^{k+j})^2 \leq 2\zeta_\ell \gamma_{\mathcal{A}}^4 \|\mathcal{G}_k\|^2 \quad \text{for every } j_2 + 1 \leq j \leq j_3 + 1. \tag{42}$$

From (42) and (35) we conclude that (34) holds for every $j \in \{j_1, \cdots, j_3\}$. Finally, by setting $j_s = j_3$ and restart the process for j_3 justified in (38) and repeating the same argument, it can be shown that (34) holds for every $j \geq j_1$. Recall that $j_1 \in \{r_\ell, \cdots, r_\ell + \Theta_\ell + 1\}$. Therefore (34) holds for every $j \geq r_{\ell+1}$ and the proof is finished. □

In the next lemma, we investigate both of the cases $\delta_{\sup} < 2\delta_{\inf}$ and $\delta_{\sup} \geq 2\delta_{\inf}$.

Lemma 3.4 *Let $\{u_k\}_k$ be the sequence generated by Algorithm 1 for* (QP)*. Then there exists a positive integer m depending on $\delta_{\inf}$ and $\delta_{\sup}$ such that we have*

$$\|\mathcal{G}_{k+m}\| \leq \frac{1}{2}\|\mathcal{G}_k\| \quad \textit{for all } k \geq 0, \tag{43}$$

or equivalently,

$$\|u_{k+m} - u^*\| \leq \frac{1}{2}\|u_k - u^*\| \quad \textit{for all } k \geq 0, \tag{44}$$

for all initial iterates $u_{-1}, u_0 \in \mathcal{H}$ with $u_{-1} \neq u_0$.

Proof If $\delta_{\sup} < 2\delta_{\inf}$, then $\gamma_{\mathcal{A}} < 1$ and, by (13) in the proof of Theorem 3.1, we have

$$\|\mathcal{G}_{k+1}\| \leq \gamma_{\mathcal{A}}\|\mathcal{G}_k\| \quad \text{for every } k \geq 0.$$

Therefore, (43) follows for the choice of $m := \left\lceil \frac{-\log 2}{\log \gamma_{\mathcal{A}}} \right\rceil$.

Now, we consider the case in which $\delta_{\sup} \geq 2\delta_{\inf}$. In this case we have for $\rho_{\mathcal{A}}$ that $\frac{1}{2} \leq \rho_{\mathcal{A}} < 1$. First we decompose the interval $[\delta_{\inf}, \delta_{\sup}]$ into the finite family of intervals $\{I_i\}_i^{n_\eta^u}$ defined by (14) with a fixed length $\eta \in]0, \frac{\delta_{\inf}}{2}[\subset]0, \rho_{\mathcal{A}}\delta_{\inf}]$. Then due to (17) and (18), we have for every $k \geq 0$

$$G(k, n_\eta^u) = \sum_{i=1}^{n_\eta^u} (g_i^k)^2 = \|\mathcal{G}_k\|^2,$$

where $(g_i^k)^2$ is defined by (16). Moreover due to (19) in the proof of Lemma 3.1, there exists an integer $n_\eta^l > 0$ such that for every ℓ with $1 \leq \ell \leq n_\eta^l$, we have

$$(g_\ell^{k+j})^2 \leq \rho_{\mathcal{A}}^{2j} (g_\ell^k)^2 \quad \text{for every } j \geq 0 \text{ and } k \geq 0.$$

By summing over all ℓ with $1 \leq \ell \leq n_\eta^l$, we obtain

$$G(k+j, n_\eta^l) \leq \rho_{\mathcal{A}}^{2j} G(k, n_\eta^l) \leq \rho_{\mathcal{A}}^{2j} \|\mathcal{G}_k\|^2 \quad \text{for every } j \geq 0 \text{ and } k \geq 0.$$

Now for the choice of $r_{n_\eta^l} := \left\lceil \frac{\log \zeta_{n_\eta^l}}{2 \log \rho_{\mathcal{A}}} \right\rceil$ for any given $\zeta_{n_\eta^l} > 0$, we have

$$G(k+j, n_\eta^l) \leq \zeta_{n_\eta^l} \|\mathcal{G}_k\|^2 \quad \text{for every } j \geq r_{n_\eta^l} \text{ and } k \geq 0,$$

and thus, by choosing $\zeta_{n_\eta^l} := \frac{1}{4}(1+2\gamma_{\mathcal{A}}^4)^{-(n_\eta^u - n_\eta^l)}$ we are in the position to use Lemma 3.3. By using this lemma we have for ℓ with $n_\eta^l \le \ell \le n_\eta^u - 1$ that

$$\zeta_{\ell+1} = (1+2\gamma_{\mathcal{A}}^4)\zeta_\ell = \frac{1}{4}(1+2\gamma_{\mathcal{A}}^4)^{\ell+1-n_\eta^u}, \quad \text{and} \quad r_{\ell+1} = r_\ell + \Theta_\ell + 1,$$

where $\Theta_\ell = \Theta(\zeta_\ell, r_\ell)$ has been defined as in Lemma 3.2. To be more precise, by applying Lemma 3.3 once, for the first iteration, we obtain

$$G(k+j, n_\eta^l + 1) \le \zeta_{n_\eta^l+1}\|\mathcal{G}_k\|^2 = (1+2\gamma_{\mathcal{A}}^4)\zeta_{n_\eta^l}\|\mathcal{G}_k\|^2 = \frac{1}{4}(1+2\gamma_{\mathcal{A}}^4)^{1-(n_\eta^u-n_\eta^l)}\|\mathcal{G}_k\|^2$$

for all $j \ge r_{n_\eta^l+1} := r_{n_\eta^l} + \Theta_{n_\eta^l} + 1$. Applying this lemma repeatedly we conclude after $(n_\eta^u - n_\eta^l) - 1$ iterations that

$$\|\mathcal{G}_{k+j}\|^2 = G(k+j, n_\eta^u) \le \zeta_{n_\eta^u}\|\mathcal{G}_k\|^2 = \frac{1}{4}\|\mathcal{G}_k\|^2 \quad \text{for all } j \ge r_{n_\eta^u}.$$

By putting $m = r_{n_\eta^u}$, (43) holds.
Moreover, the equivalence of (43) and (44) is justified due to the fact that, similarly to (8) for $\mathcal{G}_k$, it can easily be shown that

$$(u_{k+1} - u^*) = \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})(u_k - u^*) \quad \text{for all } k = 0, 1, 2, \ldots. \tag{45}$$

Hence, the same machinery can be used to derive (44) and this completes the proof. □

Proof of Theorem 3.2. We need only to consider the case in which for every $k \ge 0$ we have $u_k \ne u^*$. In this case, we will show that $u_k \to u^*$ R-linearly. Due to (45) and with a similar argument as in (9), we can write for every $k \ge 0$ that

$$\begin{aligned} \|u_{k+1} - u^*\|^2 &= \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k}\right)^2 d(E_\lambda(u_k - u^*), (u_k - u^*)) \\ &\le \gamma_{\mathcal{A}}^2 \int_{\sigma(\mathcal{A})} d(E_\lambda(u_k - u^*), (u_k - u^*)) = \gamma_{\mathcal{A}}^2 \left\|u_k - u^*\right\|^2. \end{aligned} \tag{46}$$

Moreover, due to (44) in Lemma 3.4, we obtain

$$\|u_{jm} - u^*\| \le (\frac{1}{2})^j \|u_0 - u^*\| \quad \text{for all } j \ge 0, \tag{47}$$

where m has been defined in Lemma 3.4. Now for every $k \ge 0$, there exists an integer j such that $jm \le k < (j+1)m$. Therefore, it follows that $k - jm < m$ and $j \ge \frac{k}{m} > \frac{k}{m} - 1$. Using (46) and (47), we obtain

$$\begin{aligned} \|u_k - u^*\| &\le \gamma_{\mathcal{A}}^m \|u_{jm} - u^*\| \le \gamma_{\mathcal{A}}^m (\frac{1}{2})^j \|u_0 - u^*\| \le \gamma_{\mathcal{A}}^m (\frac{1}{2})^{\frac{k}{m}} \|u_0 - u^*\| \\ &= c_1 c_2^k \|u_0 - u^*\| \qquad \text{for all } k \ge 0, \end{aligned}$$

where $c_1 := \gamma_{\mathcal{A}}^m$ and $c_2 := (\frac{1}{2})^{\frac{1}{m}} < 1$, and this completes the proof. □

Remark 3.1 If $\sigma(\mathcal{A})$ is finite, we can infer that $\sigma(\mathcal{A}) = \{\lambda_i : i = 1, \ldots, m\}$ with $\lambda_{i+1} > \lambda_i$ for $i = 1, \ldots, m-1$, $\lambda_1 = \delta_{\inf}$, and $\lambda_m = \delta_{sup}$. Then for every arbitrary $\eta > 0$ and partitioning $\{I_i\}_{i=1}^{n_\eta^u}$ of $[\delta_{\inf}, \delta_{\sup}]$, we obtain for $k \geq 0$ that

$$\begin{aligned}\|\mathcal{G}_k\|^2 &= \int_{\delta_{\inf}}^{\delta_{\sup}} \left[\prod_{p=0}^{k-1}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} = \sum_{i=1}^{n_\eta^u}\int_{I_i}\left[\prod_{p=0}^{k-1}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0} \\ &= \sum_{i=1}^{m}\left[\prod_{p=0}^{k-1}\left(\frac{\alpha_p - \lambda_i}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}}\mathcal{G}_0\|^2 = \sum_{i=1}^{m}(g_i^k)^2,\end{aligned} \tag{48}$$

where $(g^k)_i^2 := \left[\prod_{p=0}^{k-1}\left(\frac{\alpha_p-\lambda_i}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}}\mathcal{G}_0\|^2$. Then the statements of Lemma 3.1 is true for $n_\eta^l = 1$. Further, Lemma 3.4 and Theorem 3.2 are applicable. Moreover, in the proof of Lemma 3.2, similarly to (48), all the integrations are replaced by finite sum and it follows that $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2}\}$. See [3,2] for more details.

Remark 3.2 Note that, due to Theorem 3.1, the numerical behaviour of Algorithm 1 is strongly depending on $\sigma(\mathcal{A})$. In fact, this relation can be explained based on the value of the spectral condition number $\kappa(\mathcal{A}) := \|\mathcal{A}\|\|\mathcal{A}^{-1}\| = \frac{\delta_{sup}}{\delta_{\inf}}$. It can be seen that $\gamma_{\mathcal{A}} = \kappa(\mathcal{A}) - 1$ and $\rho_{\mathcal{A}} = 1 - \frac{1}{\kappa(\mathcal{A})} < 1$. Further, depending on the value of $\kappa(\mathcal{A})$, we can summarize the following cases:

1. $\kappa(\mathcal{A}) < 2$: In this case, due to Theorem 3.1, Algorithm 1 is Q-linearly convergent with the rate $\gamma_{\mathcal{A}} < 1$. Moreover, from (13), we infer that the sequence $\{\|\mathcal{G}_k\|\}_k$ is monotone decreasing.
2. $\kappa(\mathcal{A}) \geq 2$: This case is more delicate. Recall from (17) that for every fixed $\eta \in]0, \frac{\delta_{\inf}}{2}[$, and $k \geq 1$, we have $\|\mathcal{G}_{k+1}\|^2 = \sum_{i=1}^{n_\eta^u}(g_i^{k+1})^2$ where the values $(g_i^{k+1})^2$ with $i = 1, \ldots, n_\eta^u$ are defined by
$$(g_i^{k+1})^2 = \int_{I_i}\left(1 - \frac{\lambda}{\alpha_k}\right)^2\left[\prod_{p=0}^{k-1}\left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d_{\lambda,0}. \tag{49}$$
Due to Lemma 3.1, there exists an index $n_\eta^l \geq 1$ such that the sequences $\{|g_i^k|\}_k$ with $i = 1, \ldots, n_\eta^l$ are Q-linearly monotonically decreasing with factor $\rho_{\mathcal{A}} < 1$. Therefore it remains only to consider the values of $|g_i^k|$ for $i = n_\eta^l + 1, \ldots, n_\eta^u$. From (49), it can be shown that for every interval I_i with $\alpha_k \in I_i$ it holds that $|g_i^{k+1}| \leq \frac{\eta}{\alpha_k}|g_i^k| < \frac{1}{2}|g_i^k|$. On the other hand, if for an interval I_i it holds that $a_i > 2\alpha_k$, then we obtain $|g_i^{k+1}| > |g_i^k|$. Further, for the last interval $I_{n_\eta^u}$, we have $\frac{|g_{n_\eta^u}^{k+1}|}{|g_{n_\eta^u}^k|} \leq \kappa(\mathcal{A}) - 1$. These facts explain the potential nonmonotonic behaviour of the sequence $\{\|\mathcal{G}_k\|\}_k$ and its dependence on $\kappa(A)$.

Remark 3.3 (Preconditioning) Due to Remark 3.2, the convergence of Algorithm 1 depends strongly on $\kappa(\mathcal{A})$. Analogously to the case of the conjugate gradient methods, the problem (QP) can, by using an appropriate uniformly positive, self-adjoint, and continuous operator $\mathcal{C} : \mathcal{H} \to \mathcal{H}$, be transformed to the following equivalent problem

$$\min_{z\in\mathcal{H}} \tilde{\mathcal{F}}(z) := \frac{1}{2}(\tilde{\mathcal{A}}z, z) - (\tilde{b}, z),$$

where $\tilde{\mathcal{A}} := \mathcal{C}^{-\frac{1}{2}}\mathcal{A}\mathcal{C}^{-\frac{1}{2}}$, $\tilde{b} := \mathcal{C}^{-\frac{1}{2}}b$ and $z := \mathcal{C}^{\frac{1}{2}}u$. Clearly, $\sigma(\tilde{\mathcal{A}}) = \sigma(\mathcal{C}^{-1}\mathcal{A})$ and, as a consequence, the spectrum of $\tilde{\mathcal{A}}$ is completely determined by $\mathcal{C}$ and $\mathcal{A}$. Thus, the operator $\mathcal{C}$ can be chosen such that the application of Algorithm 1 yields faster convergence. In [8], preconditioning has been studied for Algorithm 1 in the case of the Euclidean space $\mathbb{R}^n$. For the case of infinite-dimensional Hilbert spaces, preconditioning methods have been studied for the conjugate gradient methods. Among them we can mention [9,10,11,12].

Remark 3.4 Let a function $\mathcal{F}$ of the form (QP) with $\kappa(\mathcal{A}) < 2$ and a proper lower semicontinuous convex (possibly nonsmooth) $\Phi : \mathcal{H} \to \mathbb{R}$ be given, and consider

$$\min_{u\in\mathcal{H}} \mathcal{E}(u) := \mathcal{F}(u) + \Phi(u). \tag{50}$$

For the solution $u^* \in \mathcal{H}$ to (50), the first-order optimality conditions can be expressed as

$$\mathbf{u}^* = \operatorname{Prox}_{\bar{\alpha}\Phi}(\mathbf{u}^* - \bar{\alpha}\mathcal{G}(\mathbf{u}^*)) \quad \text{for each } \bar{\alpha} > 0, \tag{51}$$

where $\operatorname{Prox}_{\bar{\alpha}\Phi} : \mathcal{H} \to \mathcal{H}$ stands for the proximal operator, see e.g, [13, Cor. 27.3]. Then, the sequence $\{\mathbf{u}_k\}_k$, generated by the following proximal gradient method

$$\mathbf{u}_{k+1} = \operatorname{Prox}_{\frac{1}{\alpha_k}\Phi}(\mathbf{u}_k - \frac{1}{\alpha_k}\mathcal{G}(\mathbf{u}_k)) \quad \text{with } \alpha_k \in \{\alpha_k^{BB1}, \alpha_k^{BB2}\}, \tag{52}$$

converges Q-linearly to the solution u^* of (QP) with the rate $\gamma_{\mathcal{A}} < 1$. More precisely, due to the fact the proximal operator is nonexpansive and using the same argument as in the proof of Theorem 3.1 and (46), we obtain

$$\begin{aligned} \|\mathbf{u}_{k+1} - \mathbf{u}^*\| &= \|\operatorname{Prox}_{\frac{1}{\alpha_k}\Phi}(\mathbf{u}_k - \frac{1}{\alpha_k}\mathcal{G}(\mathbf{u}_k)) - \operatorname{Prox}_{\frac{1}{\alpha_k}\Phi}(\mathbf{u}^* - \frac{1}{\alpha_k}\mathcal{G}(\mathbf{u}^*))\| \\ &\le \|\frac{1}{\alpha_k}(\alpha_k\mathcal{I} - \mathcal{A})(\mathbf{u}_k - \mathbf{u}^*)\| \le \gamma_{\mathcal{A}}\|(\mathbf{u}_k - \mathbf{u}^*)\|. \end{aligned} \tag{53}$$

Note that, Φ can be chosen as the indicator function of a convex set or any convex functional enhancing sparsity.

4 Neumann Optimal Control for the Linear Wave Equation

In this section, we investigate the applicability of Algorithm 1 for an optimal control problem governed by linear wave equation. We consider

$$\min_{u\in L^2(\Sigma_c)} J(u,y) := \frac{\alpha_1}{2}\|y - y_d\|^2_{L^2(Q)} + \frac{\alpha_2}{2}\|y(T) - z_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Sigma_c)}, \tag{54}$$

$$\text{subject to } \begin{cases} y_{tt} - \Delta y = f & \text{in } Q, \\ \partial_\nu y = u & \text{on } \Sigma_c, \\ y = 0 & \text{on } \Sigma_0, \\ y(0) = y_0^1, \quad y_t(0) = y_0^2 & \text{on } \Omega, \end{cases} \tag{55}$$

where α_1, α_2, and β are positive constants, the desired state y_d and the desired final state z_d are smooth enough, $Q :=]0,T[\times\Omega$, $\Sigma_c :=]0,T[\times\Gamma_c$, $\Sigma_0 :=]0,T[\times\Gamma_0$, and $\Omega \in \mathbb{R}^n$ is a bounded domain with the smooth boundary $\partial\Omega := \overline{\Gamma_c} \cup \overline{\Gamma_0}$. Further,

two disjoint components Γ_c, Γ_0 are relatively open in $\partial\Omega$.
Before investigating the optimal control problem, we recall some useful results for equation (55). The operator $A : L^2(\Omega) \supset \mathcal{D}(A) \to L^2(\Omega)$ defined by $Ah = -\Delta h$ is a positive self-adjoint operator with $\mathcal{D}(A) := \{h \in H^2(\Omega), h|_{\Gamma_0} = \partial_\nu h|_{\Gamma_c} = 0\}$. Thus, we define the spaces $H^\alpha_{\Gamma_0}(\Omega) := \mathcal{D}(A^{\frac{\alpha}{2}})$ for $0 \le \alpha \le 1$, and by $(H^\alpha_{\Gamma_0}(\Omega))^*$ we denote the corresponding dual space. These spaces are used throughout this section. We use the following notion of solution [14].

Definition 4.1 (Very weak solution) Let $T > 0$, and

$$(y^1_0, y^2_0, u, f) \in L^2(\Omega) \times (H^1_{\Gamma_0}(\Omega))^* \times L^2(\Sigma_c) \times L^2(0, T; (H^1_{\Gamma_0}(\Omega))^*)$$

be given. A function $y \in L^2(Q)$ is referred to as the very weak solution of (55), if the following inequality holds

$$\begin{aligned}&\langle f, \varphi\rangle_{(L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*),L^2(0,T;H^1_{\Gamma_0}(\Omega)))} =\\ &(g, y)_{L^2(Q)} + (y^1_0, \varphi_t(0))_{L^2(\Omega)} - \langle y^2_0, \varphi(0)\rangle_{((H^1_{\Gamma_0}(\Omega))^*,H^1_{\Gamma_0}(\Omega))} - (u, \varphi)_{L^2(\Sigma_0)}\end{aligned} \tag{56}$$

for all $g \in L^2(Q)$, where $\varphi(g) \in C^0([0, T]; H^1_{\Gamma_0}(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ is the weak solution of the following backward in time problem

$$\begin{cases}\varphi_{tt} - \Delta\varphi = g & \text{in } Q,\\ \partial_\nu\varphi = 0 & \text{on } \Sigma_c,\\ \varphi = 0 & \text{on } \Sigma_0,\\ \varphi(T) = 0, \quad \varphi_t(T) = 0 & \text{on } \Omega.\end{cases}$$

We have the following existence and regularity results from [15,16] for (55).

Lemma 4.1 *For every*

$$(y^1_0, y^2_0, u, f) \in L^2(\Omega) \times (H^1_{\Gamma_0}(\Omega))^* \times L^2(\Sigma_c) \times L^2(0, T; (H^1_{\Gamma_0}(\Omega))^*),$$

equation (55) *admits a unique very weak solution* $y(y^1_0, y^2_0, u, f)$ *in the space*

$$C^0([0, T]; L^2(\Omega)) \cap C^1([0, T]; (H^1_{\Gamma_0}(\Omega))^*),$$

satisfying

$$\begin{aligned}&\|y\|_{C^0([0,T];L^2(\Omega))} + \|y_t\|_{C^0([0,T];(H^1_{\Gamma_0}(\Omega))^*)}\\ &\le c\left(\|y^1_0\|_{L^2(\Omega)} + \|y^2_0\|_{(H^1_{\Gamma_0}(\Omega))^*} + \|u\|_{L^2(\Sigma_c)} + \|f\|_{L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*)}\right),\end{aligned} \tag{57}$$

where the constant c_1 *is independent of* y^1_0*,* y^2_0*,* u*, and* f*. Moreover, the solution operator* $L : L^2(\Sigma_c) \to C^0([0, T]; H^{\frac{1}{2}}_{\Gamma_0}(\Omega)) \cap C^1([0, T]; H^{-\frac{1}{2}}(\Omega))$ *defined by* $u \mapsto y(0, 0, u, 0)$ *is bounded. Furthermore, the mapping*

$$\Pi : L^2(\Omega) \times (H^1_{\Gamma_0}(\Omega))^* \times L^2(0, T; (H^1_{\Gamma_0}(\Omega))^*) \to C^0([0, T]; L^2(\Omega)) \cap C^1([0, T]; (H^1_{\Gamma_0}(\Omega))^*),$$

defined by $(y^1_0, y^2_0, f) \mapsto y(y^1_0, y^2_0, 0, f)$ *is continuous.*

By considering the following continuous embeddings

$$i_1 : C^0([0,T]; H^{\frac{1}{2}}_{\Gamma_0}(\Omega)) \hookrightarrow L^2(Q), \qquad i_2 : C^0([0,T]; L^2(\Omega)) \hookrightarrow L^2(Q),$$

and the continuous operator $\delta_T : C^0([0,T]; L^2(\Omega)) \to L^2(\Omega)$ defined by $y \mapsto y(T)$, we can rewrite the problem (54)-(55) as the following linear least squares problem

$$\min_{u\in\mathcal{H}} \mathcal{F}(u) := \frac{1}{2}\|\mathcal{L}u - \psi\|^2_{\mathcal{X}} + \frac{\beta}{2}\|u\|^2_{\mathcal{H}}, \tag{LS}$$

where $\mathcal{H} := L^2(\Sigma_c)$, $\mathcal{X} := L^2(Q) \times L^2(\Omega)$, and the linear operator $\mathcal{L} : \mathcal{H} \to \mathcal{X}$ and $\psi \in \mathcal{X}$ are defined as follows

$$\mathcal{L}u := \begin{pmatrix} \alpha_1(i_1 \circ L)(u) \\ \alpha_2(\delta_T \circ L)(u) \end{pmatrix}, \quad \psi := \begin{pmatrix} \alpha_1 y_d - \alpha_1(i_2 \circ \Pi)(y_0^1, y_0^2, f) \\ \alpha_2 z_d - \alpha_2(\delta_T \circ \Pi)(y_0^1, y_0^2, f) \end{pmatrix}. \tag{58}$$

Then, the problem (54)-(55), can be also rewritten in the form of (QP), where $\mathcal{A} := \mathcal{L}^*\mathcal{L} + \beta\mathcal{I}$ with $\mathcal{L}^* : \mathcal{X}^* \to \mathcal{H}$, and $b := \mathcal{L}^*\psi$. In addition, due to the fact that the operator $\mathcal{A}$ is uniformly positive, bounded, and self-adjoint, the existence and uniqueness of the solution to the problem (54)-(55) can be justified due to the fact that $\mathcal{A}$ has a bounded inverse.
Now assume that $u^* \in \mathcal{H}$ is the optimal solution of the optimal control problem (54)-(55). Then, the first-order optimality condition can be expressed as

$$(\mathcal{L}^*\mathcal{L} + \beta\mathcal{I})u^* = \mathcal{L}^*\psi, \tag{59}$$

where the operator $\mathcal{L}$ and the function ψ were defined in (58). Moreover, it can be shown (see [17,18,19]) that (59) is equivalent to the condition $\beta u^* = p^*$ on Σ_c, where $p^* \in C^1([0,T]; L^2(\Omega)) \cap C^0([0,T]; H^1_{\Gamma_0}(\Omega))$ is the weak solution of the following linear wave equation

$$\begin{cases} p^*_{tt} - \Delta p^* = -\alpha_1(y^* - y_d) & \text{in } Q, \\ \partial_\nu p^* = 0 & \text{on } \Sigma_c, \\ p^* = 0 & \text{on } \Sigma_0, \\ p^*(T) = 0, \quad p^*_t(T) = \alpha_2(y^*(T) - z_d) & \text{on } \Omega, \end{cases}$$

and $y^* = y(y_0^1, y_0^2, f, u^*)$ is the very weak solution of (55).

5 Numerical Experiments

Here we report on numerical results for optimal control problems related to the linear wave equation and to the nonlinear viscous Burgers' equation. Algorithm 1 will be used with the following step-size strategies:

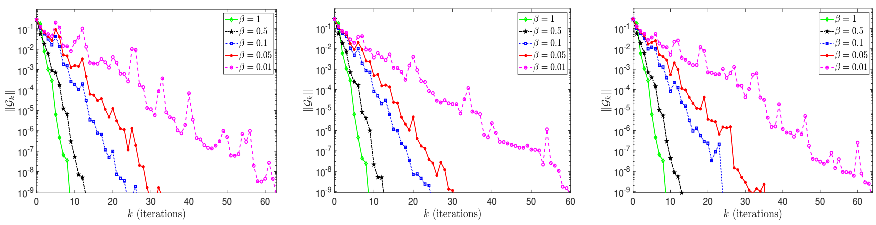
BB1: $\alpha_k := \alpha_k^{BB1}$ for every $k \geq 1$,
BB2: $\alpha_k := \alpha_k^{BB2}$ for every $k \geq 1$,
ABB: $\alpha_k = \alpha_k^{BB1}$ for odd $k \geq 1$ and $\alpha_k = \alpha_k^{BB2}$ for even $k \geq 1$,

and for different values of control cost parameter β. The value β in all the optimal control problems of the previous section has a direct influence on the spectral condition number of $\mathcal{A}_{u^*}^{\mathcal{F}}$ corresponding to $\mathcal{F}$. To be more precise, as the value of β increases, the value of $\kappa(\mathcal{A}_{u^*}^{\mathcal{F}})$ is getting smaller. Therefore, as discussed in Remarks 3.2, one expects slower convergence for smaller values of β. For each example, we illustrate the convergence of the sequence $\{\|\mathcal{G}_k\|\}_k$ for different values of β. We have chosen $u_{-1} = 0$ and $u_0 := -\mathcal{G}(0)$ as the initial iterates. All computations were done on the MATLAB platform.

Example 5.1 (Neumann optimal control for the linear wave equation) In this example, we deal with problem (54)-(55). The spatial domain $\Omega :=]0,1[^2$ was discretized by a conforming linear finite element scheme using continuous piecewise linear basis functions over a uniform triangulation with 32768 cells. Further, for the temporal discretization we used the Crank-Nicolson time stepping with 157 equidistant nodes ($\Delta t = \frac{1}{156} \approx 0.0064$). Here we set $T = \alpha_1 = \alpha_2 = 1$, $y_0^1(x) = \sin(\pi x_1)\sin(\pi x_2)$, $y_0^2(x) = 0$, $f(t,x) = \pi^2 \sin(\pi x_1 t)\sin(\pi x_2 t)$, $z_d(x) = 0$, and

$$y_d(x,t) = \begin{cases} -x_1 & \text{for } x_1 < 0.5, \\ x_1 & \text{for } x_1 \geq 0.5, \end{cases}$$

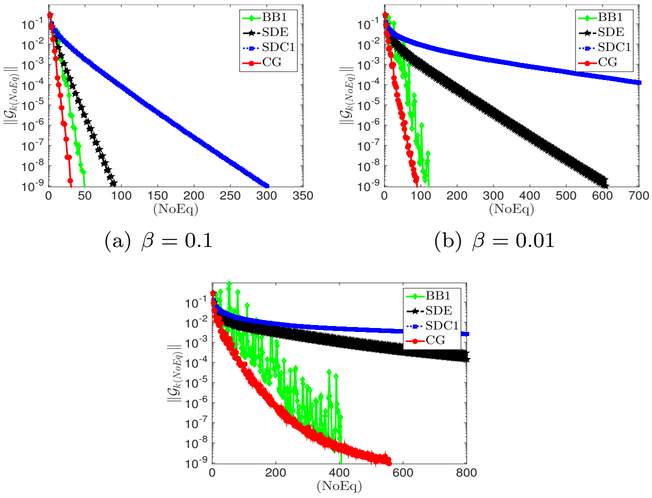
where $x := (x_1, x_2) \in \Omega$. The Neumann control is applied on the subset $\Gamma_c \subset \partial\Omega$ given by $\{(1,x_2) : x_2 \in]0,1[\} \cup \{(x_1,1) : x_1 \in]0,1[\}$. In Figure 1, we report the behaviour of the gradient norm for different values of β. For each choice of the



(a) BB1 (b) BB2 (c) ABB

Fig. 1 Example 5.1: Convergence of $\{\|\mathcal{G}_k\|\}_k$ for $\beta = 1, 0.5, 0.1, 0.05, 0.01$

step-size strategy, decreasing the value of β implies that the number of required iterations becomes larger and, thus, the convergence is getting slower. This is due to the fact that there is a trade-off between the magnitude of β and the value of $\kappa(\mathcal{A})$ where $\mathcal{A} = \mathcal{L}^*\mathcal{L} + \beta\mathcal{I}$ with $\mathcal{L}$ specified in the previous section. More precisely, $\kappa(\mathcal{A}) = \frac{\beta+\delta_{\sup}}{\beta+\delta_{\inf}}$ with $\delta_{\inf} := \inf(\sigma(\mathcal{L}^*\mathcal{L}))$ and $\delta_{\sup} := \sup(\sigma(\mathcal{L}^*\mathcal{L}))$. This behaviour is clearly illustrated in Figure 1. As can be seen from Figure 1, the convergence for the cases $\beta = 1$ and $\beta = 0.5$ is Q-linear. For these cases we might conjecture that $\kappa(\mathcal{A}) < 2$ with a smaller value of convergence rate $\gamma_{\mathcal{A}}$ for $\beta = 1$ compared to $\beta = 0.5$. However, for the rest of the cases, nonmonotonic behaviour is exhibited in the sequences $\{\|\mathcal{G}_k\|\}_k$ which corresponds to $\kappa(\mathcal{A}) \geq 2$. Clearly, as β decreases, the nonmonotonic behaviour is getting stronger. As discussed in Remarks 3.2 if $\kappa(\mathcal{A})$ becomes larger, then the changes in the decreasing components $|g_i^k|$ are getting smaller compared to the nondecreasing components. This explains why a decrease in the value β leads to an increase in nonmonotonicity. Figure 2 reports on the



(a) $\beta = 0.1$

(b) $\beta = 0.01$

(c) $\beta = 0.001$

Fig. 2 Example 5.1: Convergence of $\{\|\mathcal{G}\|_{k(NoEq)}\}_{NoEq}$ for different Algorithms and values of β

performance of four different algorithms including: 1. BB1, 2. Conjugate gradient method (CG) without preconditioning, 3. Steepest descent method with exact line search (SDE), and 4. Steepest descent direction with constant step-size 1 (SDC1). Here, the spatial discretization was achieved with 8192 cells and with $\Delta t = 0.01$ for the temporal discretization. The norm of the gradient $\|\mathcal{G}_{k(NoEq)}\|$ with respect to the number $(NoEq)$ of linear wave equation solves is plotted in Figure 2. The algorithm was terminated as soon as either $\|\mathcal{G}\|_{k(NoEq)} \leq 10^{-9}$ or $NoEq > 1200$ held. Indeed, every multiplication $\mathcal{A}v = (\mathcal{L}^*\mathcal{L} + \beta\mathcal{I})v$ for $v \in \mathcal{H}$ requires solving a forward- and a backward-in-time linear wave equation (see (58)). Therefore, all algorithms need solving two linear wave equations at every iteration (gradient evaluation), that is, $k = 2 \times NoEq$. As can be seen from Figure 2, BB1 and CG have better performance compared to SDE and SDC1. In the cases $\beta = 0.1, 0.01$, CG converges slightly faster than BB1. As β gets larger, we can see that the performance of BB1 is getting closer to the performance of CG. For the case $\beta = 0.001$, algorithm BB1, though exhibiting stronger nonmonotonic behaviour, reached the termination ($\|\mathcal{G}_k\| \leq 10^{-9}$) faster than the other algorithms.

In the next example, we results of Algorithm 1 when applied to a nonlinear optimal control problem which does not have quadratic form.

Example 5.2 (Distributed optimal control for the Burgers equation) We consider the following optimal control problem

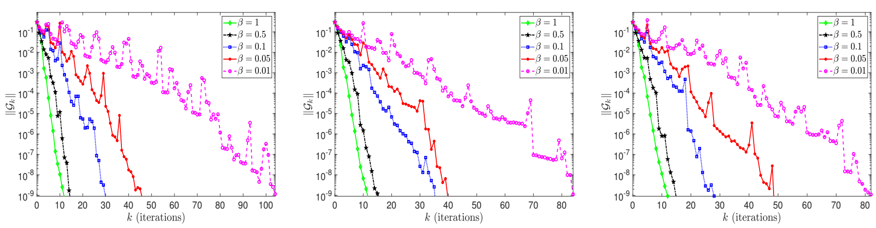
$$\min_{u\in L^2(\hat{Q})} J(y,u) := \frac{1}{2}\|y\|^2_{L^2(Q)} + \frac{1}{2}\|y(T)\|^2_{L^2(0,1)} + \frac{\beta}{2}\|u\|^2_{L^2(\hat{Q})}, \tag{60}$$

$$\text{subject to} \begin{cases} y_t - \vartheta y_{xx} + y y_x = Bu, & (t,x)\in Q,\\ y(t,0) = y(t,1) = 0, & t \in \,]0,T[\,,\\ y(0,x) = y_0(x), & x\in \,]0,1[\,. \end{cases} \tag{61}$$

where $\beta > 0$, $\vartheta = 0.01$, $T = 1$, $Q := \,]0,T[\,\times\,]0,1[\,$, and $\hat{Q} := \,]0,T[\,\times \hat{\Omega}$, with $\hat{\Omega} := \,]0.1, 0.4[\, \subset \,]0,1[\,$. The variables $y(t) = y(t,x)$ and $u(t) = u(t,x)$ denote the state and control, respectively. The extension-by-zero operator $B \in \mathcal{L}(L^2(\hat{\Omega}), L^2(0,1))$ is defined by

$$(Bu)(x) = \begin{cases} u(x), & x\in\hat{\Omega},\\ 0, & x\in \,]0,1[\,\setminus\hat{\Omega}. \end{cases}$$

It is wellknown from e.g., [20,21] that, for every pair $(y_0,u)\in L^2(0,1)\times L^2(\hat{Q})$, equation (61) admits a unique weak solution $y(y_0,u)\in W(0,T)$ where $W(0,T) := \{\phi : \phi \in L^2(0,T;H^1_0(0,1)), \phi_t \in L^2(0,T;H^{-1})\}$. Therefore the control-to-state operator $u\in\mathcal{H}\mapsto y(u)\in W(0,T)$ is well-defined. By setting $\mathcal{H} := L^2(\hat{Q})$, and rewriting (60)-(61) in reduced form, we obtain a nonconvex unconstrained optimization problem posed in the Hilbert space $\mathcal{H} := L^2(\hat{Q})$ and thus the algorithm (1) is applicable. We report the performance of Algorithm (1) for different values of β and $y_0(x) = 5\exp(-20(x-0.5)^2)$.
The spatial discretization was carried out by the standard Galerkin method based on piecewise linear basis functions with mesh-size $h = 0.004$. For temporal discretization, we used the implicit Euler method with step-size $\Delta t = 0.008$. Moreover, the resulting nonlinear systems after the temporal discretization were solved by Newton's method with tolerance $\epsilon_n = 10^{-13}$. Figure 3 shows the convergence of $\{\|\mathcal{G}_k\|\}_k$ corresponding to Example 5.2 for different values of β. As can be seen from Figure 3, despite the nonlinearity similar observations as in Example 5.1 also hold for this example.



(a) BB1 (b) BB2 (c) ABB

Fig. 3 Example 5.2: Convergence of $\{\|\mathcal{G}_k\|\}_k$ for $\beta = 1, 0.5, 0.1, 0.05, 0.01$

6 Conclusions

Based on spectral analysis for self-adjoint operators, convergence results for the Barzilai-Borwein (BB) method were obtained for quadratic functions in a general Hilbert spaces.

As observed for Example 5.2, Algorithm 1 converges even for a more general class of infinite-dimensional optimization problems than strictly convex quadratic ones. In fact, the results of this paper, can be extended to R-linear local convergence for a twice continuously Frećhet-differentiable function $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ whose second derivative is Liptschitz continuous in a neighbourhood of a strong local minimum.

Inspired by the BB step-sizes, a significant amount of investigations has been carried out for the design and analysis of spectral gradient methods for finite-dimensional problems, see e.g., [22,4,23,24,25,26,27,28]. The analysis of these methods is based on the behaviour of the eigenvalues of the Hessian matrix. We believe that, by spectral analysis and using similar arguments as in this manuscript, the convergence results of these methods can be extended to problems posed in infinite-dimensional Hilbert spaces such as PDE-constrained optimization problems. An interesting next step consists in investigating the mesh-independent principle for these spectral gradient methods. As an outlook, it would be interesting to further investigate Remark 3.4 for the case that $\kappa(\mathcal{A}) \geq 2$ and to see whether the sequence of iterations (52) is convergent without incorporating any step-size strategy.

Acknowledgements The work of Karl Kunisch was supported the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research program.

References

1. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. 8(1), 141–148 (1988)
2. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. 13(3), 321–326 (1993)
3. 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)
4. Dai, Y.H., Fletcher, R.: On the asymptotic behaviour of some new gradient methods. Math. Program. 103(3, Ser. A), 541–559 (2005)
5. Fletcher, R.: On the Barzilai-Borwein method. In: Optimization and control with applications, Appl. Optim., vol. 96, pp. 235–256. Springer, New York (2005)
6. Hall, B.C.: Quantum theory for mathematicians, Graduate Texts in Mathematics, vol. 267. Springer, New York (2013)
7. Werner, D.: Funktionalanalysis, extended edn. Springer-Verlag, Berlin (2000)
8. Molina, B., Raydan, M.: Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. Numer. Algorithms 13(1-2), 45–60 (1996)
9. Axelsson, O., Karátson, J.: On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space. Numer. Funct. Anal. Optim. 23(3-4), 285–302 (2002)
10. Axelsson, O., Karátson, J.: Mesh independent superlinear PCG rates via compact-equivalent operators. SIAM J. Numer. Anal. 45(4), 1495–1516 (2007)
11. Herzog, R., Sachs, E.: Superlinear convergence of Krylov subspace methods for self-adjoint problems in Hilbert space. SIAM J. Numer. Anal. 53(3), 1304–1324 (2015)
12. Mardal, K.A., Winther, R.: Preconditioning discretizations of systems of partial differential equations. Numer. Linear Algebra Appl. 18(1), 1–40 (2011)

13. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces, second edn. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017)
14. Lions, J.L., Magenes, E.: Non-homogeneous boundary value problems and applications. Vol. I. Springer-Verlag, New York-Heidelberg (1972). Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin (1971)
15. Lasiecka, I., Triggiani, R.: Sharp regularity theory for second order hyperbolic equations of Neumann type. I. L_2 nonhomogeneous data. Ann. Mat. Pura Appl. (4) 157, 285–367 (1990).
16. Lasiecka, I., Triggiani, R.: Regularity theory of hyperbolic equations with nonhomogeneous Neumann boundary conditions. II. General boundary data. J. Differential Equations 94(1), 112–164 (1991)
17. Kröner, A., Kunisch, K., Vexler, B.: Semismooth Newton methods for optimal control of the wave equation with control constraints. SIAM J. Control Optim. 49(2), 830–858 (2011)
18. Lions, J.L.: Optimal control of systems governed by partial differential equations. Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin (1971)
19. Mordukhovich, B.S., Raymond, J.P.: Neumann boundary control of hyperbolic equations with pointwise state constraints. SIAM J. Control Optim. 43(4), 1354–1372 (electronic) (2004/05)
20. Tröltzsch, F., Volkwein, S.: The SQP method for control constrained optimal control of the Burgers equation. ESAIM Control Optim. Calc. Var. 6, 649–674 (2001)
21. Volkwein, S.: Distributed control problems for the Burgers equation. Comput. Optim. Appl. 18(2), 115–140 (2001)
22. Curtis, F.E., Guo, W.: Handling nonpositive curvature in a limited memory steepest descent method. IMA J. Numer. Anal. 36(2), 717–742 (2016)
23. De Asmundis, R., di Serafino, D., Hager, W.W., Toraldo, G., Zhang, H.: An efficient gradient method using the Yuan steplength. Comput. Optim. Appl. 59(3), 541–563 (2014)
24. De Asmundis, R., di Serafino, D., Riccio, F., Toraldo, G.: On spectral properties of steepest descent methods. IMA J. Numer. Anal. 33(4), 1416–1435 (2013)
25. 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)
26. Fletcher, R.: A limited memory steepest descent method. Math. Program. 135(1-2, Ser. A), 413–436 (2012)
27. Yuan, Y.x.: A new stepsize for the steepest descent method. J. Comput. Math. 24(2), 149–156 (2006)
28. Zheng, Y., Zheng, B.: A new modified Barzilai-Borwein gradient method for the quadratic minimization problem. J. Optim. Theory Appl. 172(1), 179–186 (2017)