myjournal 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 Due to simplicity, computational cheapness, and efficiency, the Barzilai and Borwein (BB) gradient method has received a significant amount of attention in different fields of optimization. In the first part of this paper, based on spectral analysis, R-linear global convergence for the BB-method is proven for strictly convex quadratic problems posed in infinite-dimensional Hilbert spaces. Then this result is strengthened to R-linear local convergence for a class of twice continuously Frećhet-differentiable functions. In the second part, aiming at problems governed by partial differential equations (PDE), the mesh-independent principle is investigated for the BB-method. The applicability of these results is demonstrated for three different types of PDE-constrained optimization problems. Numerical experiments illustrate the theoretical results.

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

Mathematics Subject Classification (2000) 49J20 · 90C48 · 65K05 · 90C52 · 49M25

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 [10] 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. Despite the simplicity and cheapness, this method has exhibited a surprisingly efficient numerical behaviour. This stimulated a significant amount of research. In the original work [10], the authors established R-superlinear convergence for two-dimensional strictly convex quadratic problems. Later, Raydan [70] and Dai and Liao [25] proved, respectively, global convergence and R-linear convergence rate of the BB-method for any finite-dimensional strictly convex quadratic problem.

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

Behzad Azmi
Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences
Altenbergerstraße 69, A-4040 Linz, Austria
E-mail: behzad.azmi@ricam.oeaw.ac.at

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

One of the important feature of this method is the nonmonotonicity in the values of the objective function and gradient norm. To preserve this feature, some authors [34,71] managed to prove the global convergence of the BB-method for finite-dimensional unconstrained optimization problems based on the nonmonotone line search techniques introduced in [33]. A deep analysis of the asymptotic behaviour of the BB-method was given in [22,28]. 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). Since then, inspired by the BB-method, many authors designed and analysed several step-length rules for the gradient method by investigating the role and behaviour of the eigenvalues of the Hessian matrix, rather than the decrease of the function values see e.g., [20,22,26,27,29, 47,85,87]. Due to simplicity and efficiency, the BB step-sizes have been being widely used in various fields of mathematical optimization and applications, including nonsmooth optimization [12,13,46,63,64,69,81, 83], inverse problems [18,32,54,60,65,72,73,79,84] constrained optimization [17,30,44,45,61,86].

In this work we aim to 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 are 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 this respect, we mention [7,8,9] in which the BB-method was efficiently employed in the context of the model predictive control for PDEs. In view of the above discussion, we are motivated to study the BB-method for a more general class of problems, namely, unconstrained problems posed in infinite-dimensional Hilbert spaces. Here we focus on the following unconstrained optimization problem

$$\min_{u\in\mathcal{H}} \mathcal{F}(u), \tag{1}$$

where $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ is a twice continuously Fréchet differentiable function defined on an abstract Hilbert space $\mathcal{H}$ with the inner product $(\cdot,\cdot)$ and its associated norm $\|\cdot\|$. The Barzilai-Borwein iterations for solving (1) are defined by

$$u_{k+1} = u_k - \frac{1}{\alpha_k}\mathcal{G}_k, \tag{2}$$

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

$$\alpha_k^{BB1} := \frac{(\mathcal{S}_{k-1}, \mathcal{Y}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{S}_{k-1})}, \quad \text{or} \quad \alpha_k^{BB2} := \frac{(\mathcal{Y}_{k-1}, \mathcal{Y}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{Y}_{k-1})}, \tag{3}$$

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. 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, as a first contribution, we analyse the convergence of Algorithm 1. Inspired by the result in [25] and based on the spectral theorem, we establish the R-linear global convergence of Algorithm 1 when it is applied to strictly convex quadratic problems defined by bounded uniformly positive self-adjoint operators. Then this result will be extended to a local convergence result for twice continuously Fréchet differentiable functions.

As the second contribution, we analyse the mesh independence principle (MIP) for Algorithm 1. This important property roughly states that the algorithm shows a similar convergence behaviour for the infinite-dimensional problem and its finite-dimensional approximations (discretized problems), independent of the mesh size. This concept of MIP was initially introduced in [1] for Newton's method. Since then, MIP was

Algorithm 1 BB-gradient

Require: Let one of the following initial conditions be satisfied:
- C1: Initial iterates $u_0, u_1 \in \mathcal{H}$ with $u_0 \neq u_1$ have been given.
- C2: An initial iterate $u_1 \in \mathcal{H}$ and an initial step-size α_1 with $\alpha_1 > 0$ have been given.

1: Set $k = 1$.
2: If $\|\mathcal{G}_k\| = 0$ stop.
3: If $k = 1$ and C2 holds go to Step 4, otherwise choose α_k equal 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.

studied for many different optimization algorithms and problem formulations. From these, we can mention generalized equations [2,3,4], Newton methods [48,80], SQP methods [77], shape design problems [57], constrained Gauss-Newton methods [38], gradient projection methods [52], quasi-Newton methods [49,50, 51], and semi-smooth Newton methods [40,41]. The convergence analysis of Algorithm 1 will show that, depending on the spectrum of the Hessian, the sequence $\{\|\mathcal{G}_k\|\}_k$ can be nonmonotone. This is the main reason which distinguishes our analysis from that in [49,50,51].

Our theoretical framework is supported by three optimizations problems with partial differential equations as constraints, including linear elliptic (Poisson equation), second-order linear hyperbolic (wave equation), and semilinear parabolic equations (viscous Burger equation). We show that our results are applicable to these problems and report our numerical experience for them.

The rest of paper is organized as follows: In Section 2, 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. Relying on this analysis, the local convergence of a class of nonlinear functions is discussed. Section 3 is devoted for developing the mesh-independent principle for Algorithm 1. In Section 4, the PDE-constrained optimal control problems alluded to above are investigated. Finally, Section 5 presents the numerical results.

2 Convergence Analysis

In this section, we are concerned with the convergence analysis of Algorithm 1. The section is divided in two parts. The first part deals with strictly convex quadratic problems defined by bounded self-adjoint operators. In particular, the case in which the operator is a compact perturbation of the identity will be treated in more detail. Strictly convex quadratic problems are of great importance, not only in their own right, but also as a model to study the behaviour of the algorithm for twice continuously Fréchet-differentiable functions in a neighbourhood of strong minima. In the second part, relying on the analysis of the first part, we discuss the local convergence of Algorithm 1 for twice continuously Fréchet-differentiable functions with Lipschitz continuous second derivatives.

2.1 Quadratic Functions

2.1.1 General Case

In this subsection, 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 $b \in \mathcal{H}$. In this case $\mathcal{G}_k := \mathcal{G}(u_k) = \mathcal{A}u_k - b$ and it can easily be shown that

$$\alpha_k^{BB1} = \frac{(\mathcal{S}_{k-1}, \mathcal{A}\mathcal{S}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{S}_{k-1})} = \frac{(\mathcal{G}_{k-1}, \mathcal{A}\mathcal{G}_{k-1})}{(\mathcal{G}_{k-1}, \mathcal{G}_{k-1})}, \tag{4}$$

$$\alpha_k^{BB2} = \frac{(\mathcal{S}_{k-1}, \mathcal{A}^2\mathcal{S}_{k-1})}{(\mathcal{S}_{k-1}, \mathcal{A}\mathcal{S}_{k-1})} = \frac{(\mathcal{G}_{k-1}, \mathcal{A}^2\mathcal{G}_{k-1})}{(\mathcal{G}_{k-1}, \mathcal{A}\mathcal{G}_{k-1})}, \tag{5}$$

where we have used that $\mathcal{S}_k = -\frac{1}{\alpha_k}\mathcal{G}_k$. 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 (4), (5), and the fact that

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

with $\bar{\mathcal{G}}_{k-1} := \mathcal{A}^{\frac{1}{2}}\mathcal{G}_{k-1}$, we infer that $\alpha_k^{BB1}, \alpha_k^{BB2} \in \mathcal{W}(\mathcal{A})$ for all $k \geq 1$. 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

$$\alpha_k^{BB1}, \alpha_k^{BB2} \in [\delta_{\inf}, \delta_{\sup}] \quad \text{for all } k \geq 1. \tag{6}$$

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 strict 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 [37,82], 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

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

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

$$f(\mathcal{A}) = \int_{\sigma(\mathcal{A})} f(\lambda) \, dE_\lambda, \tag{7}$$

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

$$(f(\mathcal{A})x, y) = \int_{\sigma(\mathcal{A})} f(\lambda) d(E_\lambda x, y), \tag{8}$$

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 (2) we have

$$\mathcal{G}_{k+1} = \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k \quad \text{for all } k = 1, 2, \ldots. \tag{9}$$

For $\mathcal{G}_1 \in \mathcal{H}$ we find

$$\mathcal{G}_1 = \int_{\sigma(\mathcal{A})} dE_\lambda \, \mathcal{G}_1, \quad \text{and} \quad \|\mathcal{G}_1\|^2 = \int_{\sigma(\mathcal{A})} d(E_\lambda \mathcal{G}_1, \mathcal{G}_1).$$

Using (7) and (9), we have

$$\mathcal{G}_2 = \frac{1}{\alpha_1}(\alpha_1 - \mathcal{A})\mathcal{G}_1 = \int_{\sigma(\mathcal{A})} \frac{1}{\alpha_1}(\alpha_1 - \lambda) \, dE_\lambda \mathcal{G}_1,$$

and, in a similar manner, we obtain

$$\mathcal{G}_k = \int_{\sigma(\mathcal{A})} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right) \right] dE_\lambda \mathcal{G}_1 \quad \text{for every } k = 1, 2, \ldots.$$

where $\prod_{p=1}^{0} = 1$. Moreover, we can write for $k = 1, 2, \ldots$

$$\begin{aligned} \|\mathcal{G}_{k+1}\|^2 &= \left(\frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k, \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})\mathcal{G}_k \right) \\ &= \left(\frac{1}{\alpha_k^2}(\alpha_k \mathcal{I} - \mathcal{A})^2 \mathcal{G}_k, \mathcal{G}_k \right) = \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k} \right)^2 d(E_\lambda \mathcal{G}_k, \mathcal{G}_k). \end{aligned} \tag{10}$$

Similarly, we have

$$\begin{aligned} \|\mathcal{G}_k\|^2 &= \left(\left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p} \right) \right] \mathcal{G}_1, \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p} \right) \right] \mathcal{G}_1 \right) \\ &= \left(\left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p} \right)^2 \right] \mathcal{G}_1, \mathcal{G}_1 \right) = \int_{\sigma(\mathcal{A})} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1). \end{aligned} \tag{11}$$

We define $\gamma_{\mathcal{A}} := \frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}$ and $\rho_{\mathcal{A}} := \frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\sup}}$. 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 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 (10), we have for $k \geq 1$ that

$$\|\mathcal{G}_{k+1}\|^2 = \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k} \right)^2 d(E_\lambda \mathcal{G}_k, \mathcal{G}_k). \tag{12}$$

Since $\delta_{\sup} < 2\delta_{\inf}$, it follows for every $k \geq 1$ and $\lambda \in \sigma(\mathcal{A})$ that

$$\left| \frac{\alpha_k - \lambda}{\alpha_k} \right|^2 \leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}} \right)^2 = \gamma_{\mathcal{A}}^2 < 1. \tag{13}$$

Using (12) and (13), we obtain

$$\|\mathcal{G}_{k+1}\|^2 \leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}} \right)^2 \int_{\sigma(\mathcal{A})} d(E_\lambda \mathcal{G}_k, \mathcal{G}_k) = \gamma_{\mathcal{A}}^2 \|\mathcal{G}_k\|^2 \quad \text{for every } k \geq 1. \tag{14}$$

Therefore, we can conclude that

$$\|\mathcal{G}_{k+1}\|^2 \leq \gamma_{\mathcal{A}}^{2k} \|\mathcal{G}_1\|^2 \quad \text{for every } k \geq 1,$$

and this completes the proof. □

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

Theorem 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 subsection. 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 $1 \leq i \leq n_\eta^u$, and

$$b_i := \begin{cases} \delta_{\inf} + i\eta & \text{for } 1 \leq i \leq n_\eta^u - 1, \\ \delta_{\sup} & \text{for } i = n_\eta^u, \end{cases}$$

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

$$I_i = \begin{cases} [a_i, b_i) & \text{for } 1 \leq i \leq n_\eta^u - 1, \\ [a_{n_\eta^u}, b_{n_\eta^u}] & \text{for } i = n_\eta^u. \end{cases} \tag{15}$$

By construction it is clear that $|I_i| \leq \eta$ for every $i = 1, \dots, n_\eta^u$, and

$$\sigma(\mathcal{A}) \subseteq [\delta_{\inf}, \delta_{\sup}] = \bigcup_{i=1}^{n_\eta^u} I_i. \tag{16}$$

For $i = 1, \cdots, n_\eta^u$, we define

$$(g_i^{k+1})^2 := \int_{I_i} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1), \tag{17}$$

and attain

$$\begin{aligned} \|\mathcal{G}_{k+1}\|^2 &= \int_{\sigma(\mathcal{A})} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = \int_{\delta_{\inf}}^{\delta_{\sup}} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\ &= \sum_{i=1}^{n_\eta^u} \int_{I_i} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = \sum_{i=1}^{n_\eta^u} (g_i^{k+1})^2. \end{aligned} \tag{18}$$

Moreover, we define

$$G(k, \ell) := \sum_{i=1}^{\ell} (g_i^k)^2 \quad \text{for every } k \geq 1 \text{ and } 1 \leq \ell \leq n_\eta^u, \tag{19}$$

where n_η^u is defined with respect to an interval length $\eta > 0$, and g_i^k is defined in (17). Then it is clear that

$$G(k, n_\eta^u) = \sum_{i=1}^{n_\eta^u} (g_i^k)^2 = \|\mathcal{G}_k\|^2 \quad \text{for every } k \geq 1.$$

In the following lemma we show that there exists an index n_η^l such that the sequences $\{g_i^k\}_k$ with $1 \leq i \leq n_\eta^l$ converge to zero Q-linearly as k tends to infinity.

Lemma 1 *For every $\eta \in (0, \rho_{\mathcal{A}} \delta_{\inf}]$, there exists a positive integer n_η^l with $1 \leq n_\eta^l \leq n_\eta^u$ such that for every $1 \leq i \leq n_\eta^l$, 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 $\eta \leq \rho_{\mathcal{A}} \delta_{\inf}$. Moreover, for every $\lambda \in I_i$ with $1 \leq i \leq n_\eta^l$ and every $p \geq 1$ we have the following two cases:

1. If $\alpha_p - \lambda \geq 0$, then we have
$$\left| \frac{\alpha_p - \lambda}{\alpha_p} \right| = \frac{\alpha_p - \lambda}{\alpha_p} \leq \rho_{\mathcal{A}} < 1.$$
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
$$\left| \frac{\alpha_p - \lambda}{\alpha_p} \right| = \frac{\lambda - \alpha_p}{\alpha_p} \leq \frac{(1 + \rho_{\mathcal{A}}) \delta_{\inf} - \delta_{\inf}}{\delta_{\inf}} = \rho_{\mathcal{A}}.$$

Therefore, we obtain

$$\begin{aligned} (g_i^{k+1})^2 &= \int_{I_i} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \leq \int_{I_i} \left(\frac{\alpha_k - \lambda}{\alpha_k} \right)^2 \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\ &\leq \rho_{\mathcal{A}}^2 \int_{I_i} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = \rho_{\mathcal{A}}^2 (g_i^k)^2. \end{aligned} \tag{20}$$

This concludes the proof. □

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

Lemma 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

$$G(k+j,\ell) \le \overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j \ge \overline{r} \tag{21}$$

holds for some positive $\overline{r} \in \mathbb{N}$ 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

$$(g_{\ell+1}^{k+j})^2 > 2\overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}, \tag{22}$$

we will show that

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

Due to (17), we have for every $k \ge 1$ that

$$\begin{aligned}
(g_{\ell+1}^{k+\overline{r}+1})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=1}^{k+\overline{r}} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \int_{I_{\ell+1}} \left[\prod_{p=k}^{k+\overline{r}} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&\le \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}}\right)^{2(\overline{r}+1)} \int_{I_{\ell+1}} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \gamma_{\mathcal{A}}^{2(\overline{r}+1)} (g_{\ell+1}^{k})^2 \le \gamma_{\mathcal{A}}^{2(\overline{r}+1)} \|\mathcal{G}_k\|^2.
\end{aligned} \tag{23}$$

Due to Algorithm 1, for every $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ we have one of the cases $\alpha_{k+j} = \alpha_{k+j}^{BB1}$ or $\alpha_{k+j} = \alpha_{k+j}^{BB2}$. Further, using (9), the fact that $\mathcal{A}$ is self-adjoint, and the spectral property (8), we have for every $k \ge 1$ and $q = 0, 1, 2$, that

$$\begin{aligned}
(\mathcal{G}_k, \mathcal{A}^q \mathcal{G}_k) &= \left(\left[\prod_{j=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_1, \mathcal{A}^q \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)\right] \mathcal{G}_1\right) \\
&= \left(\mathcal{A}^q \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \mathcal{A}}{\alpha_p}\right)^2\right] \mathcal{G}_1, \mathcal{G}_1\right) = \int_{\sigma(\mathcal{A})} \lambda^q \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda^q \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1).
\end{aligned} \tag{24}$$

Now, by using (4), (5), and (24), we can write for $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ that

$$\alpha_{k+j+1}^{BB1} = \frac{(\mathcal{G}_{k+j}, \mathcal{A}\mathcal{G}_{k+j})}{\mathcal{G}_{k+j}, \mathcal{G}_{k+j})} = \frac{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda \left[\prod_{p=1}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1)}{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \left[\prod_{p=1}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1)}, \tag{25}$$

and

$$\alpha_{k+j+1}^{BB2} = \frac{(\mathcal{G}_{k+j}, \mathcal{A}^2\mathcal{G}_{k+j})}{\mathcal{G}_{k+j}, \mathcal{A}\mathcal{G}_{k+j})} = \frac{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda^2 \left[\prod_{p=1}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1)}{\int_{\bigcup_{i=1}^{n_\eta^u} I_i} \lambda \left[\prod_{p=1}^{k+j-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1)}. \tag{26}$$

Moreover, due to (17) and (21), we have

$$\begin{aligned}\int_{\bigcup_{i=1}^{\ell} I_i}\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)&=\sum_{i=1}^{\ell}\int_{I_i}\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)\\ =\sum_{i=1}^{\ell}(g_i^{k+j})^2=G(k+j,\ell)&\le\overline{\zeta}\|\mathcal{G}_k\|^2\quad\text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\end{aligned}\tag{27}$$

For every $\lambda\in\bigcup_{i=1}^{\ell} I_i$, we have $\lambda\le a_{\ell+1}$. Thus, by (27), we can write

$$\begin{aligned}\int_{\bigcup_{i=1}^{\ell} I_i}\lambda\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)&\le a_{\ell+1}\int_{\bigcup_{i=1}^{\ell} I_i}\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)\\ =a_{\ell+1}G(k+j,\ell)&\le a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2\quad\text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\end{aligned}\tag{28}$$

From (25) and (27), we obtain

$$\frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}}\le\alpha_{k+j+1}^{BB1}\le\delta_{\sup}\quad\text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\},\tag{29}$$

where $\mathcal{Z}:=\int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)$. From (26), (28), and the fact that $\lambda\ge a_{\ell+1}$ for every $\lambda\in\bigcup_{i=\ell+1}^{n_\eta^u} I_i$, it follows that

$$\begin{aligned}\frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}}=\frac{a_{\ell+1}^2\mathcal{Z}}{a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2+\alpha_{\ell+1}\mathcal{Z}}&\le\frac{a_{\ell+1}\int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\lambda\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)}{a_{\ell+1}\overline{\zeta}\|\mathcal{G}_k\|^2+\int_{\bigcup_{i=\ell+1}^{n_\eta^u} I_i}\lambda\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)}\\ \le\alpha_{k+j+1}^{BB2}\le\delta_{\sup}&\quad\text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\end{aligned}\tag{30}$$

Now, using the fact that

$$\mathcal{Z}\ge\int_{I_{\ell+1}}\left[\prod_{p=1}^{k+j-1}\left(\frac{\alpha_p-\lambda}{\alpha_p}\right)^2\right]d(E_\lambda\mathcal{G}_1,\mathcal{G}_1)=(g_{\ell+1}^{k+j})^2,$$

and by (22), (29), and (30), we infer that for a chosen $\alpha_{k+j+1}=\alpha_{k+j+1}^{BB1}$ or $\alpha_{k+j+1}=\alpha_{k+j+1}^{BB2}$ that

$$\frac{2}{3}a_{\ell+1}=\frac{a_{\ell+1}\mathcal{Z}}{\frac{1}{2}\mathcal{Z}+\mathcal{Z}}\le\frac{a_{\ell+1}\mathcal{Z}}{\frac{1}{2}(g_{\ell+1}^{k+j})^2+\mathcal{Z}}\le\frac{a_{\ell+1}\mathcal{Z}}{\overline{\zeta}\|\mathcal{G}_k\|^2+\mathcal{Z}}\le\alpha_{k+j+1}\le\delta_{\sup}\quad\text{for all } j\in\{\overline{r},\cdots,\overline{r}+\Theta\}.\tag{31}$$

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 (6) 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 (31) and using the fact that $\lambda\le b_{\ell+1}\le a_{\ell+1}+\eta$ for $\lambda\in I_{\ell+1}$, we obtain

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

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

$$\left|1 - \frac{\lambda}{\alpha_{k+j+1}}\right| \leq c < 1. \tag{32}$$

Finally, by using (17) and (32) we obtain for every $j \in \{\overline{r}, \cdots, \overline{r} + \Theta\}$ that

$$\begin{aligned}
(g_{\ell+1}^{k+j+2})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=1}^{k+j+1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \int_{I_{\ell+1}} \left|1 - \frac{\lambda}{\alpha_{k+j+1}}\right|^2 \left[\prod_{p=1}^{k+j} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&\leq c^2 \int_{I_{\ell+1}} \left[\prod_{p=1}^{k+j} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = c^2 (g_{\ell+1}^{k+j+1})^2.
\end{aligned} \tag{33}$$

Using (23), (33), 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 *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 $k \geq 1$, there exist $r_\ell \in \mathbb{N}$ and $\zeta_\ell \in \mathbb{R}_+$ such that the condition*

$$G(k+j, \ell) \leq \zeta_\ell \|\mathcal{G}_k\|^2 \quad \textit{for all } j \geq r_\ell \tag{34}$$

holds. Then we show that for the choice of

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

with $\Theta_\ell := \Theta(\zeta_\ell, r_\ell)$ defined as in Lemma 2, we have

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

Proof First, observe that

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

Therefore, using (34) we only need to show that for every $j \geq r_{\ell+1}$

$$(g_{\ell+1}^{k+j})^2 \leq 2\gamma_{\mathcal{A}}^4 \zeta_\ell \|\mathcal{G}_k\|^2. \tag{35}$$

Due to Lemma 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

$$(g_{\ell+1}^{k+j})^2 \leq 2\zeta_\ell \|\mathcal{G}_k\|^2 \quad \text{for all } j_1 \leq j \leq j_2, \tag{36}$$

and

$$(g_{\ell+1}^{k+j_2+1})^2 > 2\zeta_\ell \|\mathcal{G}_k\|^2. \tag{37}$$

Note that if this case does not arise, clearly, (35) holds for all $j \geq j_s = j_1$ and since $\gamma_{\mathcal{A}} \geq 1$ the proof is finished. Further, we can write

$$(g_{\ell+1}^{k+j+1})^2 > 2\zeta_\ell \|\mathcal{G}_k\|^2 \quad \text{for all } j_2 \leq j \leq j_3 - 2, \tag{38}$$

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

$$(g_{\ell+1}^{k+j_3})^2 \leq 2\zeta_\ell \|\mathcal{G}_k\|^2. \tag{39}$$

Existence of such an index is justified using Lemma 2 for $\overline{r} = j_2$ and $\overline{\zeta} = \zeta_\ell$. Now, by (38) and using the same argument as in the proof of Lemma 2, where we have shown that from (22) implies (31), 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 2 we infer that

$$(g_{\ell+1}^{k+j+3})^2 \leq c^2 (g_{\ell+1}^{k+j+2})^2 \quad \text{for every } j_2 \leq j \leq j_3 - 2, \tag{40}$$

where $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2} + \frac{\eta}{\delta_{\inf}}\} < 1$. Finally, using (17) and (36), we have for $r = 1, 2$

$$\begin{aligned}
(g_{\ell+1}^{k+j_2+r})^2 &= \int_{I_{\ell+1}} \left[\prod_{p=1}^{k+j_2+r-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \int_{I_{\ell+1}} \left[\prod_{p=k+j_2}^{k+j_2+r-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] \left[\prod_{p=1}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&= \int_{I_{\ell+1}} \left[\prod_{p=1}^{r} \left(\frac{\alpha_{k+j_2+p-1} - \lambda}{\alpha_{k+j_2+p-1}} \right)^2 \right] \left[\prod_{p=1}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\
&\leq \left(\frac{\delta_{\sup} - \delta_{\inf}}{\delta_{\inf}} \right)^{2r} \int_{I_{\ell+1}} \left[\prod_{p=1}^{k+j_2-1} \left(\frac{\alpha_p - \lambda}{\alpha_p} \right)^2 \right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = \gamma_{\mathcal{A}}^{2r} (g_{\ell+1}^{k+j_2})^2.
\end{aligned} \tag{41}$$

Now since $c < 1$ and $\gamma_{\mathcal{A}} \geq 1$ due the fact that $\delta_{\sup} \geq 2\delta_{\inf}$, we obtain from (36), (40), and (41) that

$$(g_{\ell+1}^{k+j+3})^2 \leq \gamma_{\mathcal{A}}^4 (g_{\ell+1}^{k+j_2})^2 \leq 2\zeta_\ell \gamma_{\mathcal{A}}^4 \|\mathcal{G}_k\|^2 \quad \text{for every } j_2 - 2 \leq j \leq j_3 - 2, \tag{42}$$

and as a consequence, we obtain

$$(g_{\ell+1}^{k+j})^2 \leq 2\zeta_\ell \gamma_{\mathcal{A}}^4 \|\mathcal{G}_k\|^2 \quad \text{for every } j_2 + 1 \leq j \leq j_3 + 1. \tag{43}$$

From (43) and (36) we conclude that (35) 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 (39) and repeating the same argument, it can be shown that (35) holds for every $j \geq j_1$. Recall that $j_1 \in \{r_\ell, \cdots, r_\ell + \Theta_\ell + 1\}$. Therefore (35) 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 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*

$$\|\mathcal{G}_{k+m}\| \leq \frac{1}{2} \|\mathcal{G}_k\| \quad \textit{for all } k \geq 1, \tag{44}$$

or equivalently,

$$\|u_{k+m} - u^*\| \leq \frac{1}{2} \|u_k - u^*\| \quad \textit{for all } k \geq 1, \tag{45}$$

for all initial iterates $u_0, u_1 \in \mathcal{H}$ with $u_0 \neq u_1$ in the condition C1, or every initial iterate $u_1 \in \mathcal{H}$ and every initial step-size $\alpha_1 > 0$ in the condition C2.

Proof If $\delta_{\sup} < 2\delta_{\inf}$, then $\gamma_{\mathcal{A}} < 1$ and, by (14) in the proof of Lemma 1, we have

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

Therefore, (44) 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 (15) with a fixed length $\eta \in (0, \frac{\delta_{\inf}}{2}) \subset (0, \rho_{\mathcal{A}}\delta_{\inf}]$. Then due to (18) and (19), we have for every $k \geq 1$

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

where $(g_i^k)^2$ is defined by (17). Moreover due to (20) in the proof of Lemma 1, there exists an integer $n_\eta^l > 0$ such that for every ℓ with $1 \leq \ell \leq n_\eta^l$, we have

$$(g_\ell^{k+j})^2 \leq \rho_{\mathcal{A}}^{2j} (g_\ell^k)^2 \quad \text{for every } j \geq 0 \text{ and } k \geq 1.$$

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

$$G(k+j, n_\eta^l) \leq \rho_{\mathcal{A}}^{2j} G(k, n_\eta^l) \leq \rho_{\mathcal{A}}^{2j} \|\mathcal{G}_k\|^2 \quad \text{for every } j \geq 0 \text{ and } k \geq 1.$$

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

$$G(k+j, n_\eta^l) \leq \zeta_{n_\eta^l} \|\mathcal{G}_k\|^2 \quad \text{for every } j \geq r_{n_\eta^l} \text{ and } k \geq 1,$$

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. By using this lemma we have for ℓ with $n_\eta^l \leq \ell \leq n_\eta^u - 1$ 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 2. To be more precise, by applying Lemma 3 once, for the first iteration, we obtain

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

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

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

By putting $m = r_{n_\eta^u}$, (44) holds.

Moreover, the equivalence of (44) with (45) is justified due the fact that, similarly to (9) for $\mathcal{G}_k$, it can easily be shown that

$$(u_{k+1} - u^*) = \frac{1}{\alpha_k}(\alpha_k \mathcal{I} - \mathcal{A})(u_k - u^*) \quad \text{for all } k = 0, 1, 2, \ldots. \tag{46}$$

Hence, the same machinery can be used to derive (45) and this completes the proof. □

Proof of Theorem 2. We need only to consider the case in which for every $k \geq 0$ we have $u_k \neq u^*$. In this case, we will show that $u_k \to u^*$ R-linearly. Due to (46) and with a similar argument as in (10), we can write

$$\begin{aligned} \|u_{k+1} - u^*\|^2 &= \int_{\sigma(\mathcal{A})} \left(\frac{\alpha_k - \lambda}{\alpha_k}\right)^2 d(E_\lambda(u_k - u^*), (u_k - u^*)) \\ &\leq \gamma_{\mathcal{A}}^2 \int_{\sigma(\mathcal{A})} d(E_\lambda(u_k - u^*), (u_k - u^*)) = \gamma_{\mathcal{A}}^2 \left\|u_k - u^*\right\|^2 \quad \text{for every } k \geq 1. \end{aligned} \tag{47}$$

Moreover, due to (45) in Lemma 4, we obtain

$$\|u_{jm+1} - u^*\| \leq (\frac{1}{2})^j \|u_1 - u^*\| \quad \text{for all } j \geq 0, \tag{48}$$

where m has been defined in Lemma 4. Now for every $k \geq 1$, there exists an integer j such that $1 + jm \leq k < 1 + (j+1)m$. Therefore, it follows that $k - (jm+1) < m$ and $j \geq \frac{k}{m} - 1$. Using (47) and (48), we obtain

$$\begin{aligned}\|u_k - u^*\| &\leq \gamma_{\mathcal{A}}^m \|u_{jm+1} - u^*\| \leq \gamma_{\mathcal{A}}^m (\frac{1}{2})^j \|u_1 - u^*\| \leq \gamma_{\mathcal{A}}^m (\frac{1}{2})^{\frac{k}{m}-1} \|u_1 - u^*\| \\ &= c_1 c_2^k \|u_1 - u^*\| \qquad \text{for all } k \geq 1,\end{aligned}$$

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

Remark 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 $k \geq 1$ that

$$\begin{aligned}\|\mathcal{G}_{k+1}\|^2 &= \int_{\delta_{\inf}}^{\delta_{\sup}} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) = \sum_{i=1}^{n_\eta^u} \int_{I_i} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1) \\ &= \sum_{i=1}^{m} \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \delta_{\inf}}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}} \mathcal{G}_1\|^2 = \sum_{i=1}^{m} (g_i^{k+1})^2,\end{aligned} \tag{49}$$

where $(g^{k+1})_i^2 := \left[\prod_{p=1}^{k} \left(\frac{\alpha_p - \delta_{\inf}}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}} \mathcal{G}_1\|^2$. Then the statements of Lemma 1 is true for $n_\eta^l = 1$. Further, Lemma 4 and Theorem 2 are applicable. Moreover, in the proof of Lemma 2, similarly to (49), all the integrations are replaced by finite sum and it follows that $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2}\}$. See [25,70] for more details.

Remark 2 Note that, due to Theorem 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 1, Algorithm 1 is Q-linearly convergent with the rate $\gamma_{\mathcal{A}} < 1$. Moreover, from (14), we infer that the sequence $\{\|\mathcal{G}_k\|\}_k$ in monotone decreasing.
2. $\kappa(\mathcal{A}) \geq 2$: This case is more delicate. Recall from (18) 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

$$(g_i^{k+1})^2 = \int_{I_i} \left(1 - \frac{\lambda}{\alpha_k}\right)^2 \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda}{\alpha_p}\right)^2\right] d(E_\lambda \mathcal{G}_1, \mathcal{G}_1). \tag{50}$$

   Due to Lemma 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 (50), 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 (Preconditioning) Due to Remark 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 consequnce, 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 [67], 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 [5,6,31,39, 66].

2.1.2 The Case of a Positive Compact Perturbation

In many situations of practical importance, we are faced with problems of the form (QP), in which $\mathcal{A} : \mathcal{H} \to \mathcal{H}$ is a compact perturbation of the identity. Therefore, it is of interest to consider this case separately. Here, we have

$$\mathcal{A} := \mathcal{T} + \beta\mathcal{I} \tag{poco}$$

with a positive self-adjoint compact operator $\mathcal{T} : \mathcal{H} \to \mathcal{H}$ and a constant $\beta > 0$. In this subsection we show how the special form of the spectrum of operators with the form (poco) allows us to simplify the proof of Theorems 1 and 2 of the previous section. For $\mathcal{T}$ as above we have $\sigma(\mathcal{T}) = \{0\} \cup \Sigma_{\mathcal{T}}$, where $\Sigma_{\mathcal{T}} := \{\delta_i : i \in N\}$ admits an enumeration for a countable set N. This set contains an ordered sequence of nonzero pairwise distinct eigenvalues, i.e, $\delta_{i+1} < \delta_i$ for $i \in N$. Moreover $\delta_1 = \|\mathcal{T}\|$ and for every $i \in N$, we have $\dim(\ker(\delta_i - \mathcal{T})) < \infty$ where $\ker(L) := \{u \in \mathcal{H} : Lu = 0\}$ for a given linear operator $L : \mathcal{H} \to \mathcal{H}$. Further, if N is infinite, $\delta_n \to 0$ and N can be taken to be $\mathbb{N} := \{1, 2, 3, \dots\}$. Then we have

$$\mathcal{T} = \sum_{i \in N} \delta_i E_{\{\delta_i\}},$$

where $E_{\{\delta_i\}}$ is the orthogonal projection to the space $\ker(\mathcal{T} - \delta_i)$. In the case that N is finite, the convergence can be proven as explained in Remark 1, therefore we assume here that N is infinite. Then, due to spectral mapping Theorem, we have

$$\sigma(\mathcal{A}) = \{\beta\} \cup \{\lambda_i : \lambda_i = \delta_i + \beta \text{ for } i \in N\}, \tag{51}$$

where β is a cluster point of the spectrum with $\lambda_i \to \beta$, and $\lambda_{i+1} < \lambda_i$ for ever $i \geq 1$. Then, for every measurable function $f : \sigma(\mathcal{A}) \to \mathbb{R}$, the operator $f(\mathcal{A})$ is defined by

$$f(\mathcal{A}) = \int_{\sigma(\mathcal{A})} f(\lambda)\, dE_\lambda = \sum_{i=1}^{\infty} f(\lambda_i) E_{\{\lambda_i\}} + f(\beta) E_{\{\beta\}},$$

where $E_{\{\beta\}}$ is the orthogonal projection to the space $\ker(\mathcal{T})$. Note that $E_{\{\beta\}} = 0$ unless zero is an eigenvalue of $\mathcal{T}$. For convenience in notation, we denote $\lambda_0 = \beta$. Then we have

$$f(\mathcal{A}) = \sum_{i=0}^{\infty} f(\lambda_i) E_{\{\lambda_i\}}, \tag{52}$$

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

$$(f(\mathcal{A})x, y) = \int_{\sigma(\mathcal{A})} f(\lambda) d(E_\lambda x, y) = \sum_{i=0}^{\infty} f(\lambda_i)(E_{\{\lambda_i\}} x, y). \tag{53}$$

Due to structure of $\mathcal{A}$ and the definition of $\delta_{\inf}$ and $\delta_{\sup}$, we have

$$\delta_{\inf} = \lambda_0 = \beta, \quad \delta_{\sup} = \lambda_1 = \|\mathcal{T}\| + \beta. \tag{54}$$

Moreover, due to (52) and (53), for $\mathcal{G}_1 \in \mathcal{H}$ we can write

$$\mathcal{G}_1 = \sum_{i=0}^{\infty} E_{\{\lambda_i\}} \mathcal{G}_1 \quad \text{and} \quad \|\mathcal{G}_1\|^2 = \sum_{i=0}^{\infty} (E_{\{\lambda_i\}} \mathcal{G}_1, \mathcal{G}_1) = \sum_{i=0}^{\infty} \|E_{\{\lambda_i\}} \mathcal{G}_1\|^2.$$

Using (9) and (52), we have

$$\mathcal{G}_2 = \frac{1}{\alpha_1}(\alpha_1 - \mathcal{A})\mathcal{G}_1 = \sum_{i=0}^{\infty} \frac{1}{\alpha_1}(\alpha_1 - \lambda_i) E_{\{\lambda_i\}} \mathcal{G}_1,$$

and, in a similar manner by induction, we obtain for every $k \geq 1$ that

$$\mathcal{G}_k = \sum_{i=0}^{\infty} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda_i}{\alpha_p} \right) \right] E_{\{\lambda_i\}} \mathcal{G}_1 \quad \text{and} \quad \|\mathcal{G}_k\|^2 = \sum_{i=0}^{\infty} \left[\prod_{p=1}^{k-1} \left(\frac{\alpha_p - \lambda_i}{\alpha_p} \right)^2 \right] \|E_{\{\lambda_i\}} \mathcal{G}_1\|^2. \tag{55}$$

For every $i \geq 0$ and $k \geq 1$ we define

$$(g_i^k)^2 := \left[\prod_{p=1}^{k-1}\left(\frac{\alpha_p - \lambda_i}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}}\mathcal{G}_1\|^2 = \left[\prod_{p=1}^{k-1}\left(\frac{\alpha_p - \lambda_i}{\alpha_p}\right)^2\right] (g_i^1)^2, \tag{56}$$

and conclude that

$$\|\mathcal{G}_k\|^2 = \sum_{i=0}^{\infty}(g_i^k)^2 \quad \text{and} \quad (g_i^{k+1})^2 = \left(\frac{\alpha_k - \lambda_i}{\alpha_k}\right)^2 (g_i^k)^2. \tag{57}$$

Using (6) and (55), we can write for every $k \geq 1$ and any chosen $\alpha_p = \alpha_p^{BB1}$ or $\alpha_p = \alpha_p^{BB2}$ with $p = 1, 2, \ldots, k$ that

$$\|\mathcal{G}_{k+1}\|^2 = \sum_{i=0}^{\infty}\left[\prod_{p=1}^{k}\left(\frac{\alpha_p - \lambda_i}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}}\mathcal{G}_1\|^2 \leq \gamma_{\mathcal{A}}^2 \sum_{i=0}^{\infty}\left[\prod_{p=1}^{k-1}\left(\frac{\alpha_p - \lambda_i}{\alpha_p}\right)^2\right] \|E_{\{\lambda_i\}}\mathcal{G}_1\|^2 = \gamma_{\mathcal{A}}^2\|\mathcal{G}_k\|^2. \tag{58}$$

Hence, analogously to Theorem 1, we can conclude that for (QP) with an operator $\mathcal{A}$ of the form (poco), Algorithm 1 is Q-linearly convergent, provided that $\delta_{\sup} < 2\delta_{\inf}$. Next we consider the general case. In a similar manner as in the previous subsubsection, we define

$$G(k, \ell) := \sum_{i \in \{0\} \cup \{i : i \geq \ell\}} (g_i^k)^2 \quad \text{for every } \ell, k \geq 1.$$

Then, by (55) we have

$$G(k, 1) = \sum_{i \in \{0\} \cup \{i : i \geq 1\}} (g_i^k)^2 = \sum_{i=1}^{\infty}(g_i^k)^2 = \|\mathcal{G}_k\|^2 \quad \text{for every } k \geq 1. \tag{59}$$

First, analogously to Lemma 1, we will show that that there exists an index n^u depending on $\delta_{\inf}$ and $\delta_{\sup}$ such that the sequences $\{g_i^k\}_k$ with $i \geq n^u$ converge to zero Q-linearly.

Lemma 5 *There exists a positive integer n^u such that for any $i \in \{0\} \cup \{i : i \geq n^u\}$, the sequences $\{g_i^k\}_k$ converge to zero Q-linearly with the factor $\rho_{\mathcal{A}}$ as k tends to infinity and we also have*

$$\lim_{k \to \infty} G(k, n^u) = 0. \tag{60}$$

Proof Since $\{\lambda_i\}_i$ is a positive and decreasing sequence and $\lambda_i \to \beta = \delta_{\inf}$, there exists a positive integer n^u such that for every $i \geq n^u$ we have $\lambda_i - \delta_{\inf} \leq \rho_{\mathcal{A}}\delta_{\inf}$. For every $i \in \{0\} \cup \{i : i \geq n^u\}$ and $k \geq 1$, we have the following two cases:

1. If $\alpha_k - \lambda_i \geq 0$ then we have
$$\left|\frac{\alpha_k - \lambda_i}{\alpha_k}\right| = \left(1 - \frac{\lambda_i}{\alpha_k}\right) \leq \rho_{\mathcal{A}} < 1.$$
2. If $\alpha_k - \lambda_i < 0$ then we have $\alpha_k \in [\delta_{\inf}, \lambda_i)$ and thus
$$\left|\frac{\alpha_k - \lambda_i}{\alpha_k}\right| = \frac{\lambda_i - \alpha_k}{\alpha_k} \leq \frac{\lambda_i - \delta_{\inf}}{\delta_{\inf}} \leq \rho_{\mathcal{A}}.$$

Therefore, by using (56) and (57), we can infer for every $k \geq 1$ and $i \in \{0\} \cup \{i : i \geq n^u\}$ that

$$(g_i^{k+1})^2 = \left(\frac{\alpha_k - \lambda_i}{\alpha_k}\right)^2 (g_i^k)^2 \leq \rho_{\mathcal{A}}^2 (g_i^k)^2 \leq \cdots \leq \rho_{\mathcal{A}}^{2k}(g_i^1)^2. \tag{61}$$

Now due to (61), we conclude that the sequences $\{g_i^k\}_k$ converge to zero Q-linearly for all $i \in \{0\} \cup \{i : i \geq n^u\}$. Moreover, using (61) and summing up for every $i \in \{0\} \cup \{i : i \geq n^u\}$ we obtain for every $k \geq 1$ that

$$G(k+1, n^u) = \sum_{i \in \{0\} \cup \{i : i \geq n^u\}} (g_i^{k+1})^2 \leq \rho_{\mathcal{A}}^{2k} \sum_{i \in \{0\} \cup \{i : i \geq n^u\}} (g_i^1)^2 = \rho_{\mathcal{A}}^{2k}\|\mathcal{G}_1\|^2.$$

Thus (60) follows. □

Next, we consider results which are analogous to Lemmas 2 and 3 for the special case (poco).

Lemma 6 *For each integer ℓ with $1 < \ell \leq n^u$, and $k \geq 1$, the following property holds:*
If the condition

$$G(k+j,\ell) \leq \overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j \geq \overline{r} \tag{62}$$

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

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

where $\Theta = \Theta(\overline{\zeta}, \overline{r})$ is defined as in Lemma 2 with $c := \max\{\rho_{\mathcal{A}}, \frac{1}{2}\}$.

Proof Similarly to the proof of Lemma 2, we assume that

$$(g_{\ell-1}^{k+j})^2 > 2\overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for all } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}, \tag{63}$$

and we show that

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

By Algorithm 1, for every $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ we have either $\alpha_{k+j} = \alpha_{k+j}^{BB1}$ or $\alpha_{k+j} = \alpha_{k+j}^{BB2}$. Using (4), (5), and (53), we can write for every $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ that

$$\alpha_{k+j+1}^{BB1} = \frac{\sum_{i=0}^{\infty}(g_i^{k+j})^2\lambda_i}{\sum_{i=0}^{\infty}(g_i^{k+j})^2}, \text{ and } \quad \alpha_{k+j+1}^{BB2} = \frac{\sum_{i=0}^{\infty}(g_i^{k+j})^2\lambda_i^2}{\sum_{i=0}^{\infty}(g_i^{k+j})^2\lambda_i}. \tag{64}$$

Using (62) and (64), we obtain for every $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ that

$$\frac{\lambda_{\ell-1}\sum_{i=1}^{\ell-1}(g_i^{k+j})^2}{\overline{\zeta}\|\mathcal{G}_k\|^2 + \sum_{i=1}^{\ell-1}(g_i^{k+j})^2} \leq \alpha_{k+j+1}^{BB1} \leq \delta_{\sup}. \tag{65}$$

Moreover, using (62) and the fact that $\lambda_{\ell-1} \geq \lambda_i$ for every $i \in \{0\} \cup \{i : i \geq \ell\}$, we obtain

$$\sum_{i\in\{0\}\cup\{i:i\geq\ell\}} (g_i^{k+j})^2\lambda_i \leq \lambda_{\ell-1}G(k+j,\ell) \leq \lambda_{\ell-1}\overline{\zeta}\|\mathcal{G}_k\|^2 \quad \text{for every } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}. \tag{66}$$

Then by using (62) and (64), we have for every $j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}$ that

$$\begin{aligned}\frac{\lambda_{\ell-1}\sum_{i=1}^{\ell-1}(g_i^{k+j})^2}{\overline{\zeta}\|\mathcal{G}_k\|^2 + \sum_{i=1}^{\ell-1}(g_i^{k+j})^2} = \frac{\lambda_{\ell-1}^2\sum_{i=1}^{\ell-1}(g_i^{k+j})^2}{\lambda_{\ell-1}(\overline{\zeta}\|\mathcal{G}_k\|^2 + \sum_{i=1}^{\ell-1}(g_i^{k+j})^2)} &\leq \frac{\lambda_{\ell-1}\sum_{i=1}^{\ell-1}(g_i^{k+j})^2\lambda_i}{\lambda_{\ell-1}\overline{\zeta}\|\mathcal{G}_k\|^2 + \sum_{i=1}^{\ell-1}(g_i^{k+j})^2\lambda_i} \\ &\leq \alpha_{k+j+1}^{BB2} \leq \delta_{\sup}.\end{aligned} \tag{67}$$

From (63), (65), (67), and the fact that $\sum_{i=1}^{\ell-1}(g_i^k)^2 \geq (g_{\ell-1}^k)^2$, it follows, with a computations similar to those in the proof of Lemma 2 which leads to (31), for a chosen $\alpha_{k+j+1} = \alpha_{k+j+1}^{BB1}$ or $\alpha_{k+j+1} = \alpha_{k+j+1}^{BB2}$ that

$$\frac{2}{3}\lambda_{\ell-1} \leq \alpha_{k+j+1} \leq \delta_{\sup} \quad \text{for all } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}. \tag{68}$$

Moreover, considering separately the cases $\frac{\lambda_{\ell-1}}{\alpha_{k+j+1}} < 1$ and $\frac{\lambda_{\ell-1}}{\alpha_{k+j+1}} \geq 1$, we obtain due to (68) that

$$\left|1 - \frac{\lambda_{\ell-1}}{\alpha_{k+j+1}}\right| \leq c < 1 \quad \text{for all } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}, \tag{69}$$

where $c = \max\{\frac{1}{2}, \rho_{\mathcal{A}}\}$. Using (57) and (69) we conclude that

$$|g_{\ell-1}^{k+j+2}| = \left|1 - \frac{\lambda_{\ell-1}}{\alpha_{k+j+1}}\right| |g_{\ell-1}^{k+j+1}| \leq c|g_{\ell-1}^{k+j+1}| \quad \text{for all } j \in \{\overline{r}, \cdots, \overline{r}+\Theta\}. \tag{70}$$

Finally, by (57) and (70) 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. □

Next since we have

$$G(k+j,\ell-1) = G(k+j,\ell) + (g_{\ell-1}^{k+j})^2 \quad \text{for every } k,j \geq 1,$$

with a similar argument as in the proof of Lemma 3, the following lemma can be proven.

Lemma 7 *Let $\delta_{\sup} \geq 2\delta_{\inf}$. Moreover, assume that for any integer ℓ with $1 < \ell \leq n^u$, and $k \geq 1$, there exist positive numbers r_ℓ and ζ_ℓ such that (62) holds for $\overline{r} = r_\ell$ and $\overline{\zeta} = \zeta_\ell$. Then for the choice of $\zeta_{\ell-1} := (1+2\gamma_{\mathcal{A}}^4)\zeta_\ell$ and $r_{\ell-1} := r_\ell + \Theta_\ell + 1$, we have*

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

Finally by using Lemma 5 and 7, we can prove Lemma 4 for the special case in which $\mathcal{A}$ is a compact perturbation of the identity. For the case of $\delta_{sup} < 2\delta_{inf}$ the proof is similar to the proof of Lemma 4. Here we give some hints about the other case, namely, $\delta_{sup} \geq 2\delta_{inf}$. First, due to Lemma 5, for $\zeta_{n^u} := \frac{1}{4}(1+4\gamma_{\mathcal{A}}^2)^{1-n^u}$ there exists an integer $r_{n^u} > 0$ such that

$$G(k,n^u) \leq \zeta_{n^u}\|\mathcal{G}_k\|^2 \quad \text{for all } k \geq r_{n^u},$$

where $n^u = n^u(\delta_{\inf},\delta_{\sup}) > 0$ is defined in Lemma 5. Then, similarly to the proof of Lemma 4, by induction and using Lemma 7, we have for ℓ with $2 \leq \ell \leq n^u$ that

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

Therefore, due to (59), (44) holds for $m := r_1$.

Now we can conclude that Theorems 1 and 2 hold for (poco).

Remark 4 The explanations in Remark 2 also apply for operators $\mathcal{A}$ of the form (poco). Due to (57), the case $\kappa(\mathcal{A}) < 2$ is analogous to that in Remark 2. For the case $\kappa(\mathcal{A}) \geq 2$, we can gain more information since the spectrum of $\sigma(\mathcal{A})$ is discrete. Due to the right equality in (57) we have

$$|g_i^{k+1}| = \left|\frac{\alpha_k - \lambda_i}{\alpha_k}\right| |g_i^k|. \tag{71}$$

Therefore, for $i = 0,1,\ldots$ we obtain $g_i^{k+1} = 0$ if $\alpha_k = \lambda_i$. Moreover, due to Lemma 5, there exists an index $n^u \geq 1$ such that the sequences $\{|g_i^k|\}_k$ with $i \in \{0\} \cup \{i : i \geq n^u\}$ are Q-linearly convergent with factor $\rho_{\mathcal{A}} < 1$. Therefore, we need only to consider the values of $|g_i^k|$ with $i = 1,\ldots,n^u - 1$. From (71), for any λ_i close to α_k we have a significant reduction and $|g_i^{k+1}| \ll |g_i^k|$, while for $\lambda_i > 2\alpha_k$, we obtain $|g_i^{k+1}| > |g_i^k|$. These facts clarify the potential nonmonotonic behaviour in the sequence $\{\|\mathcal{G}_k\|\}_k$. In fact, for α_k close to λ_1, the coefficients g_i decrease in modulus, but the changes in g_i with $i \in \{0\} \cup \{i : i \geq n^u\}$ are negligible provided that $\kappa(\mathcal{A})$ is large. Furthermore, small values of α_k, tend to diminish the components $|g_i^k|$ for small $i \neq 0$ and thus, enhance the relative contribution of components for large i.

2.2 General Objective Function

In this section, we will prove the local R-linear convergence of Algorithm 1, in the case that this algorithm is applied for finding a local minimum $u^* \in \mathcal{H}$ of a not necessarily quadratic function $\mathcal{F} : \mathcal{H} \to \mathbb{R}$. More precisely, $\mathcal{F}$ is twice continuously Fréchet-differentiable at u^* with Lipschitz continuous second derivative $\mathcal{F}''$ in a neighbourhood of $u^* \in \mathcal{H}$. Then if we identify the first derivative $\mathcal{F}'$ by its corresponding representation $\mathcal{G}$, we have the following first-order optimality condition

$$\mathcal{G}(u^*) = 0 \quad \text{in } \mathcal{H}. \tag{EP}$$

Due to the continuity of the bilinear map $\mathcal{F}''(u^*)$, the exists a positive constant $\delta_{\sup}$ such that

$$\mathcal{F}''(u^*)(v,u) \leq \delta_{\sup}\|v\|\|u\| \quad \text{for all } u,v \in \mathcal{H}. \tag{72}$$

Moreover, we assume that the continuous bilinear map $\mathcal{F}''(u^*)$ is uniformly positive, that is

$$\delta_{\inf}\|v\|^2 \leq \mathcal{F}''(u^*)(v,v) \leq \delta_{\sup}\|v\|^2 \quad \text{for all } v \in \mathcal{H}, \tag{73}$$

where $\delta_{\sup} \geq \delta_{\inf} > 0$. Then, due to the Riesz representation theorem, there exists a unique self-adjoint bounded operator $\mathcal{A}^{\mathcal{F}}_{u^*}$ (see [19][Theorem 2.2, page 31]) such that

$$\mathcal{F}''(u^*)(v,u) = (\mathcal{A}^{\mathcal{F}}_{u^*}v, u) \quad \text{for all } v, u \in \mathcal{H}.$$

Similarly to the analysis of [24,62], the R-linearly convergence result is proven by comparing the sequences $\{u_k\}_k$ and $\{\hat{u}_k\}_k$ which are generated by Algorithm 1 applied to, respectively, $\mathcal{F}$ and its second-order Taylor approximation $\hat{\mathcal{F}}$ defined by

$$\hat{\mathcal{F}}(u) = \mathcal{F}(u^*) + \frac{1}{2}(\mathcal{A}^{\mathcal{F}}_{u^*}(u-u^*), u-u^*). \tag{74}$$

Throughout this section, all notations with the accent " ˆ " are related to the quadratic approximation (74). For instance with $\hat{\mathcal{G}}(\cdot)$ and $\hat{\alpha}_k$, we denote the gradient and the step-sizes of Algorithm 1 applied to $\hat{\mathcal{F}}$, respectively.

Since $\mathcal{F}'' : \mathcal{H} \to \mathcal{L}(\mathcal{H}, \mathcal{L}(\mathcal{H}, \mathbb{R}))$ is locally Lipschitz continuous and $\mathcal{F}''(u^*) : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is continuous and uniformly positive, there exist a ball $\mathcal{B}_\tau(u^*)$ centered at u^* with a radius τ, positive constants $\alpha_{\inf}$, $\alpha_{\sup}$ depending on τ, and L such that

$$\|\mathcal{G}(u) - \mathcal{A}^{\mathcal{F}}_{u^*}(u-u^*)\| \leq L\|u-u^*\|^2 \quad \text{for all } u \in \mathcal{B}_\tau(u^*), \tag{$L1$}$$

and

$$\alpha_{\inf}\|v\|^2 \leq \mathcal{F}''(u)(v,v) \leq \alpha_{\sup}\|v\|^2 \quad \text{for all } v \in \mathcal{H} \text{ and } u \in \mathcal{B}_\tau(u^*). \tag{$L2$}$$

Moreover, due to the mean value theorem we have

$$\alpha_{\inf} \leq \alpha_k^{BB1}, \alpha_k^{BB2} \leq \alpha_{\sup}, \tag{75}$$

provided that u_k and u_{k-1} belong to $\mathcal{B}_\tau(u^*)$. Moreover if the iterations of Algorithm 1 applied to $\hat{\mathcal{F}}$ lie in $\mathcal{B}_\tau(u^*)$, we will also have

$$\alpha_{\inf} \leq \delta_{\inf} \leq \hat{\alpha}_k^{BB1}, \hat{\alpha}_k^{BB2} \leq \delta_{\sup} \leq \alpha_{\sup}. \tag{76}$$

Further, by the fundamental theorem of calculus, we infer that

$$\alpha_{\inf}\|u-u^*\| \leq \|\mathcal{G}(u)\| = \|\mathcal{G}(u) - \mathcal{G}(u^*)\| \leq \alpha_{\sup}\|u-u^*\| \quad \text{for all } u \in \mathcal{B}_\tau(u^*). \tag{77}$$

In the next lemma we study the distance of the sequences $\{u_k\}_k$ and $\{\hat{u}_k\}_k$.

Lemma 8 *Let u^* be a local minimizer of $\mathcal{F}$ with $\mathcal{F} \in C^2(\mathcal{H}, \mathbb{R})$ and assume that $L1$ and $L2$ hold for a radius τ and constants $\alpha_{\inf}$ and $\alpha_{\sup}$, and for the bilinear form $\mathcal{F}''(u^*)$ estimate* (73) *holds with the constants $\delta_{\sup}$ and $\delta_{\inf}$. Further, let $\{u_j\}_j$ be a sequence generated by Algorithm 1 applied to $\mathcal{F}$, and $\{\hat{u}_j^k\}_j$ be the sequence generated by Algorithm 1 applied to the quadratic approximation* (74) *of $\mathcal{F}$ at u^* with an initial iterate u_k and an initial step-size α_k with $k \geq 1$. Then for any fixed positive integer m, there exist positive constants $\eta \leq \tau$ and λ such that the following property holds:*

If $u_k \in \mathcal{B}_\eta(u^*)$, $\alpha_k \in [\alpha_{\inf}, \alpha_{\sup}]$, and if for some $\ell \in \{0, \ldots, m\}$, the following condition holds

$$\|\hat{u}_j^k - u^*\| \geq \frac{1}{2}\|u_k - u^*\| \quad \textit{for all } j \in \{0, \ldots, \max\{0, \ell-1\}\}, \tag{78}$$

then we have

$$u_{k+j} \in \mathcal{B}_\tau(u^*) \quad \textit{and} \quad \|u_{k+j} - \hat{u}_j^k\| \leq \lambda\|u_k - u^*\|^2 \tag{79}$$

for all $j \in \{0, \ldots, \ell\}$.

Proof The proof is given in Appendix A.1. □

In the next theorem, we present the main result of this section which is the local R-linearly convergence of Algorithm 1 applied to twice continuously Fréchet differentiable objective functions.

Theorem 3 *Let u^* be a local minimizer of a twice continuously Fréchet differentiable function $\mathcal{F}$, with a locally Lipschitz continuous second-derivative. Further suppose that the bilinear mapping $\mathcal{F}''(u^*)$ satisfies estimate* (73) *for constants $\delta_{\sup}$ and $\delta_{\inf}$. Then there exist positive constants ζ, λ_1, λ_2, and $\theta < 1$ such that the sequence $\{u_k\}_k$, generated by Algorithm 1, satisfies*

$$\|u_k - u^*\| \leq \lambda_1 \theta^k \|u_1 - u^*\| \quad \text{for all } k \geq 1, \tag{80}$$

and

$$\|\mathcal{G}_k\| \leq \lambda_2 \theta^k \|\mathcal{G}_1\| \quad \text{for all } k \geq 1, \tag{81}$$

for all initial iterates $u_0, u_1 \in \mathcal{B}_\zeta(u^*) \in \mathcal{H}$ with $u_0 \neq u_1$.

Proof The assumptions on $\mathcal{F}$ imply that $L1$ and $L2$ are satisfied for a radius τ and constants $\alpha_{\inf}$ and $\alpha_{\sup}$. The proof relies on Lemma 4 and Lemma 8 in an essential manner. By Lemma 4, which we use for the sequences $\{\hat{u}_j^k\}_j$, for every initial iterate $\hat{u}_0^k := u_k$ and initial step-size $\hat{\alpha}_0^k := \alpha_k$ with

$$\alpha_{\inf} \leq \alpha_k = \hat{\alpha}_0^k \leq \alpha_{\sup}, \tag{82}$$

we have

$$\|\hat{u}_m^k - u^*\| \leq \frac{1}{2}\|u_k - u^*\|. \tag{83}$$

Given the constants $\eta \leq \tau$ and λ from Lemma 8 we define $\zeta := \min\{\eta, \tau_1\}$, where τ_1 is chosen such that $c_2 := \frac{1}{2} + \lambda\tau_1 < 1$. Then, due to Lemma 8, for the fixed integer m, if $u_k \in \mathcal{B}_\zeta(u^*)$, if α_k satisfies (82), and if

$$\|\hat{u}_j^k - u^*\| \geq \frac{1}{2}\|u_k - u^*\| \quad \text{for all } j \in \{0, \ldots, \max\{0, \ell - 1\}\} \text{ with } \ell \leq m, \tag{84}$$

then we have

$$u_{k+j} \in \mathcal{B}_\tau(u^*) \text{ and } \|u_{k+j} - \hat{u}_j^k\| \leq \lambda\|u_k - u^*\|^2 \quad \text{for all } j \in \{1, \cdots, \ell\}. \tag{85}$$

Next we show by induction that there exists a subsequence of indices $\{k_i\}_i$ with $k_1 = 1$, for which we have

$$k_{i+1} - k_i \leq m \quad \text{and} \quad \|u_{k_{i+1}} - u^*\| \leq c_2\|u_{k_i} - u^*\|, \tag{86}$$

for all $i = 1, 2, \ldots$.

For any $u_0, u_1 \in \mathcal{B}_\zeta(u^*) \subset \mathcal{B}_\tau(u^*)$ and $k_1 = 1$, due to $L2$ we obtain

$$\alpha_{\inf} \leq \alpha_1 = \hat{\alpha}_0^1 \leq \alpha_{\sup}.$$

Due to Lemma 4 and (83), there exists a smallest integer $j_1 \leq m$ such that

$$\|\hat{u}_{j_1}^{k_1} - u^*\| \leq \frac{1}{2}\|\hat{u}_0^{k_1} - u^*\| = \frac{1}{2}\|u_{k_1} - u^*\|. \tag{87}$$

Defining $k_2 := k_1 + j_1 > k_1$, and using (85) and (87), we have

$$\begin{aligned} \|u_{k_2} - u^*\| = \|u_{k_1+j_1} - u^*\| &\leq \|u_{k_1+j_1} - \hat{u}_{j_1}^{k_1}\| + \|\hat{u}_{j_1}^{k_1} - u^*\| \\ &\leq \lambda\|u_{k_1} - u^*\|^2 + \frac{1}{2}\|\hat{u}_0^{k_1} - u^*\| \\ &\leq \lambda\tau_1\|u_{k_1} - u^*\| + \frac{1}{2}\|u_{k_1} - u^*\| \leq c_2\|u_{k_1} - u^*\|, \end{aligned} \tag{88}$$

and hence (86) follows for $i = 1$. By (88) and the fact that $u_{k_1} = u_1 \in \mathcal{B}_\zeta(u^*)$, it follows that $u_{k_2} \in \mathcal{B}_\zeta(u^*) \subset \mathcal{B}_\tau(u^*)$. Together with the inclusion in (85) we obtain that $u_k \in \mathcal{B}_\tau(u^*)$ for all $k \in \{0, 1, \ldots, k_2\}$.

To carry out the induction step we assume that for an index k_i we have $u_{k_i} \in \mathcal{B}_\zeta(u^*)$ and, $u_k \in \mathcal{B}_\tau(u^*)$ for all $k \in \{0, 1, \ldots, k_i\}$. We will show that there exists an index $k_{i+1} > k_i$ with $k_{i+1} - k_i \leq m$ such that $u_{k_{i+1}} \in \mathcal{B}_\zeta(u^*)$, $u_k \in \mathcal{B}_\tau(u^*)$ for all $k \in \{0, 1, \ldots, k_{i+1}\}$, and (86) holds.

Since $u_{k_i}, u_{k_i-1} \in \mathcal{B}_\tau(u^*)$ we have $\alpha_{k_i} \in [\alpha_{\inf}, \alpha_{\sup}]$. Moreover, due to (83), there is an integer $j_i \leq m$ with the property that

$$\|\hat{u}_{j_i}^{k_i} - u^*\| \leq \frac{1}{2}\|\hat{u}_0^{k_i} - u^*\| = \frac{1}{2}\|u_{k_i} - u^*\|.$$

Due to (85), by defining $k_{i+1} = k_i + j_i > k_i$ and using the similar argument as in (88), we can show that (86) holds and, consequently, we have $u_{k_{i+1}} \in \mathcal{B}_\zeta(u^*)$, and $u_k \in \mathcal{B}_\tau(u^*)$ for all $k \in \{0, 1, \ldots, k_{i+1}\}$.

Now, due to (163), there is a positive constant c_1 such that

$$\|u_{k+j} - u^*\| \leq c_1 \|u_k - u^*\| \quad \text{for all } j \in \{1, \ldots, m\}, \tag{89}$$

where c_1 depends only on m and the constants $\alpha_{\sup}$ and $\alpha_{\inf}$ which have been defined in $L2$. Further, for every $k \geq 1$, there exists an integer $i \geq 1$ such that $k_i \leq k < k_{i+1}$ with $k \leq k_i + m - 1$ and $k_i \leq m(i-1)+1$. Therefore, $i \geq \frac{k}{m}$ and also by (89), we obtain

$$\|u_k - u^*\| \leq c_1 \|u_{k_i} - u^*\| \leq c_1 (c_2)^{i-1} \|u_{k_1} - u^*\| \leq c_1 (c_2)^{\frac{k}{m}-1} \|u_{k_1} - u^*\|.$$

By setting $\theta := (c_2)^{\frac{1}{m}} < 1$, and $\lambda_1 := \frac{c_1}{c_2}$, we can conclude (80).

We turn to verification of (81). By using the fact that for every $k \in \mathbb{N}$ the sequence $\{u_k\}_k$ lies in $\mathcal{B}_\tau(u^*)$, the property (77), and (80), we obtain

$$\|\mathcal{G}_k\| \leq \alpha_{\sup} \|u_k - u^*\| \leq \alpha_{\sup} \lambda_1 \theta^k \|u_1 - u^*\| \leq \frac{\alpha_{\sup} \lambda_1}{\alpha_{\inf}} \theta^k \|\mathcal{G}_1\|.$$

By setting $\lambda_2 := \frac{\alpha_{\sup} \lambda_1}{\alpha_{\inf}}$ we complete the proof. □

3 Mesh Independence Principle

In this section, we investigate finite-dimensional approximations of Algorithm 1. More specifically we investigate the dependence of the iteration count of the algorithm to achieve a desired accuracy of the residue under finite-dimensional approximations. We note that our objective here is not to estimate the error between the solutions of the discretized problem and continuous one.

Thus let $\{\mathcal{H}^h\}_h$ be a family of finite-dimensional Hilbert spaces indexed by some real number $h > 0$, and endowed with inner products and their associated norms denoted by $(\cdot, \cdot)_h$ and $\|\cdot\|_h$, respectively. Let $\mathcal{G}^h : \mathcal{H}^h \to \mathcal{H}^h$ denote continuous nonlinear mappings which will be required to approximate $\mathcal{G}$ in a sense to be made precise in Assumption A2 below. We then consider the family of problems:

$$\text{Find } u^{*h} \in \mathcal{H}^h \text{ such that} \quad \mathcal{G}^h(u^{*h}) = 0. \tag{EP^h}$$

Throughout this section we pose the following assumption:

A0: The assumptions of Theorem 3 in Section 2.2 hold and we denote by $\{u_k\}_k$ the sequence generated by Algorithm 1 which enjoys the properties asserted in Theorem 3.

In particular, it is assumed that $\|u_0 - u^*\|$ and $\|u_1 - u^*\|$ are sufficiently small ($< \zeta$ with ζ defined in Theorem 3) unless $\mathcal{F}$ is a strictly convex quadratic function. For the case of strictly convex quadratic functions, u_0 and u_1 can be chosen from the whole of $\mathcal{H}$.

To describe the family of approximating sequences we choose $u_0^h, u_1^h \in \mathcal{H}^h$ and update u_k^h, for $k = 1, \ldots$ by

$$u_{k+1}^h = u_k^h - \frac{1}{\alpha_k^h} \mathcal{G}_k^h, \tag{90}$$

where $\mathcal{G}_k^h := \mathcal{G}^h(u_k^h)$ and the step-size α_k^h is chosen according to either

$$\alpha_k^{BB1,h} := \frac{(\mathcal{S}_{k-1}^h, \mathcal{Y}_{k-1}^h)_h}{(\mathcal{S}_{k-1}^h, \mathcal{S}_{k-1}^h)_h}, \text{ or } \alpha_k^{BB2,h} := \frac{(\mathcal{Y}_{k-1}^h, \mathcal{Y}_{k-1}^h)_h}{(\mathcal{S}_{k-1}^h, \mathcal{Y}_{k-1}^h)_h}. \tag{91}$$

Here we have set $\mathcal{S}_{k-1}^h := u_k^h - u_{k-1}^h$ and $\mathcal{Y}_{k-1}^h := \mathcal{G}_k^h - \mathcal{G}_{k-1}^h$. We should point out that the inner product on $\mathcal{H}^h$ will typically reflect the norm on $\mathcal{H}$. It should not be thought of as the canonical inner-product in $\mathbb{R}^{N(h)}$.

Let us now formulate some additional notation and assumptions that we require for the main result of this section. Suppose that $\{\mathbb{P}^h\}_h$ is a family of linear 'prolongation' operators

$$\mathbb{P}^h : \mathcal{H}^h \to \mathcal{H}.$$

We use the following notion of convergence in the space $\mathcal{H}$. A sequence $u^h \in \mathcal{H}^h$ is $\mathcal{H}$-convergent to $u \in \mathcal{H}$ if

$$\lim_{h\downarrow 0} \|\mathbb{P}^h u^h - u\| = 0.$$

We have to assume that the discrete inner products approximate the original one in the following sense:

A1: If $u^h \overset{\mathcal{H}}{\to} u$ and $z^h \overset{\mathcal{H}}{\to} z$ for $u, z \in \mathcal{H}$, then

$$\lim_{h\downarrow 0} (u^h, z^h)_h = (u, z). \tag{92}$$

Moreover we need the following approximation property of $\mathcal{G}$ by the family $\mathcal{G}^h$.

A2: Suppose that $\mathcal{G}(u^*) = 0$. Then, if $u^h \overset{\mathcal{H}}{\to} u$ with u in a neighborhood of u^*, then

$$\mathcal{G}^h(u^h) \overset{\mathcal{H}}{\to} \mathcal{G}(u). \tag{93}$$

Remark 5 In applications it can occur that the convergence specified in (93) requires additional regularity of u and $\mathcal{G}(u)$. In this case one assumes the existence of a subspace $\mathcal{W}$ in $\mathcal{H}$ of more regular functions, and one needs to assure that the limit of the iterations remains in $\mathcal{W}$. In this case Assumption A2 is replaced by A2' below. For details we refer to [49], for instance.

A2': There is $u^* \in \mathcal{W}$ with $\mathcal{G}(u^*) = 0$, such that $\mathcal{G}$ is well-defined for all $u \in \mathcal{W}$ sufficiently near u^* with respect to the $\mathcal{H}$-norm. Moreover, if $u \in \mathcal{W}$ with $\|u - u^*\|$ sufficiently small and $u^h \overset{\mathcal{H}}{\to} u$, then $\mathcal{G}(u) \in \mathcal{W}$ and

$$\mathcal{G}^h(u^h) \overset{\mathcal{H}}{\to} \mathcal{G}(u).$$

Theorem 4 *Suppose that Assumptions A0-A2 hold. Moreover, let $u_i^h \overset{\mathcal{H}}{\to} u_i$ for $i = 0, 1$ with $u_1 \neq u_0$ and $u_1^h \neq u_0^h$. Then for any $k' \geq 1$, we have*

$$\lim_{h\downarrow 0} \max_{1\leq k\leq k'} \|\mathbb{P}^h u_k^h - u_k\| = 0. \tag{94}$$

Proof Using (90) and the triangle inequality we obtain

$$\|\mathbb{P}^h u_{k+1}^h - u_{k+1}\| \leq \|\mathbb{P}^h u_k^h - u_k\| + \left|\frac{1}{\alpha_k^h} - \frac{1}{\alpha_k}\right| \|\mathbb{P}^h \mathcal{G}_k^h\| + \left|\frac{1}{\alpha_k}\right| \|\mathbb{P}^h \mathcal{G}_k^h - \mathcal{G}_k\|, \tag{95}$$

for every $k \geq 1$. Then, proceeding by induction, using (92) and (93), and passing the limit in (91) and (95), it can be shown that (94) is true for every $k' \geq 1$. □

The termination condition for EP^h is based on the norm of the gradients for the approximated and the original problem. Thus for $\epsilon > 0$ the iteration is terminated according to

$$\|\mathcal{G}_k^h\|_h < \epsilon, \quad \text{and} \quad \|\mathcal{G}_k\| < \epsilon, \tag{96}$$

where ϵ is a sufficiently small positive number. In order to investigate the behaviour of convergence of the approximated problem with respect to the original problem, we consider the following quantities:

$$k^*(\epsilon) := \min\{k \in \mathbb{N} : \|\mathcal{G}_k\| < \epsilon\}, \qquad k_h^*(\epsilon) := \min\{k \in \mathbb{N} : \|\mathcal{G}_k^h\|_h < \epsilon\},$$

where $k^*(\epsilon)$ and $k_h^*(\epsilon)$ are the smallest iteration numbers for which the norm of corresponding gradients is less than ϵ. In the following we study the relation between $k^*(\epsilon)$ and $k_h^*(\epsilon)$.

Theorem 5 *Suppose that Assumptions A0-A2 hold. Further, let $u_i^h \overset{\mathcal{H}}{\to} u_i$ for $i = 0, 1$ with $u_1 \neq u_0$ and $u_1^h \neq u_0^h$. Then for any given numbers $\epsilon > 0$ and $\delta > 0$, there exists a number $h_{\delta,\epsilon} > 0$ such that*

$$k^*(\epsilon + \delta) \leq k_h^*(\epsilon) \leq k^*(\epsilon) \tag{97}$$

for every $h \in (0, h_{\delta,\epsilon}]$.

Proof Due to (94) and A2, we have for every k that

$$\mathcal{G}_k^h \xrightarrow{\mathcal{H}} \mathcal{G}_k \tag{98}$$

and by A1, we obtain

$$\lim_{h\downarrow 0} \|\mathcal{G}_k^h\|_h = \|\mathcal{G}_k\|. \tag{99}$$

Now, we show that $\|\mathcal{G}_k^h\|_h < \epsilon$ for a sufficiently small $h > 0$, provided that $\|\mathcal{G}_k\| < \epsilon$ holds for an iterate k. Since $\|\mathcal{G}_{k^*(\epsilon)}\| < \epsilon$, there exists a positive number $\zeta := \zeta(\epsilon)$ such that $\|\mathcal{G}_{k^*(\epsilon)}\| + \zeta < \epsilon$. Moreover, due to (99), there exists a positive number $h_\epsilon > 0$ such that for every $h \in (0, h_\epsilon]$ we have

$$\Big| \|\mathcal{G}_{k^*(\epsilon)}^h\|_h - \|\mathcal{G}_{k^*(\epsilon)}\| \Big| \le \zeta. \tag{100}$$

Hence, for every $h \in (0, h_\epsilon]$, we obtain

$$\|\mathcal{G}_{k^*(\epsilon)}^h\|_h = \|\mathcal{G}_{k^*(\epsilon)}\| + \|\mathcal{G}_{k^*(\epsilon)}^h\|_h - \|\mathcal{G}_{k^*(\epsilon)}\| \le \|\mathcal{G}_{k^*(\epsilon)}\| + \zeta < \epsilon,$$

and, thus, we have

$$k_h^*(\epsilon) \le k^*(\epsilon) \quad \text{for every } h \in (0, h_\epsilon],$$

which implies the second inequality in (97). Now assume that $\delta > 0$ be given. Then due to (99) we have

$$\lim_{h\downarrow 0} \max_{1\le k<k^*(\delta+\epsilon)} \Big| \|\mathcal{G}_k^h\|_h - \|\mathcal{G}_k\| \Big| = 0. \tag{101}$$

By the definition of $k^*(\delta+\epsilon)$, we have

$$\|\mathcal{G}_k\| \ge \delta + \epsilon \quad \text{for all } k < k^*(\delta+\epsilon). \tag{102}$$

Moreover due to (101), there exists a positive number h_δ such that

$$\Big| \|\mathcal{G}_k^h\|_h - \|\mathcal{G}_k\| \Big| \le \max_{1\le k'<k^*(\delta+\epsilon)} \Big| \|\mathcal{G}_{k'}^h\|_h - \|\mathcal{G}_{k'}\| \Big| \le \delta \quad \text{for all } h \in (0, h_\delta] \text{ and } k < k^*(\delta+\epsilon). \tag{103}$$

Using (102) and (103) we infer for every $h \in (0, h_\delta]$ and $k < k^*(\delta+\epsilon)$ that

$$\|\mathcal{G}_k^h\|_h \ge \|\mathcal{G}_k\| - \delta \ge \delta + \epsilon - \delta = \epsilon,$$

and, thus, $k_h^*(\epsilon) \ge k^*(\delta+\epsilon)$ for every $h \in (0, h_\delta]$. Now for the choice of $h_{\delta,\epsilon} := \min\{h_\delta, h_\epsilon\}$, the relation (97) holds for every $h \in (0, h_{\delta,\epsilon}]$ and we are finished with the proof.

□

Theorem 6 *Suppose that Assumptions A0-A2 hold. Further assume that $u_i^h \xrightarrow{\mathcal{H}} u_i$ for $i = 0, 1$ with $u_1 \ne u_0$ and $u_1^h \ne u_0^h$. Then for each $\epsilon > 0$ there exists $h_\epsilon > 0$ such that*

$$k^*(\epsilon) - \ell \le k_h^*(\epsilon) \le k^*(\epsilon) \quad \textit{for every } h \in (0, h_\epsilon],$$

where the integer $\ell > 0$ is independent of h and ϵ.

Proof Theorem 3 implies R-linear convergence of $u_k \to u^*$. It can be shown as in the proof of Theorem 3 that there exist a positive integer m, positive numbers $c_2 < 1$ and $\zeta \le \tau$, and a subsequence of indices $\{k_i\}_i \in \mathbb{N}$ with $k_1 = 1$, for which we have

$$u_k \in \mathcal{B}_\zeta(u^*) \quad \text{for every } k \ge k_1, \tag{104}$$

and

$$k_{i+1} - k_i \le m \quad \text{and} \quad \|u_{k_{i+1}} - u^*\| \le c_2 \|u_{k_i} - u^*\|, \text{ for all } i \ge 1. \tag{105}$$

Moreover, as mentioned in the proof of Theorem 3, there exists a number $c_1 > 0$ such that

$$\|u_{k+j} - u^*\| \le c_1 \|u_k - u^*\| \quad \text{for all } j \in \{1, \dots, m\} \text{ and any } k \ge k_1. \tag{106}$$

Let us first denote the integer q^* as the smallest integer for which $c_2^{q^*} < \frac{\alpha_{\inf}}{c_1\alpha_{\sup}}$ holds. The existence of such q^* is guaranteed since $c_2 < 1$. Next, we show for every $k \geq k_1$ that there exists a positive integer $i^+(k) \leq m(q^*+1)-1 =: \ell$ such that

$$\|\mathcal{G}_{k+i^+(k)}\| < \|\mathcal{G}_k\|. \tag{107}$$

For every $k \geq k_1$, the exists an index i such that $k_i \leq k < k_{i+1}$. Due to (77), (105), (106), and the definition of q^*, we obtain

$$\begin{aligned}\|\mathcal{G}_{k_{i+q^*+1}}\| &\leq \alpha_{\sup}\|u_{k_{i+q^*+1}} - u^*\| \leq \alpha_{\sup}c_2^{q^*}\|u_{k_{i+1}} - u^*\| \leq \alpha_{\sup}c_1c_2^{q^*}\|u_k - u^*\| \\ &\leq \frac{\alpha_{\sup}c_1c_2^{q^*}}{\alpha_{\inf}}\|\mathcal{G}_k\| < \|\mathcal{G}_k\|.\end{aligned}$$

By setting $i^+(k) := k_{i+q^*+1} - k$, we have $i^+(k) \leq \ell$ and we are finished with the verification of (107).

Now, due to the definition of $k^*(\epsilon)$, we have $\|\mathcal{G}_k\| \geq \epsilon$ for every $k < k^*(\epsilon)$. We will next show that for every $k^*(\epsilon) \geq \ell$ that

$$\|\mathcal{G}_k\| > \epsilon \quad \text{for every } k < k^*(\epsilon) - \ell. \tag{108}$$

Suppose on contrary that there exists an index $\bar{k} < k^*(\epsilon) - \ell$ with $\|\mathcal{G}_{\bar{k}}\| = \epsilon$. Then due to (107) there exists an integer $i^+(\bar{k}) \leq \ell$ such that we have $\|\mathcal{G}_{\bar{k}+i^+(\bar{k})}\| < \|\mathcal{G}_{\bar{k}}\| = \epsilon$ with $\bar{k} + i^+(\bar{k}) \leq \bar{k} + \ell < k^*(\epsilon)$, and this contradicts the definition of $k^*(\epsilon)$. Hence, (108) holds.

Due to (108), for $k < k^*(\epsilon) - \ell$ there exist strictly positive numbers $\{\delta_k\}_k$ such that $\|\mathcal{G}_k\| = \epsilon + \delta_k$ for $k < k^*(\epsilon) - \ell$. By setting $0 < \delta := \min\{\delta_k : k < k^*(\epsilon) - \ell\}$, we obtain

$$\|\mathcal{G}_k\| \geq \epsilon + \delta \quad \text{for every } k < k^*(\epsilon) - \ell.$$

Therefore we conclude that $k^*(\epsilon) - \ell \leq k^*(\epsilon + \delta)$. Due to Theorem 5, for $\epsilon > 0$ and $\delta > 0$, there exists a number $h_\epsilon > 0$ such that we have

$$k^*(\epsilon) - \ell \leq k^*(\epsilon + \delta) \leq k_h^*(\epsilon) \leq k^*(\epsilon) \quad \text{for every } h \in (0, h_\epsilon]. \tag{109}$$

This concludes the proof. □

Remark 6 In the case of quadratic functions (QP), due to Lemma 4, inequality (109) holds for $\ell = m$ and all initial iterates $u_0, u_1 \in \mathcal{H}$ with $u_0 \neq u_1$. In particular, if also $\delta_{\sup} < 2\delta_{\inf}$, then (109) holds for $\ell = 1$.

Remark 7 In general, the sequence $\{\|\mathcal{G}_k\|\}_k$ corresponding to Algorithm 1 is not monotonically decreasing. This is the reason why we have to introduce ℓ in Theorem 6 which can possibly be larger than 1. In the case that $\delta_{\sup} < 2\delta_{\inf}$, $\{\|\mathcal{G}_k\|\}_k$ is monotone decreasing and as a consequence $\ell = 1$.

4 Application to Optimal Control Problems with PDEs

In this section, we will apply Algorithm 1 to optimal control problems which are governed by three types of partial differential equations, including an elliptic, a hyperbolic, and a parabolic problem. We introduce these problems in reminder of this section. For the sake of brevity, finite-dimensional approximation is only discussed for the elliptic case.

4.1 Dirichlet Optimal Control for the Poisson Equation

4.1.1 Continuous Problem

In this subsection, we consider the following elliptic Dirichlet boundary control problem

$$\min_{u \in L^2(\Gamma)} J(u,y) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Gamma)}, \tag{110}$$

$$\text{subject to } \begin{cases} -\Delta y = f & \text{in } \Omega, \\ y = u & \text{on } \Gamma, \end{cases} \tag{111}$$

on an open convex bounded polygonal set $\Omega \subset \mathbb{R}^2$ with boundary denoted by $\Gamma := \partial\Omega$. We assume that $f, y_d \in L^2(\Omega)$ and $\beta > 0$. Then, for a given $(u, f) \in L^2(\Gamma) \times L^2(\Omega)$, the solution $y(u, f) \in L^2(Q)$ of (111) exists in a very weak sense and it satisfies the following variational equation

$$(y, -\Delta\varphi)_{L^2(\Omega)} + (u, \partial_\nu\varphi)_{L^2(\Gamma)} = (\varphi, f)_{L^2(\Omega)} \quad \text{for all } \varphi \in H^2(\Omega) \cup H_0^1(\Omega).$$

The corresponding solution operator defined by $(u, f) \mapsto y(u, f)$ is a continuous operator from $L^2(\Gamma) \times L^2(\Omega)$ to $L^2(\Omega)$. See e.g., [35,36]. Moreover, the linear operators $\mathcal{L} : L^2(\Gamma) \to L^2(\Omega)$ defined by $u \mapsto y(u, 0)$, and $\Pi : L^2(\Omega) \to L^2(\Omega)$ defined by $f \mapsto y(f, 0)$ are continuous. Then by defining $\mathcal{X} := L^2(\Omega)$, $\mathcal{H} := L^2(\Gamma)$ and $\psi := -\Pi f + y_d$, we can express the optimal control problem (110)-(111) as the following linear least squares problem

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

By a short computation, it can be shown that the problem (110)-(111) can be written in the form of (QP), where $\mathcal{A} := \mathcal{L}^*\mathcal{L} + \beta\mathcal{I}$ with $\mathcal{L}^* : \mathcal{X} \to \mathcal{H}$ defined as the adjoint operator of $\mathcal{L}$, and $b := \mathcal{L}^*\psi$. Clearly, the operator $\mathcal{A}$ is uniformly positive, bounded, and self-adjoint on the Hilbert space $\mathcal{H}$ and thus the existence and uniqueness of the solution to the problem (110)-(111) can be obtained due the fact that $\mathcal{A}$ has a bounded inverse.

Remark 8 According to [11][Theorem 4.2], for each pair $(f, u) \in L^2(\Omega) \times L^2(\Gamma)$, the solution $y(f, u)$ to (111) belongs to the space $H^{\frac{1}{2}}(\Omega)$, which is continuously and compactly embedded to $L^2(\Omega)$. Therefore the linear operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}$ is compact, and we conclude that $\mathcal{A}$ has the form (poco).

For every $u \in \mathcal{H}$, the derivative of $\mathcal{F}$ at u in direction $\delta u \in \mathcal{H}$ can be expressed by

$$\mathcal{F}'(u)\delta u = (\mathcal{L}^*(\mathcal{L}u - \psi) + \beta u, \delta u), \tag{112}$$

and the gradient of $\mathcal{F}$ at u is identified by $\mathcal{G}(u) = \mathcal{L}^*(\mathcal{L}u - \psi) + \beta u$. Alternatively, if we consider the solution $p(u) \in H^2(\Omega) \cap H_0^1(\Omega)$ of the adjoint equation

$$\begin{cases} -\Delta p = y(u, f) - y_d & \text{in } \Omega, \\ p = 0 & \text{on } \Gamma, \end{cases} \tag{113}$$

where $y(u, f) \in L^2(\Omega)$ is the solution of (111), then the directional derivative (112) and the corresponding gradient $\mathcal{G}$ at point u can be rewritten as

$$\mathcal{F}'(u)\delta u = (\partial_\nu p(u) + \beta u, \delta u) \text{ for all } \delta u \in \mathcal{H}, \text{ and } \quad \mathcal{G}(u) = \partial_\nu p(u) + \beta u \text{ in } \mathcal{H}. \tag{114}$$

For the global minimizer $u^* \in \mathcal{H}$ to (LS), the first-order optimality condition can be expressed as

$$(\mathcal{L}^*\mathcal{L} + \beta\mathcal{I})u^* = \mathcal{L}^*\psi, \tag{115}$$

which can be rewritten, equivalently, as the following systems of equations

$$\begin{cases} y^* = y(u^*, f) & \text{in } L^2(\Omega), \\ \partial_\nu p^* = -\beta u^* & \text{in } L^2(\Gamma), \\ -\Delta p^* = y^* - y_d & \text{in } L^2(\Omega), \text{with } p^* = 0 \text{ on } \Gamma. \end{cases}$$

4.1.2 Discretized Problem

For the discretization of (110)-(111), we use finite elements. Let us consider the regular family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\overline{\Omega}$ with $\overline{\Omega} = \cup_{T\in\mathcal{T}_h} T$ and the mesh-size defined by $h := \max\{\text{diam}(T) : T \in \mathcal{T}_h\}$. Let $\{x_j\}_{1\le j\le N(h)}$ be the nodes which lies on the boundary with the numbering which starts at the origin in the counterclockwise and $x_{N(h)+1} = x_1$. Then we define the space of discretized control by

$$\mathcal{H}^h := \{u^h \in C(\Gamma) : u^h|_{[x_j, x_{j+1}]} \in \mathcal{P}^1 \text{ for } j = 1, \ldots, N(h)\},$$

and, we consider the space $V^h \subset H^1(\Omega)$ defined by

$$V^h := \{y^h \in C(\bar{\Omega}) : y^h|_T \in \mathcal{P}^1 \text{ for every } T \in \mathcal{T}_h\},$$

where $\mathcal{P}^1$ is the space of polynomials of degree less than or equal to 1. Further we set $V_0^h := V^h \cap H_0^1(\Omega)$. The space $\mathcal{H}^h$ is formed by the restriction of the functions of V^h to $\partial\Omega$. Clearly, we have $\mathcal{H}^h \subset \mathcal{H}$ and, as a result, the finite-dimensional space $\mathcal{H}^h$ is endowed with the inner product and the norm introduced by the space $\mathcal{H} = L^2(\Gamma)$. Then, naturally, the prolongation operator $\mathbb{P}^h : \mathcal{H}^h \to \mathcal{H}$ is defined to be the canonical injection operator i.e., $\mathbb{P}^h(u^h) = u^h$ for every $u^h \in \mathcal{H}^h$. Let us consider the orthogonal projection operator $\Pi^h : \mathcal{H} \to \mathcal{H}^h$ defined by

$$(\Pi^h v, u^h)_{\mathcal{H}} = (v, u^h)_{\mathcal{H}} \text{ for all } u^h \in \mathcal{H}^h.$$

It satisfies the following estimate

$$\|u - \Pi^h u\|_{\mathcal{H}} \leq c h^{\frac{1}{2}} \|u\|_{H^{\frac{1}{2}}(\Gamma)}, \tag{116}$$

for every $u \in H^{\frac{1}{2}}(\Gamma)$, see, e.g., [11,16]. For every $u \in \mathcal{H}$ we consider the unique discrete solution $y^h(u) \in V^h$ satisfying

$$\begin{cases} (\nabla y^h, \nabla \phi^h) = (f, \phi^h) \text{ for all } \phi^h \in V^h, \\ y^h|_\Gamma = \Pi^h u. \end{cases} \tag{117}$$

Then we can define the discrete objective function in $\mathcal{H}$ by

$$J^h(u, y^h(u)) := \frac{1}{2}\|y^h(u) - y_d\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Gamma)}. \tag{118}$$

The finite-dimensional approximation of (110)-(111) can be expressed as

$$\min_{u^h \in \mathcal{H}^h} \mathcal{F}^h(u^h) = \min_{u^h \in \mathcal{H}^h} J^h(u^h, y^h(u^h)). \tag{119}$$

Existence of a solution to (119) follows by similar arguments as for the continuous problem. Given $u \in \mathcal{H}$ we consider the adjoint state $p^h(u) \in V_0^h$ as the solution of

$$(\nabla p^h(u), \nabla \psi^h)_{L^2(\Omega)} = (y^h(u) - y_d, \psi^h)_{L^2(\Omega)} \text{ for all } \psi^h \in V_0^h. \tag{120}$$

In order to compute the gradient of $\mathcal{F}^h$, analogously to the expression (114), we need to characterise a discrete normal derivative $\partial_\nu^h p^h(u)$. For every $u \in \mathcal{H}$, similarly to [16][Proposition 4.2], $\partial_\nu^h p^h(u) \in \mathcal{H}^h$ is characterized as the unique solution of the following variational problem

$$(\partial_\nu^h p^h(u), \varphi^h)_{\mathcal{H}} = (\nabla p^h(u), \nabla \varphi^h)_{L^2(\Omega)} - (y^h(u) - y_d, \varphi^h)_{L^2(\Omega)} \text{ for all } \varphi^h \in V^h,$$

where $p^h(u) \in V_0^h$ is the solution of (120). Next, we prove the following useful estimate.

Lemma 9 *There exists a constant c depending on f and y_d, and independent of h such that*

$$\|\partial_\nu p(u) - \partial_\nu^h p^h(v)\|_{\mathcal{H}} \leq c\left(\|u - v\|_{\mathcal{H}} + h^{\frac{1}{2}}(1 + \|v\|_{\mathcal{H}})\right) \text{ for all } u, v \in \mathcal{H}. \tag{121}$$

Proof This proof is based on the results from [16], where $u \in L^\infty(\Omega)$ was used in the context of semilinear elliptic equation. First, using a similar argument as in [11,16], one can show that

$$\|y(u) - y^h(u)\|_{\mathcal{H}} \leq c(1 + \|u\|_{\mathcal{H}})^2 h^{\frac{1}{2}}, \tag{122}$$

where the constant c depends on f. From (122), it follows that

$$\|y(u) - y^h(v)\|_{\mathcal{H}} \leq c\left(\|u - v\|_{\mathcal{H}} + h^{\frac{1}{2}}(1 + \|u\|_{\mathcal{H}})\right) \text{ for all } u, v \in \mathcal{H},$$

Next, we show that

$$\|\partial_\nu p(u) - \partial_\nu^h p^h(u)\|_{\mathcal{H}} \leq c h^{\frac{1}{2}}(1 + \|u\|_{\mathcal{H}}) \text{ for all } u \in \mathcal{H}. \tag{123}$$

Recall that $p(u) \in H^2(\Omega) \cap H_0^1(\Omega)$ and therefore $\partial_\nu p(u) \in H^{\frac{1}{2}}(\Gamma)$. For the left hand-side of (123) we obtain

$$\|\partial_\nu p(u) - \partial_\nu^h p^h(u)\|^2_{\mathcal{H}} = \int_\Gamma \left|\partial_\nu p(u) - \Pi^h \partial_\nu p(u)\right|^2 d\mathcal{S} + \int_\Gamma \left|\Pi^h \partial_\nu p(u) - \partial_\nu^h p^h(u)\right|^2 d\mathcal{S} =: I_1 + I_2. \tag{124}$$

The last term can be equivalently be expressed as

$$I_2 = \int_\Gamma (\partial_\nu p(u) - \partial_\nu^h p^h(u))(\Pi^h \partial_\nu p(u) - \partial_\nu^h p^h(u)) d\mathcal{S}. \tag{125}$$

Let $w^h \in V^h$ be the solution of the following variational equation

$$\begin{cases} (\nabla w^h, \nabla \phi^h) = 0 \text{ for all } \phi^h \in V^h, \\ w^h|_\Gamma = \Pi^h \partial_\nu p(u) - \partial_\nu^h p^h(u). \end{cases} \tag{126}$$

Then, by referring to [14][Lemma 3.2], we have the following estimate for (126)

$$\|w^h\|_{H^1(\Omega)} \le c\|\Pi^h \partial_\nu p(u) - \partial_\nu^h p^h(u)\|_{H^{\frac{1}{2}}(\Gamma)}, \tag{127}$$

with a constant c independent of h. Using the definition of $\partial_\nu^h p^h(u)$ and Green formula for $\partial_\nu p(u)$, we obtain

$$(\partial_\nu p(u) - \partial_\nu^h p^h(u), \phi^h)_{\mathcal{H}} = (\nabla(p(u) - p^h(u)), \nabla \phi^h)_{L^2(\Omega)} + (y^h(u) - y(u), \phi^h)_{L^2(\Omega)} \tag{128}$$

for every $\phi^h \in V^h$. Using (125), (126), and (128), we find

$$I_2 = (\nabla(p(u) - p^h(u)), \nabla w^h)_{L^2(\Omega)} + (y^h(u) - y(u), w^h)_{L^2(\Omega)}.$$

Moreover, we have

$$(\nabla p^h(u), \nabla w^h)_{L^2(\Omega)} = (\nabla \mathcal{I}_h p(u), \nabla w^h)_{L^2(\Omega)} = 0, \tag{129}$$

where $\mathcal{I}_h \in \mathcal{L}(C(\overline{\Omega}), V_0^h)$ stands for the classical interpolation operator, see e.g., [15]. Due to (129) and the definition of w^h from (126), we obtain

$$I_2 = (\nabla(p(u) - \mathcal{I}_h p(u)), \nabla w^h)_{L^2(\Omega)} + (y^h(u) - y(u), w^h)_{L^2(\Omega)}. \tag{130}$$

Using (127), the interpolation estimate, and the following inverse estimate (see e.g., [11])

$$\|u^h\|_{H^{\frac{1}{2}}(\Gamma)} \le Ch^{-\frac{1}{2}} \|u^h\|_{\mathcal{H}} \text{ for all } u^h \in \mathcal{H}^h,$$

we infer that

$$\begin{aligned} |(\nabla(p(u) - \mathcal{I}_h p(u)), \nabla w^h)_{L^2(\Omega)}| &\le \|\nabla(p(u) - \mathcal{I}_h p(u))\|_{L^2(\Omega)} \|w^h\|_{H^1(\Omega)} \\ &\le ch\|p(u)\|_{H^2(\Omega)} \|w^h|_\Gamma\|_{H^{\frac{1}{2}}(\Gamma)} \\ &\le ch^{\frac{1}{2}}(1 + \|u\|_{\mathcal{H}}) \|w^h|_\Gamma\|_{\mathcal{H}} \\ &\le ch^{\frac{1}{2}}(1 + \|u\|_{\mathcal{H}}) \sqrt{I_2}, \end{aligned} \tag{131}$$

where the constant c from the third line of (131) depends also on y_d. Moreover, due to (122), we can write

$$|(y^h(u) - y(u), w^h)_{L^2(\Omega)}| \le \|y^h(u) - y(u)\|_{L^2(\Omega)} \|w^h\|_{L^2(\Omega)} \le ch^{\frac{1}{2}}(1 + \|u\|_{\mathcal{H}}) \sqrt{I_2}. \tag{132}$$

From (130), (131), and (132), it follows that

$$I_2 \le ch(1 + \|u\|_{\mathcal{H}})^2. \tag{133}$$

Further, using (116) we obtain

$$I_1 \le ch\|\partial_\nu p(u)\|^2_{H^{\frac{1}{2}}(\partial\Omega)} \le ch\|p(u)\|^2_{H^2(\Omega)} \le ch(1 + \|u\|_{\mathcal{H}})^2. \tag{134}$$

Now, from (124), (133), and (134), we conclude (123). Finally, using (123) we can write that

$$\begin{aligned} \|\partial_\nu p(u) - \partial_\nu^h p^h(v)\|_{\mathcal{H}} &\le \|\partial_\nu p(u) - \partial_\nu p(v)\|_{\mathcal{H}} + \|\partial_\nu p(v) - \partial_\nu^h p^h(v)\|_{\mathcal{H}} \\ &\le c\|p(u) - p(v)\|_{H^2(\Omega)} + ch^{\frac{1}{2}}(1 + \|v\|_{\mathcal{H}}) \\ &\le c\left(\|u - v\|_{\mathcal{H}} + h^{\frac{1}{2}}(1 + \|v\|_{\mathcal{H}})\right), \end{aligned}$$

for every $u, v \in \mathcal{H}$ and we are finished with the verification of (121). □

Now we are in the position in which we can verify the assumptions A1-A2 of Section 3. A1 follows from the definition of $\mathcal{H}^h$ and $\mathbb{P}^h$. To verify A2, assume that $u^h \overset{\mathcal{H}}{\to} u$ with $u^h \in \mathcal{H}^h$. Similarly to (114), the directional derivative and its corresponding gradient of $\mathcal{F}^h$ of the discretized problem (119) at point u^h can be rewritten as

$$\mathcal{F}^{h\prime}(u^h)\delta u^h = (\partial_\nu^h p^h(u^h) + \beta u^h, \delta u^h) \text{ for all } \delta u^h \in \mathcal{H}^h, \text{ and } \quad \mathcal{G}^h(u^h) = \partial_\nu^h p^h(u^h) + \beta u^h \text{ in } \mathcal{H}^h. \quad (135)$$

Then by (114), (135), and (121), we obtain

$$\|\mathcal{G}(u) - \mathbb{P}^h\mathcal{G}^h(u^h)\|_{\mathcal{H}} \le \|\partial_\nu p(u) - \partial_\nu^h p^h(u^h)\|_{\mathcal{H}} + \beta\|u - u^h\|_{\mathcal{H}} \le (c+\beta)\|u - u^h\|_{\mathcal{H}} + ch^{\frac{1}{2}}(1 + \|u^h\|_{\mathcal{H}}). \quad (136)$$

Hence, $\mathcal{G}^h(u^h) \overset{\mathcal{H}}{\to} \mathcal{G}(u)$ follows by sending h to zero in (136).

Remark 9 Due the fact that $\partial_\nu p(u) \in H^{\frac{1}{2}}(\Gamma)$ for every $u \in \mathcal{H}$, using (114) and Step 4 in Algorithm 1, it is easy to see that for every $u_0, u_1 \in H^{\frac{1}{2}}(\Gamma)$, the sequence $\{u_k\}_k$ stays in the space $H^{\frac{1}{2}}(\Gamma)$. Moreover, for given $y_d, f \in L^{p^*}(\Omega)$ with $p^* > 2$, we have $y \in W^{1,\overline{p}}(\Omega)$ and $p(y, y_d) \in W^{2,\overline{p}}(\Omega)$ for $\overline{p} \in (2, p^*]$ depending on Ω, see e.g., [16][Theorem 3.4]. Hence, $\{u_k\}_k \subset W^{1-\frac{1}{\overline{p}},\overline{p}}(\Gamma) \subset C(\Gamma)$ provided that $u_0, u_1 \in W^{1-\frac{1}{\overline{p}},\overline{p}}(\Gamma)$. In this case u_0^h, u_1^h can be chosen as $\mathcal{I}_h^\Gamma u_0, \mathcal{I}_h^\Gamma u_1 \in \mathcal{H}^h$ where $\mathcal{I}_h^\Gamma \in \mathcal{L}(W^{1-\frac{1}{\overline{p}},\overline{p}}(\Gamma), \mathcal{H}^h)$ is the standard interpolation operator.

4.2 Neumann Optimal Control for the Linear Wave Equation

Let us consider the optimal control problem

$$\min_{u \in L^2(\Sigma_c)} J(u, y) := \frac{\alpha_1}{2}\|y - y_d\|^2_{L^2(Q)} + \frac{\alpha_2}{2}\|y(T) - z_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|u\|^2_{L^2(\Sigma_c)}, \quad (137)$$

$$\text{subject to } \begin{cases} y_{tt} - \Delta y = f & \text{in } Q, \\ \partial_\nu y = u & \text{on } \Sigma_c, \\ y = 0 & \text{on } \Sigma_0, \\ y(0) = y_0^1, \quad y_t(0) = y_0^2 & \text{on } \Omega, \end{cases} \quad (138)$$

where α_1, α_2, and β are positive constants, the desired state y_d and the desired finial 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 (138). 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 subsection. We use the following notion of solution [58,59].

Definition 1 (Very weak solution) Let $T > 0$, and $(y_0^1, y_0^2, u, f) \in L^2(\Omega) \times (H^1_{\Gamma_0}(\Omega))^* \times L^2(\Sigma_c) \times L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*)$ be given. A function $y \in L^2(Q)$ is referred to as the very weak solution of (138), if the following inequality holds

$$\begin{aligned} &\langle f, \varphi\rangle_{(L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*),L^2(0,T;H^1_{\Gamma_0}(\Omega)))} = \\ &\quad (g, y)_{L^2(Q)} + (y_0^1, \varphi_t(0))_{L^2(\Omega)} - \langle y_0^2, \varphi(0)\rangle_{((H^1_{\Gamma_0}(\Omega))^*, H^1_{\Gamma_0}(\Omega))} - (u, \varphi)_{L^2(\Sigma_0)} \end{aligned} \quad (139)$$

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 [55,56,75] for the solution of (138).

Lemma 10 *For every* $(y_0^1, y_0^2, u, f) \in L^2(\Omega) \times (H^1_{\Gamma_0}(\Omega))^* \times L^2(\Sigma_c) \times L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*)$*, equation* (138) *admits a unique very weak solution* $y(y_0^1, y_0^2, u, f)$ *in the space* $C^0([0,T];L^2(\Omega)) \cap C^1([0,T];(H^1_{\Gamma_0}(\Omega))^*)$ *satisfying*

$$\begin{aligned} &\|y\|_{C^0([0,T];L^2(\Omega))} + \|y_t\|_{C^0([0,T];(H^1_{\Gamma_0}(\Omega))^*)} \\ &\qquad \le c\left(\|y_0^1\|_{L^2(\Omega)} + \|y_0^2\|_{(H^1_{\Gamma_0}(\Omega))^*} + \|u\|_{L^2(\Sigma_c)} + \|f\|_{L^2(0,T;(H^1_{\Gamma_0}(\Omega))^*)}\right), \end{aligned} \tag{140}$$

where the constant c_1 *is independent of* y_0^1*,* y_0^2*,* 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* $(y_0^1, y_0^2, f) \mapsto y(y_0^1, y_0^2, 0, f)$ *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 optimal control problem (137)-(138) in the form (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

$$\mathcal{L}u := \begin{pmatrix} \alpha_1(i_1 \circ L)(u) \\ \alpha_2(\delta_T \circ L)(u) \end{pmatrix}, \quad \psi := \begin{pmatrix} \alpha_1 y_d - \alpha_1(i_2 \circ \Pi)(y_0^1, y_0^2, f) \\ \alpha_2 z_d - \alpha_2(\delta_T \circ \Pi)(y_0^1, y_0^2, f) \end{pmatrix}. \tag{141}$$

Similarly to the previous subsection, the optimal control problem (137)-(138), 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 the fact that the operator $\mathcal{A}$ is uniformly positive, bounded, and self-adjoint, the existence and uniqueness of the solution to optimal control problem (137)-(138) can be justified due the fact that $\mathcal{A}$ has a bounded inverse.

Remark 10 In the optimal control problem (137)-(138), the operator $\mathcal{A}$ is a compact perturbation of the identity, since $\mathcal{L} : \mathcal{H} \to \mathcal{H}$ is compact. Indeed, due to [74][Corollary 5.], the continuous embedding from the space $C^0([0,T];H^{\frac{1}{2}}_{\Gamma_0}(\Omega)) \cap C^1([0,T];H^{-\frac{1}{2}}(\Omega))$ to the space $C^0([0,T];L^2(\Omega))$ is compact and this implies the compactness of $\delta_T \circ L$ and $i_1 \circ L$. Therefore, due to (141), $\mathcal{L}$ is compact with respect to the product topology $L^2(Q) \times L^2(\Omega)$.

Now assume that $u^* \in \mathcal{H}$ is the optimal solution of the optimal control problem (137)-(138). Then, the first-order optimality condition (EP) can be expressed as (115) where the operator $\mathcal{L}$ and the function ψ were defined in (141). Moreover, it can be shown (see [53,58,68]) that (115) 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 (138).

4.3 Distributed Optimal Control for the Burgers Equation

Here we consider the following optimal control problem which consists of minimizing the performance index

$$J(y,u) := \frac{\alpha_1}{2}\|y - y_d\|^2_{L^2(Q)} + \frac{\alpha_2}{2}\|y(T) - z_d\|^2_{L^2(0,1)} + \frac{\beta}{2}\|u\|^2_{L^2(\hat{Q})}, \tag{142}$$

subject to the Burgers equation with homogeneous Dirichlet boundary condition.

$$\begin{cases} y_t - \vartheta y_{xx} + y y_x = Bu + f, \quad (t,x) \in Q, \\ y(t,0) = y(t,1) = 0, \quad t \in (0,T), \\ y(0,x) = y_0(x), \quad x \in (0,1). \end{cases} \tag{143}$$

where ϑ, β α_1, α_2, and T are positive constants, $y(t) = y(t,x), u(t) = u(t,x)$, $Q := (0,T) \times (0,1)$, and $\hat{Q} = (0,T) \times \hat{\Omega}$ where $\hat{\Omega}$ is an open subset of $(0,1)$. Moreover, $y_0 \in L^2(0,1)$, $f \in L^2(0,T;H^{-1}(0,1))$, the desired states y_d and z_d are smooth enough, and the extension operator $B \in \mathcal{L}(L^2(\hat{\Omega}), L^2(0,1))$ is defined by

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

Considering the space

$$W(0,T) := \{\phi : \phi \in L^2(0,T;H_0^1(0,1)), \phi_t \in L^2(0,T;H^{-1}(0,1))\}.$$

as the space of solutions, we have the following notion of weak solution.

Definition 2 Let $(y_0, u, f) \in L^2(0,1) \times L^2(\hat{Q}) \times L^2(0,T;H^{-1}(0,1))$ be given. Then, a function $y \in W(0,T)$ is referred as a weak solution to (143) if $y(0) = y_0$ is satisfied in $L^2(0,1)$ and for almost every $t \in (0,T)$, the following equality

$$\langle y_t(t), \varphi\rangle_{H^{-1},H_0^1} + \vartheta(y(t),\varphi)_{H_0^1} + b(y(t),y(t),\varphi) = \langle Bu(t) + f, \varphi\rangle_{H^{-1},H_0^1} \quad \text{for all } \varphi \in H_0^1(0,1)$$

holds, where the continuous trilinear form $b : H_0^1(0,1) \times H_0^1(0,1) \times H_0^1(0,1) \to \mathbb{R}_+$ is defined as

$$b(\varphi,\psi,\phi) = \int_0^1 \varphi\psi_x\phi dx \quad \text{for all } \varphi,\psi,\phi \in H_0^1(0,1).$$

It is known that, for every triple $(y_0, u, f) \in L^2(0,1) \times L^2(\hat{Q}) \times L^2(0,T;H^{-1}(0,1))$, equation (143) admits a unique weak solution $y(y_0,u,f) \in W(0,T)$ and for this weak solution we have the following estimate

$$\|y\|_{W(0,T)} \leq C\left(\|y_0\|_{L^2(\Omega)} + \|u\|_{L^2(\hat{Q})} + \|f\|_{L^2(0,T;H^{-1}(0,1))}\right)^2, \tag{144}$$

where the constant C depends only on T and ϑ. Now, by setting $X = W(0,T) \times \mathcal{H}$ with $\mathcal{H} := L^2(\hat{Q})$, and $Y := L^2(0,T;H^{-1}(0,1)) \times L^2(0,1)$, we define $e : X \to Y$ by

$$e(y,u) := \begin{pmatrix} y_t - \vartheta y_{xx} + yy_x - Bu - f \\ y(0) - y_0 \end{pmatrix}.$$

The mapping $e : X \to Y$ consists of a sum of continuous linear terms and a continuous bilinear term. Hence it can be shown that it is infinitely Fréchet differentiable. Moreover due to the unique solvability of (143), for every $u \in \mathcal{H}$ there exists a unique element $y = y(u) \in W(0,T)$ satisfying $e(y(u),u) = 0$ and estimate (144) holds. Therefore the control-to-state $u \in \mathcal{H} \mapsto y(u) \in W(0,T)$ is well-defined. Then we can rewrite the optimal control problem (142)-(143) in the following form

$$\min_{u\in\mathcal{H}} \mathcal{F}(u) = \min_{u\in\mathcal{H}} J(y(u),u) = \min_{(y,u)\in X} \{J(y,u) : \text{ subject to } e(y,u) = 0\}. \tag{145}$$

Further, due to estimate (144) and the compact embedding from the space $W(0,T)$ to the space $L^2(Q)$, it follows from standard subsequential limit arguments that the optimal control problems (142)-(143) admits a solution, see e.g., [76,78]. Before dealing with the optimality conditions, we refer to the following linearized Burgers equation at $y \in W(0,T)$ and its corresponding backward in time adjoint equation

$$\begin{cases} q_t - \vartheta q_{xx} + (yq)_x = \phi & (t,x) \in Q, \\ q(t,0) = q(t,1) = 0 & t \in (0,T), \\ q(0) = q_0 & x \in (0,1), \end{cases} \tag{146}$$

$$\begin{cases} -p_t - \vartheta p_{xx} - yp_x = \psi & (t,x) \in Q, \\ p(t,0) = p(t,1) = 0 & t \in (0,T), \\ p(T,x) = p_T & x \in (0,1). \end{cases} \tag{147}$$

It can be shown that for every pairs (ϕ, q_0) and (ψ, p_T) in the space Y, the solution operators $\mathcal{S}^y_{lin} : Y \to W(0,T)$ of (146), and $\mathcal{S}^y_{adj} : Y \to W(0,T)$ of (147) defined by $(\phi, q_0) \mapsto v$ and $(\psi, p_T) \mapsto p$, respectively, are well-defined and continuous. See e.g., [76,78].

Due to the definitions of $\mathcal{S}^{y(u)}_{lin}$ and $e_y(y(u), u)$, we can infer that $e_y^{-1}(y(u), u) = \mathcal{S}^{y(u)}_{lin}$ and, as consequence, $e_y(y, u)$ is continuously invertible. In addition, since e is infinitely continously Fréchet differentiable [43], the implicit function theorem implies that the control-to-state operator $u \mapsto y(u)$ is infinitely continuously Fréchet differentiable and its Fréchet derivatives of all orders are Lipschitz continuous on bounded sets. Now we are in the position to derive the first-order optimality conditions. First, by using the implicit function theorem, the first derivative of the mapping $u \mapsto y(u)$ at u in direction of an arbitrary $\delta u \in \mathcal{H}$ is given by

$$y'(u)\delta u = -e_y^{-1}(x)e_u(x)\delta u, \tag{148}$$

where $x := (y(u), u) \in X$. Then, by the chain rule we obtain

$$\mathcal{F}'(u)\delta u = (\mathcal{G}(u), \delta u) = ((y'(u))^* J_y(x) + J_u(x), \delta u),$$

where $(y'(u))^*$ stands for the adjoint operator of $y'(u)$. Since δu is arbitrary, the first-order optimality condition (EP) can be written as

$$\mathcal{G}(u^*) = J_u(x^*) - e_u^*(x)e_y^{-*}(x)J_y(x^*) = 0. \tag{149}$$

where $x^* := (y(u^*), u^*)$. Moreover, by setting $(p^*, \bar{p}) := -e_y^{-*}(x)J_y(x^*)$ with $(p^*, \bar{p}) \in Y^*$ and $\bar{p} = p^*(0)$, the first-order optimality condition (149) can be expressed as the following system of differential equations

$$\begin{cases} B^* p^* = \beta u^* \quad \text{ in } L^2(\hat{Q}), \\ p^* = \mathcal{S}^{y(u^*)}_{adj}(-\alpha_1(y(u^*) - y_d), -\alpha_2((y(u^*))(T) - z_d)). \end{cases}$$

Next, we compute the second derivative of $\mathcal{F}$. Let $(\delta u, \delta v) \in \mathcal{H} \times \mathcal{H}$ be arbitrary, then using the implicit functions theorem, the second derivative of the operator $u \mapsto y(u)$ from $\mathcal{H}$ to $W(0,T)$ can be written as

$$y''(u)(\delta u, \delta v) = -e_y^{-1}(x)e_{yy}(x)(y'(u)\delta u, y'(u)\delta v). \tag{150}$$

Now, by using the chain rule and (150) as in [42,43], we obtain

$$\begin{aligned} \mathcal{F}''(u)(\delta u, \delta v) &= \langle J_{yy}(x)y'(u)\delta u, y'(u)\delta v\rangle \\ &\quad + \langle -e_y^{-*}(x)J_y(x), e_{yy}(x)(y'(u)\delta u, y'(u)\delta v)\rangle + \langle J_{uu}(x)\delta u, \delta v\rangle_{\mathcal{H}}. \end{aligned} \tag{151}$$

Furthermore, due to the first estimate in (151) and the fact that $J : \mathcal{H} \times W(0,T) \to \mathbb{R}$ and the control-to-state operator are infinitely Fréchet differentiable, it follows, clearly, that $\mathcal{F}'' : \mathcal{H} \to \mathcal{L}(\mathcal{H}, \mathcal{L}(\mathcal{H}, \mathbb{R}))$ is locally Lipschitz continuous. Then, the uniformly positiveness of $\mathcal{F}''(u^*)$ can be expressed as

$$\begin{aligned} \mathcal{F}''(u^*)(v,v) :&= \alpha_1\|q^*\|^2_{L^2(Q)} + \alpha_2\|q^*(T)\|^2_{L^2(0,1)} + (p^*, 2q^*q^*_x)_{L^2(Q)} + \beta\|v\|^2_{\mathcal{H}} \\ &\geq \delta_{\inf}\|v\|^2_{\mathcal{H}} \quad \text{ for all } v \in \mathcal{H}, \end{aligned} \tag{152}$$

where $\delta_{\inf} > 0$, $p^* := \mathcal{S}^{y(u^*)}_{adj}(-\alpha_1(y(u^*) - y_d), -\alpha_2((y(u^*))(T) - z_d)$ and $q^* := \mathcal{S}^{y(u^*)}_{lin}(v, 0)$.

Remark 11 Clearly, the only term in (152) that can spoil the uniformly positiveness of $\mathcal{F}''(u^*)$ is the term involving p^*. This term originates from the nonlinear convection term in the state equation. Since

$$\left|(p^*, 2q^*q^*_x)_{L^2(Q)}\right| \leq c\|p^*\|_{L^2(0,T;L^\infty(0,1))}\|q^*\|^2_{W(0,T)}$$

for a constant c, the uniformly positiveness of $\mathcal{F}''(u^*)$ holds, provided that $\|p^*\|_{L^2(0,T;L^\infty(0,1))}$ is small enough. Indeed, for $p^* = 0$, inequality (152) holds for $\delta_{\inf} := \beta$. For instance, by setting $y_d = z_d = f = 0$, inequality (152) holds for every initial function y_0 with sufficiently small $\|y_0\|_{L^2(0,1)}$.

5 Numerical Experiments

In order to validate our theoretical findings in the previous sections, we report numerical results corresponding to the optimal control problems introduced in the previous section. We investigate the application of Algorithm 1 with respect to different strategies for selecting step-sizes and different choices of the discretization parameter h, the control cost parameter β, and the tolerance ϵ in the termination condition (96). For Algorithm 1, we consider the cases:

BB1: $\alpha_k := \alpha_k^{BB1}$ for every $k \geq 1$.
BB2: $\alpha_k := \alpha_k^{BB2}$ for every $k \geq 1$.
ABB: $\alpha_k := \begin{cases} \alpha_k^{BB1} \text{ for odd } k \geq 1, \\ \alpha_k^{BB2} \text{ for even } k \geq 1. \end{cases}$

The last case, which is known as the alternating strategy, has already been introduced by e.g., [21,34] in the context of finite-dimensional unconstrained optimization. Moreover, [23] reports numerical results for the case of finite-dimensional bound-constrained optimization problems which show that projected ABB works somewhat better than projected BB1. According to (51) 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 its has been discussed in Remarks 2 and 4, one expects a larger total number of iterations for a smaller value of β and a fixed tolerance ϵ. Moreover, according to Remark 6, the number ℓ depends on the behaviour (monotonicity versus nonmonotonicity) of $\{\|\mathcal{G}_k\|\}_k$, and consequently also on $\kappa(\mathcal{A}_{u^*}^{\mathcal{F}})$. Hence, the smaller β is chosen, the larger the value of ℓ is expected to be. We report the total number of iteration of the optimization Algorithm for different levels of discretization, or equivalently, different values of mesh-sizes. Then, for every example and fixed tolerance ϵ, ℓ is reported as the maximum of the pairwise differences of $k_h(\epsilon)$ for different choices of h. We have chosen $u_0 = 0$ and $\alpha_0 = 1$ ($u_1 := -\mathcal{G}(0)$) as the initial iterates. All computations were done in the MATLAB platform.

Example 1 (Dirichlet optimal control for the Poisson equation) We consider the problem introduced in Subsection 4.1 which is posed on the domain $\Omega := (0,1)^2$. For the discretization a uniform mesh was generated by triangulation. Then over this mesh, the discretization was done by a conforming linear finite element scheme using continuous piecewise linear basis functions as described in Subsubsection 4.1.2. We set $f(x) = 10\sin(\pi(x_1 + x_2))$ and $y_d(x) = (x_1^2 + x_2^2)^{\frac{1}{3}}$ where $x := (x_1, x_2) \in \Omega$. Table 1 shows the number of required iterations $k_h^*(\epsilon)$ for different step-size strategies, and different values of β, ϵ, and the mesh-size h. From Table 1, it can be observed that:

1. For every fixed h, ϵ, and choice of step-size, decreasing in the value of β implies that the number of required iterations $k_h^*(\epsilon)$ becomes larger and, thus, the convergence is getting slower. This is in accordance with 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 Subsubsection 4.1.1. More precisely, $\kappa(\mathcal{A}) = \frac{\beta+\delta_{\sup}}{\beta+\delta_{\inf}}$ with $\delta_{\inf} := \inf(\sigma(\mathcal{L}^*\mathcal{L}))$ and $\delta_{\sup} := \sup(\sigma(\mathcal{L}^*\mathcal{L}))$. Hence a larger value of β yields a smaller value of $\kappa(\mathcal{A})$. That is as expected from the theory, for a larger β Algorithm 1 requires fewer iterations $k_h^*(\epsilon)$ for every fixed h and ϵ. This behaviour is clearly illustrated in Figure 1 which depicts the convergence of $\|\mathcal{G}_k^h\|_h$ for the choice of BB1 step-sizes, $h = 2^{-9}\sqrt{2}$, and different values of β. As can be seen from Figure 1, the convergence for the cases $\beta = 0.5$ and $\beta = 0.2$ 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 $\beta = 0.5$ compared to $\beta = 0.2$. However, for the rest of the cases, nonmonotonic behaviour occurs, which corresponds to $\kappa(\mathcal{A}) \geq 2$. Apparently, as β decreases, the nonmonotonic behaviour in the sequences $\{\|\mathcal{G}_k\|\}_k$ and, consequently, in $\{\|\mathcal{G}_k^h\|_h\}_k$ becomes stronger. As discussed in Remarks 2 and 4, if $\kappa(\mathcal{A})$ becomes larger, then the changes in the decreasing components $|g_i^k|$ (for instance $i \in \{0\} \cup \{i : i \geq n^u\}$) are getting smaller compared to the nondecreasing components. This explains why a decrease in the value β leads to an increase in nonmonotonicity.
2. Mesh-independence can be observed from Table 1. More precisely, we can see that for every fixed β, ϵ, and step-size strategy, the iterations $k_h(\epsilon)$ stay almost constant and do not change as the discretezation levels changes. Moreover, for $\beta = 0.2$, $\beta = 0.05$, and $\beta = 0.01$ we can state that $\ell \approx 1$, $\ell \approx 3$, and $\ell \approx 6$, respectively. This is also due to the dependence of the spectrum of $\mathcal{A} = \mathcal{L}^*\mathcal{L} + \beta\mathcal{I}$ on the magnitude of β (see Remark 6).

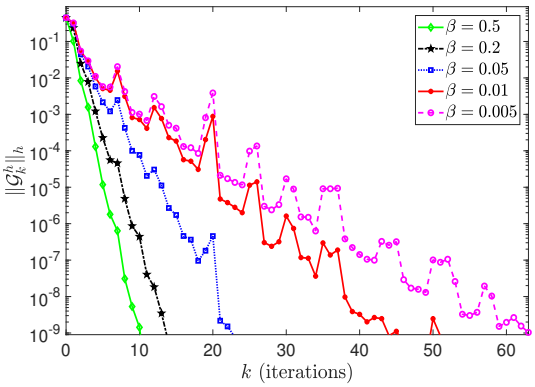

Fig. 1 Convergence of $\|\mathcal{G}_k^h\|_h$ for Algorithm 1 applied to Example 1 with BB1 step-sizes, $h = 2^{-9}\sqrt{2}$, and for different choices of β

To further study the behaviour of Algorithms 1, we consider Table 2 which summarizes the values of $\|\mathcal{G}_k^h\|_h$ for the choice of $\beta = 0.01$, at the iterations $k = 37, \ldots, 45$, and different levels of discretization. It can be seen that in any case the sequence $\{\|\mathcal{G}_k^h\|_h\}_k$ has a nonmonotonic behaviour. For every case the members of $\{\|\mathcal{G}_k^h\|_h\}_k$ at which the monotonicity of the sequence is violated, are indicated by bold type. With the superscript star we denote the members corresponding to $k_h^*(\epsilon)$ with $\epsilon = 1e-8$.

Example 2 (Neumann optimal control the for the linear wave equation) In this example, we deal with the optimal control problem (137)-(138). Here, the spatial discretization has been done similarly to the previous example on the domain $\Omega := (0,1)^2$ with the mesh-size h. Further, for the temporal discretization of the state equation we used a Petrov-Galerkin scheme based on continuous piecewise linear basis functions for the trial space and piecewise constant test functions. By doing so, the resulting discretized system is equivalent to the system first discretized in space followed by the Crank-Nicolson time stepping method with a step-size Δt. Since the temporal test functions have been chosen to be piecewise constant, it is natural to also discretize the adjoint equation and also control by these functions. This implies that the approximated gradient is consistent with both continuous functional and the discrete functional. Here we set $T = 1$, $y_0^1(x) = \sin(\pi x_1)\sin(\pi x_2)$, $y_0^2(x) = 0$, $f(t,x) = \pi^2 \sin(\pi x_1 t)\sin(\pi x_2 t)$, $z_d(x) = 0$, and

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

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 2, we report the behaviour of the gradient norm for Example 2 for the choice of BB2 step-sizes, $(\Delta t, h) = (0.0064, 2^{-7}\sqrt{2})$, and for different values of β. To illustrate the mesh-independence, we reported the values of $k^*_{\Delta t,h}(\epsilon)$ for different levels of temporal and spatial discretization. As it is reported in Table 3, we decreased the mesh-size h and step-size Δt simultaneously. Clearly, similar observations as in the previous example are also valid for this example, with the difference that here for $\beta = 0.5$, and $\beta = 0.05$, we have $\ell \approx 0$ and $\ell \approx 6$, respectively.

Example 3 (Distributed optimal control for the Burgers equation) We consider the optimal control problem (142)-(143) posed on the interval $(0,1)$. The spatial discretization was done by the standard Galerkin method based on piecewise linear basis functions with mesh-size h. For temporal discretization, we used the implicit Euler method with a step-size denoted by Δt. Moreover, the resulting nonlinear systems after the temporal discretization were solved by Newton's method with the tolerance $\epsilon_n = 10^{-13}$. Here the control acts on the open interval $\hat{\Omega} = (0.1, 0.4)$. Moreover we set $\vartheta = 0.01$, $y_0(x) = 5\exp(-20(x-0.5)^2)$, and $y_d(t,x) = z_d(x) = f(t,x) = 0$. Similarly to the previous example, we compute the values of $k^*_{\Delta t,h}(\epsilon)$ for different levels of temporal and spatial discretization. These results are gathered in Table 4. Further, Figure 3 shows the convergence of Algorithm 1 applied to Example 3, for the choice of ABB step-sizes, $(\Delta t, h) = (2^{-7}, 2^{-8})$, and different values of β. As can be seen from Table 4 and Figures 3, despite the nonlinearity the observations 1 and 2 from Example 1 hold also true for this example.

The number of required iteration $k_h^*(\epsilon)$							
$\beta = 0.2$							
BB1	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	3	3	3	3	3	3
	$1e-4$	6	6	6	6	6	6
	$1e-6$	9	9	9	9	9	9
	$1e-8$	12	13	13	13	13	13
BB2	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	3	3	3	3	3	3
	$1e-4$	6	6	6	6	6	6
	$1e-6$	9	9	9	9	9	9
	$1e-8$	11	12	12	12	12	12
ABB	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	3	3	3	3	3	3
	$1e-4$	6	6	6	6	6	6
	$1e-6$	9	9	9	9	9	9
	$1e-8$	12	12	13	13	13	13
$\beta = 0.05$							
BB1	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	4	4	4	4
	$1e-4$	9	9	9	9	10	10
	$1e-6$	14	16	16	16	16	16
	$1e-8$	21	21	21	21	21	21
BB2	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	4	4	4	4
	$1e-4$	9	9	9	9	9	9
	$1e-6$	14	15	15	15	15	15
	$1e-8$	19	21	21	21	22	22
ABB	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	4	4	4	4
	$1e-4$	9	9	9	9	9	9
	$1e-6$	14	14	15	16	16	16
	$1e-8$	18	21	21	21	21	21
$\beta = 0.01$							
BB1	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	4	5	5	5
	$1e-4$	16	16	16	16	16	16
	$1e-6$	24	28	27	27	27	27
	$1e-8$	38	39	38	40	38	39
BB2	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	5	5	5	5
	$1e-4$	13	15	15	15	15	15
	$1e-6$	26	26	30	31	31	32
	$1e-8$	39	44	43	45	45	44
ABB	ϵ \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
	$1e-2$	4	4	5	5	5	5
	$1e-4$	15	15	16	16	16	16
	$1e-6$	24	24	28	28	28	28
	$1e-8$	43	38	43	43	43	40

Table 1 Numerical results for Example 1

A Appendix

A.1 Proof of Proposition 8

For every $k \geq 1$, we consider the sequence $\{\hat{\alpha}_j^k\}_j$ associated to $\{\hat{u}_j^k\}_j$, which is defined by

$$\hat{\alpha}_j^k := \begin{cases} \hat{\alpha}_j^{BB1k} & \text{if } \alpha_{k+j} = \alpha_{k+j}^{BB1}, \\ \hat{\alpha}_j^{BB2k} & \text{if } \alpha_{k+j} = \alpha_{k+j}^{BB2}. \end{cases} \tag{153}$$

The value of $\|\mathcal{G}_k^h\|$ at an iteration k							
	k \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
BB1	37	$1.29e-8$	$1.59e-8$	$3.96e-8$	$5.25e-7$	$1.90e-7$	$1.98e-7$
	38	$9.22e-9^*$	$1.05e-8$	$4.26e-9^*$	$1.03e-7$	$9.69e-9^*$	$1.39e-8$
	39	$\mathbf{2.93e-8}$	$5.17e-9^*$	$3.78e-9$	$5.93e-8$	$3.88e-9$	$8.75e-9^*$
	40	$\mathbf{1.05e-7}$	$1.38e-9$	$2.38e-9$	$7.21e-9^*$	$3.25e-9$	$3.46e-9$
	41	$1.43e-8$	$1.08e-9$	$\mathbf{6.90e-9}$	$4.30e-9$	$2.02e-9$	$1.90e-9$
	42	$\mathbf{1.46e-8}$	$\mathbf{1.69e-9}$	$4.72e-9$	$\mathbf{5.76e-9}$	$2.66e-9$	$1.28e-9$
	43	$7.26e-10$	$\mathbf{1.26e-8}$	$3.74e-9$	$1.12e-9$	$2.47e-9$	$6.74e-9$
	44	$6.24e-10$	$1.55e-9$	$2.63e-10$	$8.49e-10$	$8.46e-10$	$\mathbf{1.84e-8}$
	45	$4.09e-10$	$1.43e-9$	$1.74e-10$	$7.33e-10$	$\mathbf{1.11e-9}$	$4.10e-10$
	k \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
BB2	37	$5.94e-8$	$2.91e-8$	$7.59e-8$	$3.38e-7$	$7.83e-8$	$2.40e-7$
	38	$2.52e-8$	$2.71e-8$	$4.59e-8$	$2.19e-7$	$6.21e-8$	$1.95e-7$
	39	$4.97e-9^*$	$1.78e-8$	$1.28e-8$	$2.08e-7$	$5.32e-8$	$3.46e-8$
	40	$4.06e-9$	$\mathbf{2.87e-8}$	$\mathbf{1.38e-7}$	$1.71e-7$	$4.72e-8$	$\mathbf{2.61e-7}$
	41	$3.52e-9$	$\mathbf{6.00e-8}$	$4.12e-8$	$6.06e-8$	$1.77e-8$	$4.37e-8$
	42	$\mathbf{4.08e-9}$	$2.13e-8$	$2.63e-8$	$\mathbf{1.13e-6}$	$\mathbf{2.12e-7}$	$1.40e-8$
	43	$3.36e-9$	$1.23e-8$	$1.02e-9^*$	$9.80e-8$	$6.14e-8$	$1.00e-8$
	44	$2.38e-9$	$1.88e-9^*$	$7.07e-10$	$4.43e-8$	$1.64e-8$	$2.19e-9^*$
	45	$2.67e-9$	$1.78e-9$	$4.10e-10$	$9.39e-9^*$	$7.97e-9^*$	$2.00e-9$
	k \ h	$2^{-5}\sqrt{2}$	$2^{-6}\sqrt{2}$	$2^{-7}\sqrt{2}$	$2^{-8}\sqrt{2}$	$2^{-9}\sqrt{2}$	$2^{-10}\sqrt{2}$
ABB	37	$5.10e-8$	$1.57e-8$	$2.93e-8$	$3.44e-8$	$3.83e-8$	$2.55e-8$
	38	$2.70e-8$	$7.44e-9^*$	$1.96e-8$	$2.49e-8$	$2.94e-8$	$1.55e-8$
	39	$2.37e-8$	$5.79e-9$	$\mathbf{3.70e-8}$	$\mathbf{4.67e-8}$	$2.20e-8$	$1.15e-8$
	40	$\mathbf{2.50e-8}$	$4.61e-9$	$\mathbf{7.85e-8}$	$\mathbf{1.50e-7}$	$\mathbf{9.20e-8}$	$2.21e-9^*$
	41	$\mathbf{1.50e-7}$	$\mathbf{1.25e-8}$	$\mathbf{8.04e-8}$	$9.71e-8$	$\mathbf{1.81e-7}$	$\mathbf{5.17e-8}$
	42	$1.31e-8$	$\mathbf{1.39e-8}$	$1.06e-8$	$1.08e-8$	$1.17e-8$	$3.26e-8$
	43	$9.26e-9^*$	$8.82e-9$	$2.66e-9^*$	$6.73e-9^*$	$6.90e-9^*$	$1.36e-8$
	44	$8.76e-9$	$1.06e-9$	$1.75e-9$	$1.90e-9$	$6.11e-9$	$2.33e-9$
	45	$4.43e-9$	$9.25e-10$	$7.39e-10$	$1.29e-9$	$5.62e-9$	$1.90e-9$

Table 2 The values of $\|\mathcal{G}_k^h\|_h$ related to Example 1 for $\beta = 0.01$ and iterations $k = 37, \ldots, 45$

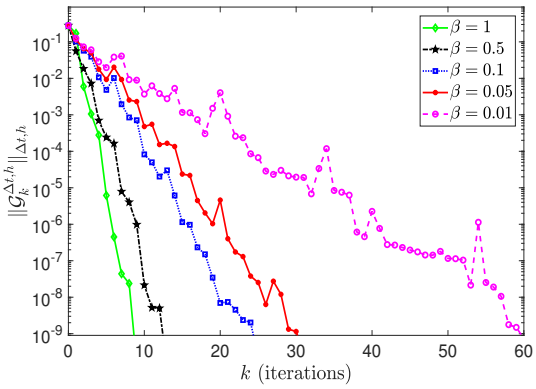

Fig. 2 Convergence of $\|\mathcal{G}_k^{\Delta t,h}\|_{\Delta t,h}$ for Algorithm 1 applied to Example 2 with BB2 step-sizes, $(\Delta t, h) = (0.0064, 2^{-7}\sqrt{2})$, and for different choices of β

for all $j \geq 1$. We will show by induction that for every $q \in \{0, \ldots, m\}$, there exist positive constants λ_q and η_q such that

$$\begin{cases} \text{If } u_k \in \mathcal{B}_{\eta_q}(u^*),\ \alpha_k \in [\alpha_{\inf}, \alpha_{\sup}], \text{ and if for some } \ell \in \{0, \ldots, q\}, \text{ the property (78) holds,} \\ \text{then we have } u_{k+j} \in \mathcal{B}_\tau(u^*), \text{ and } \quad \|u_{k+j} - \hat{u}_j^k\| \leq \lambda_q \|u_k - u^*\|^2 \text{ for all } j \in \{0, \ldots, \ell\}. \end{cases} \tag{P_q}$$

For the case that $q = \ell = 0$ and the choice of $\eta_0 = \tau$ and arbitrary $\lambda_0 > 0$, property (P_q) holds clearly since $\hat{u}_0^k = u_k$. For the case that $q = 1$, by using (75) and (77), we have

$$\|u_{k+1} - u^*\| \leq \|u_k - u^*\| + \|\mathcal{S}_k\| \leq \|u_k - u^*\| + \frac{1}{|\alpha_k|}\|\mathcal{G}_k\| \leq (1 + \frac{\alpha_{\sup}}{\alpha_{\inf}})\|u_k - u^*\|,$$

where $\mathcal{S}_k := u_{k+1} - u_k$. Hence, for $\eta_1 := \frac{\tau}{1+\frac{\alpha_{\sup}}{\alpha_{\inf}}}$, we obtain $u_{k+1} \in \mathcal{B}_\tau(u^*)$. In the case $q = 1$ we have either $\ell = 0$ or $\ell = 1$. For $\ell = 0$, (79) holds trivially. Therefore, we need to investigate (79) for $\ell = 1$. By $L1$ and using the facts that

The number of required iteration $k^*_{\Delta t,h}(\epsilon)$						
$\beta = 0.5$						
BB1	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	3	3	3	3	3
	$1e-4$	7	7	7	7	7
	$1e-6$	9	9	9	9	9
	$1e-8$	11	11	11	11	11
BB2	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	3	3	3	3	3
	$1e-4$	7	7	7	7	7
	$1e-6$	9	9	9	9	9
	$1e-8$	11	11	11	11	11
ABB	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	3	3	3	3	3
	$1e-4$	7	7	7	7	7
	$1e-6$	9	9	9	9	9
	$1e-8$	11	11	11	11	11
$\beta = 0.05$						
BB1	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	7	7	7	7	7
	$1e-4$	14	14	14	14	14
	$1e-6$	21	21	21	24	24
	$1e-8$	28	29	28	29	29
BB2	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	5	5	5	5	5
	$1e-4$	12	15	15	15	15
	$1e-6$	19	21	21	21	21
	$1e-8$	24	25	25	26	26
ABB	ϵ \ $(\Delta t, h)$	$(0.01, 2^{-4}\sqrt{2})$	$(0.04, 2^{-5}\sqrt{2})$	$(0.016, 2^{-6}\sqrt{2})$	$(0.0064, 2^{-7}\sqrt{2})$	$(0.0026, 2^{-8}\sqrt{2})$
	$1e-2$	5	7	7	7	7
	$1e-4$	14	14	14	14	13
	$1e-6$	21	22	24	23	23
	$1e-8$	28	31	34	29	28

Table 3 Numerical results for Example 2

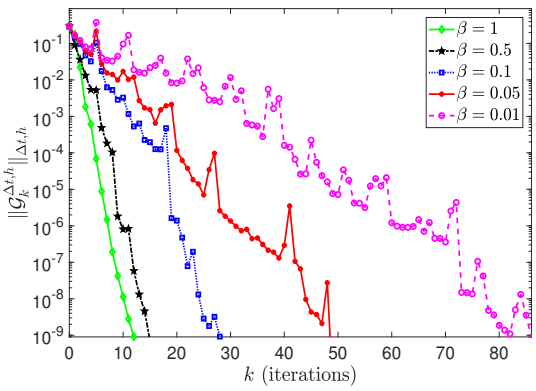

Fig. 3 Convergence of $\|\mathcal{G}_k^{\Delta t,h}\|_{\Delta t,h}$ for Algorithm 1 applied to Example 3 with ABB step-sizes, $(\Delta t, h) = (2^{-7}, 2^{-8})$, and for different choices of β

$u_k = \hat{u}_0^k \in \mathcal{B}_\tau(u^*)$ and $\alpha_k = \hat{\alpha}_0^k$, we can infer that

$$\begin{aligned}\|u_{k+1} - \hat{u}_1^k\| &\le \|u_k - \frac{1}{\alpha_k}\mathcal{G}(u_k) - (\hat{u}_0^k - \frac{1}{\hat{\alpha}_0^k}\hat{\mathcal{G}}(\hat{u}_0^k))\| \\ &\le \frac{1}{|\alpha_k|}\|\mathcal{G}(u_k) - \hat{\mathcal{G}}(\hat{u}_0^k)\| \\ &\le \frac{1}{\alpha_{\inf}}\|\mathcal{G}(u_k) - \mathcal{A}_{u^*}^{\mathcal{F}}(u_k - u^*)\| \le \frac{L}{\alpha_{\inf}}\|u_k - u^*\|^2,\end{aligned}$$

where $\hat{\mathcal{G}}(u) = \mathcal{A}_{u^*}^{\mathcal{F}}(u - u^*)$. This ends the justification of the induction basis by choosing $\lambda_1 := \frac{L}{\alpha_{\inf}}$ and $\eta_1 := \frac{\tau}{1+\frac{\alpha_{\sup}}{\alpha_{\inf}}}$.

The number of required iteration $k^*_{\Delta t,h}(\epsilon)$						
$\beta = 0.5$						
BB1	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	4	4	4	4	4
	$1e-4$	8	8	9	9	9
	$1e-6$	11	11	11	11	11
	$1e-8$	13	14	14	14	14
BB2	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	4	4	4	4	4
	$1e-4$	8	8	8	8	8
	$1e-6$	10	11	11	11	11
	$1e-8$	14	14	14	14	14
ABB	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	4	4	4	4	4
	$1e-4$	8	8	9	9	9
	$1e-6$	11	12	10	10	10
	$1e-8$	13	13	14	14	14
$\beta = 0.05$						
BB1	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	12	12	12	12	12
	$1e-4$	25	24	23	25	24
	$1e-6$	36	32	32	33	32
	$1e-8$	43	40	44	42	39
BB2	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	9	9	9	9	9
	$1e-4$	23	21	24	23	23
	$1e-6$	31	29	33	34	35
	$1e-8$	35	41	41	38	40
ABB	ϵ \ $(\Delta t, h)$	$(2^{-4}, 2^{-5})$	$(2^{-5}, 2^{-6})$	$(2^{-6}, 2^{-7})$	$(2^{-7}, 2^{-8})$	$(2^{-8}, 2^{-9})$
	$1e-2$	13	13	13	9	9
	$1e-4$	24	26	21	21	21
	$1e-6$	30	32	29	31	35
	$1e-8$	36	38	42	44	40

Table 4 Numerical results for Example 3

Now, let p be an integer with $2 \leq p < m$ such that Property (P_q) holds for $q = p$ and, constants λ_p and η_p. We will show that this property holds for $q = p+1$, a positive constant $\lambda_{p+1} \geq \lambda_p$, and for the choice of

$$\eta_{p+1} := \min\left\{\frac{1}{4\lambda_p}, \tau\left(1 + \frac{\alpha_{\sup}}{\alpha_{\inf}}\right)^{-(p+1)}\right\}, \tag{154}$$

where due to (154), we obtain $\eta_{p+1} \leq \eta_p$.

Now assume that $u_k \in \mathcal{B}_{\eta_{p+1}}(u^*)$ and $\alpha_k \in [\alpha_{\inf}, \alpha_{\sup}]$. First we investigate Property (P_q) for $q = p+1$ and $\ell \leq p$. That is, we assume that (78) holds for any given $\ell \leq p$ and we show that (79) holds. In this case, since $\eta_{p+1} \leq \eta_p$, we can use the induction hypothesis (Property (P_q) for $q = p$) and conclude, for every $j \in \{0, \ldots, \ell\}$ and $\lambda_{p+1} \geq \lambda_p$, that

$$u_{k+j} \in \mathcal{B}_\tau(u^*) \quad \text{and} \quad \|u_{k+j} - \hat{u}_j^k\| \leq \lambda_p \|u_k - u^*\|^2 \leq \lambda_{p+1}\|u_k - u^*\|^2, \tag{155}$$

and, thus, (79) holds. In the remainder of the proof, we consider the case $\ell = p+1$. In this case $u_k \in \mathcal{B}_{\eta_{p+1}}(u^*)$, $\alpha_k \in [\alpha_{\inf}, \alpha_{\sup}]$, and

$$\|\hat{u}_j^k - u^*\| \geq \frac{1}{2}\|u_k - u^*\| \quad \text{for all } j \in \{0, \ldots, p\}, \tag{156}$$

and we need to verify that $u_{k+j} \in \mathcal{B}_\tau(u^*)$ for $j = \{1, \ldots, p+1\}$ and

$$\|u_{k+j+1} - \hat{u}_{j+1}^k\| \leq \lambda_{p+1}\|u_k - u^*\|^2 \text{ for all } j \in \{0, \ldots, p+1\}. \tag{157}$$

First, suppose that $u_{k+j} \in \mathcal{B}_\tau(u^*)$ for $j = 1, 2, \ldots, p$. By (75) and (77), we have

$$\begin{aligned}\|u_{k+p+1} - u^*\| &\leq \|u_{k+p} - u^*\| + \|\mathcal{S}_{k+p}\| \\ &\leq \|u_{k+p} - u^*\| + \frac{1}{|\alpha_{k+p}|}\|\mathcal{G}_{k+p}\| \leq (1 + \frac{\alpha_{\sup}}{\alpha_{\inf}})\|u_{k+p} - u^*\|,\end{aligned}$$

and, in a similar manner, it can be shown by induction that

$$\|u_{k+p+1} - u^*\| \leq (1 + \frac{\alpha_{\sup}}{\alpha_{\inf}})^{p+1}\|u_k - u^*\|.$$

Therefore, due to the definition of η_{p+1}, it follows that $u_\ell \in \mathcal{B}_\tau(u^*)$ for every $\ell \in \{1, 2, \dots, p+1\}$ and any $u_k \in \mathcal{B}_{\eta_{p+1}}(u^*)$ and $\alpha_k \in [\alpha_{\inf}, \alpha_{\sup}]$. It remains to verify (157). In fact, due to (156), the induction hypothesis, and the fact that $\eta_{q+1} \le \eta_q$, (157) holds for any arbitrary $\lambda_{p+1} \ge \lambda_p$ and $j \le p$. Hence, it suffices to show that

$$\|u_{k+p+1} - \hat{u}^k_{p+1}\| \le \lambda_{p+1}\|u_k - u^*\|^2 \tag{158}$$

for some $\lambda_{p+1} \ge \lambda_p$.

By using (2) and the triangle inequality, we obtain

$$\begin{aligned}\|u_{k+p+1} - \hat{u}^k_{p+1}\| &\le \|u_{k+p} - \frac{1}{\alpha_{k+p}}\mathcal{G}(u_{k+p}) - (\hat{u}^k_p - \frac{1}{\hat{\alpha}^k_p}\hat{\mathcal{G}}(\hat{u}^k_p))\| \\ &\le \|u_{k+p} - \hat{u}^k_p\| + \frac{1}{|\hat{\alpha}^k_p|}\|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| \\ &+ \left|\frac{1}{\alpha_{k+p}} - \frac{1}{\hat{\alpha}^k_p}\right| \|\mathcal{G}(u_{k+p})\|.\end{aligned} \tag{159}$$

From now on, we define c as a positive generic constant which depends only on τ, $\alpha_{\inf}$, $\alpha_{\sup}$ and m , but not on k, and the choice of α_k and $u_k \in \mathcal{B}_\tau(u^*)$. We shall show that the following inequalities hold

$$\frac{1}{|\hat{\alpha}^k_p|}\|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| \le c\|u_k - u^*\|^2, \tag{160}$$

$$\left|\frac{1}{\alpha_{k+p}} - \frac{1}{\hat{\alpha}^k_p}\right| \|\mathcal{G}(u_{k+p})\| \le c\|u_k - u^*\|^2. \tag{161}$$

Verification of inequality (160): First, by adding and subtracting $\hat{\mathcal{G}}(u_{k+p})$, using the triangle inequality, $L1$ and $L2$, we obtain

$$\begin{aligned}\|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| &\le \|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(u_{k+p})\| + \|\hat{\mathcal{G}}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| \\ &\le \|\mathcal{G}(u_{k+p}) - \mathcal{A}^{\mathcal{F}}_{u^*}(u_{k+p} - u^*)\| + \|\mathcal{A}^{\mathcal{F}}_{u^*}(u_{k+p} - \hat{u}^k_p)\| \\ &\le L\|u_{k+p} - u^*\|^2 + \alpha_{\sup}\|u_{k+p} - \hat{u}^k_p\| \le c\|u_k - u^*\|^2,\end{aligned} \tag{162}$$

where $\hat{\mathcal{G}}(u) = \mathcal{A}^{\mathcal{F}}_{u^*}(u - u^*)$. In the last estimate, we have used the induction hypothesis and the fact that

$$\begin{aligned}\|u_{k+p} - u^*\| &\le \|u_k - u^*\| + \sum_{i=0}^{p-1}\|\mathcal{S}_{k+i}\| \le \left(1 + \sum_{i=1}^{p}\left(\frac{\alpha_{\sup}}{\alpha_{\inf}}\right)^i\right)\|u_k - u^*\| \\ &\le \left(1 + \sum_{i=1}^{m}\left(\frac{\alpha_{\sup}}{\alpha_{\inf}}\right)^i\right)\|u_k - u^*\|.\end{aligned} \tag{163}$$

Now by using (162) and (76), we can infer that

$$\frac{1}{|\hat{\alpha}^k_p|}\|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| \le \frac{1}{\alpha_{\inf}}\|\mathcal{G}(u_{k+p}) - \hat{\mathcal{G}}(\hat{u}^k_p)\| \le c\|u_k - u^*\|^2.$$

Verification of inequality (161): Here we need only to show that

$$\left|\frac{1}{\alpha_{k+p}} - \frac{1}{\hat{\alpha}^k_p}\right| \le c\|u_k - u^*\|. \tag{164}$$

Then, thanks to (77), (163), and (164), we obtain

$$\left|\frac{1}{\alpha_{k+p}} - \frac{1}{\hat{\alpha}^k_p}\right| \|\mathcal{G}(u_{k+p})\| \le c\alpha_{\sup}\|u_k - u^*\|\|u_{k+p} - u^*\| \le c\|u_k - u^*\|^2.$$

which implies (161).

Due to (153) we have only these two cases :

1. $\hat{\alpha}^k_p = \hat{\alpha}^{BB1,k}_p$ and $\alpha_{k+p} = \alpha^{BB1}_{k+p}$.
2. $\hat{\alpha}^k_p = \hat{\alpha}^{BB2,k}_p$ and $\alpha_{k+p} = \alpha^{BB2}_{k+p}$.

We investigate the first case. Due to (3), we have

$$\frac{1}{\alpha_{k+p}} = \frac{(\mathcal{S}_{k+p-1}, \mathcal{S}_{k+p-1})}{(\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1})}, \text{ and } \frac{1}{\hat{\alpha}^k_p} = \frac{(\hat{\mathcal{S}}^k_{p-1}, \hat{\mathcal{S}}^k_{p-1})}{(\hat{\mathcal{S}}^k_{p-1}, \hat{\mathcal{Y}}^k_{p-1})}. \tag{165}$$

Due to (156) and the induction hypothesis i.e., property (P_q) for $q = p$, we have

$$\|\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}\| \leq \|u_{k+p} - \hat{u}^k_p\| + \|u_{k+p-1} - \hat{u}^k_{p-1}\| \leq 2\lambda_p \|u_k - u^*\|^2, \tag{166}$$

and, as a consequence, we obtain

$$\begin{aligned} \left| \|\mathcal{S}_{k+p-1}\|^2 - \|\hat{\mathcal{S}}^k_{p-1}\|^2 \right| &\leq \left| 2(\mathcal{S}_{k+p-1}, \mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}) - \|\hat{\mathcal{S}}^k_{p-1} - \mathcal{S}_{k+p-1}\|^2 \right| \\ &\leq c\|u_k - u^*\|^3. \end{aligned} \tag{167}$$

Further, by (76), (156), we have

$$\begin{aligned} \|\hat{\mathcal{S}}^k_{p-1}\| &= \frac{1}{|\hat{\alpha}^k_{p-1}|}\|\hat{\mathcal{G}}^k_{p-1}\| \geq \frac{1}{\alpha_{\sup}}\|\mathcal{A}^{\mathcal{F}}_{u^*}(\hat{u}^k_{p-1} - u^*)\| \geq \frac{\alpha_{\inf}}{\alpha_{\sup}}\|(\hat{u}^k_{p-1} - u^*)\| \\ &\geq \frac{\alpha_{\inf}}{2\alpha_{\sup}}\|(\hat{u}^k_0 - u^*)\| = \frac{\alpha_{\inf}}{2\alpha_{\sup}}\|(u_k - u^*)\|. \end{aligned} \tag{168}$$

From (167) and (168), it follows that

$$\left| 1 - \frac{\|\mathcal{S}_{k+p-1}\|^2}{\|\hat{\mathcal{S}}^k_{p-1}\|^2} \right| \leq c\|u_k - u^*\|. \tag{169}$$

Now observe that

$$\begin{aligned} (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1}) - (\hat{\mathcal{S}}^k_{p-1}, \hat{\mathcal{Y}}^k_{p-1}) &= (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1} - \hat{\mathcal{Y}}^k_{p-1}) + (\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}, \hat{\mathcal{Y}}^k_{p-1}) \\ &= (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1} - \hat{\mathcal{Y}}^k_{p-1}) + (\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}, \mathcal{A}^{\mathcal{F}}_{u^*}\hat{\mathcal{S}}^k_{p-1}). \end{aligned} \tag{170}$$

Using (166) and $L2$, we obtain

$$\begin{aligned} &\left| (\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}, \mathcal{A}^{\mathcal{F}}_{u^*}\hat{\mathcal{S}}^k_{p-1}) \right| \\ &= \left| (\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}, \mathcal{A}^{\mathcal{F}}_{u^*}\mathcal{S}_{k+p-1}) - (\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1}, \mathcal{A}^{\mathcal{F}}_{u^*}(\mathcal{S}_{k+p-1} - \hat{\mathcal{S}}^k_{p-1})) \right| \\ &\leq c\|u_k - u^*\|^3, \end{aligned} \tag{171}$$

and, by (162) and the induction hypothesis, we have

$$\left| (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1} - \hat{\mathcal{Y}}^k_{p-1}) \right| \leq \|\mathcal{S}_{k+p-1}\|(\|\mathcal{G}_{k+p} - \hat{\mathcal{G}}^k_p\| + \|\mathcal{G}_{k+p-1} - \hat{\mathcal{G}}^k_{p-1}\|) \leq c\|u_k - u^*\|^3. \tag{172}$$

Hence, using (170), (171), and (172), we have

$$\left| (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1}) - (\hat{\mathcal{S}}^k_{p-1}, \hat{\mathcal{Y}}^k_{p-1}) \right| \leq c\|u_k - u^*\|^3. \tag{173}$$

Moreover, by using $L2$, (75), (77), and the facts that $u_{k+p}, u_{k+p-1} \in \mathcal{B}_\tau(u^*)$ and $\alpha_k \leq \alpha_{\sup}$ for all $k \geq 1$, we can write that

$$\begin{aligned} (\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1}) &= (\mathcal{S}_{k+p-1}, \mathcal{G}_{k+p} - \mathcal{G}_{k+p-1}) \\ &\geq \alpha_{\inf}\|\mathcal{S}_{k+p-1}\|^2 = \alpha_{\inf}\left|\frac{1}{\alpha_{k+p-1}}\right|^2 \|\mathcal{G}_{k+p-1}\|^2 \\ &\geq \frac{\alpha_{\inf}}{\alpha^2_{\sup}}\|\mathcal{G}_{k+p-1}\|^2 = \frac{\alpha_{\inf}}{\alpha^2_{\sup}}\|\mathcal{G}(u_{k+p-1}) - \mathcal{G}(u^*)\|^2 \\ &\geq \frac{\alpha^3_{\inf}}{\alpha^2_{\sup}}\|u_{k+p-1} - u^*\|^2. \end{aligned} \tag{174}$$

Further, by $L2$, the definition of η_{p+1} in (154), (156) and (P_q) with $q = p$, we have

$$\begin{aligned} \|u_{k+p-1} - u^*\|^2 &\geq \frac{1}{2}\|\hat{u}^k_{p-1} - u^*\|^2 - \|u_{k+p-1} - \hat{u}^k_{p-1}\|^2 \\ &\geq \frac{1}{8}\|\hat{u}^k_0 - u^*\|^2 - \lambda^2_p\|u_k - u^*\|^4 \\ &\geq (\frac{1}{8} - \lambda^2_p\eta^2_{p+1})\|u_k - u^*\|^2 = \frac{1}{16}\|u_k - u^*\|^2, \end{aligned} \tag{175}$$

Combining (174) and (175) we have

$$(\mathcal{S}_{k+p-1}, \mathcal{Y}_{k+p-1}) \geq \frac{\alpha^3_{\inf}}{\alpha^2_{\sup}}\|u_{k+p-1} - u^*\|^2 \geq \frac{\alpha^3_{\inf}}{16\alpha^2_{\sup}}\|u_k - u^*\|^2. \tag{176}$$

From (173) and (176) we can write

$$\left|1-\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}\right|\leq c\|u_k-u^*\|. \tag{177}$$

Now, observe that by (165)

$$\begin{aligned}\left|\frac{1}{\alpha_{k+p}}-\frac{1}{\hat{\alpha}^k_p}\right| &\leq \left|\frac{(\mathcal{S}_{k+p-1},\mathcal{S}_{k+p-1})}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}-\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{S}}^k_{p-1})}{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}\right|\\ &= \frac{1}{|\hat{\alpha}^k_p|}\left|1-\left(\frac{(\mathcal{S}_{k+p-1},\mathcal{S}_{k+p-1})}{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{S}}^k_{p-1})}\right)\left(\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}\right)\right|\\ &\leq \frac{1}{\alpha_{\inf}}\left|1-\left(\frac{(\mathcal{S}_{k+p-1},\mathcal{S}_{k+p-1})}{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{S}}^k_{p-1})}\right)\left(\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}\right)\right|\\ &= \frac{1}{\alpha_{\inf}}\left|\phi_1(1-\phi_2)+\phi_2\right|\leq\frac{1}{\alpha_{\inf}}(|\phi_1|+|\phi_2|+|\phi_1\phi_2|),\end{aligned} \tag{178}$$

where

$$\phi_1:=1-\frac{(\mathcal{S}_{k+p-1},\mathcal{S}_{k+p-1})}{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{S}}^k_{p-1})}\quad\text{and}\quad\phi_2:=1-\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}. \tag{179}$$

By (169), (177), (178), and (179), we can infer that estimate (164) holds for the case that $\alpha_{k+p}=\alpha^{BB1}_{k+p}$ and $\hat{\alpha}^k_p=\hat{\alpha}^{BB1,k}_p$ are chosen.
Now we deal with the second case, i.e., $\alpha_{k+p}=\alpha^{BB2}_{k+p}$ and $\hat{\alpha}^k_p=\hat{\alpha}^{BB2,k}_p$. First due to (3), we have

$$\frac{1}{\alpha_{k+p}}=\frac{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}{(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1})},\ \text{and}\ \frac{1}{\hat{\alpha}^k_p}=\frac{(\hat{\mathcal{S}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}{(\hat{\mathcal{Y}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})}.$$

By using the fact that $u_{k+p},u_{k+p-1},\hat{u}^k_{p-1},\hat{u}^k_p\in\mathcal{B}_\tau(u^*)$, and the hypothesis of induction which is applicable due (156), we can write

$$\|\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1}\|\leq\|\mathcal{G}_{k+p}-\hat{\mathcal{G}}^k_p\|+\|\mathcal{G}_{k+p-1}-\hat{\mathcal{G}}^k_{p-1}\|\leq c\|u_k-u^*\|^2. \tag{180}$$

In addition, by using (77), (163), and the triangle inequality we obtain

$$\begin{aligned}\|\mathcal{Y}_{k+p-1}\|=\|\mathcal{G}_{k+p}-\mathcal{G}_{k+p-1}\|&\leq\|\mathcal{G}_{k+p}-\mathcal{G}(u^*)\|+\|\mathcal{G}_{k+p-1}-\mathcal{G}(u^*)\|\\ &\leq\alpha_{\sup}\left(\|u_{k+p}-u^*\|+\|u_{k+p-1}-u^*\|\right)\leq c\|u_k-u^*\|.\end{aligned} \tag{181}$$

From (180), (181), we deduce

$$\begin{aligned}\left|\|\mathcal{Y}_{k+p-1}\|^2-\|\hat{\mathcal{Y}}^k_{p-1}\|^2\right|&\leq\left|(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1})+(\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1},\hat{\mathcal{Y}}^k_{p-1})\right|\\ &\leq\|\mathcal{Y}_{k+p-1}\|\|\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1}\|+\|\hat{\mathcal{Y}}^k_{p-1}\|\|\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1}\|\\ &\leq c\|u_k-u^*\|^3,\end{aligned} \tag{182}$$

where in the last line we have used the fact that

$$\|\hat{\mathcal{Y}}^k_{p-1}\|\leq\|\mathcal{Y}_{k+p-1}-\hat{\mathcal{Y}}^k_{p-1}\|+\|\mathcal{Y}_{k+p-1}\|. \tag{183}$$

Furthermore, by using $L1$, (76), (168), and (183), we obtain

$$\|\hat{\mathcal{Y}}^k_{p-1}\|=\|\mathcal{A}^{\mathcal{F}}_{u^*}\hat{\mathcal{S}}^k_{p-1}\|\geq\alpha_{\inf}\|\hat{\mathcal{S}}^k_{p-1}\|\geq\frac{\alpha^2_{\inf}}{2\alpha_{\sup}}\|(u_k-u^*)\|, \tag{184}$$

and, as a consequence, it follows from (182) and (184) that

$$\left|1-\frac{\|\mathcal{Y}_{k+p-1}\|^2}{\|\hat{\mathcal{Y}}^k_{p-1}\|^2}\right|\leq c\|u_k-u^*\|. \tag{185}$$

Now similarly to the case for $BB1$, by (75) we can write

$$\begin{aligned}
\left|\frac{1}{\alpha_{k+p}}-\frac{1}{\hat{\alpha}_p^k}\right| &= \left|\frac{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}{(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1})}-\frac{(\hat{\mathcal{S}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}{(\hat{\mathcal{Y}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}\right| \\
&= \frac{1}{|\alpha_{k+p}|}\left|1-\left(\frac{(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1})}{(\hat{\mathcal{Y}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}\right)\left(\frac{(\hat{\mathcal{S}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}\right)\right| \\
&\le \frac{1}{\alpha_{\inf}}\left|1-\left(\frac{(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1})}{(\hat{\mathcal{Y}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}\right)\left(\frac{(\hat{\mathcal{S}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}\right)\right| \\
&= \frac{1}{\alpha_{\inf}}\left|\phi_1(1-\phi_2)+\phi_2\right| \le \frac{1}{\alpha_{\inf}}(|\phi_1|+|\phi_2|+|\phi_1\phi_2|),
\end{aligned} \tag{186}$$

where

$$\phi_1 := 1-\frac{(\mathcal{Y}_{k+p-1},\mathcal{Y}_{k+p-1})}{(\hat{\mathcal{Y}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)} \quad \text{and} \quad \phi_2 := 1-\frac{(\hat{\mathcal{S}}_{p-1}^k,\hat{\mathcal{Y}}_{p-1}^k)}{(\mathcal{S}_{k+p-1},\mathcal{Y}_{k+p-1})}. \tag{187}$$

By (177), (185), (186), and (187), we can infer that (164) holds for the case $BB2$.

Hence, we are finished with the verification of (161). Now from (159), (160), and (161), estimate (158) follows and, thus, the property (P_q) holds for $q = p+1$. Since m is fixed and finite, we can choose λ and η independent of k and ℓ, and, thus the proof is complete.

References

1. Allgower, E.L., Böhmer, K., Potra, F.A., Rheinboldt, W.C.: A mesh-independence principle for operator equations and their discretizations. SIAM J. Numer. Anal. **23**(1), 160–169 (1986). DOI 10.1137/0723011. URL `http://dx.doi.org/10.1137/0723011`
2. Alt, W.: Mesh-independence of the Lagrange-Newton method for nonlinear optimal control problems and their discretizations. Ann. Oper. Res. **101**, 101–117 (2001). DOI 10.1023/A:1010912305365. URL `http://dx.doi.org/10.1023/A:1010912305365`. Optimization with data perturbations, II
3. Argyros, I.K.: A mesh-independence principle for nonlinear operator equations and their discretizations under mild differentiability conditions. Computing **45**(3), 265–268 (1990). URL `https://doi.org/10.1007/BF02250637`
4. Argyros, I.K.: On a mesh-independence principle for operator equations and the secant method. Acta Math. Hungar. **60**(1-2), 7–19 (1992). URL `https://doi.org/10.1007/BF00051753`
5. 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). DOI 10.1081/NFA-120006694. URL `http://dx.doi.org/10.1081/NFA-120006694`
6. Axelsson, O., Karátson, J.: Mesh independent superlinear PCG rates via compact-equivalent operators. SIAM J. Numer. Anal. **45**(4), 1495–1516 (2007). DOI 10.1137/06066391X. URL `http://dx.doi.org/10.1137/06066391X`
7. Azmi, B., Kunisch, K.: On the Stabilizability of the Burgers Equation by Receding Horizon Control. SIAM J. Control Optim. **54**(3), 1378–1405 (2016). DOI 10.1137/15M1030352. URL `http://dx.doi.org/10.1137/15M1030352`
8. Azmi, B., Kunisch, K.: Receding horizon control for the stabilization of the wave equation. Discrete Contin. Dyn. Syst. **38**(2), 449–484 (2018). URL `https://doi.org/10.3934/dcds.2018021`
9. Azmi, B., Kunisch, K., Boulanger, A.C.: On the semi-global stabilizability of the Korteweg-de Vries equation via model predictive control. ESAIM: Control, Optimisation and Calculus of Variations (2018). URL `https://doi.org/10.1051/cocv/2017001`. (in press)
10. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**(1), 141–148 (1988). DOI 10.1093/imanum/8.1.141. URL `http://dx.doi.org/10.1093/imanum/8.1.141`
11. Berggren, M.: Approximations of very weak solutions to boundary-value problems. SIAM J. Numer. Anal. **42**(2), 860–877 (2004). URL `https://doi.org/10.1137/S0036142903382048`
12. Bonettini, S., Benfenati, A., Ruggiero, V.: Scaling techniques for ϵ-subgradient methods. SIAM J. Optim. **26**(3), 1741–1772 (2016). URL `https://doi.org/10.1137/14097642X`
13. Bonettini, S., Loris, I., Porta, F., Prato, M., Rebegoldi, S.: On the convergence of a linesearch based proximal-gradient method for nonconvex optimization. Inverse Problems **33**(5), 055005, 30 (2017). URL `https://doi.org/10.1088/1361-6420/aa5bfd`
14. Bramble, J.H., Pasciak, J.E., Schatz, A.H.: The construction of preconditioners for elliptic problems by substructuring. I. Math. Comp. **47**(175), 103–134 (1986). URL `https://doi.org/10.2307/2008084`
15. Brenner, S.C., Scott, L.R.: The mathematical theory of finite element methods, *Texts in Applied Mathematics*, vol. 15. Springer-Verlag, New York (1994). URL `https://doi.org/10.1007/978-1-4757-4338-8`
16. Casas, E., Raymond, J.P.: Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. SIAM J. Control Optim. **45**(5), 1586–1611 (2006). URL `https://doi.org/10.1137/050626600`
17. Chen, P., Gui, C.: Linear convergence analysis of the use of gradient projection methods on total variation problems. Comput. Optim. Appl. **54**(2), 283–315 (2013). URL `https://doi.org/10.1007/s10589-011-9412-4`

18. Chow, S.N., Yin, K., Zhou, H.M., Behrooz, A.: Solving inverse source problems by the Orthogonal Solution and Kernel Correction Algorithm (OSKCA) with applications in fluorescence tomography. Inverse Probl. Imaging **8**(1), 79–102 (2014). URL https://doi.org/10.3934/ipi.2014.8.79
19. Conway, J.B.: A course in functional analysis, *Graduate Texts in Mathematics*, vol. 96, second edn. Springer-Verlag, New York (1990)
20. Curtis, F.E., Guo, W.: Handling nonpositive curvature in a limited memory steepest descent method. IMA J. Numer. Anal. **36**(2), 717–742 (2016). URL https://doi.org/10.1093/imanum/drv034
21. Dai, Y., Yuan, J., Yuan, Y.X.: Modified two-point stepsize gradient methods for unconstrained optimization. Comput. Optim. Appl. **22**(1), 103–109 (2002). URL https://doi.org/10.1023/A:1014838419611
22. Dai, Y.H., Fletcher, R.: On the asymptotic behaviour of some new gradient methods. Math. Program. **103**(3, Ser. A), 541–559 (2005). URL https://doi.org/10.1007/s10107-004-0516-9
23. Dai, Y.H., Fletcher, R.: Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. Numer. Math. **100**(1), 21–47 (2005). URL https://doi.org/10.1007/s00211-004-0569-y
24. 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). DOI 10.1093/imanum/drl006. URL http://dx.doi.org/10.1093/imanum/drl006
25. 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). DOI 10.1093/imanum/22.1.1. URL http://dx.doi.org/10.1093/imanum/22.1.1
26. 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). URL https://doi.org/10.1007/s10589-014-9669-5
27. van den Doel, K., Ascher, U.: The chaotic nature of faster gradient descent methods. J. Sci. Comput. **51**(3), 560–581 (2012). URL https://doi.org/10.1007/s10915-011-9521-3
28. Fletcher, R.: On the Barzilai-Borwein method. In: Optimization and control with applications, *Appl. Optim.*, vol. 96, pp. 235–256. Springer, New York (2005). URL https://doi.org/10.1007/0-387-24255-4_10
29. Fletcher, R.: A limited memory steepest descent method. Math. Program. **135**(1-2, Ser. A), 413–436 (2012). URL https://doi.org/10.1007/s10107-011-0479-6
30. Gao, B., Liu, X., Chen, X., Yuan, Y.x.: A New First-Order Algorithmic Framework for Optimization Problems with Orthogonality Constraints. SIAM J. Optim. **28**(1), 302–332 (2018). URL https://doi.org/10.1137/16M1098759
31. Glowinski, R., Lapin, S.: Iterative solution of linear variational problems in Hilbert spaces: some conjugate gradients success stories. In: Conjugate gradient algorithms and finite element methods, Sci. Comput., pp. 223–245. Springer, Berlin (2004)
32. Goldfarb, D., Wen, Z., Yin, W.: A curvilinear search method for p-harmonic flows on spheres. SIAM J. Imaging Sci. **2**(1), 84–109 (2009). URL https://doi.org/10.1137/080726926
33. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. **23**(4), 707–716 (1986). URL https://doi.org/10.1137/0723046
34. Grippo, L., Sciandrone, M.: Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. Comput. Optim. Appl. **23**(2), 143–169 (2002). URL https://doi.org/10.1023/A:1020587701058
35. Grisvard, P.: Elliptic problems in nonsmooth domains, *Monographs and Studies in Mathematics*, vol. 24. Pitman (Advanced Publishing Program), Boston, MA (1985)
36. Grisvard, P.: Singularities in boundary value problems, *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*, vol. 22. Masson, Paris; Springer-Verlag, Berlin (1992)
37. Hall, B.C.: Quantum theory for mathematicians, *Graduate Texts in Mathematics*, vol. 267. Springer, New York (2013). DOI 10.1007/978-1-4614-7116-5. URL http://dx.doi.org/10.1007/978-1-4614-7116-5
38. Heinkenschloss, M.: Mesh independence for nonlinear least squares problems with norm constraints. SIAM J. Optim. **3**(1), 81–117 (1993). DOI 10.1137/0803005. URL http://dx.doi.org/10.1137/0803005
39. 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). DOI 10.1137/140973050. URL http://dx.doi.org/10.1137/140973050
40. Hintermüller, M., Tröltzsch, F., Yousept, I.: Mesh-independence of semismooth Newton methods for Lavrentiev-regularized state constrained nonlinear optimal control problems. Numer. Math. **108**(4), 571–603 (2008). URL https://doi.org/10.1007/s00211-007-0134-6
41. Hintermüller, M., Ulbrich, M.: A mesh-independence result for semismooth Newton methods. Math. Program. **101**(1, Ser. B), 151–184 (2004). URL https://doi.org/10.1007/s10107-004-0540-9
42. Hinze, M., Kunisch, K.: Second order methods for optimal control of time-dependent fluid flow. SIAM J. Control Optim. **40**(3), 925–946 (2001). URL https://doi.org/10.1137/S0363012999361810
43. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints, *Mathematical Modelling: Theory and Applications*, vol. 23. Springer, New York (2009)
44. Huang, Y., Liu, H.: Smoothing projected Barzilai-Borwein method for constrained non-Lipschitz optimization. Comput. Optim. Appl. **65**(3), 671–698 (2016). URL https://doi.org/10.1007/s10589-016-9854-9
45. Jiang, B., Dai, Y.H.: A framework of constraint preserving update schemes for optimization on Stiefel manifold. Math. Program. **153**(2, Ser. A), 535–575 (2015). URL https://doi.org/10.1007/s10107-014-0816-7
46. Jiang, B., Liu, Y.F., Wen, Z.: L_p-norm regularization algorithms for optimization over permutation matrices. SIAM J. Optim. **26**(4), 2284–2313 (2016). URL https://doi.org/10.1137/15M1048021
47. Kalousek, Z.e.: Steepest descent method with random step lengths. Found. Comput. Math. **17**(2), 359–422 (2017). URL https://doi.org/10.1007/s10208-015-9290-8
48. Karátson, J.: Characterizing mesh independent quadratic convergence of Newton's method for a class of elliptic problems. SIAM J. Math. Anal. **44**(3), 1279–1303 (2012). URL https://doi.org/10.1137/100817589
49. Kelley, C.T., Sachs, E.W.: Quasi-Newton methods and unconstrained optimal control problems. SIAM J. Control Optim. **25**(6), 1503–1516 (1987). DOI 10.1137/0325083. URL http://dx.doi.org/10.1137/0325083
50. Kelley, C.T., Sachs, E.W.: Approximate quasi-Newton methods. Math. Programming **48**(1, (Ser. B)), 41–70 (1990). URL https://doi.org/10.1007/BF01582251

51. Kelley, C.T., Sachs, E.W.: Mesh independence of Newton-like methods for infinite-dimensional problems. J. Integral Equations Appl. **3**(4), 549–573 (1991). URL https://doi.org/10.1216/jiea/1181075649
52. Kelley, C.T., Sachs, E.W.: Mesh independence of the gradient projection method for optimal control problems. SIAM J. Control Optim. **30**(2), 477–493 (1992). DOI 10.1137/0330029. URL http://dx.doi.org/10.1137/0330029
53. 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). DOI 10.1137/090766541. URL http://dx.doi.org/10.1137/090766541
54. Kutyniok, G., Mehrmann, V., Petersen, P.C.: Regularization and numerical solution of the inverse scattering problem using shearlet frames. J. Inverse Ill-Posed Probl. **25**(3), 287–309 (2017). URL https://doi.org/10.1515/jiip-2015-0048
55. 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). DOI 10.1007/BF01765322. URL http://dx.doi.org/10.1007/BF01765322
56. 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). DOI 10.1016/0022-0396(91)90106-J. URL http://dx.doi.org/10.1016/0022-0396(91)90106-J
57. Laumen, M.: Newton's mesh independence principle for a class of optimal shape design problems. SIAM J. Control Optim. **37**(4), 1070–1088 (1999). URL https://doi.org/10.1137/S0363012996303529
58. 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)
59. 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 P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181
60. Liu, C.S.: A dynamical Tikhonov regularization for solving ill-posed linear algebraic systems. Acta Appl. Math. **123**, 285–307 (2013). URL https://doi.org/10.1007/s10440-012-9766-3
61. Liu, H., Li, X.: Modified subspace Barzilai-Borwein gradient method for non-negative matrix factorization. Comput. Optim. Appl. **55**(1), 173–196 (2013). URL https://doi.org/10.1007/s10589-012-9507-6
62. Liu, W., Dai, Y.H.: Minimization algorithms based on supervisor and searcher cooperation. J. Optim. Theory Appl. **111**(2), 359–379 (2001). DOI 10.1023/A:1011986402461. URL http://dx.doi.org/10.1023/A:1011986402461
63. Lu, Z.: Iterative hard thresholding methods for l_0 regularized convex cone programming. Math. Program. **147**(1-2, Ser. A), 125–154 (2014). URL https://doi.org/10.1007/s10107-013-0714-4
64. Lu, Z.: Iterative reweighted minimization methods for l_p regularized unconstrained nonlinear programming. Math. Program. **147**(1-2, Ser. A), 277–307 (2014). URL https://doi.org/10.1007/s10107-013-0722-4
65. Maaß, P., Strehlow, R.: An iterative regularization method for nonlinear problems based on Bregman projections. Inverse Problems **32**(11), 115013, 20 (2016). URL https://doi.org/10.1088/0266-5611/32/11/115013
66. Mardal, K.A., Winther, R.: Preconditioning discretizations of systems of partial differential equations. Numer. Linear Algebra Appl. **18**(1), 1–40 (2011). DOI 10.1002/nla.716. URL http://dx.doi.org/10.1002/nla.716
67. Molina, B., Raydan, M.: Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. Numer. Algorithms **13**(1-2), 45–60 (1996). DOI 10.1007/BF02143126. URL http://dx.doi.org/10.1007/BF02143126
68. 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). DOI 10.1137/S0363012903431177. URL http://dx.doi.org/10.1137/S0363012903431177
69. Prieto, K., Dorn, O.: Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. Inverse Problems **33**(1), 014001, 28 (2017). URL https://doi.org/10.1088/0266-5611/33/1/014001
70. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. **13**(3), 321–326 (1993). DOI 10.1093/imanum/13.3.321. URL http://dx.doi.org/10.1093/imanum/13.3.321
71. Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. **7**(1), 26–33 (1997). DOI 10.1137/S1052623494266365. URL http://dx.doi.org/10.1137/S1052623494266365
72. Rebegoldi, S., Bautista, L., Blanc-Féraud, L., Prato, M., Zanni, L., Plata, A.: A comparison of edge-preserving approaches for differential interference contrast microscopy. Inverse Problems **33**(8), 085009, 24 (2017). URL https://doi.org/10.1088/1361-6420/aa790a
73. Saxenhuber, D., Ramlau, R.: A gradient-based method for atmospheric tomography. Inverse Probl. Imaging **10**(3), 781–805 (2016). URL https://doi.org/10.3934/ipi.2016021
74. Simon, J.: Compact sets in the space $L^p(0,T;B)$. Ann. Mat. Pura Appl. (4) **146**, 65–96 (1987). DOI 10.1007/BF01762360. URL http://dx.doi.org/10.1007/BF01762360
75. Triggiani, R.: Global exact controllability on $H^1_{\Gamma_0}(\Omega) \times L_2(\Omega)$ of semilinear wave equations with Neumann $L_2(0,T;L_2(\Gamma_1))$-boundary control. In: Control theory of partial differential equations, *Lect. Notes Pure Appl. Math.*, vol. 242, pp. 273–336. Chapman & Hall/CRC, Boca Raton, FL (2005). DOI 10.1201/9781420028317.ch19. URL http://dx.doi.org/10.1201/9781420028317.ch19
76. 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). DOI 10.1051/cocv:2001127. URL http://dx.doi.org/10.1051/cocv:2001127
77. Volkwein, S.: Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces. SIAM J. Control Optim. **38**(3), 767–785 (2000). DOI 10.1137/S0363012998334468. URL http://dx.doi.org/10.1137/S0363012998334468
78. Volkwein, S.: Distributed control problems for the Burgers equation. Comput. Optim. Appl. **18**(2), 115–140 (2001). DOI 10.1023/A:1008770404256. URL http://dx.doi.org/10.1023/A:1008770404256
79. Wang, C., Zhou, T.: On iterative algorithms for quantitative photoacoustic tomography in the radiative transport regime. Inverse Problems **33**(11), 115006, 25 (2017). URL https://doi.org/10.1088/1361-6420/aa89c5

80. Weiser, M., Schiela, A., Deuflhard, P.: Asymptotic mesh independence of Newton's method revisited. SIAM J. Numer. Anal. **42**(5), 1830–1845 (2005). URL https://doi.org/10.1137/S0036142903434047
81. Wen, Z., Yin, W., Goldfarb, D., Zhang, Y.: A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation. SIAM J. Sci. Comput. **32**(4), 1832–1857 (2010). URL https://doi.org/10.1137/090747695
82. Werner, D.: Funktionalanalysis, extended edn. Springer-Verlag, Berlin (2000)
83. Wu, L., Sun, Z.: A new spectral method for ℓ_1-regularized minimization. Inverse Probl. Imaging **9**(1), 257–272 (2015). URL https://doi.org/10.3934/ipi.2015.9.257
84. Yang, Y., Dong, B., Wen, Z.: Randomized algorithms for high quality treatment planning in volumetric modulated arc therapy. Inverse Problems **33**(2), 025007, 22 (2017). URL https://doi.org/10.1088/1361-6420/aa5175
85. Yuan, Y.x.: A new stepsize for the steepest descent method. J. Comput. Math. **24**(2), 149–156 (2006)
86. Zhao, N., Yang, Q., Liu, Y.: Computing the generalized eigenvalues of weakly symmetric tensors. Comput. Optim. Appl. **66**(2), 285–307 (2017). URL https://doi.org/10.1007/s10589-016-9865-6
87. 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). URL https://doi.org/10.1007/s10957-016-1008-9