ON REGULARIZED STRUCTURE-EXPLOITING QUASI-NEWTON METHODS FOR INVERSE PROBLEMS

RAPHAEL KUESS AND ANDREA WALTHER

ABSTRACT. This paper introduces a regularized, structure-exploiting Powell-Symmetric-Broyden (RSE-PSB) method under modified secant conditions for solving ill-posed inverse problems in both infinite dimensional and finite dimensional settings. The approximation of the symmetric, yet potentially indefinite, second-order term, which is neglected by standard Levenberg-Marquardt (LM) approaches, integrates regularization and structure exploitation directly within the Quasi-Newton (QN) framework, leveraging the strengths of QN and LM methods, Tikhonov-type regularization, and structure exploitation. We establish local Q-linear convergence via the bounded deterioration principle and prove local Q-super-linear convergence under the assumption that the initial error is a Hilbert-Schmidt operator. Furthermore, we present a globalization strategy in the discretized setting based on the dynamic control of the regularization parameter. Hence, this approach stabilizes the ill-posed problem while ensuring global convergence, addressing even the choice of an appropriate regularization parameter. Finally, we discuss a numerical example based on a PDE-driven parameter identification problem in piezoelectricity, relevant to industrial sensor and actuator applications.

1. INTRODUCTION

Inverse problems usually aim to determine an unknown cause that leads to a known effect or consequence. They occur in numerous real-world applications. As the unknown cause often appears in a state space system that serves as the underlying mathematical model, several types of inverse problems can be categorized.

- Inverse source problems: The objective is to determine an unknown source term in a PDE based on observational data. This kind of problem appears often in fields such as physics and biomedical imaging, where identifying the origin of signals or forces is critical for understanding the underlying phenomena.
- Inverse boundary value problems: The task is to determine unknown boundary conditions from measurements or observations. Such problems often arise in geophysics and engineering where characterization of boundary interactions is essential for modeling the behavior, such as in heat transfer analysis in which surface temperatures and fluxes need to be identified.
- Parameter identification problems: The goal is to reconstruct unknown parameters within a mathematical model such that they best fit the observed or measured data. This type of problem frequently arises in engineering and materials science, where, e.g., material parameters such as diffusion coefficients must be estimated from experimental data.

Note that this categorization should not be understood too strictly, as combinations also often occur, e.g., one wants to identify a parameter and a source term. To tackle these types of problems, an associated, well-posed forward problem is needed, i.e., a mathematical model that is uniquely solvable and stable, meaning that small changes in the cause lead to small changes in the effect. In contrast, inverse problems are often ill-posed, meaning that solutions may not exist, may not be unique, or may not depend continuously on the observed data. As a result, inverse problems are highly sensitive to measurement errors and noise. Therefore, regularization techniques are required to stabilize the problem and achieve well-posedness. Specifically, inverse problems aim at solving an operator equation

$$F(x) = y^\delta \tag{1.1}$$

Date: December 30, 2025.
2020 *Mathematics Subject Classification.* 49M15, 90C53, 47J06, 65J20, 46N10.
Key words and phrases. Newton-type methods, iterative regularization, globalization, convergence analysis, inverse problems.

to determine x for given data y^δ. These problems are in practice often approached via iterative regularization methods that refine an approximate solution from an initial guess. For that purpose, they apply regularization iteratively in order to stabilize the problem and to achieve a reduction of the effects of noise. If we assume that we have access to the first derivative of the forward operator F, the Levenberg-Marquardt method (LM), i.e.,

$$(F'(x_k)^* F'(x_k) + \alpha_k I)\, s_k = -F'(x_k)^* \left(F(x_k) - y^\delta\right), \tag{1.2}$$

where

$$x_{k+1} := x_k + s_k, \tag{1.3}$$

is a well-known example and widely used approach based on linearizing the operator equation to be solved and iteratively applying Tikhonov regularization. Therein, $\alpha_k > 0$ for all $k \in \mathbb{N}$ denotes the regularization parameter. In general, despite its effectiveness for a wide range of inverse problems, this method is sensitive to the deviation of the initial guess, but can be globalized, see e.g., [6]. The LM method can also be viewed as a QN method applied to

$$\frac{1}{2}\left\|F(x) - y^\delta\right\|_Y^2 + \frac{\alpha_k}{2}\|x - x_k\|_X^2, \tag{1.4}$$

where x_k is the last obtained iterate or the linearization point, respectively. From this perspective, the approximation of the Hessian in the LM method may be insufficiently accurate as the approximation error is only small if F is linear or the residual is sufficiently small, i.e., we have a good approximation and a good initial guess. In our setting F can also be non-linear. Additionally, in real-world applications obtaining a suitable initial guess can be challenging. Moreover, second-order derivatives can be interpreted as measures of cross-sensitivity, providing valuable insight into the interactions between different components of the object to be reconstructed. This information is particularly important in multi-parameter identification problems. Since Newton-type methods utilize gradient information to efficiently approximate the Hessian, they typically exhibit fast and robust convergence while inherently capturing approximated cross-sensitivity information. Thus, we want to derive an approximation for the second order term $(F''(x_k)(\cdot))^* \left(F(x_k) - y^\delta\right)$, using Quasi-Newton (QN) methods. This can then be seen as a "corrected" LM method or a regularized structure-exploiting (RSE) QN approach, respectively. Motivated by finite dimensional Hilbert spaces, we assume that the second Fréchet derivative of the functional defined by (1.4) is symmetric at the minimizer. Since the first-order term in the Hessian operator is symmetric, we obtain that the second-order term must be symmetric at the minimizer, but not necessarily positive. Thus, we need QN methods that converge without assuming positivity of the operators, while still forcing symmetry. One method satisfying these requirements is the symmetric-rank-1 update, which needs a well-definedness condition to prevent the denominator from being zero. Hence, this method can be numerically unstable. To overcome this drawback, we want to focus on a rank-2 update, satisfying the restrictions, namely the Powell-symmetric Broyden update. Furthermore, the key property for achieving a super-linear convergence rate in infinite dimensional Hilbert spaces is to have a QN method that can be expressed as the solution to a variational problem. Broyden's method can be used according to [2], but it is not guaranteed to produce symmetric operators. It was shown in [30] that the PSB method has this property as well, which again motivates the choice of this method.

Related Work. QN methods for solving well-posed optimization problems have been extensively studied over the last five decades. Among the most prominent examples are the Broyden-Fletcher-Goldfarb-Shanno (BFGS), the symmetric rank-one (SR1), the Powell-symmetric Broyden (PSB) and the Davidon-Fletcher-Powell (DFP) methods, which all belong to the class of symmetric secant updates considered, e.g., by Dennis [9] in 1971. As a matter of fact, exploiting the structure of the problem and developing theory for structured secant methods to construct a more efficient approximation to the Hessian can be traced back to Dennis and Walker [11] in 1981 and has since evolved, see e.g., [10, 12, 21, 31].
A comprehensive study of structure-exploiting QN methods in finite dimensional spaces is given in [14]. In this paper, local and super-linear convergence of QN methods from the convex Broyden class is proved with partially known Hessian matrices. Additionally, the application to least squares problems is addressed.

In [20], the analysis of structure exploitation in finite dimensional least squares problems is further investigated in detail. Furthermore, in [7] structure exploitation of QN methods in the context of compact representations, especially of BFGS, is discussed.
Regarding inverse problems, structure-exploiting QN methods have been applied to finite dimensional settings, see e.g., [19], where an electromagnetic inverse problem was decomposed in a data discrepancy term and a regularization term. Similarly, [26] extends this decomposition to infinite dimensional Hilbert spaces using L-BFGS updates.
Unstructured QN methods for infinite dimensional problems have also been explored. For example, [30] derives QN formulas such as BFGS, DFP, SR1, and PSB via variational frameworks, while [32, 33] analyze their local convergence via bounded deterioration principles and discuss the application on well-posed inverse problems. Further insights into Hilbert-space applications and convergence are given in [4, 5].
The publications that are closest to this article are [25] and [23]. Firstly, in [25] a Kantorovich theorem and local convergence of a structured PSB method in Hilbert spaces are discussed. Secondly, in [23] a globalization approach for unstructured bounded QN Hessian approximations in a finite dimensional setting is presented as well as the compact representation of QN matrices, particularly the compact representation of the PSB method is given. However, our setting differs slightly and has weaker assumptions for convergence than in [25], and [23] does not consider possibly unbounded QN Hessian approximations. Additionally, in [34], a hybrid Gauss–Newton structured BFGS scheme for non-linear least squares problems is proposed, and global convergence of the resulting method is established. The approach combines curvature information from the Gauss–Newton model with a BFGS-type update in order to better capture the structure inherent in least squares formulations. A comprehensive analysis of iterative regularization methods for nonlinear ill-posed operator equations, including Broyden's method, stability issues and regularization aspects of quasi-Newton updates can be found in [22].

Contribution. Consequently, the first aim of this paper is to extend convergence results of Newton-type methods in Hilbert spaces and to generalize the previous theory by introducing a Tikhonov-type regularizer. The latter is used additively in the target functional. By deriving a modified secant equation in the Hilbert space setting, we propose the regularized structure-exploiting PSB (RSE-PSB) method. A key aspect is the specific choice of the PSB update, which allows maintaining symmetry of the approximation of the second-order term $(F''(x_k)(\cdot))^*(F(x_k) - y^\delta)$ without requiring positive definiteness. Then, local convergence results of [25] are extended to an adjusted framework, to weaker assumptions and to the inclusion of regularization, yielding that the regularized structure-exploiting PSB method satisfies the Bounded Deterioration Principle (BDP) and thus is locally Q-linear convergent in Hilbert spaces. Furthermore, under the assumption that the initial error is a Hilbert-Schmidt operator and using the variational characterization of the PSB update as the solution to a minimum norm problem in the Hilbert-Schmidt norm, we establish local Q-super-linear convergence. Moreover, the relationship to the Levenberg-Marquardt (LM) method is discussed, interpreting the proposed method as a "corrected" LM approach that captures cross-sensitivity information.
The second aim of this paper is to address globalization of the proposed methods motivated by possibly large deviations of the initial guess to the ground truth in the application of the proposed algorithm. Therefore, we established a global convergence result for the regularized structure-exploiting PSB method by controlling the regularization parameter within a trust-region-like framework. This globalization approach is presented in the discretized setting and is based on the results of [23]. Finally, we demonstrate the efficiency of the method using a discretize-then-optimize approach with Algorithmic Differentiation (AD) on a parameter identification problem for a piezoelectric specimen.
To the best of our knowledge, this method has not been addressed in the literature so far. Previous studies have primarily focused on finite dimensional cases or on structure exploitation via Hessian approximation of the data discrepancy term without the specific regularized framework presented here.

Structure of the paper. The second section deals with Newton-type methods in a general Hilbert space setting, providing the necessary preliminaries and auxiliary results for the PSB method, including the Bounded Deterioration Principle. In the third section, the regularized structure-exploiting PSB (RSE-PSB) method is introduced. Therein, the derivation of a modified secant equation is given. Subsequently, the local RSE-PSB algorithm is analyzed, establishing both local Q-linear and Q-super-linear convergence to a minimizer. In the fourth section, a globalization approach in a finite dimensional environment is provided to account for larger deviations of the initial estimate to the ground truth. Then, the Global Regularized structure-exploiting algorithm (GRSE) is presented and analyzed with respect to its global convergence. Numerical experiments, applied to a parameter identification problem in piezoelectricity, are shown in the fifth section to validate both local and global performance. The final section briefly summarizes the contributions of this paper.

2. Newton-type methods in Hilbert spaces

Let X and Y be separable Hilbert spaces over $\mathbb{R}$ and let $U \subseteq X$ be open and convex. Consider an objective functional $J : X \to \mathbb{R}$, which has to be minimized. Hence, Newton's method reads as

$$J''(x_k)(x_{k+1} - x_k) = -J'(x_k).$$

Here we used the Riesz isomorphism to identify the first Fréchet derivative $J'(x_k) \in X$ and the second Fréchet derivative or Hessian operator, respectively, $J''(x_k) \in L(X)$, where $L(X)$ denotes the space of linear and bounded operators from X to X. As in general the second Fréchet derivatives are numerically very costly to compute, approximations of the Hessian operator may be used. For this purpose, we introduce the Newton-type method

$$B_k^J s_k = -J'(x_k) \tag{2.1}$$

with $B_k^J \in L(X)$. If B_k^J is chosen such that it approximates $J''(x_k)$, then we obtain a Newton-type method and if it equals $J''(x_k)$ then we obtain Newton's method. To discuss convergence, we use the norm in X to define

$$\mu_k := \max\{\|x^* - x_k\|, \|x^* - x_{k+1}\|\}. \tag{2.2}$$

Furthermore, if not indicated otherwise, the standard operator norm is used for operators in $L(X)$ throughout this paper. Motivated by Lemma 3.2 in [33] we state the local Q-linear convergence result.

Theorem 2.1 (Local Q-linear convergence of Newton-type methods)**.** *Let $x^* \in U$ be a minimizer of J and let B_0^J be boundedly invertible. Furthermore, suppose J'' is Lipschitz continuous with Lipschitz constant $L_{J''} > 0$. Let $\nu \in (0, 1)$ be arbitrarily given and assume that B_0^J and x_0 satisfy*

$$\|x_0 - x^*\| \le \xi, \quad \|B_0^J - J''(x^*)\| \le a, \tag{2.3}$$

where $\xi = \xi(\nu)$ and $a = a(\nu)$ are constants dependent on ν. If the operator sequence $\{B_k^J\}_{k\in\mathbb{N}}$ is chosen such that (2.1) *and the bounded deterioration principle (BDP), i.e.,*

$$\|B_{k+1}^J - J''(x^*)\| \le \|B_k^J - J''(x^*)\| + c\mu_k, \tag{2.4}$$

are satisfied, where $c > 0$ is a constant, then the sequence $\{x_k\}$ is well-defined and converges Q-linearly to x^*. Furthermore, the sequences $\{\|B_k^J\|\}$ and $\{\|(B_k^J)^{-1}\|\}$ are uniformly bounded.

Proof. By assumption, there exists $b > 0$ such that

$$\|(B_0^J)^{-1}\| \le b. \tag{2.5}$$

Using induction we will show for all $k \in \mathbb{N}$ that B_k^J is boundedly invertible and that

$$\|B_k^J - J''(x^*)\| \le a + \frac{c\xi(1 - \nu^k)}{1 - \nu}. \tag{2.6}$$

Since U is open and convex, and $x^* \in U$, for any $\nu \in (0, 1)$, one can choose $\xi = \xi(\nu) > 0$ and $a = a(\nu) > 0$ satisfying the following inequalities

$$b\left(\frac{L_{J''}\xi}{2} + a + \frac{c\xi}{1 - \nu}\right) < \nu, \tag{2.7}$$

$$2ab + \frac{c\xi}{1 - \nu} b < 1. \tag{2.8}$$

For the base case of the induction, we will now show that B_1^J is boundedly invertible and that

$$\|B_1^J - J''(x^*)\| \le a + \frac{c\xi(1 - \nu^1)}{1 - \nu} = a + c\xi.$$

Assumptions (2.3) and (2.4) yield

$$\|B_1^J - J''(x^*)\| \le a + c\mu_0. \tag{2.9}$$

By the mean value theorem and the Lipschitz continuity of J'' we have that

$$\begin{aligned}
\|x_1 - x^*\| &= \|x_1 - x_0 + x_0 - x^*\| = \| - (B_0^J)^{-1} J'(x_0) + x_0 - x^*\| \\
&= \left\|(B_0^J)^{-1} \left\{-[J'(x_0) - J'(x^*) - J''(x^*)(x_0 - x^*)] + [B_0^J - J''(x^*)](x_0 - x^*)\right\}\right\| \\
&\le \left\|(B_0^J)^{-1}\right\| \left\{\|J'(x_0) - J'(x^*) - J''(x^*)(x_0 - x^*)\| + \left\|[B_0^J - J''(x^*)](x_0 - x^*)\right\|\right\} \\
&\le b\,\|J'(x_0) - J'(x^*) - J''(x^*)(x_0 - x^*)\| + ab\,\|x_0 - x^*\| \\
&\le b\left\|\int_0^1 (J''(x^* + \tau(x_0 - x^*)) - J''(x^*))(x_0 - x^*)\,\mathrm{d}\tau\right\| + ab\,\|x_0 - x^*\| \\
&\le b\left(\frac{L_{J''}\xi}{2} + a\right)\|x_0 - x^*\| \le b\left(\frac{L_{J''}\xi}{2} + a + \frac{c\xi}{1-\nu}\right)\|x_0 - x^*\| < \nu\|x_0 - x^*\|.
\end{aligned} \tag{2.10}$$

Hence, using inequality (2.9) we obtain

$$\|B_1^J - J''(x^*)\| \le a + c\xi. \tag{2.11}$$

This yields that

$$\|B_1^J - B_0^J\| \le \|B_1^J - J''(x^*)\| + \|B_0^J - J''(x^*)\| \le 2a + c\xi.$$

Consequently,

$$\left\|I - (B_0^J)^{-1}B_1^J\right\| \le \left\|(B_0^J)^{-1}\right\|\|B_1^J - B_0^J\| \le b(2a + c\xi) \le 2ab + \frac{c\xi}{1-\nu}b < 1.$$

Applying Neumann series theorem (Lemma 14.2 in [8]) yields that $(B_0^J)^{-1}B_1^J$ is bijective. Since B_0^J is boundedly invertible, B_1^J is bijective. Using boundedness of B_1^J and applying the Bounded Inverse Theorem (Theorem 5.6 in [8]) yields that B_1^J is boundedly invertible. For the induction hypothesis we assume that B_k^J is boundedly invertible and that

$$\|B_k^J - J''(x^*)\| \le a + \frac{c\xi(1-\nu^k)}{1-\nu} \tag{2.12}$$

for an arbitrarily fixed $k \in \mathbb{N}$. Using (2.4) we have

$$\|B_{k+1}^J - J''(x^*)\| \le \|B_k^J - J''(x^*)\| + c\mu_k \le a + c\left(\frac{\xi(1-\nu^k)}{1-\nu} + \mu_k\right), \tag{2.13}$$

for this k. Similar to above, we obtain by the mean value theorem, the induction hypothesis and the Lipschitz continuity of J'' that

$$\begin{aligned}
\|x_{k+1} - x^*\| &\le \left\|(B_k^J)^{-1}\right\| \left\{\|J'(x_k) - J'(x^*) - J''(x^*)(x_k - x^*)\| + \left\|[B_k^J - J''(x^*)](x_k - x^*)\right\|\right\} \\
&\le b\,\|J'(x_k) - J'(x^*) - J''(x^*)(x_k - x^*)\| + b\left(a + \frac{c\xi(1-\nu^k)}{1-\nu}\right)\|x_k - x^*\| \\
&\le b\left\{\left\|\int_0^1 (J''(x^* + \tau(x_k - x^*)) - J''(x^*))(x_k - x^*)\,\mathrm{d}\tau\right\| + \left(a + \frac{c\xi}{1-\nu}\right)\|x_k - x^*\|\right\} \\
&\le b\left(\frac{L_{J''}\xi}{2} + a + \frac{c\xi}{1-\nu}\right)\|x_k - x^*\| < \nu\|x_k - x^*\|.
\end{aligned} \tag{2.14}$$

By the induction hypothesis and the Lipschitz continuity of J'' it follows by repeatedly using the mean value theorem that

$$\|x_k - x^*\| < \nu^k\|x_0 - x^*\| < \nu^k\xi \tag{2.15}$$

and hence

$$\|x_{k+1} - x^*\| < \nu^{k+1}\|x_0 - x^*\| \quad \text{and} \quad \mu_k < \nu^k\xi. \tag{2.16}$$

Using (2.13) we obtain

$$\|B_{k+1}^J - J''(x^*)\| \leq a + c\xi\left(\frac{1-\nu^k}{1-\nu} + \nu^k\right) = a + c\xi\sum_{i=0}^{k-1}\nu^i + c\nu^k\xi = a + \frac{c\xi(1-\nu^{k+1})}{1-\nu} \tag{2.17}$$

and

$$\|B_{k+1}^J\| \leq \|J''(x^*)\| + a + \frac{c\xi}{1-\nu} < b < \infty. \tag{2.18}$$

Therefore,

$$\|B_{k+1}^J - B_0^J\| \leq \|B_{k+1}^J - J''(x^*)\| + \|B_0^J - J''(x^*)\| \leq 2a + \frac{c\xi(1-\nu^{k+1})}{1-\nu} \leq 2a + \frac{c\xi}{1-\nu}.$$

Consequently,

$$q := \left\|I - (B_0^J)^{-1}B_{k+1}^J\right\| \leq \left\|(B_0^J)^{-1}\right\|\|B_{k+1}^J - B_0^J\| \leq 2ab + \frac{c\xi}{1-\nu}b < 1. \tag{2.19}$$

By the same argument as above it follows that B_{k+1}^J is boundedly invertible. Using now the Neumann series representation, we reformulate

$$B_{k+1}^J = B_0^J\big((B_0^J)^{-1}B_{k+1}^J\big) = B_0^J\big(I - (I - (B_0^J)^{-1}B_{k+1}^J)\big).$$

Since $q < 1$, the Neumann series theorem yields

$$\big((B_0^J)^{-1}B_{k+1}^J\big)^{-1} = \sum_{m=0}^{\infty}\big(I - (B_0^J)^{-1}B_{k+1}^J\big)^m.$$

Consequently,

$$(B_{k+1}^J)^{-1} = \big((B_0^J)^{-1}B_{k+1}^J\big)^{-1}(B_0^J)^{-1} = \left(\sum_{m=0}^{\infty}\big(I - (B_0^J)^{-1}B_{k+1}^J\big)^m\right)(B_0^J)^{-1}.$$

Taking norms and using the geometric series gives

$$\|(B_{k+1}^J)^{-1}\| \leq \|(B_0^J)^{-1}\|\sum_{m=0}^{\infty}q^m = \frac{\|(B_0^J)^{-1}\|}{1-q} \leq \frac{b}{1-q}.$$

Using the previously derived bound (2.19) we obtain

$$\|(B_{k+1}^J)^{-1}\| \leq \frac{b}{1 - 2ab - \frac{c\xi}{1-\nu}b},$$

where $1 - 2ab - \frac{c\xi}{1-\nu}b > 0$ due to inequality (2.8). Hence, $\{\|B_k^J\|\}$ and $\{\|(B_k^J)^{-1}\|\}$ are uniformly bounded. Therefore, $\{x_k\} \subset U$ converges Q-linearly to x^*, since $\nu \in (0,1)$. □

To satisfy the conditions (2.7) and (2.8), the choice of $\xi > 0$ and $a > 0$ to can be done using the subsequent Lemma 2.2 or Lemma 2.3.

Lemma 2.2. Let $\nu \in (0,1)$, $c > 0$ and $b > 0$ be fixed arbitrarily and $L_{J''} > 0$ be given. Define

$$\xi_{\max} := \xi_{\max}(\nu) = \min\left\{\frac{1-\nu}{bc},\ \frac{\nu}{b\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right)}\right\}$$

and

$$a_{\max} := a_{\max}(\xi(\nu)) = \min\left\{\frac{1}{2}\left(\frac{1}{b} - \frac{c\xi}{1-\nu}\right),\ \frac{\nu}{b} - \xi\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right)\right\}.$$

Then, for any $\xi = \xi(\nu)$ satisfying $0 < \xi < \xi_{\max}$ and any $a = a(\xi(\nu))$ satisfying $0 < a < a_{\max}$, the two inequalities (2.7) and (2.8) hold.

Proof. Since $\nu \in (0,1)$ and $b, c, L_{J''} > 0$, we have that

$$\frac{1-\nu}{bc} > 0 \quad \text{and} \quad \frac{\nu}{b\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right)} > 0.$$

Hence, $\xi_{\max} > 0$. Let now $\xi \in (0, \xi_{\max})$ be fixed arbitrarily. By the definition of $\xi_{\max}(\nu)$, the two strict inequalities

$$\xi < \frac{1-\nu}{bc} \quad \text{and} \quad \xi < \frac{\nu}{b\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right)} \tag{2.20}$$

hold. The first inequality in (2.20) is equivalent to

$$\frac{c\xi}{1-\nu} < \frac{1}{b},$$

and thus,

$$A(\xi) := \frac{1}{2}\left(\frac{1}{b} - \frac{c\xi}{1-\nu}\right) > 0.$$

From the second inequality in (2.20), we directly have

$$\xi\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right) < \frac{\nu}{b} \quad \Longleftrightarrow \quad B(\xi) := \frac{\nu}{b} - \xi\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right) > 0.$$

Consequently,

$$a_{\max} = \min\{A(\xi), B(\xi)\} > 0.$$

As $a < A(\xi)$, we have

$$a < \frac{1}{2}\left(\frac{1}{b} - \frac{c\xi}{1-\nu}\right) \quad \Longleftrightarrow \quad b\left(2a + \frac{c\xi}{1-\nu}\right) < 1,$$

which is precisely inequality (2.8). Furthermore, since $a < B(\xi)$, we obtain

$$ab < \nu - b\left(\frac{L_{J''}}{2}\xi + \frac{c\xi}{1-\nu}\right) \quad \Longleftrightarrow \quad b\left(\frac{L_{J''}}{2}\xi + a + \frac{c\xi}{1-\nu}\right) < \nu.$$

This is exactly inequality (2.7). Thus, both (2.7) and (2.8) hold for any $a \in (0, a_{\max})$ and $\xi \in (0, \xi_{\max})$. ☐

Lemma 2.3. Let $\nu \in (0,1),\ c > 0$ and $b > 0$ be fixed arbitrarily and $L_{J''} > 0$ be given. Define

$$a_{\max} := a_{\max}(\nu) = \min\left\{\frac{1}{2b}, \frac{\nu}{b}\right\}$$

and

$$\xi_{\max} := \xi_{\max}(a(\nu)) = \min\left\{\frac{1-\nu}{bc}(1-2ab),\ \frac{\frac{\nu}{b} - a}{\frac{L_{J''}}{2} + \frac{c}{1-\nu}}\right\}.$$

Then, for any $a = a(\nu)$ satisfying $0 < a < a_{\max}$ and any $\xi = \xi(a(\nu))$ satisfying $0 < \xi < \xi_{\max}$, the two inequalities (2.7) and (2.8) hold.

Proof. Since $\nu \in (0,1)$ and $b > 0$, we have that $0 < a < \min\{\frac{1}{2b}, \frac{\nu}{b}\}$. Using $\nu \in (0,1),\ c > 0,\ b > 0,\ L_{J''} > 0$, we obtain with $2ab < 1$ that both terms in the definition of $\xi_{\max}(a)$ are positive. Now, let $0 < \xi < \xi_{\max}(a)$ be fixed arbitrarily. By definition of $\xi_{\max}(a)$, we have the two strict inequalities

$$\xi < \frac{1-\nu}{bc}(1-2ab) \quad \text{and} \quad \xi < \frac{\frac{\nu}{b} - a}{\frac{L_{J''}}{2} + \frac{c}{1-\nu}}. \tag{2.21}$$

From the first inequality in (2.21) we obtain by multiplying $\frac{bc}{1-\nu}$ and adding $2ab$ exactly inequality (2.8). From the second inequality in (2.21) we obtain

$$\xi\left(\frac{L_{J''}}{2} + \frac{c}{1-\nu}\right) < \frac{\nu}{b} - a \quad \Longleftrightarrow \quad b\left(\frac{L_{J''}}{2}\xi + a + \frac{c\xi}{1-\nu}\right) < \nu,$$

which is precisely inequality (2.7). Hence, both inequalities (2.7) and (2.8) hold for any $\xi \in (0, \xi_{\max}(a(\nu)))$ and any $a \in (0, a_{\max}(\nu))$. ☐

Secant updates are often used to define QN methods. Hence, setting

$$\zeta_k := J'(x_{k+1}) - J'(x_k), \tag{2.22}$$

we introduce for J the secant condition

$$B_{k+1}^J s_k = \zeta_k, \tag{2.23}$$

where B_{k+1}^J has to be chosen such that it approximates $J''(x_k)$. Furthermore, QN methods frequently require a suitable outer product, which in finite dimensional spaces is represented by the matrix xy^T. In infinite dimensional Hilbert spaces, we need the definition of the outer product, which can be achieved by the dyadic product, see [5] or [25].

Definition 2.4 (Dyadic product)**.** Let X and Y be Hilbert spaces. Then the dyadic product $\otimes : X \times Y \to L(Y, X)$ is a rank-1 operator defined for $x \in X$, $y \in Y$ by

$$(x \otimes y)z = \langle y, z\rangle_Y x \quad \text{for all } z \in Y.$$

We now recall some properties of this operator, see [5] or [25].

Proposition 2.5. Let X and Y be Hilbert spaces and $x \in X$ and $y \in Y$. Denote by $L(X, Y)$ the space of bounded linear operators $T : X \to Y$ and $L(X) := L(X, X)$. Then it holds that:

- The operator $\otimes : X \times Y \to L(Y, X)$ is bilinear.
- $\|x \otimes y\| = \|x\|\|y\|$.
- If $Y = X$, then for all selfadjoint operators $A \in L(X)$ it holds that $A \circ (x \otimes y) \circ A = (Ax) \otimes (Ay)$, where $\circ$ denotes the concatenation of operators or functions.
- For all $x_1, x_2, y_1, y_2 \in X$, $(x_1 \otimes y_1) \circ (x_2 \otimes y_2) = \langle y_1, x_2\rangle(x_1 \otimes y_2)$.
- $P = I - \frac{s \otimes s}{\langle s,s\rangle}$ is an orthogonal projector, i.e., $\|P\| = 1$ for $X \neq \text{span}(s)$.

Since we will use the PSB update,

$$B_{k+1}^J = B_k^J + \frac{(\zeta_k - B_k^J s_k) \otimes s_k + s_k \otimes (\zeta_k - B_k^J s_k)}{\langle s_k, s_k\rangle} - \frac{\langle \zeta_k - B_k^J s_k, s_k\rangle s_k \otimes s_k}{\langle s_k, s_k\rangle^2}, \tag{2.24}$$

we now state the following auxiliary result motivated by [25].

Lemma 2.6 (BDP for the PSB method)**.** Let $x^* \in U$ be a minimizer of J and J'' be Lipschitz continuous with Lipschitz constant $L_{J''} > 0$. Then, the PSB method satisfies the BDP

$$\left\|B_{k+1}^J - J''(x^*)\right\| \leq \left\|B_k^J - J''(x^*)\right\| + \tilde{c}\mu_k. \tag{2.25}$$

Proof. Similarly to Theorem 2.1 in [25] we use the orthogonal projector defined in Proposition 2.5 to reformulate the distance between B_{k+1}^J and $J''(x^*)$. Due to Theorem 2.1 in [25], we have that

$$\begin{aligned}
B_{k+1}^J - J''(x^*) =& B_k^J - J''(x^*) + \frac{(\zeta_k - B_k^J s_k) \otimes s_k + s_k \otimes (\zeta_k - B_k^J s_k)}{\langle s_k, s_k\rangle} - \frac{\langle \zeta_k - B_k^J s_k, s_k\rangle s_k \otimes s_k}{\langle s_k, s_k\rangle^2} \\
=& P(B_k^J - J''(x^*))P - \frac{(s_k \otimes s_k)J''(x^*) + J''(x^*)(s_k \otimes s_k)}{\langle s_k, s_k\rangle} \\
&+ \frac{(s_k \otimes s_k)J''(x^*)(s_k \otimes s_k)}{\langle s_k, s_k\rangle^2} + \frac{(\zeta_k \otimes s_k) + (s_k \otimes \zeta_k)}{\langle s_k, s_k\rangle} - \frac{(s_k \otimes \zeta_k)(s_k \otimes s_k)}{\langle s_k, s_k\rangle^2} \\
=& P(B_k^J - J''(x^*))P + \frac{s_k \otimes (\zeta_k - J''(x^*)s_k)}{\langle s_k, s_k\rangle} P + \frac{(\zeta_k - J''(x^*)s_k) \otimes s_k}{\langle s_k, s_k\rangle}.
\end{aligned} \tag{2.26}$$

Using the triangle inequality and Proposition 2.5 we have

$$\left\|B_{k+1}^J - J''(x^*)\right\| \leq \left\|B_k^J - J''(x^*)\right\| + 2\frac{\|\zeta_k - J''(x^*)s_k\|}{\|s_k\|}. \tag{2.27}$$

By the fundamental theorem of calculus in Banach spaces and the Lipschitz continuity of J'' on U we obtain

$$\begin{aligned}\|\zeta_k - J''(x^*)s_k\| &= \|J'(x_{k+1}) - J'(x_k) - J''(x^*)s_k\| = \left\|\int_0^1 (J''(x_k+\tau s_k) - J''(x^*))s_k \,\mathrm{d}\tau\right\| \\ &\le \|s_k\| \int_0^1 \|J''(x_k+\tau s_k) - J''(x^*)\| \,\mathrm{d}\tau \\ &\le L_{J''}\|s_k\| \left(\int_0^1 \tau\|x_{k+1}-x^*\|\,\mathrm{d}\tau + \int_0^1 (1-\tau)\|x_k - x^*\|\,\mathrm{d}\tau\right) \le L_{J''}\|s_k\|\mu_k.\end{aligned}$$

With this, inequality (2.27) simplifies to

$$\left\|B_{k+1}^J - J''(x^*)\right\| \le \left\|B_k^J - J''(x^*)\right\| + 2L_{J''}\mu_k. \tag{2.28}$$

Therefore, the PSB update satisfies the BDP. □

3. A regularized structure-exploiting PSB method

For an operator F, we denote the adjoint with F^* and the dual space of X with X^*. Furthermore, we assume that $F \in C^2(U,Y)$ is well-defined. Let y^δ be the data contaminated with noise up to a noise level $\delta > 0$, which might not be in the range of F and $\left\|y - y^\delta\right\| \le \delta$, where y is the possibly non-unique projection, i.e., an element in the range of F with minimal distance to y^δ. This noise level is frequently dictated by the measurement process, which yields the assumption that $\delta > 0$ is fixed arbitrarily. Since the measurement data is contaminated with noise, the direct inversion of the forward operator is ineffective as it tends to produce results that deviate significantly from the exact solution. Hence, we want to approach the inverse problem via an optimization problem, where we define

$$J : X \to \mathbb{R}, \qquad J(x) := \frac{1}{2}\left\|F(x) - y^\delta\right\|_Y^2, \tag{3.1}$$

and want to find the minimizer of the objective functional J. Computing the first Fréchet derivative yields for all $v \in U$

$$dJ(x)v = (F'(x)v, F(x) - y^\delta)_Y = (v, F'(x)^*(F(x) - y^\delta))_X, \tag{3.2}$$

where we used the Hilbert space adjoint $F'(x)^* \in L(Y,X)$, see Chapter 16 in [8]. Hence, we identify

$$J'(x) = dJ(x) = F'(x)^*(F(x) - y^\delta) \in X.$$

Computing the second Fréchet derivative yields for all $v, w \in U$

$$\begin{aligned} d(dJ(x)v)w = d(dJ(x)w)v &= ((F''(x)w)v, F(x) - y^\delta)_Y + (F'(x)w, F'(x)v)_Y \\ &= (v, (F''(x)w)^*(F(x) - y^\delta))_X + (F'(x)^*F'(x)w, v)_X \\ &= (v, F'(x)^*F'(x)w + (F''(x)w)^*(F(x) - y^\delta))_X. \end{aligned} \tag{3.3}$$

Thus,

$$d^2J(x)w = F'(x)^*F'(x)w + (F''(x)w)^*(F(x) - y^\delta) \in X$$

and

$$J''(x) = d^2J(x) \in L(X).$$

We assume that $J''(x)$ is a compact operator for all $x \in U$. Compactness of the Hessian is a realistic assumption for a large class of ill-posed inverse problems, e.g., PDE-based inverse problems or inverse problems involving integral operators. In various inverse problems, requiring $J''(x)$ to be a compact operator is closely related to assuming that the forward operator F is "smoothing". Whenever F maps the parameter space into a more regular space (typically through a PDE solution operator or an integral operator), the Fréchet derivative $F'(x)$ inherits this smoothing property. Since the second derivative of the data misfit involves compositions of $F'(x)$ and $F''(x)$, the same smoothing mechanism typically ensures that $J''(x)$ is compact. However, the assumption might be restrictive in some applications, e.g., tomographic reconstruction problems or deblurring/inverse convolution problems with non-smoothing kernels. As we want to use Newton-type methods, we consider the classic Newton method obtained by employing the Taylor approximation

$$0 = J'(x_k) + J''(x_k)(x_{k+1} - x_k),$$

yielding

$$J''(x_k)(x_{k+1} - x_k) = -J'(x_k). \tag{3.4}$$

Since $J''(x)$ is a compact operator for all $x \in U$ and X, Y can be infinite dimensional, $J''(x)$ might not be continuously invertible. Therefore, we introduce the Tikhonov-type regularizer

$$\mathcal{R}_{\alpha_k} : X \to \mathbb{R},\ x \mapsto \frac{\alpha_k}{2}\|x - x_k\|_X^2 \tag{3.5}$$

with a suitable regularization parameter sequence $\{\alpha_k\} \subset \mathbb{R}^+$, where x_k is the last obtained iterate or the linearization point, respectively. Hence, we obtain the regularized target functional

$$\mathcal{J}_k : X \to \mathbb{R}, \quad \mathcal{J}_k(x) := J(x) + \mathcal{R}_{\alpha_k}(x). \tag{3.6}$$

Since $\mathcal{R}'_{\alpha_k}(x_k) = 0$, i.e., $\mathcal{J}'_k(x_k) = J'(x_k)$, and $\mathcal{R}''_{\alpha_k} = \alpha_k I$, applying Newton's method to minimize $\mathcal{J}_k$ defined in (3.6) yields

$$F'(x_k)^* F'(x_k) s_k + (F''(x_k)s_k)^* \left(F(x_k) - y^\delta\right) + \alpha_k s_k = -F'(x_k)^* \left(F(x_k) - y^\delta\right). \tag{3.7}$$

In contrast, the LM method

$$\left(F'(x_k)^* F'(x_k) + \alpha_k I\right) s_k = -F'(x_k)^* \left(F(x_k) - y^\delta\right) \tag{3.8}$$

neglects the second order term $(F''(x_k)(\cdot))^* \left(F(x_k) - y^\delta\right)$. Hence, the approximation of the Hessian by the LM method may be too inaccurate as the approximation error is small only if F is linear or the residual is sufficiently small, i.e., we have a good approximation and initial guess. Furthermore, if we assume that we have access to the first derivative of the forward operator F, secant based QN updates to approximate $(F''(x_k)(\cdot))^* \left(F(x_k) - y^\delta\right)$ are accessible as well. Additionally, there is no necessity to approximate the full Hessian operator, but rather only the second order term involving F''. Thus, we want to derive an approximation for $(F''(x_k)(\cdot))^* \left(F(x_k) - y^\delta\right)$. For this purpose, we consider the Taylor approximation of F given by

$$\begin{aligned} F(x_{k+1}) &= F(x_k) + F'(x_k)s_k + \mathcal{O}(\|s_k\|^2) \\ \Leftrightarrow\ F'(x_k)s_k &= F(x_{k+1}) - F(x_k) + \mathcal{O}(\|s_k\|^2). \end{aligned} \tag{3.9}$$

By defining

$$r_k := F(x_k) - y^\delta, \tag{3.10}$$

we conclude

$$\begin{aligned} F'(x_k)^*(r_{k+1} - r_k) &= F'(x_k)^*(F(x_{k+1}) - y^\delta - F(x_k) + y^\delta) = F'(x_k)^*(F(x_{k+1}) - F(x_k)) \\ &= F'(x_k)^* F'(x_k) s_k + \mathcal{O}(\|s_k\|^2). \end{aligned} \tag{3.11}$$

Then, exploiting the Taylor approximation of J', i.e.,

$$\begin{aligned} J'(x_{k+1}) = F'(x_{k+1})^* r_{k+1} &= F'(x_k)^* r_k + (F''(x_k)s_k)^* r_k + F'(x_k)^* F'(x_k) s_k + \mathcal{O}(\|s_k\|^2) \\ \Leftrightarrow\ (F''(x_k)s_k)^* r_k &= F'(x_{k+1})^* r_{k+1} - F'(x_k)^* r_k - F'(x_k)^* F'(x_k) s_k + \mathcal{O}(\|s_k\|^2) \end{aligned}$$

and using (3.11) yields

$$\begin{aligned} &(F''(x_k)s_k)^* r_k = (F'(x_{k+1}) - F'(x_k))^* r_{k+1} + F'(x_k)^*(r_{k+1} - r_k) - F'(x_k)^* F'(x_k) s_k + \mathcal{O}(\|s_k\|^2) \\ \Leftrightarrow\ &(F''(x_k)s_k)^* r_k = (F'(x_{k+1}) - F'(x_k))^* r_{k+1} + \mathcal{O}(\|s_k\|^2). \end{aligned}$$

Consequently, we will consider the secant equation

$$(F''(x_k)s_k)^* r_k = (F'(x_{k+1}) - F'(x_k))^* r_{k+1}. \tag{3.12}$$

Hence, this equation defines the step s_k to be taken. We now analyze the relationship between the standard secant equation used for unstructured QN methods and the modified secant equation (3.12).

Remark 3.1. Assume that the secant equation for J, i.e.,

$$J''(x_k)s_k = J'(x_{k+1}) - J'(x_k) \tag{3.13}$$

is used to define s_k. Hence,

$$\begin{aligned} J''(x_k)s_k &= J'(x_{k+1}) - J'(x_k) \\ &= F'(x_{k+1})^* \left(F(x_{k+1}) - y^\delta\right) - F'(x_k)^* \left(F(x_k) - y^\delta\right) \\ &= \left(F'(x_{k+1}) - F'(x_k)\right)^* \left(F(x_{k+1}) - y^\delta\right) + F'(x_k)^* \left(F(x_{k+1}) - F(x_k)\right). \end{aligned}$$

By the Taylor theorem in Banach spaces and the Lipschitz continuity of F' we know

$$F(x_{k+1}) - F(x_k) = F'(x_k)s_k + E_k s_k \;\text{ with }\; \|E_k\| \le L_{F'}\|s_k\| \le 2L_{F'}\mu_k, \tag{3.14}$$

which yields

$$J''(x_k)s_k = \left(F'(x_{k+1}) - F'(x_k)\right)^* \left(F(x_{k+1}) - y^\delta\right) + F'(x_k)^*F'(x_k)s_k + F'(x_k)^*E_k s_k$$

and thus

$$\mathcal{J}_k''(x_k)s_k = \left(F'(x_{k+1}) - F'(x_k)\right)^* \left(F(x_{k+1}) - y^\delta\right) + F'(x_k)^*F'(x_k)s_k + F'(x_k)^*E_k s_k + \alpha_k s_k.$$

With identity (3.3) we obtain

$$(F''(x_k)s_k)^* r_k = \left(F'(x_{k+1}) - F'(x_k)\right)^* r_{k+1} + F'(x_k)^*E_k s_k. \tag{3.15}$$

Equation (3.15) describes the relation to the modified secant condition (3.12), as equation (3.15) consists of the same components as equation (3.12) but with the additional error term $F'(x_k)^*E_k s_k$ on the right-hand side.

Remark 3.2. Define for s_k fulfilling (3.13)

$$y_k := \left(F'(x_{k+1}) - F'(x_k)\right)^* r_{k+1} \tag{3.16}$$

and

$$C(x) := F'(x)^*F'(x). \tag{3.17}$$

For ζ_k as defined in (2.22), we obtain due to Remark 3.1 that

$$\zeta_k = y_k + C(x_k)s_k + F'(x_k)^*E_k s_k. \tag{3.18}$$

Hence, equation (3.18) yields that B_k^J can be expressed as

$$B_k^J = C(x_k) + A_k + F'(x_k)^*E_k, \tag{3.19}$$

where $A_k \approx (F''(x_{k-1})(\cdot))^* r_{k-1}$ has to be determined such that

$$A_k s_{k-1} = y_{k-1} \tag{3.20}$$

holds.

Since F is twice continuously Fréchet differentiable, $J''(x^*) \in L(X) \setminus \{0\}$ is symmetric. Therefore, $F'(x^*)^*F'(x^*)$ is symmetric and $(F''(x^*)(\cdot))^* \left(F(x^*) - y^\delta\right)$ is symmetric but not necessarily positive. The Powell-symmetric Broyden update forces symmetry, without assuming positivity of the operators, while being numerically more stable than the symmetric-rank-1 update, as the PSB update does not need a well-definedness condition to prevent the denominator from being zero. Consequently, we will use the PSB update. For the derivation of our theory we will employ the following assumptions:

Assumptions 3.3.

A1 The operator F' is Lipschitz continuous with Lipschitz constant $L_{F'} > 0$ and the operator C defined in (3.17) is Lipschitz continuous with Lipschitz constant $L_C > 0$.

A2 The operator $J' : U \subset X \to X$ is Lipschitz continuously Fréchet differentiable in U, i.e., J'' is Lipschitz continuous with Lipschitz constant $L_{J''} > 0$.

A3 The operator F' is bounded on U by $M > 0$.

Note that J' being Lipschitz continuously Fréchet differentiable implies that $\mathcal{J}_k'$ is Lipschitz continuously Fréchet differentiable for all $k \in \mathbb{N}$ since $\mathcal{R}_{\alpha_k} \in C^\infty(X, \mathbb{R})$ for all $k \in \mathbb{N}$. Furthermore, if there exists a minimizer x^* of (3.6) in a convex and compact set $D \subset U$, there exists a projection y in the range of F with minimal distance to y^δ, due to F being continuous, i.e., $F(D)$ is compact. Note, that in general this y might not be unique. Using the operator C defined by (3.17) in Remark 3.2, the regularized structure-exploiting PSB (RSE-PSB) update is given by

$$B_k := C(x_k) + A_k + \alpha_k I, \tag{3.21}$$

where A_k is updated with the PSB update rule, i.e.,

$$A_{k+1} = A_k + \frac{(y_k - A_k s_k) \otimes s_k + s_k \otimes (y_k - A_k s_k)}{\langle s_k, s_k \rangle} - \frac{\langle y_k - A_k s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2}. \tag{3.22}$$

Remark 3.4. From Remark 3.2 we deduce that for the secant condition (2.23) B_k^J is given by

$$B_k^J = B_k + F'(x_k)^* E_k - \alpha_k I. \tag{3.23}$$

Using our update (3.21) together with (3.22), we just approximate a part of the Hessian operator, namely $(F''(x_k)(\cdot))^* \left(F(x_k) - y^\delta\right)$ using the PSB update with a modified secant condition.

Remark 3.5. To verify the BDP for the RSE-PSB update, we need an additional assumption on the regularization parameter. Since F' is bounded by $M > 0$ and E_k is bounded by $L_{F'}\|s_k\|$ for all $k \in \mathbb{N}$ due to (3.14), we know

$$\|F'(x_k)^* E_k\| \leq M L_{F'} \|s_k\|.$$

We will now assume that there exists $\hat{c}, \tilde{c} > 0$ such that the regularization parameters satisfies

$$\begin{cases} |\alpha_{k+1} - \alpha_k| \leq \hat{c}\mu_k & \forall k \in \mathbb{N} \\ \|\alpha_{k+1} I - F'(x_k)^* E_k\| \leq \tilde{c}\mu_k & \forall k \in \mathbb{N} \end{cases} \tag{3.24}$$

with μ_k defined as in (2.2). A sufficient condition for the regularization parameter to satisfy the second inequality in (3.24) is the assumption that there exists some c' such that

$$\alpha_{k+1} \leq c' \mu_k \qquad \forall k \in \mathbb{N} \tag{3.25}$$

and choosing $\tilde{c} = c' + 2ML_{F'}$. The first inequality in (3.24) can be viewed as the necessity that the regularization parameters form a converging sequence, where additionally the speed of convergence is dictated by μ_k or the iteration step, since $\|s_k\| \leq 2\mu_k$. Note that the sufficient condition (3.25) can be verified a-priori.

Lemma 3.6 (BDP for the RSE-PSB method)**.** Let $x^* \in U$, Assumptions 3.3 hold and let the regularization parameters satisfy condition (3.24). Then, the RSE-PSB method, i.e., B_k as in (3.21), where A_k is updated according to (3.22) using (1.3) and (3.16), satisfies the following BDP

$$\|B_{k+1} - J''(x^*)\| \leq \|B_k - J''(x^*)\| + c\mu_k \tag{3.26}$$

with $c = 2L_C + 2L_{J''} + \hat{c} + 4ML_{F'} > 0$. Furthermore, B_k^{-1} exists and $\{\|B_k\|\}$ as well as $\{\|B_k^{-1}\|\}$ are bounded.

Proof. Due to Remark 3.2 and Remark 3.4 we conclude that

$$\zeta_k - B_k^J s_k = y_k - A_k s_k$$

and therefore

$$\begin{aligned} B_{k+1} &= C(x_{k+1}) + A_{k+1} + \alpha_{k+1} I \\ &= C(x_{k+1}) + \alpha_{k+1} I + A_k + \frac{(y_k - A_k s_k) \otimes s_k + s_k \otimes (y_k - A_k s_k)}{\langle s_k, s_k \rangle} - \frac{\langle y_k - A_k s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2} \\ &= C(x_{k+1}) - C(x_k) + \alpha_{k+1} I - F'(x_k)^* E_k \\ &\quad + B_k^J + \frac{(\zeta_k - B_k^J s_k) \otimes s_k + s_k \otimes (\zeta_k - B_k^J s_k)}{\langle s_k, s_k \rangle} - \frac{\langle \zeta_k - B_k^J s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2} \\ &= C(x_{k+1}) - C(x_k) + \alpha_{k+1} I - F'(x_k)^* E_k + B_{k+1}^J, \end{aligned}$$

where the last equality is obtained using identity (2.24). Due to the Lipschitz continuity of C and condition (3.24), we obtain

$$\left\|B_{k+1} - B_{k+1}^J\right\| \le L_C\|s_k\| + \tilde{c}\mu_k. \tag{3.27}$$

Now, using Lemma 2.6 and using specifically inequality (2.28) yields

$$\begin{aligned}\|B_{k+1} - J''(x^*)\| - \left\|B_{k+1} - B_{k+1}^J\right\| &\le \left\|B_{k+1} - J''(x^*) - B_{k+1} + B_{k+1}^J\right\| \\ &= \left\|B_{k+1}^J - J''(x^*)\right\| \\ &\le \left\|B_k^J - J''(x^*)\right\| + 2L_{J''}\mu_k.\end{aligned}$$

By using identity (3.23), the condition(3.24) and the estimates (3.27) as well as (3.14), we obtain

$$\begin{aligned}\|B_{k+1} - J''(x^*)\| &\le \left\|B_k^J - J''(x^*)\right\| + \left\|B_{k+1} - B_{k+1}^J\right\| + 2L_{J''}\mu_k \\ &\le \left\|B_k^J - J''(x^*)\right\| + (2L_C + 2L_{J''} + \tilde{c})\mu_k \\ &\le \|B_k - J''(x^*)\| + \|F'(x_k)^*E_k - \alpha_{k+1}I\| + |\alpha_{k+1} - \alpha_k| + (2L_C + 2L_{J''} + \tilde{c})\mu_k \\ &\le \|B_k - J''(x^*)\| + (2L_C + 2L_{J''} + \hat{c} + 2\tilde{c})\mu_k.\end{aligned} \tag{3.28}$$

□

Motivated by Theorem 2.1, we now prove local Q-linear convergence of the regularized structure-exploiting PSB method to a minimizer of (3.1).

Theorem 3.7 (Local Q-linear convergence of the RSE-PSB method)**.** *Let $x^* \in U$ be a minimizer of* (3.1)*, Assumptions 3.3 hold, $\{\alpha_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^+$ satisfy the condition* (3.24) *for some $\hat{c} > 0$ and $\tilde{c} > 0$. Let μ_k be defined as in* (2.2)*. Furthermore, let B_0 be boundedly invertible with bound $b > 0$, $\nu \in (0,1)$ be arbitrarily given and assume that B_0 as well as x_0 satisfy*

$$\|x_0 - x^*\| < \xi_{\max}, \quad \|B_0 - J''(x^*)\| < a_{\max}, \tag{3.29}$$

where $\xi_{\max}$ and $a_{\max}$ are chosen according to Lemma 2.2 or Lemma 2.3 and $c = 2L_C + 2L_{J''} + \hat{c} + 2\tilde{c} > 0$. Then, the sequence $\{x_k\}$ obtained by the RSE-PSB method, i.e., B_k as in (3.21)*, where A_k is updated according to* (3.22) *using* (1.3) *and* (3.16)*, is well-defined and converges Q-linearly to x^*. Furthermore, the sequences $\{\|B_k\|\}$ and $\{\|(B_k)^{-1}\|\}$ are uniformly bounded.*

Proof. For any $\xi = \xi(\nu)$ satisfying $0 < \xi < \xi_{\max}$ and any $a = a(\nu)$ satisfying $0 < a < a_{\max}$, where $\xi_{\max}$ and $a_{\max}$ are chosen according to Lemma 2.2 or Lemma 2.3, the two inequalities (2.7) and (2.8) hold. Hence, we can use Lemma 3.6 to obtain the BDP

$$\|B_{k+1} - J''(x^*)\| \le \|B_k - J''(x^*)\| + c\mu_k, \tag{3.30}$$

where $c = 2L_C + 2L_{J''} + \hat{c} + 2\tilde{c} > 0$. The rest of the proof is analogous to the proof of Theorem 2.1, keeping in mind that $\mathcal{J}_k'(x_k) = J'(x_k)$. □

We now prove local super-linear convergence of the regularized structure-exploiting PSB method. Preliminarily, we define

$$Z_* := (F''(x^*)(\cdot))^*(F(x^*) - y^\delta). \tag{3.31}$$

Theorem 3.8 (Local Q-super-linear convergence of the RSE-PSB method)**.** *Let the assumptions of Theorem 3.7 be satisfied. Let F'' be Lipschitz continuous in the first component with respect to the Hilbert–Schmidt norm, i.e., there exists $L_{F''} > 0$ such that for all $x, \tilde{x} \in B_{\xi_{\max}}(x^*)$ it holds that*

$$\|F''(x)(\cdot) - F''(\tilde{x})(\cdot)\|_{HS} \le \xi_{\max} L_{F''}\|x - \tilde{x}\|.$$

Suppose that F is Lipschitz continuous with respect to the Hilbert–Schmidt norm as well and that

$$A_0 - Z_* \tag{3.32}$$

is a Hilbert–Schmidt operator. Then, the sequence $\{x_k\}$ obtained by the RSE-PSB method, i.e., B_k as in (3.21)*, where A_k is updated according to* (3.22) *using* (1.3) *and* (3.16)*, is well-defined and converges Q-super-linearly to x^*. Furthermore, the sequences $\{\|B_k\|\}$ and $\{\|(B_k)^{-1}\|\}$ are uniformly bounded.*

Proof. First we apply Theorem 3.7 to obtain that the sequence $\{x_k\}$ converges Q-linearly to x^*. We want to apply the Dennis–Moré condition for generalized equations in Banach spaces (see Theorem 3 in [13]) to obtain super-linear convergence of $\{x_k\}$ to x^*. Using Remark 3.1, the fundamental theorem of calculus in Banach spaces, the Lipschitz continuity of J'' on U and condition (3.24) yields

$$\begin{aligned}\|(B_{k+1} - J''(x^*))s_k\| &= \|J'(x_{k+1}) - J'(x_k) - J''(x^*)s_k + (\alpha_{k+1}I - F'(x_{k+1})^*E_{k+1})s_k\| \\ &\leq \left\| \int_0^1 (J''(x_k + \tau s_k) - J''(x^*))s_k \,\mathrm{d}\tau \right\| + (\tilde{c} + \hat{c})\mu_{k+1}\|s_k\| \\ &\leq \|s_k\| \int_0^1 \|J''(x_k + \tau s_k) - J''(x^*)\| \,\mathrm{d}\tau + (\tilde{c} + \hat{c})\mu_{k+1}\|s_k\| \\ &\leq L_{J''}\|s_k\| \left(\int_0^1 \tau \|x_{k+1} - x^*\| \,\mathrm{d}\tau + \int_0^1 (1-\tau)\|x_k - x^*\| \,\mathrm{d}\tau \right) \\ &\quad + (\tilde{c} + \hat{c})\mu_{k+1}\|s_k\| \\ &\leq L_{J''}\|s_k\|\mu_k + (\tilde{c} + \hat{c})\mu_{k+1}\|s_k\|.\end{aligned}$$

Therefore,

$$\begin{aligned}\|(B_k - J''(x^*))s_k\| &\leq \|(B_{k+1} - B_k)s_k\| + \|(B_{k+1} - J''(x^*))s_k\| \\ &\leq \|B_{k+1} - B_k\|\|s_k\| + (L_{J''}\mu_k + (\tilde{c} + \hat{c})\mu_{k+1})\|s_k\|.\end{aligned} \tag{3.33}$$

Using the Lipschitz continuity of C and the condition (3.24) on the regularization parameter yields

$$\begin{aligned}\lim_{k\to\infty} \frac{\|(B_{k+1} - J''(x^*))s_k\|}{\|s_k\|} &\leq \lim_{k\to\infty} (\|B_{k+1} - B_k\| + L_{J''}\mu_k + (\tilde{c} + \hat{c})\mu_{k+1}) \\ &\leq \lim_{k\to\infty} (\|A_{k+1} - A_k\| + L_C\|s_k\| + L_{J''}\mu_k + \hat{c}\mu_k + (\tilde{c} + \hat{c})\mu_{k+1}).\end{aligned} \tag{3.34}$$

The last four terms tend to 0 as k increases to infinity. Due to Corollary 16.9 in [27] we know that the operator norm is less than or equal to the Hilbert-Schmidt norm, i.e.,

$$\|A_{k+1} - A_k\| \leq \|A_{k+1} - A_k\|_{HS}. \tag{3.35}$$

Therefore, we now prove that

$$\lim_{k\to\infty} \|A_{k+1} - A_k\|_{HS} = 0. \tag{3.36}$$

Consider the convex set

$$D_k := \{T \in L(X, Y) \mid T \text{ is symmetric and } Ts_k = y_k\}.$$

From similar arguments as in the proof of Theorem 4.8 in [2], we observe that D_k is closed. By construction, the PSB update for approximating $(F''(x_k)(\cdot))^*(F(x_k) - y^\delta)$ satisfies the secant condition (3.20) and thus $A_{k+1} \in D_k$. Using Proposition 4.3 in [30], we have that the PSB update (2.24) is the solution to the variational problem

$$\min_{T \in D_k} \|T - A_k\|_{HS}. \tag{3.37}$$

Arguing as in the proof of Theorem 4.8 in [2], we obtain that A_{k+1} is the projection of A_k onto the closed convex set D_k, which is firmly non-expansive, see Proposition 4.16 in [3], meaning that for every $T \in D_k$ one has

$$\|P_{D_k}(A_k) - P_{D_k}(T)\|_{HS}^2 + \|(I - P_{D_k})(A_k) - (I - P_{D_k})(T)\|_{HS}^2 \leq \|A_k - T\|_{HS}^2,$$

and therefore, for all $T \in D_k$,

$$\|A_{k+1} - T\|_{HS}^2 + \|A_{k+1} - A_k\|_{HS}^2 \leq \|A_k - T\|_{HS}^2. \tag{3.38}$$

Define

$$T_k := \int_0^1 (F''(x_k + ts_k)(\cdot))^*(F(x_{k+1}) - y^\delta) \,\mathrm{d}t. \tag{3.39}$$

Since taking the adjoint is a linear and continuous operation, we obtain by applying the fundamental theorem of calculus that

$$T_k s_k = (F'(x_{k+1}) - F'(x_k))^*(F(x_{k+1}) - y^\delta) = y_k. \tag{3.40}$$

Hence, there exists a $K > 0$ such that

$$\begin{aligned}
\|T_k - Z_*\|_{HS} &\le \int_0^1 \|(F''(x_k + ts_k)(\cdot))^*(F(x_{k+1}) - F(x^*))\|_{HS} dt \\
&\quad + \int_0^1 \|(F''(x_k + ts_k)(\cdot) - F''(x^*)(\cdot))^*(F(x^*) - y^\delta)\|_{HS} dt \\
&\le M_{F''}\|F(x_{k+1}) - F(x^*)\|_{HS} + \delta_{HS}\int_0^1 \|F''(x_k + ts_k)(\cdot) - F''(x^*)(\cdot)\|_{HS} dt \\
&\le M_{F''}\|x_{k+1} - x^*\| + \frac{\delta_{HS}\xi_{\max}L_{F''}}{2}(\|x_{k+1} - x^*\| + \|x_k - x^*\|) \\
&\le K\mu_k, \qquad (3.41)
\end{aligned}$$

which implies that $T_k \in D_k$ and $\|T_k - Z_*\|_{HS}$ converges to zero, since x_k converges to x^*. As $A_0 - Z_*$ is a Hilbert–Schmidt operator, it follows inductively that $A_k - Z_*$ is a Hilbert–Schmidt operator for all $k \in \mathbb{N}$, as only a rank-2 operator is added to obtain $A_{k+1} - Z_*$. Similarly to Lemma 3.6 we use the orthogonal projector defined in Proposition 2.5 to reformulate the distance between A_{k+1} and Z_* and obtain

$$A_{k+1} - Z_* = P(A_k - Z_*)P + \frac{s_k \otimes (y_k - Z_* s_k)}{\langle s_k, s_k\rangle}P + \frac{(y - Z_* s_k) \otimes s_k}{\langle s_k, s_k\rangle}.$$

Using the triangle inequality in the Hilbert-Schmidt norm and Proposition 2.5 we have

$$\|A_{k+1} - Z_*\|_{HS} \le \|A_k - Z_*\|_{HS} + 2\frac{\|y_k - Z_* s_k\|_{HS}}{\|s_k\|}. \tag{3.42}$$

By the fundamental theorem of calculus and the Lipschitz continuity of F'' on U we obtain similarly to inequality (3.41)

$$\|y_k - Z_*\|_{HS} \le K\|s_k\|\mu_k.$$

Therefore,

$$\|A_{k+1} - Z_*\|_{HS} \le \|A_k - Z_*\|_{HS} + 2K\mu_k. \tag{3.43}$$

Due to the linear convergence of x_k to x^*, we know

$$\|x_{k+1} - x^*\| \le \nu\|x_k - x^*\|$$

for all $k \in \mathbb{N}$, where $\nu \in (0, 1)$. Thus,

$$\|A_{k+1} - Z_*\|_{HS} \le \|A_k - Z_*\|_{HS} + 2K\|x_k - x^*\| \tag{3.44}$$

yielding that for all $m > n$, we have

$$\begin{aligned}
\|A_m - Z_*\|_{HS} &\le \|A_n - Z_*\|_{HS} + 2K\sum_{k=n}^{m-1}\|x_k - x^*\| \le \|A_n - Z_*\|_{HS} + 2K\sum_{k=n}^{\infty}\nu^k\|x_0 - x^*\| \\
&\le \|A_n - Z_*\|_{HS} + 2K\frac{\nu^n}{1-\nu}\|x_0 - x^*\|. \qquad (3.45)
\end{aligned}$$

Consequently, $\|A_k - Z_*\|_{HS}$ is a Cauchy sequence, and thus convergent. Furthermore, we know that T_k defined in (3.39) converges to Z_*. Therefore, $\|A_k - T_k\|_{HS}$ and $\|A_{k+1} - T_k\|_{HS}$ converge to the same limit. Due to (3.38) we have that

$$\lim_{k\to\infty}\|A_{k+1} - A_k\|_{HS}^2 \le \lim_{k\to\infty}\left(\|A_k - T_k\|_{HS}^2 - \|A_{k+1} - T_k\|_{HS}^2\right) = 0$$

and hence,

$$\lim_{k\to\infty}\|A_{k+1} - A_k\|_{HS} = 0,$$

which concludes the proof. □

As $\{\|B_k\|\}_{k\in\mathbb{N}}$ is a bounded sequence, we know that

$$\|A_k\| \leq \|B_k\| + \|C(x_k)\| + \alpha_k$$

yielding that $\{\|A_k\|\}_{k\in\mathbb{N}}$ is a bounded sequence due to the boundedness of F' on D and condition (3.24). Applying the regularized structure-exploiting PSB method can also be interpreted as applying a gradient step to

$$\hat{\mathcal{J}}_k(x) := \frac{1}{2}\left\|F'(x_k)(x-x_k) + F(x_k) - y^\delta\right\|^2 + \frac{1}{2}\langle x - x_k, A_k(x-x_k)\rangle + \frac{\alpha_k}{2}\|x-x_k\|^2. \tag{3.46}$$

The first term reflects the idea of Newton's method, namely to linearize the non-linear operator equation $F(x) = y^\delta$ around an approximate solution x_k, yielding the linearized equation

$$F'^*(x_k)(x-x_k) = y^\delta - F(x_k).$$

Note, that the linearized equation is still ill-posed. The second term of (3.46) is bounded due to $\{\|A_k\|\}_{k\in\mathbb{N}}$ being bounded and can therefore be interpreted as a term that controls the distance to the point around which the linearization of the non-linear operator equation was performed. Lastly, the third term of (3.46) is the standard iterated Tikhonov regularization term used for deriving the LM method. Furthermore, the LM update can be obtained by applying a gradient step to

$$\widetilde{\mathcal{J}}_k(x) := \frac{1}{2}\left\|F'^*(x_k)(x-x_k) + F(x_k) - y^\delta\right\|^2 + \frac{\alpha_k}{2}\|x-x_k\|^2.$$

For the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the regularized structure-exploiting PSB method in Hilbert spaces, i.e., B_k as in (3.21), where A_k is updated according to (3.22) using (1.3) and (3.16), one obtains

$$J(x^*) = \lim_{k\to\infty} \mathcal{J}_k(x_{k+1}) = \lim_{k\to\infty} \hat{\mathcal{J}}_k(x_{k+1}) = \lim_{k\to\infty} \widetilde{\mathcal{J}}_k(x_{k+1}) + \frac{1}{2}\langle x_{k+1} - x_k, A_k(x_{k+1}-x_k)\rangle.$$

4. A GLOBALIZATION APPROACH

In order to computationally solve the inverse problem, one can follow the optimize-then-discretize approach, meaning that we have to discretize the proposed methods after formulating them in function spaces. On the other hand, one can use the discretize-then-optimize approach where we discretize the problem setting, i.e., the forward operator and the corresponding spaces, first and optimize afterwards. In the latter approach, easy access to the first derivative of the forward operator F is provided by Algorithmic Differentiation (AD). Developed over the last few decades, this method offers an outstanding way to provide derivative information for a given code segment [18]. The central concept is that the computation of a discretized operator can be decomposed into a finite sequence of elementary operations such as addition, multiplication, and elementary function calls. By calculating the derivatives with respect to the arguments of these operations, which can be easily computed, one has the necessary tool to systematically apply the chain rule in order to arrive at the derivatives of the entire sequence of operations with respect to the input variables. An advantage of this method is that the derivative information provided is as accurate as possible in a computationally measurable sense, i.e., to machine precision. Based on the starting point, a distinction can be drawn between the forward mode and the reverse mode of AD. In our context, the forward mode is similar to a sensitivity approach, while the reverse mode is a discrete analogue of the adjoint-based calculation of gradients. As any Newton-type method is highly dependent on the quality of the first derivative in terms of consistency and high precision [17], we explicitly recommend using AD tools in the discretized framework.

For the discretize-then-optimize approach we have $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, where $m, n \in \mathbb{N}$. Consequently, we are interested in regularized non-linear least-squares problems with the forward function $F : \mathbb{R}^n \to \mathbb{R}^m$ being twice continuously differentiable. We define analogously to (3.12) the secant condition

$$y_k = (F'(x_{k+1}) - F'(x_k))^T r_{k+1} \tag{4.1}$$

and

$$C(x_k) := F'(x_k)^T F'(x_k) \tag{4.2}$$

$$Z(x_k) := F''(x_k)^T \left(F(x_k) - y^\delta\right). \tag{4.3}$$

Next, we discuss the PSB update in the finite dimensional setting, which belongs to the general class of symmetric secant updates for QN methods (cf. [14]) and can be represented by

$$A_{k+1} = A_k + \frac{(y_k - A_k s_k)v_k^T + v_k(y_k - A_k s_k)^T}{v_k^T s_k} - \frac{(y_k - A_k s_k)^T s_k}{(v_k^T s_k)^2} v_k v_k^T. \tag{4.4}$$

Therein, the vector $v_k \in \mathbb{R}^n$ is a scaling vector that can be a function of s_k, y_k, and A_k. In the PSB case, we set $v_k = s_k$.

Corollary 4.1 (Local convergence and convergence rates)**.** Let the assumptions of Theorem 3.7 be satisfied. Then, the sequence $\{x_k\}$ obtained by the regularized structure-exploiting PSB method, i.e., B_k as in (3.21), where A_k is updated according to (4.4) using (1.3) and (4.1), is well-defined and converges Q-super-linearly to x^*. Furthermore, the sequences $\{\|B_k\|\}$ and $\{\|(B_k)^{-1}\|\}$ are uniformly bounded.

Proof. As $\mathbb{R}^n$ is a finite dimensional separable Hilbert space for all $n \in \mathbb{N}$, Theorem 3.7 guarantees that the sequence $\{x_k\}_{k\in\mathbb{N}}$ obtained by the regularized structure-exploiting PSB method is well-defined and converges linearly to x^*. Since all operators are now finite-rank, they are Hilbert-Schmidt operators, especially $A_0 - Z(x^*)$. Furthermore, all norms on a finite dimensional vector space are equivalent, yielding that Assumption A2 holds with respect to the Hilbert–Schmidt norm. Then, Theorem 3.8 implies Q-super-linear convergence to x^*. □

On the one hand, it is necessary to regularize the ill-posed problem and thus obtain a well-posed problem, while on the other hand, one is usually interested in obtaining a globally convergent algorithm that is resistant to the quality of the initial guess. Moreover, a common difficulty in solving inverse problems in practice is the specific choice of the regularization parameter. Therefore, we want to assess whether we can address these two issues by controlling the regularization parameter in such a way that it leads to globalization.

To obtain global convergence, QN methods are usually combined with a line search or trust region method. Kanzow and Steck [23] have already conducted extensive studies on this topic, which serve as a basis for us. They have developed a globalization strategy in which the Hessian approximation is regularized and the regularization parameter is controlled by combining some of the respective advantages of line search and trust region methods. Therefore, our framework uses the results of [23] to improve robustness by regularization to stabilize the solution process for a better handling of poor initial guesses.

Remark 4.2. Note that this globalization is also applicable for the LM method by simply setting $A_k = 0$ for all $k \in \mathbb{N}_0$.

We now proceed with analyzing the convergence of Algorithm 1 both in the PSB case and in the LM case.

Algorithm 1 Global regularized structure-exploiting QN (GRSE)

Require: $A_0 \in \mathbb{R}^{n\times n}; \alpha_0, \eta > 0; x_0 \in \mathbb{R}^n; g_0 = \|\nabla J(x_0)\|; y^\delta \in \mathbb{R}^m; c, p, \theta \in (0,1), \sigma > 1, k = 0$

while $g_k \geq \eta$ **do**

Evaluate the forward function $F(x_k)$, compute its derivative $F'(x_k)$, the residual $r_k = F(x_k) - y^\delta$ and the gradient $\nabla J(x_k) = F'(x_k)^T r_k$.

Solve for s_k:

$$\left(F'(x_k)^T F'(x_k) + A_k + \alpha_k I\right) s_k = -\nabla J(x_k) \tag{4.5}$$

Compute $\text{pred}_k = \frac{\alpha_k}{2}\|s_k\|^2 - \frac{1}{2}\nabla J(x_k)^T s_k,\ \text{ared}_k = J(x_k) - J(x_k + s_k)$ and $\rho_k = \frac{\text{ared}_k}{\text{pred}_k}$

if Eq. (4.5) admits no solution or $\rho_k \leq c$ or $\text{pred}_k \leq p g_k \|s_k\|$: **then**

Set $\alpha_{k+1} = \sigma\alpha_k$, $x_{k+1} = x_k$ and $A_{k+1} = A_k$ (unsuccessful step)

else

Set $x_{k+1} = x_k + s_k$, $\alpha_{k+1} = \theta\alpha_k$ and compute $g_{k+1} = \|\nabla J(x_{k+1})\|$ (successful step)

Compute A_{k+1} according to (4.4) with (4.1) and $v_k = s_k$.

end if

$k = k + 1$

end while

Theorem 4.3 (Global Convergence)**.** *Let $V \subset X$ be bounded and let the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 be a subset of V, where the termination criterion is neglected, i.e., $\eta = 0$. Let J be bounded from below, C be bounded, and Z be Lipschitz continuous on V. Then, for the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 it holds that*

$$\liminf_{k\to\infty} \|\nabla J(x_k)\| = 0.$$

Furthermore, if ∇J is uniformly continuous on V, then it holds for the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 that

$$\lim_{k\to\infty} \|\nabla J(x_k)\| = 0.$$

Proof. Preliminarily, we prove that Algorithm 1 performs infinitely many successful steps. Therefore, we assume the contrary, namely that there exists $k_0 \in \mathbb{N}$ such that all steps with index $k \geq k_0$ are unsuccessful. This yields that $\alpha_k \to \infty$ for $k \to \infty$ as well as $x_k = x_{k_0}$ for all $k \geq k_0$. Consequently,

$$C(x_k) + A_k = F'(x_k)^T F'(x_k) + A_k = F'(x_{k_0})^T F'(x_{k_0}) + A_{k_0} = C(x_{k_0}) + A_{k_0}$$

for all $k \geq k_0$. Thus, for sufficiently large $k \geq k_0$ we know that $C(x_{k_0}) + A_{k_0} + \alpha_k I$ is invertible and that

$$\lim_{k\to\infty} \frac{(C(x_{k_0}) + A_{k_0} + \alpha_k I)\, s}{\|(C(x_{k_0}) + A_{k_0} + \alpha_k I)\, s\|} = \frac{s}{\|s\|}.$$

This implies that $s_k \to 0$ for $k \to \infty$ and

$$\lim_{k\to\infty} \frac{s_k}{\|s_k\|} = -\frac{\nabla J(x_{k_0})}{\|\nabla J(x_{k_0})\|}.$$

Observing $(C(x_{k_0}) + A_{k_0} + \alpha_k I)s_k = -\nabla J(x_{k_0})$ for $k \geq k_0$ yields that

$$\lim_{k\to\infty} \alpha_k \|s_k\| = \|\nabla J(x_{k_0})\|.$$

With this, arguing analogously as in the proof of Lemma 2 in [23] leads to a contradiction. Hence, the set of indices $\mathcal{S} \subset \mathbb{N}$ of successful steps of Algorithm 1 has infinite cardinality. Now, as every step $k \in \mathcal{S}$ is successful, we have for every $k \in \mathcal{S}$ that

$$J(x_k) - J(x_{k+1}) \geq c\, \mathrm{pred}_k \geq pc\|\nabla J(x_k)\|\|s_k\|.$$

We assume for the sake of contradiction that

$$\liminf_{k\to\infty} \|\nabla J(x_k)\| > 0, \tag{4.6}$$

which yields that there exist $k_0 \in \mathbb{N}$ and $\epsilon > 0$ such that $\|\nabla J(x_k)\| \geq \epsilon$ for all $k \geq k_0$. As J is bounded from below and x_k is not updated in unsuccessful steps, we obtain

$$\infty > \sum_{k\in\mathbb{N}} (J(x_k) - J(x_{k+1})) = \sum_{k\in\mathcal{S}} (J(x_k) - J(x_{k+1})) \geq pc\epsilon \sum_{k\in\mathcal{S}, k\geq k_0} \|s_k\|.$$

This particularly means that $\sum_{k\in\mathcal{S},k\geq k_0} \|s_k\| < \infty$. Since $\sum_{k\in\mathcal{S},k<k_0} \|s_k\|$ is bounded as it is a finite sum, we conclude that there exists a constant $M_S > 0$ such that

$$\sum_{k\in\mathcal{S}} \|s_k\| = M_S < \infty.$$

Thus, $s_k \to_{\mathcal{S}} 0$ for $k \to \infty$. By assumption Z is Lipschitz continuous on V. Hence, Z is bounded on V and Lemma 2 in [15] is applicable, which yields that there exist constants $d_1, d_2 \geq 0$ such that

$$\|A_{k+1}\| \leq d_1 + d_2 \sum_{j\in\mathcal{S}} \|s_j\|. \tag{4.7}$$

For every step $k \in \mathcal{S}$ we have $(C(x_k) + A_k + \alpha_k I)s_k = -\nabla J(x_k)$. By denoting the bound of C with $M_C > 0$, we obtain

$$\begin{aligned}\|\nabla J(x_k)\| &\leq \|C(x_k)\|\|s_k\| + \|A_k\|\|s_k\| + \alpha_k\|s_k\| \\ &\leq (M_C + d_1 + d_2 M_S)\|s_k\| + \alpha_k\|s_k\|.\end{aligned}$$

Therefore, every subsequence of $\{\alpha_k\}_{k\in\mathcal{S}}$ must be unbounded, since otherwise $\|\nabla J(x_{k_l})\|$ converges to 0 for $l \to \infty$ as $s_{k_l} \to_{\mathcal{S}} 0$ for $l \to \infty$, which violates (4.6). Hence one obtains $\alpha_k \to \infty$, meaning that

Algorithm 1 also performs infinitely many unsuccessful steps. With this, arguing now analogously as in the proof of Theorem 1 in [23] yields a contradiction to (4.6). Consequently, $\liminf_{k\to\infty} \|\nabla J(x_k)\| = 0$.
If ∇J is uniformly continuous on V, arguing in the same way as in the proof of Theorem 2 in [23], it follows that $\lim_{k\to\infty} \|\nabla J(x_k)\| = 0$, since the argumentation therein does not distinguish specifically between successful and highly successful steps.
If the LM update is chosen, $\{A_k\}_{k\in\mathbb{N}}$ is a sequence of zero matrices and C is bounded on V as depicted before, yielding boundedness of $\{C(x_k) + A_k\}_{k\in\mathbb{N}}$. Hence, we can directly apply Lemma 1, Lemma 2, Theorem 1 and Theorem 2 in [23]. □

Remark 4.4. Theorem 4.3 ensures that, given any $\eta > 0$, Algorithm 1 terminates with $\|\nabla J(x_k)\| \le \eta$ after finitely many iterations.

5. NUMERICAL EXPERIMENT

In this section, we analyze our algorithms by applying them to a suitable parameter identification problem using a parameter-dependent PDE model and additional observations or measurements. Furthermore, we may also consider F as a vector-valued forward operator, such that F'' would be a 3-tensor-valued operator, which however results in a matrix-valued operator when applied to the residual. Specifically, we want to identify the material parameters of a piezoelectric specimen. Piezoelectric components are widely used in electronic devices, ranging from everyday items such as electric toothbrushes and headphones to advanced medical and industrial applications such as ultrasound imaging and fuel injectors. Therefore, an accurate characterization of the material parameters is crucial for the design and simulation of reliable sensors and actuators.
As geometry, we consider a piezoelectric ring with an outer radius of 6.35 mm, an inner radius of 2.6 mm, and a thickness of 1 mm. To reduce the dimension of the considered material parameter matrices and the underlying PDE model and thus to reduce computational effort, we exploit the inherent rotational symmetry and transform the ring into a rectangular domain by adopting cylindrical coordinates rather than Cartesian coordinates, where the z-axis is selected as the axis of rotation. Hence, we consider Ω as a rectangle with vertices $(2.6, 0)$, $(6.35, 0)$, $(6.35, 1)$, $(2.6, 1)$, where coordinates are given in mm. Consequently, we obtain a Lipschitz domain $\Omega \subset \mathbb{R}^2$ and assume that its boundary can be represented as the disjoint union $\partial\Omega := \Gamma_a \dot\cup \Gamma_0 \dot\cup \Gamma_n$, where Γ_a is the boundary segment electrically excited with a spatially independent (equally distributed) excitation signal, Γ_0 is the grounded boundary segment, and Γ_n is the remaining part of $\partial\Omega$. We introduce the state space

$$W := \left(H^2_{\mathcal{B}}(\Omega, \mathbb{C}^2) \times H^2_{0,\Gamma}(\Omega, \mathbb{C})\right),$$

where

$$\nabla := \begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial z} \end{pmatrix} \text{ and } \mathcal{B} := \begin{pmatrix} \frac{\partial}{\partial r} & 0 \\ \frac{1}{r} & 0 \\ 0 & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial z} & \frac{\partial}{\partial r} \end{pmatrix}.$$

It coincides with the solution space of the following Fourier-transformed PDE-system

$$\forall \omega \in \mathcal{W}: \ -\rho\omega^2 d_1 \hat{u} - \mathcal{B}^{\mathrm{T}}\left(d_2 c^{\mathrm{E}} \mathcal{B}\hat{u} + e^{\mathrm{T}}\nabla\hat{\phi}_0\right) = \mathcal{B}^{\mathrm{T}} e^{\mathrm{T}} \nabla\chi \quad \text{in } \Omega \tag{5.1}$$

$$-\nabla \cdot \left(e\mathcal{B}\hat{u} - \varepsilon^{\mathrm{S}}\nabla\hat{\phi}_0\right) = -\nabla \cdot \varepsilon^{\mathrm{S}}\nabla\chi \ \text{ in } \Omega \tag{5.2}$$

$$n \cdot \left(e\mathcal{B}\hat{u} - \varepsilon^{\mathrm{S}}\nabla\hat{\phi}_0\right) = n \cdot \varepsilon^{\mathrm{S}}\nabla\chi \quad \text{on } \Gamma_{\mathrm{n}} \tag{5.3}$$

$$\mathcal{N}^{\mathrm{T}}\left(d_2 c^{\mathrm{E}}\mathcal{B}\hat{u} + e^{\mathrm{T}}\nabla\hat{\phi}_0\right) = -\mathcal{N}^{\mathrm{T}} e^{\mathrm{T}}\nabla\chi \text{ on } \partial\Omega, \tag{5.4}$$

where $s = (\hat{u}, \hat{\phi}_0)$ is the solution of the system, $\rho \in \mathbb{R}^+$, $d_1 := 1 - \mathrm{i}\frac{\alpha}{\omega}$, $d_2 := 1 + \mathrm{i}\omega\beta$. We denote the space of angular frequencies with $\mathcal{W} \subset \mathbb{R}^+$, $|\mathcal{W}| < \infty$. Furthermore, n is the normal element corresponding to ∇ and $\mathcal{N}$ the normal element corresponding to $\mathcal{B}$. Additionally, we included the Rayleigh damping model, with $\alpha, \beta \in \mathbb{R}^+_0$ as Rayleigh damping parameters. The mixed Dirichlet boundary conditions, needed to model the excitation behavior, were homogenized using the Dirichlet lift ansatz with Dirichlet lift function $\chi(\omega)$, see [24]. We assume that the piezoelectric coupling parameter e is unknown and has to be reconstructed, while the elasticity parameter c^{E} and the permittivity ε^{S} are known. The piezoelectric coupling parameter can be described by the matrix

$$e := \begin{pmatrix} 0 & 0 & 0 & e_{15} \\ e_{31} & e_{31} & e_{33} & 0 \end{pmatrix}. \tag{5.5}$$

Consequently, we define the parameter space X

$$X := \Big\{ p \in \mathbb{R}^3 : \ p_1 = e_{15}, \ p_2 = e_{31}, \ p_3 = e_{33} \Big\}.$$

and the piezoelectric model operator $A_\omega : X \times W \to W^*$ for each $\omega \in \mathcal{W}$ via

$$\begin{aligned} \langle A_\omega(p,s),(v,w)\rangle_{W^*,W} := 2\pi \int_\Omega \Bigg(& - d_1 \rho \omega^2 \hat{u}^{\mathrm{T}} \overline{v} + \Big(d_2 c^{\mathrm{E}} \mathcal{B} \hat{u} + e^{\mathrm{T}} \nabla \hat{\phi}_0 \Big)^{\mathrm{T}} \overline{\mathcal{B} v} \\ & + \Big(e \mathcal{B} \hat{u} - \varepsilon^{\mathrm{S}} \nabla \hat{\phi}_0 \Big)^{\mathrm{T}} \overline{\nabla w} + (e^{\mathrm{T}} \nabla \chi)^{\mathrm{T}} \overline{\mathcal{B} v} - (\varepsilon^{\mathrm{S}} \nabla \chi)^{\mathrm{T}} \overline{\nabla w} \Bigg) r \, \mathrm{d}(r,z) \, . \end{aligned} \tag{5.6}$$

To recover information on the parameter p, we need observed data with respect to the state s and the parameter. Hence, we define the charge pulse approximation, see Chapter 5.1 in [29], $Q_\omega : \ X \times W \to \mathbb{C}$,

$$Q_\omega(p,z) = 2\pi \int_{\Gamma_a} r \left(e(\theta) \mathcal{B} \hat{u} - \varepsilon^{\mathrm{S}}(\theta) \nabla \left(\hat{\phi}_0 + \chi \right) \right) \cdot n \, \mathrm{d}(r,z) \tag{5.7}$$

for each $\omega \in \mathcal{W}$. We assume that $\|Q_\omega\|_{\mathbb{C}} > 0$, which is physically motivated. Next, we define the observation operator $O_\omega : \ X \times W \to \mathbb{R}$ for each $\omega \in \mathcal{W}$ as

$$C_\omega(p,z) = \log \left(\|Q_\omega\|_{\mathbb{C}} \right). \tag{5.8}$$

To model the inverse problem, we employ the reduced approach, meaning that we have to eliminate the model by introducing a so-called parameter-to-state map S_ω for each $\omega \in \mathcal{W}$, which maps each parameter to the corresponding solution of the underlying PDE model (5.1)-(5.4). As observed data is usually contaminated with noise, we consider noisy data $y^\delta \in \mathbb{R}^{|\mathcal{W}|}$. Then, the forward operator $F : X \to \mathbb{R}^{|\mathcal{W}|}$ is defined by

$$F(p) = \left(C_\omega(p, S_\omega(p)) \right)_{\omega \in \mathcal{W}}.$$

Consequently, we want to identify $p \in X$ such that

$$F(p) = y^\delta. \tag{5.9}$$

Using Proposition 1 in [24], we know that for each $\omega \in \mathcal{W}$, A_ω is well-defined, bounded, bijective, and continuously Fréchet differentiable on $X \times W$, S_ω is well-defined, non-linear, and continuously Fréchet differentiable on X. Due to the affine linearity of A_ω with respect to s and p, we conclude that A_ω is three times continuously Fréchet differentiable on $X \times W$. Using the implicit function theorem, S_ω is three times continuously Fréchet differentiable on X. Due to its affine linear structure in the state, we conclude that Q_ω is also three times continuously Fréchet differentiable on $X \times W$. Hence, F is Lipschitz continuously Fréchet differentiable. Since F is a mapping between finite dimensional Hilbert spaces, the Hilbert-Schmidt and symmetry assumptions are naturally satisfied.

To implement and numerically solve the inverse problem, we employ a discretize-then-optimize approach. For this purpose, the full problem setting is discretized in advance to take advantage of AD. In particular, we use a classical finite element method (FEM), realized with the finite element tool FEniCS [1] in DOLFIN version 2019.2.0.dev0, using AD via the dolfin adjoint [28] library of FEniCS in version 2019.1.0. We use $h = 150\mu$m as the FEM element size and a polynomial degree of $g = 3$. Furthermore, we use an excitation signal $\hat{\phi}^{\mathrm{e}} = 0.03$ V and true material parameters and damping parameters as in [16], where we perform any numerical realizations in kHz. Note that specifying the Dirichlet lift function χ used for stating the system (5.1)-(5.4) is not necessary and can be avoided in practice, as it is possible to directly implement mixed Dirichlet conditions in FEniCS. We consider the frequency and angular frequency domain

$$\begin{aligned} \mathcal{F} &:= \{ f \in \mathbb{N} : \ f \equiv 0 \bmod 5, \ 50 \leq \omega < 4550 \} \\ \mathcal{W} &:= \{ \omega = 2\pi f \in \mathbb{R}^+ : f \in \mathcal{F} \}. \end{aligned} \tag{5.10}$$

To incorporate potential model errors, we generate the noisy observation data using the charge pulse approximation, see Chapter 5.1 in [29], i.e., $Q_\omega : X \times W \to \mathbb{C}$,

$$Q_\omega(p,s) = 2\pi \int_\Omega r \left(e\mathcal{B}\hat{u} - \varepsilon^{\mathrm{S}} \nabla \hat{\phi} \right) \nabla \hat{\phi} \, \mathrm{d}(r,z) \tag{5.11}$$

and contaminate this simulated data additively with uniformly distributed random noise with a noise level of 5%. We employ the hyperparameters $\theta = 0.5$, $\alpha_0 = 10^{-4}$, $\sigma = 4$, $c = p = 10^{-4}$ and $\eta = 7e-5$. As initial Quasi-Newton matrices for each block, we use $F''(p_0)^* \left(F(p_0) - y^\delta \right)$, where $F''(p_0)$ is computed via AD.

For the sake of better visualization, the reconstructed parameter values have been normalized by the respective true parameter in all figures, so that convergence to the value 1 is desired. Furthermore, we assumed that the initial guess is an overestimate. In our numerical tests, underestimating the true parameter yielded similar results.

We start by focusing on the numerical results of the local method. Therefore, we choose a sufficiently small deviation of the initial guess from the true parameter of 5%, which means that the initial guess is 5% larger than the true parameter. We compare the performance during the parameter identification process of the RSE-PSB method and the LM method in Figure 1. These two methods differ in the choice of A_k, as in the RSE-PSB method A_k is computed in each step $k \in \mathbb{N}$ using the PSB method under the modified secant condition (3.12), and in the LM method A_k is kept constant at 0. As a stopping criterion, we employ a gradient threshold, i.e., the optimization process is terminated once the norm of the gradient of the objective function is smaller or equal to η. The reconstruction results of the RSE-PSB method show clear similarity to the LM method, which is to be expected at this point since the deviation of the initial guess from the ground truth is small. Consequently, the approximation error of the LM method is correspondingly small. Finally, the reconstruction results of the RSE-PSB exhibit super-linear convergence behavior, which can be recognized in Figure 2.

Note that, even with gradients obtained via algorithmic differentiation, super-linear convergence may deteriorate once the gradient norm falls below a moderate threshold such as η. This occurs because small variations in the gradient and curvature near a minimizer amplify floating-point round-off errors. As a result, the effective numerical noise floor, which is several orders of magnitude above machine precision, precludes super-linear convergence behavior.

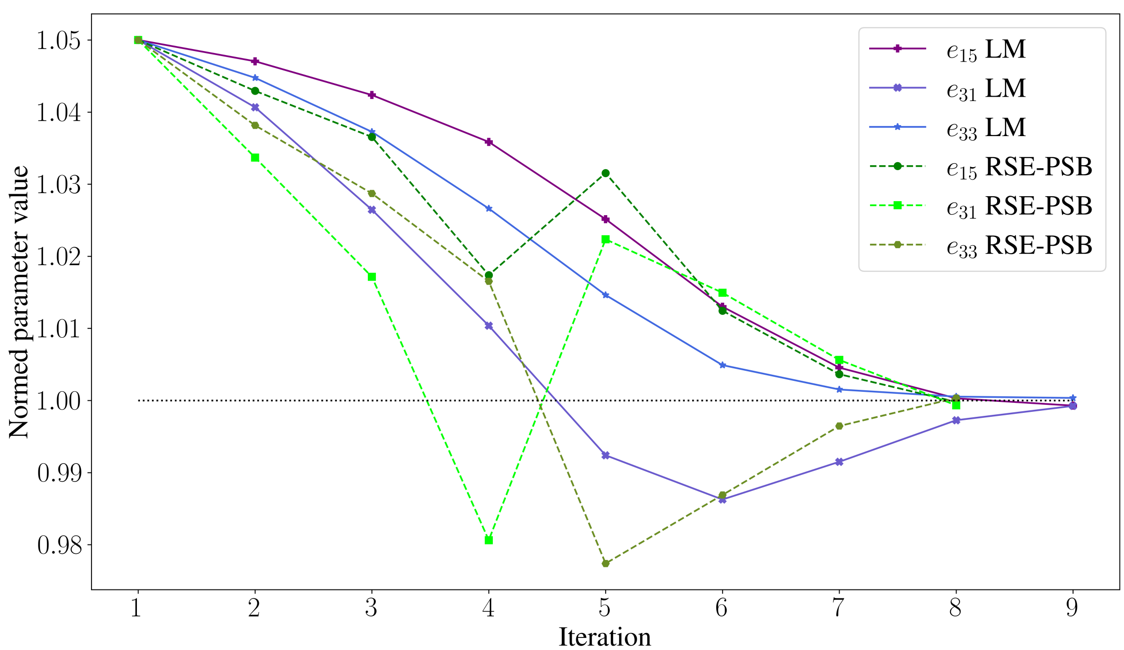

FIGURE 1. Performance of RSE-PSB and LM with 5% deviation of the initial guess from the ground truth.

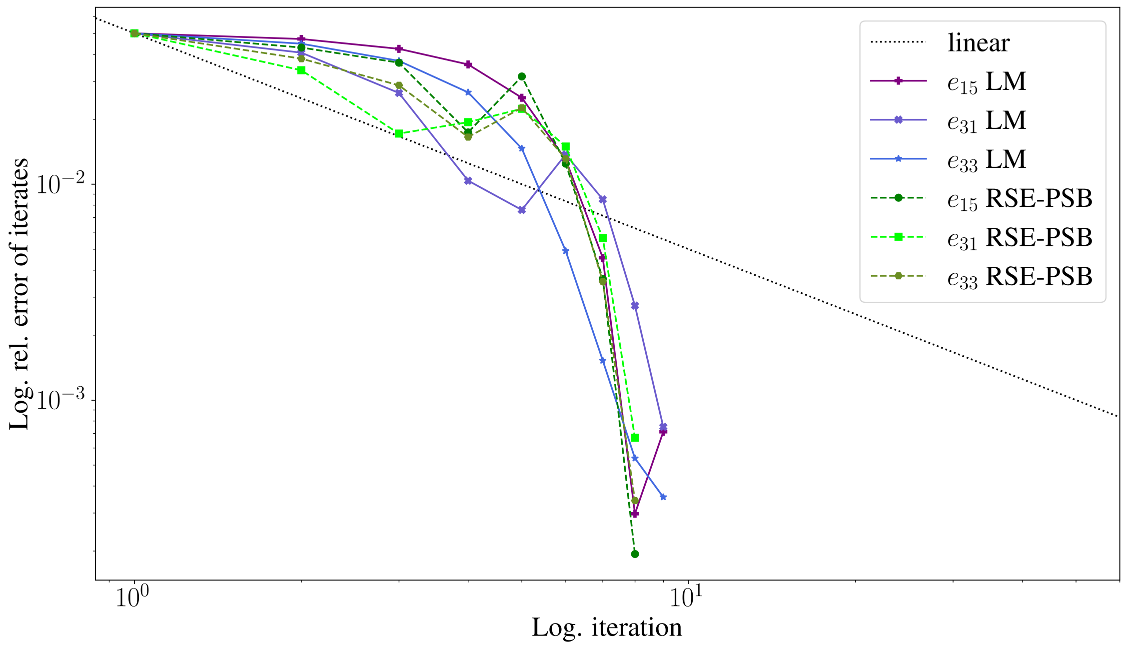

FIGURE 2. Local super-linear convergence behavior of RSE-PSB.

However, since in applications the quality of the initial guess may not be sufficient for local methods, we now focus on the numerical results of the global methods. For this purpose, we choose a deviation of the initial guess from the true parameter of 50%. We will now take a closer look at the globalized RSE-PSB method (GRSE-PSB) and the globalized LM method (G-LM). In Figure 3, we compare the performance of the GRSE-PSB method and the G-LM method during the parameter identification process. It is evident

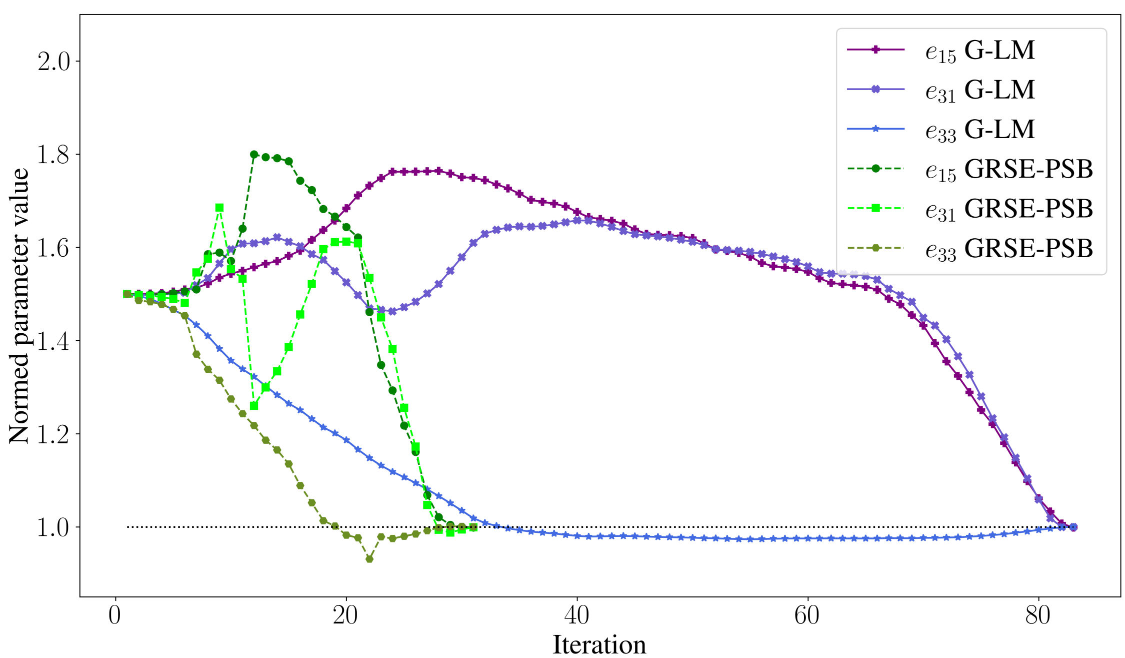

FIGURE 3. Performance of GRSE-PSB and G-LM with 50% deviation of the initial guess from the ground truth.

that the G-LM method requires roughly three times as many iteration steps as the GRSE-PSB method to satisfy the specified threshold η. This can be explained by the fact that the approximation error is large due to the large deviation of the initial guess. Nevertheless, the reconstruction results for both methods show convergence to the true parameter.
Note that with all deviations above, every iteration step remains in the feasible set, as all parameters must be positive to maintain well-posedness of the forward problem.

6. Conclusion

This paper proposed a regularized, structure-exploiting Powell-Symmetric-Broyden (RSE-PSB) method designed for solving ill-posed inverse problems in both infinite dimensional and finite dimensional settings. By deriving a modified secant condition in a Hilbert space setting and approximating the symmetric, yet potentially indefinite, second-order term $(F''(x_k)(\cdot))^*(F(x_k) - y^\delta)$, which is typically neglected in standard Levenberg-Marquardt approaches, we presented a method that overcomes challenges inherent in ill-posedness. It leverages the problem structure not only by decomposing into data discrepancy and regularization terms but also by exploiting the structure of the analytical second Fréchet derivative to approximate the symmetric second-order term. We discussed the local convergence of the method, where we proved local Q-linear convergence via the BDP and established local Q-super-linear convergence under the assumption that the initial error is a Hilbert-Schmidt operator.
To ensure robustness in practical applications, particularly when the initial guess significantly deviates from the true solution, we developed a globalization approach that employs dynamic control of the regularization parameter, ensuring global convergence and simultaneously stabilizing the ill-posed problem. Thus, the control strategy for the regularization parameter also serves to address the specific choice of the parameter. Finally, the efficiency of the method was demonstrated on a PDE-based parameter identification problem in piezoelectricity using AD.

Acknowledgement

The authors would like to thank the German Research Foundation (DFG) for financial support within the research group 5208 NEPTUN (444955436). The work on this paper was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689).

References

[1] M. Alnæs et al. "The FEniCS project version 1.5". In: *Archive of numerical software* 3 (2015).
[2] F. J. Aragón Artacho et al. "Local convergence of quasi-Newton methods under metric regularity". In: *Computational Optimization and Applications* 58.1 (2014), pp. 225–247.
[3] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer New York, 2011.
[4] B. Benahmed, B. de Malafosse, and A. Yassine. "Matrix Transformations and Quasi-Newton Methods". In: *International Journal of Mathematics and Mathematical Sciences* 2007.1 (2007), p. 025704.
[5] B. Benahmed et al. "Quasi-Newton methods in infinite-dimensional spaces and application to matrix equations". In: *Journal of Global Optimization* 49 (2011), pp. 365–379.
[6] E. H. Bergou, Y. Diouane, and V. Kungurtsev. "Convergence and complexity analysis of a Levenberg–Marquardt algorithm for inverse problems". In: *Journal of Optimization Theory and Applications* 185.3 (2020), pp. 927–944.
[7] J. J. Brust et al. "Compact representations of structured BFGS matrices". In: *Computational Optimization and Applications* 80.1 (2021), pp. 55–88.
[8] C. Clason. *Introduction to functional analysis*. Springer Nature, 2020.
[9] J. E. Dennis. *On some methods based on Broyden's secant approximation to the Hessian*. Tech. rep. Cornell University, 1971.
[10] J. E. Dennis, H. J. Martínez, and R. A. Tapia. "Convergence theory for the structured BFGS secant method with an application to nonlinear least squares". In: *Journal of Optimization Theory and Applications* 61 (1989), pp. 161–178.

[11] J. E. Dennis and H. F. Walker. “Convergence theorems for least-change secant update methods”. In: *SIAM Journal on Numerical Analysis* 18.6 (1981), pp. 949–987.

[12] J. E. Dennis and H. F. Walker. “Least-change sparse secant update methods with inaccurate secant conditions”. In: *SIAM Journal on Numerical Analysis* 22 (1985), pp. 760–778.

[13] A. L. Dontchev. “Generalizations of the Dennis–Moré theorem”. In: *SIAM Journal on Optimization* 22.3 (2012), pp. 821–830.

[14] J. R. Engels and H. J. Martínez. “Local and Superlinear Convergence for Partially Known Quasi-Newton Methods”. In: *SIAM Journal on Optimization* 1.1 (1991), pp. 42–56.

[15] R. Fletcher. “An algorithm for solving linearly constrained optimization problems”. In: *Mathematical Programming* 2 (1972), pp. 133–165.

[16] O. Friesen et al. “Estimation of piezoelectric material parameters of ring-shaped specimens”. In: *tm - Technisches Messen* (2024).

[17] R. Griesse and A. Walther. “Evaluating gradients in optimal control: continuous adjoints versus automatic differentiation”. In: *Journal of Optimization Theory and Applications* 122 (2004), pp. 63–86.

[18] A. Griewank and A. Walther. *Evaluating Derivatives*. Second edition. Society for Industrial and Applied Mathematics, 2008.

[19] E. Haber. “Quasi-Newton methods for large-scale electromagnetic inverse problems”. In: *Inverse problems* 21.1 (2004), p. 305.

[20] J. Huschens. “On the use of product structure in secant methods for nonlinear least squares problems”. In: *SIAM Journal on Optimization* 4.1 (1994), pp. 108–129.

[21] D. M. Hwang and C. T. Kelley. “Convergence of Broyden’s Method in Banach Spaces”. In: *SIAM Journal on Optimization* 2.3 (1992), pp. 505–532.

[22] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. De Gruyter, 2008.

[23] C. Kanzow and D. Steck. “Regularization of limited memory quasi-Newton methods for large-scale nonconvex minimization”. In: *Mathematical Programming Computation* 15.3 (2023), pp. 417–444.

[24] R. Kuess et al. “Identification of temperature-dependent material parameter functions in piezoelectricity”. In: *2025 International Congress on Ultrasonics, Paderborn, Germany*. AMA Service GmbH, 2025, pp. 134–137.

[25] M. Laumen. “A Kantorovich theorem for the structured PSB update in Hilbert space”. In: *Journal of Optimization Theory and Applications* 105 (2000), pp. 391–415.

[26] F. Mannel, H. O. Aggrawal, and J. Modersitzki. “A structured L-BFGS method and its application to inverse problems”. In: *Inverse Problems* 40.4 (2024), p. 045022.

[27] R. Meise and D. Vogt. *Introduction to functional analysis*. Oxford University Press, 2004.

[28] S. Mitusch, S. Funke, and J. Dokken. “dolfin-adjoint 2018.1: automated adjoints for FEniCS and Firedrake”. In: *Journal of Open Source Software* 4.38 (2019), p. 1292.

[29] V. Schulze. “Modeling and Optimization of Electrode Configurations for Piezoelectric Material”. PhD thesis. Humboldt-Universitaet zu Berlin, Mathematisch-Naturwissenschaftliche Fakultaet, 2023.

[30] R. G. Vuchkov, C. G. Petra, and N. Petra. “On the derivation of quasi-Newton formulas for optimization in function spaces”. In: *Numerical Functional Analysis and Optimization* 41.13 (2020), pp. 1564–1587.

[31] H. Yabe and N. Yamaki. “Local and superlinear convergence of structured quasi-Newton methods for nonlinear optimization”. In: *Journal of the Operations Research Society of Japan* 39.4 (1996), pp. 541–557.

[32] W. H. Yu. “A quasi-Newton approach to identification of a parabolic system”. In: *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics* 40.1 (1998), pp. 1–22.

[33] W. H. Yu. “A quasi-Newton method in infinite-dimensional spaces and its application for solving a parabolic inverse problem”. In: *Journal of Computational Mathematics* 16.4 (1998), pp. 305–318.

[34] W. Zhou and X. Chen. “Global Convergence of a New Hybrid Gauss–Newton Structured BFGS Method for Nonlinear Least Squares Problems”. In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2422–2441.