ON REGULARIZED STRUCTURE EXPLOITING QUASI-NEWTON METHODS FOR ILL-POSED PROBLEMS

RAPHAEL KUESS AND ANDREA WALTHER

ABSTRACT. Inverse problems are inherently ill-posed, leading standard optimization techniques to fail and necessitating the use of regularization. This paper introduces a regularized, structure-exploiting Powell-Symmetric-Broyden method under modified secant conditions for solving ill-posed inverse problems in both infinite dimensional and finite dimensional settings. Our approach integrates regularization and structure exploitation directly within the Quasi-Newton framework, leveraging the strengths of Quasi-Newton methods, Tikhonov-type regularization, and structure exploitation. We provide a convergence analysis demonstrating local super-linear and weakly super-linear convergence in both infinite and finite dimensional settings.
Furthermore, we discuss a globalization approach based on the dynamic control of the regularization parameter, which not only ensures global convergence but also provides an a priori method for the specific choice of the regularization parameter. The iterative adaptation of the regularization parameter within this globalization framework assures convergence even when starting far from the true solution of both the proposed method and the Levenberg-Marquardt method.
By introducing compact representations of the proposed methods, we also enable efficient computation of Hessian approximations. These representations, particularly in limited-memory forms, are well-suited for large-scale inverse problems. Finally, we discuss a numerical example based on a PDE-driven parameter identification problem, relevant to industrial 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 the PDE from observed data. This class of problems is common in fields such as physics and biomedical imaging, where identifying the origin of signals or forces is crucial to understand 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 functions that appear as parameters within a mathematical model to best fit the observed or measured data. This type of problem arises frequently in engineering and material science, where parameters such as diffusion coefficients need to 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

Date: April 4, 2025.
2020 *Mathematics Subject Classification.* 49M15, 47J06, 65J20, 46N10 .
Key words and phrases. inverse problems, iterative regularization, globalization, Newton-type methods, convergence analysis.

equation

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

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. Frequently they are local methods, depending on the quality of the initial guess.
Well-known examples are the Landweber iteration, the Levenberg-Marquardt method (LM) and the iteratively regularized Gauss-Newton method (IRGN). The Landweber iteration is a method based on gradient descent with constant stepsize. The LM and IRGN method are popular approaches based on linearizing the operator equation to be solved and iteratively applying Tikhonov regularization. Although those methods are effective for many inverse problems, they are sensitive to the deviation of the initial guess. Unfortunately, in real-world problems a good initial guess is often difficult to find or not available at all. Therefore, effective optimization techniques are required for the regularized problem. As Newton-type methods use gradient information to provide an efficient approximation of the Hessian matrix, they typically converge fast and robust. Consequently, we will take this advantage into account and additionally incorporate regularization in Newton-type methods. Since inverse problems frequently share a very similar structure, we will also exploit that structure, resulting in a regularized structure exploiting (RSE) Quasi-Newton (QN) approach.

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) method, which all belong to the class of symmetric secant updates considered, e.g., by Dennis [6] 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 [8] in 1981 and has since evolved, see e.g., [7, 9, 17, 24].
A comprehensive study of structure-exploiting QN methods in finite dimensional spaces is given in [10]. 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 [16], the analysis of structure exploitation in finite dimensional least squares problems is further investigated in detail. Furthermore, in [5] 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., [15], where an electromagnetic inverse problem was decomposed in a data discrepancy term and a regularization term. Similarly, [21] 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, [23] derives QN formulas such as BFGS, DFP, SR1, and PSB via variational frameworks, while [25, 26] analyze their local convergence via bounded deterioration principles and discusses the application on well-posed inverse problems. Further insights into Hilbert-space applications and convergence are given in [2, 3].
The publications that are closest to this article are [19] and [18]. Firstly, in [19] a Kantorovich theorem and local convergence of a structured PSB method in Hilbert spaces is discussed. Secondly, in [18] 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, we have a slightly different setting and weaker assumptions for convergence than in [19] and [18] does not consider possibly unbounded QN Hessian approximations.

Contribution. Consequently, the first aim of this paper is to extend and generalize the previous theory by introducing two Tikhonov-type regularizers used additively in the target functional and deriving a modified secant equation in the Hilbert space setting. With this, we propose a local algorithm for regularized structure exploiting PSB methods with both Tikhonov-type regularizers. Then, local convergence results of [19] are extended to an adjusted framework, to weaker assumptions and to the inclusion of regularization, yielding a local convergence result for a regularized structure exploiting PSB method in Hilbert spaces.

Additionally, we address convergence behavior in this context and contribute results on the local super-linear and weakly super-linear convergence, respectively. Moreover, the relationship to the LM and the IRGN method are discussed.
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, a globalization approach in the discretized setting is presented, relying on the results of [18]. As the PSB method in general can have unbounded Hessian approximations, we establish a globalization result for the regularized structure exploiting PSB method using results of [12] and proof techniques of [18]. Subsequently, the proposed globalized algorithm is also shown to be locally super-linear convergent. Lastly, we also give a compact representation of the proposed method.
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 and directly computing the Hessian of the regularizer. This paper aims to fill this gap by presenting a comprehensive analysis of the proposed method in infinite dimensional spaces and a globalization approach in finite dimensional spaces, contributing novel insights and methodologies to the field.

Structure of the paper. The second section deals with the regularization and the exploitation of the problem structure in QN methods in a general Hilbert space setting. Therein, the derivation of a modified secant equation is given, and two regularization methods are discussed. Subsequently, using both regularization methods, the local RSE-PSB algorithm is introduced, leading to the $\mathcal{R}^1$SE-PSB method and the $\mathcal{R}^2$SE-PSB method, which are then analyzed. In the third 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 RSE-PSB algorithm (GRSE-PSB) is presented and analyzed with respect to its convergence. Furthermore, the compact representation of the proposed methods follows. Numerical experiments, especially for the global converging methods, are shown in the fourth section. The final section briefly summarizes the contributions of this paper.

2. Regularized structure exploiting QN methods in Hilbert spaces

2.1. **Mathematical setup and notation.** Let X and Y be separable Hilbert spaces and let $U \subset X$ be open and convex. For an operator F, we denote the first Fréchet derivative with F', its adjoint with F^*, and the dual space of X with X^*. Furthermore, we assume to have a well-defined operator $F \in C^2(U, Y)$, where neither F nor its continuous Fréchet derivatives F' and F'' are necessarily continuously invertible. 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 $\|y - y^\delta\| \leq \delta$, where y is the possibly non-unique projection, i.e., an element in the range of F with minimal distance to y^δ. Since the measurement data is contaminated with noise, the direct inversion of the forward operator is ineffective as it tends to produce solutions that deviate significantly from the exact solution. Hence, we want to consider the inverse problem as an optimization problem, where we define

$$\tilde{J} : X \to \mathbb{R}, \qquad \tilde{J}(x) := \frac{1}{2}\|F(x) - y^\delta\|_Y^2, \tag{2.1}$$

and want to find x^* that minimizes $\tilde{J}$. Due to the ill-posedness of the problem, we regularize $\tilde{J}$ with an sufficiently smooth additive regularization term $\mathcal{R}_\alpha : X \to \mathbb{R}$ with regularization parameter $\alpha > 0$, yielding the regularized target functional

$$J(x) = \tilde{J}(x) + \mathcal{R}_\alpha(x) \tag{2.2}$$

with Fréchet derivatives

$$J'(x) = \tilde{J}'(x) + \mathcal{R}'_\alpha(x),$$
$$J''(x) = \tilde{J}''(x) + \mathcal{R}''_\alpha(x).$$

By including the regularization term, the problem is transformed, at least locally, into a well-posed problem to which optimization methods can be applied. For the derivation of our theory we will employ the following assumptions:

Assumptions 2.1.

M1 There exists a minimizer of (2.2) in U.
M2 There exists a minimizer of (2.2) in a convex and compact set $D \subset U$.

Assumptions 2.2.

A1 The operator F' is Lipschitz continuous with Lipschitz constant $L_{F'} > 0$ and the operator C with $C(x) := F'(x)^* F'(x)$ is Lipschitz continuous with Lipschitz constant $L_C > 0$.
A2 The operator J'' is Lipschitz continuous with Lipschitz constant $L_{J''} > 0$.
A3 The operator $J''(x^*)$ is self-adjoint and positive in the sense that

$$\langle J''(x^*)h, h\rangle \geq \lambda \|h\|^2, \quad \forall h \in X, \text{ where } \lambda > 0.$$

Assumption M2 and F being a continuous operator, implies that $F(D)$ is compact, yielding that there exists a projection y in the range of F with minimal distance to y^δ, which might not be unique in general. The compactness assumption on D can be replaced by assuming that the range of F is a closed subspace of Y, since compactness is only needed for the existence of a projection. If the range of F is a closed subspace of Y, then it would be sufficient to just assume that D is closed, bounded and convex. Furthermore, note that $J''(x^*)$ is invertible if Assumption A3 holds.

2.2. **Derivation of the secant equation.** Consider the classic Newton method obtained by employing the Taylor approximation

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

yielding

$$\tilde{J}''(x_k)(x_{k+1} - x_k) = -\tilde{J}'(x_k). \tag{2.3}$$

By computing the first and second Fréchet derivative of $\tilde{J}$ we obtain

$$\begin{aligned} \tilde{J}'(x) &= F'(x)^* \left(F(x) - y^\delta\right) \\ \tilde{J}''(x) &= F'(x)^* F'(x) + F''(x)^* \left(F(x) - y^\delta\right). \end{aligned} \tag{2.4}$$

As $\tilde{J}''(x)$ might not be continuously invertible, we regularize $\tilde{J}$ as mentioned already above. For this purpose, we introduce two Tikhonov-type regularizers

$$\mathcal{R}_\alpha^1 : \ X \to \mathbb{R}, \ x \mapsto \frac{\alpha}{2}\|x - x_k\|_X^2 \tag{2.5}$$

$$\mathcal{R}_\alpha^2 : \ X \to \mathbb{R}, \ x \mapsto \frac{\alpha}{2}\|x - x_0\|_X^2 \tag{2.6}$$

with regularization parameter $\alpha > 0$, where x_0 is the initial guess and x_k is the last obtained iterate or the linearization point, respectively. This yields for $i \in \{1, 2\}$ the regularized target functional

$$J^i : \ X \to \mathbb{R}, \quad J^i(x) := \tilde{J}(x) + \mathcal{R}_\alpha^i(x). \tag{2.7}$$

As ${\mathcal{R}^1}_\alpha'' = {\mathcal{R}^2}_\alpha'' = \alpha I$, we obtain for $i \in \{1, 2\}$

$${J^i}'(x) = \tilde{J}'(x) + {\mathcal{R}^i}_\alpha'(x) \tag{2.8}$$

$$J''(x) = {J^1}''(x) = {J^2}''(x) = \tilde{J}''(x) + \alpha I. \tag{2.9}$$

Consequently, this results in a regularized Newton method

$$\left(F'(x_k)^* F'(x_k) + F''(x_k)^* \left(F(x_k) - y^\delta\right) + \alpha_k I\right)(x_{k+1} - x_k) = -F'(x_k)^* \left(F(x_k) - y^\delta\right) \tag{2.10}$$

for the regularizer $\mathcal{R}_{\alpha_k}^1$ and in case of choosing $\mathcal{R}_{\alpha_k}^2$ we obtain

$$\begin{aligned} &\left(F'(x_k)^* F'(x_k) + F''(x_k)^* \left(F(x_k) - y^\delta\right) + \alpha_k I\right)(x_{k+1} - x_k) \\ &\qquad = -F'(x_k)^* \left(F(x_k) - y^\delta\right) + \alpha_k (x_k - x_0). \end{aligned} \tag{2.11}$$

As in general the second Fréchet derivatives of the target functional are numerically very costly to compute, usually approximations of the Hessian operator are used, leading to QN methods. Generally they do not take possible decompositions of the objective function or of its analytical second Fréchet derivative into account. In the second Fréchet derivative of the objective function (2.9) the second derivative of

the regularizer is known. Furthermore, we assume that we have access to the first derivative of the forward operator F. Then, just the term $F''(x_k)^*\left(F(x_k)-y^\delta\right)$ is difficult and expensive to compute. As a result, there is basically no necessity to approximate the full Hessian operator, but rather only the term $F''(x_k)^*\left(F(x_k)-y^\delta\right)$.
Methods using gradient information as well are the LM method

$$(F'(x_k)^*F'(x_k)+\alpha_k I)\,(x_{k+1}-x_k) = F'(x_k)^*\left(F(x_k)-y^\delta\right) \tag{2.12}$$

and the IRGN method,

$$(F'(x_k)^*F'(x_k)+\alpha_k I)\,(x_{k+1}-x_k) = -F'(x_k)^*\left(F(x_k)-y^\delta\right)+\alpha_k(x_k-x_0), \tag{2.13}$$

which are based on iteratively applying $\mathcal{R}^1_{\alpha_k}$ in the LM case and $\mathcal{R}^2_{\alpha_k}$ in the IRGN case. Due to the clear similarities to the regularized Newton methods (2.10) and (2.11), respectively, these methods can be seen as QN methods. The problem using these methods is that the approximation of the Hessian matrix may be too inaccurate as with $F'(x_k)^*F'(x_k)+\alpha_k I \approx J''(x_k)$ in (2.9) the approximation error reads as

$$\|J''(x_k)-F'(x_k)^*F'(x_k)-\alpha_k I\| = \left\|\left(\left(F(x_k)-y^\delta\right)''\right)^*\left(F(x_k)-y^\delta\right)\right\|. \tag{2.14}$$

This approximation error is only small if F is linear or the residual is sufficiently small, i.e., we have a good approximation with small noise level δ. In our setting F can also be non-linear. Additionally, in real-world applications one does neither necessarily know the noise level a-priori nor can influence the noise level, as it is dictated by the measurement process. Furthermore, both the LM and the IRGN method require exactly the same access to the Fréchet derivative of the forward operator as we assumed to have. Thus, we want to derive an approximation for $F''(x_k)^*\left(F(x_k)-y^\delta\right)$.
Define

$$s_k := x_{k+1}-x_k \tag{2.15}$$

and consider the Taylor approximation of F,

$$\begin{aligned} F(x_{k+1}) &= F(x_k)+F'(x_k)s_k+\mathcal{O}(\|s_k\|^2) \\ \Leftrightarrow\ F'(x_k)s_k &= F(x_{k+1})-F(x_k)+\mathcal{O}(\|s_k\|^2). \end{aligned} \tag{2.16}$$

By defining

$$r_k := F(x_k)-y^\delta, \tag{2.17}$$

we conclude

$$\begin{aligned} F'(x_k)^*(r_{k+1}-r_k) &= F'(x_k)^*(F(x_{k+1})-y^\delta-F(x_k)+y^\delta) = F'(x_k)^*(F(x_{k+1})-F(x_k)) \\ &= F'(x_k)^*F'(x_k)s_k+\mathcal{O}(\|s_k\|^2). \end{aligned} \tag{2.18}$$

Then, exploiting the Taylor approximation of $\tilde{J}'$, i.e.,

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

and using (2.18) yields

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

Consequently, we obtain the essential modified secant condition

$$(F''(x_k)^*r_k)s_k = (F'(x_{k+1})-F'(x_k))^*r_{k+1}. \tag{2.19}$$

Hence, this equation defines the step s_k to be taken. We now consider the relationship between the standard secant condition used for unstructured QN methods and the modified secant condition (2.19).

Remark 2.3. Assume that a secant equation holds for J, i.e., we know

$$J''(x_k)s_k = \tilde{J}''(x_k)s_k + \alpha_k s_k = \tilde{J}'(x_{k+1}) - \tilde{J}'(x_k) + \alpha_k s_k \tag{2.20}$$

for $i \in \{1,2\}$. Hence,

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

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

$$\begin{aligned} &F(x_{k+1}) - F(x_k) = F'(x_k)s_k + E_k s_k \quad \text{with} \\ &\|E_k\| \leq 2L_{F'} \max\{\|x^* - x_k\|, \|x^* - x_{k+1}\|\}, \end{aligned} \tag{2.21}$$

which yields

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

Hence, with identity (2.4) and (2.9) we obtain

$$(F''(x_k)^* r_k)s_k = (F'(x_{k+1}) - F'(x_k))^* r_{k+1} + F'(x_k)^* E_k s_k. \tag{2.22}$$

Consequently, equation (2.20) can be transformed into (2.22) for $i \in \{1,2\}$. This describes the relation to the modified secant condition (2.19), as equation (2.22) consists of the same components as equation (2.19) but with the additional error term $F'(x_k)^* E_k s_k$ on the right-hand side.

Then by defining

$$y_k := (F'(x_{k+1}) - F'(x_k))^* r_{k+1}, \tag{2.23}$$

$$\zeta_k^i := J^{i'}(x_{k+1}) - J^{i'}(x_k) \quad \text{for } i \in \{1,2\} \tag{2.24}$$

we introduce the secant condition for J^i as

$$B_{k+1}^J s_k = \zeta_k^i \quad \text{for } i \in \{1,2\} \tag{2.25}$$

where B_{k+1}^J has to be chosen such that it approximates $J''(x_k)$.

Remark 2.4. Using the operator C defined in Assumption A1, we obtain due to Remark 2.3 that

$$\zeta_k^i = \zeta_k = y_k + C(x_k)s_k + F'(x_k)^* E_k s_k + \alpha_k s_k, \tag{2.26}$$

for $i \in \{1,2\}$. Hence, using (2.26) yields that B_k^J can be expressed as

$$B_k^J = C(x_k) + A_k + \alpha_k I + F'(x_k)^* E_k, \tag{2.27}$$

where the modified QN condition

$$A_k s_{k-1} = y_{k-1}, \tag{2.28}$$

with $A_k \approx F''(x_{k-1})^* r_{k-1}$ has to be determined.

2.3. **The regularized structure exploiting PSB.** We now want to find approximations A_{k+1} to $F''(x_k)^* r_k$ satisfying (2.28). In finite dimensional spaces one suitable outer product is represented by the matrix xy^T. In infinite dimensional Hilbert spaces, QN methods require the definition of the outer product. This can be achieved by introducing the dyadic product, see [3] or [19].

Definition 2.5 (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 [3] or [19].

Proposition 2.6. Let X and Y be Hilbert spaces and $x \in X$ and $y \in Y$. Denote with $L(X,Y)$ 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 $A \circ (x \otimes y) \circ A = (Ax) \circ (Ay)$, where $\circ$ denotes the concatenation of operators or functions.
- For all $x_1, x_2, y_1, y_2 \in X$, $(x_1 \otimes y_1) \circ (x_2 \otimes y_2) = \langle y_1, x_2 \rangle (x_1 \otimes y_2)$.
- $P = I - \frac{w \otimes w}{\langle w, w \rangle}$ is an orthogonal projector, i.e., $\|P\| = 1$.

We assume that $F''(x^*)^* \left(F(x^*) - y^\delta\right)$ is self-adjoint 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. To overcome this drawback, we want to focus on a rank-2 update, satisfying the restrictions, namely the Powell-symmetric Broyden update (PSB), which is given in the unstructured case via

$$B_{k+1}^J = B_k^J + \frac{(\zeta_k^i - B_k^J s_k) \otimes s_k + s_k \otimes (\zeta_k^i - B_k^J s_k)}{\langle s_k, s_k \rangle} - \frac{\langle \zeta_k^i - B_k^J s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2} \tag{2.29}$$

for $i \in \{1, 2\}$. We define

$$\sigma_k := \max\{\|x^* - x_k\|, \|x^* - x_{k+1}\|\} \tag{2.30}$$

and state the following result motivated by [19].

Lemma 2.7. Let Assumption 2.2 and Assumption M1 be satisfied. Then, the unstructured PSB method in Hilbert spaces satisfies the bounded deterioration principle

$$\left\|B_{k+1}^J - J''(x^*)\right\| \leq \left\|B_k^J - J''(x^*)\right\| + C\sigma_k. \tag{2.31}$$

Proof. Similarly to Theorem 2.1 in [19] we use the orthogonal projector defined in Proposition 2.6 to reformulate the distance between B_{k+1}^J and $J''(x^*)$ for $i \in \{1, 2\}$. Due to Theorem 2.1 in [19], we have that

$$\begin{aligned}
B_{k+1}^J - J''(x^*) &= B_k^J - J''(x^*) + \frac{(\zeta_k^i - B_k^J s_k) \otimes s_k + s_k \otimes (\zeta_k^i - B_k^J s_k)}{\langle s_k, s_k \rangle} - \frac{\langle \zeta_k^i - B_k^J s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2} \\
&= P(B_k^J - J''(x^*))P - \frac{(s_k \otimes s_k)J''(x^*) + J''(x^*)(s_k \otimes s_k)}{\langle s_k, s_k \rangle} \\
&\quad + \frac{(s_k \otimes s_k)J''(x^*)(s_k \otimes s_k)}{\langle s_k, s_k \rangle^2} + \frac{(\zeta_k^i \otimes s_k) + (s_k \otimes \zeta_k^i)}{\langle s_k, s_k \rangle} - \frac{(s_k \otimes \zeta_k^i)(s_k \otimes s_k)}{\langle s_k, s_k \rangle^2} \\
&= P(B_k^J - J''(x^*))P + \frac{s_k \otimes (\zeta_k^i - J''(x^*)s_k)}{\langle s_k, s_k \rangle} P + \frac{(\zeta_k^i - J''(x^*)s_k) \otimes s_k}{\langle s_k, s_k \rangle}.
\end{aligned} \tag{2.32}$$

Using the triangle inequality and Proposition 2.6 we have

$$\left\|B_{k+1}^J - J''(x^*)\right\| \leq \left\|B_k^J - J''(x^*)\right\| + 2\frac{\left\|\zeta_k^i - J''(x^*)s_k\right\|}{\|s_k\|}. \tag{2.33}$$

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

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

for $i \in \{1, 2\}$. With this, inequality (2.33) simplifies to

$$\left\|B_{k+1}^J - J''(x^*)\right\| \leq \left\|B_k^J - J''(x^*)\right\| + 2L_{J''}\sigma_k. \tag{2.34}$$

□

Therefore, the unstructured PSB update satisfies the bounded deterioration principle and Theorem 4 in [3] guarantees its local linear convergence.
Using the operator C defined in Assumption A1, the structure exploiting PSB update is given by

$$B_k := C(x_k) + A_k + \alpha_k I \tag{2.35}$$

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

$$A_{k+1} = A_k + \frac{(y_k - A_k s_k) \otimes s_k + s_k \otimes (y_k - A_k s_k)}{\langle s_k, s_k \rangle} - \frac{\langle y_k - A_k s_k, s_k \rangle s_k \otimes s_k}{\langle s_k, s_k \rangle^2}. \tag{2.36}$$

Remark 2.8. Note that Remark 2.4 yields that B_k^J is given by

$$B_k^J = B_k + F'(x_k)^* E_k. \tag{2.37}$$

Using this update, we introduce the following method, which exploits the structure of the inverse problem and regularizes it.

Algorithm 1 Regularized structure exploiting PSB for inverse problems (RSE-PSB)

Require: $A_0 \in L(X), \alpha_0 > 0, \eta > 0, x_0 \in U, g_0 = \left\| \tilde{J}'(x_0) \right\|, y^\delta \in Y, \theta \in (0,1)$
  $k = 0$
  while $g_k \geq \eta$ **do**
    Evaluate the forward operator $F(x_k)$
    Compute the derivative $F'(x_k)$ of the forward operator, its adjoint and the residual $r_k = F(x_k) - y^\delta$.
    if $k > 0$ **then**
      Compute A_k using (2.15) and (2.23) according to (2.36).
    end if
    Compute B_k according to (2.35) and solve for s_k:

$$(\mathcal{R}^1\text{SE-PSB}) \quad B_k s_k = -\left(F'(x_k)^* r_k\right)$$

or

$$(\mathcal{R}^2\text{SE-PSB}) \quad B_k s_k = -\left(F'(x_k)^* r_k - \alpha_k (x_k - x_0)\right).$$

    Compute update $x_{k+1} = x_k + s_k$
    Compute $g_{k+1} = \left\| \tilde{J}'(x_{k+1}) \right\|$
    Update $\alpha_{k+1} = \theta \alpha_k$
    $k = k + 1$
  end while

As by construction we just approximate a part of the Hessian operator, namely $F''(x_k)^* \left(F(x_k) - y^\delta\right)$ using the PSB update with modified secant conditions, it is natural to ask if this algorithm shows a similar local convergence behavior to the unstructured PSB update. Furthermore, the next intuitive question is on how to improve the linear convergence behavior. Therefore, we now prove convergence and convergence rates for the regularized structure exploiting PSB algorithm in Hilbert spaces, i.e., Algorithm 1.

Theorem 2.9. *Let Assumption M1 and Assumption 2.2 be satisfied. Let F' be bounded on U by $M > 0$ and let the regularization parameters satisfy*

$$|\alpha_{k+1} - \alpha_k| \leq a \sigma_k \ \forall k \in \mathbb{N} \tag{2.38}$$

for some $a > 0$ and σ_k defined as in (2.30)*. Then, the sequence $\{x_k\}_{k \in \mathbb{N}}$ defined by Algorithm 1 is well-defined and converges linearly to x^*. Furthermore, B_k^{-1} exists and $\|B_k\|$ as well as $\left\|B_k^{-1}\right\|$ are bounded. If additionally*

- *$F''(x^*)^* \left(F(x^*) - y^\delta\right)$ is boundedly invertible, F' is Lipschitz continuous and*

$$A_0 - F''(x^*)^* \left(F(x^*) - y^\delta\right)$$

  is compact, then $\{x_k\}_{k \in \mathbb{N}}$ converges weakly super-linearly to x^*, or

- *$B_0 - J''(x^*)$ is compact, then $\{x_k\}_{k\in\mathbb{N}}$ converges super-linearly to x^*.*

Proof. Due to Remark 2.4 and Remark 2.8 we conclude that

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

and therefore

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

for $i \in \{1, 2\}$. Since C is Lipschitz continuous with Lipschitz constant L_C, F' is bounded by $M > 0$, E_k is bounded by $2L_{F'}\sigma_k$ for all $k \in \mathbb{N}$ due to (2.21) and the condition (2.38) on the regularization parameter, we obtain

$$\left\|B_{k+1} - B_{k+1}^J\right\| \leq L_C \|s_k\| + (a + 2ML_{F'})\sigma_k. \tag{2.39}$$

Now, applying Lemma 2.7 and using specifically inequality (2.34) yields

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

By using identity (2.37) and the estimates (2.39) and (2.21), we obtain

$$\begin{aligned}
\|B_{k+1} - J''(x^*)\| &\leq \|B_k + F'(x_k)^* E_k - J''(x^*)\| + (2L_C + 2L_{J''} + a + 2ML_{F'})\sigma_k \\
&\leq \|B_k - J''(x^*)\| + (2L_C + 2L_{J''} + a + 4ML_{F'})\sigma_k.
\end{aligned}$$

This means that B_{k+1} obtained by Algorithm 1 satisfies the bounded deterioration principle. Therefore, Theorem 4 in [3] guarantees that $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 is well-defined and converges linearly to x^*. Furthermore, B_k^{-1} exists and $\{\|B_k\|\}_{k\in\mathbb{N}}$ as well as $\{\left\|B_k^{-1}\right\|\}_{k\in\mathbb{N}}$ are bounded. If $B_0 - J''(x^*)$ is compact, applying Theorem 5.4 in [2] together with the generalized Sherman–Morrison–Woodbury formula, see [23], yields super-linear convergence. If $F''(x^*)^*\left(F(x^*) - y^\delta\right)$ is boundedly invertible, F' is Lipschitz continuous and $A_0 - F''(x^*)^*\left(F(x^*) - y^\delta\right)$ is compact, then due to the Lipschitz continuity of C and condition (2.38), the bounded deterioration principle is satisfied. Hence, using Theorem 5.4 in [2] and the Lipschitz continuity of C as well as condition (2.38) together with once more the generalized Sherman–Morrison– Woodbury formula, see [23], yields weak super-linear convergence.

□

We now want to discuss the context of our regularization method in the broad field of iterative regularization of inverse problems. Due to the existence of the minimizer, the minimizer has a finite norm. Then, when choosing $\mathcal{R}^2$ as regularizer, Algorithm 1 generates a sequence strongly convergent to the minimizer and as still $\alpha_k \to 0$ for $k \to \infty$, meaning that the counter position of Proposition 3.2 in [4] guarantees that there exists a x_0-minimum-norm solution to (2.2), i.e., a least squares solution with minimum norm distance to the initial guess x_0. For noisy data in general this can similarly be adopted by choosing the regularization parameter to satisfy $\alpha(\delta) \to 0$ and $\delta^2/\alpha(\delta) \to 0$ for $\delta \to 0$, see [4]. Note that in case of choosing $\mathcal{R}^1$ as regularizer, it is not necessary to obtain a minimum-norm solution as the regularization term converges to 0 for sequences converging to the minimizer.

Conversely by choosing $\mathcal{R}^2$ as regularizer, F maps the x_0-minimum-norm solution for exact data to the projection y on D, since otherwise a contradiction to y having minimal distance to y^δ would follow. This is caused by D being compact, i.e., Assumption M2 and F being a continuous operator. Therefore, it is not very restrictive to assume that there exists a x_0-minimum-norm solution $x^\dagger$ to (2.2) in D in case of choosing $\mathcal{R}^2$ as regularizer.

As $\{\|B_k\|\}_{k\in\mathbb{N}}$ is a bounded sequence on D and therefore also $\{\|A_k\|\}_{k\in\mathbb{N}}$ is a bounded sequence on D, applying the regularized structure exploiting PSB method (2.7) can also be interpreted as applying a gradient step to

$$\mathcal{J}_k^1(x) := \frac{1}{2}\left\|F'^*(x_k)(x-x_k)+F(x_k)-y^\delta\right\|^2+\frac{1}{2}\langle x-x_k, A_k(x-x_k)\rangle+\frac{\alpha_k}{2}\|x-x_k\|^2, \tag{2.40}$$

respectively to

$$\mathcal{J}_k^2(x) := \frac{1}{2}\left\|F'^*(x_k)(x-x_k)+F(x_k)-y^\delta\right\|^2+\frac{1}{2}\langle x-x_k, A_k(x-x_k)\rangle+\frac{\alpha_k}{2}\|x-x_0\|^2, \tag{2.41}$$

depending on which regularizer is chosen. The first term $\frac{1}{2}\left\|F'^*(x_k)(x-x_k)+F(x_k)-y^\delta\right\|^2$ depicts the idea of Newton methods, 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 terms $\frac{1}{2}\langle x-x_k, A_k(x-x_k)\rangle$ is bounded due to $\{\|A_k\|\}_{k\in\mathbb{N}}$ being bounded and can therefore be interpreted as penalty term that controls the distance to the point around which the linearization of the non-linear operator equation was performed. Lastly the third term is the standard iterated Tikhonov regularization term used for deriving the IRGN or LM method, respectively. Furthermore, with the condition

$$\alpha_k>0, \quad 1\leq\frac{\alpha_k}{\alpha_{k+1}}\leq r, \quad \lim_{k\to\infty}\alpha_k=0, \tag{2.42}$$

for some $r>1$, the IRGN method without penalization reads as

$$\widetilde{\mathcal{J}}_k^2(x) := \frac{1}{2}\left\|F'^*(x_k)(x-x_k)+F(x_k)-y^\delta\right\|^2+\frac{\alpha_k}{2}\|x-x_0\|^2,$$

and the LM method without penalization is given by

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

For the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1, one obtains

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

for $i\in\{1,2\}$.

3. Regularized structure exploiting QN methods in finite dimensional spaces

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 the easy access to the first derivative of the forward operator F is provided by algorithmic differentiation (AD). Developed in the last decades, this method offers an outstanding way to provide derivative information for a given code segment [14]. 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., in 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 [13], we explicitly recommend using AD tools in the discretized framework. Therefore, we use the discretize-then-optimize approach, i.e., $X=\mathbb{R}^n$ 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

$F : \mathbb{R}^n \to \mathbb{R}^m$. Hence, one obtains $C(x_k) = F'(x_k)^T F'(x_k)$, $Z(x_k) = F''(x_k)^T \left(F(x_k) - y^\delta\right)$ and the modified secant condition

$$y_k = (F'(x_{k+1}) - F'(x_k))^T r_{k+1} \tag{3.1}$$

analogously to (2.19). Next, we discuss the PSB update in the finite dimensional setting, which belongs to the general class of symmetric secant updates for QN methods [10] and can be represented by

$$A_{k+1} = A_k + \frac{(y_k - A_k s_k)v_k^T + v_k(y_k - A_k s_k)^T}{v_k^T s_k} - \frac{(y_k - A_k s_k)^T s_k}{(v_k^T s_k)^2} v_k v_k^T. \tag{3.2}$$

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 $v_k = s_k$ for the update routine.

Corollary 3.1. Let Assumption M1 and Assumption 2.2 be satisfied. Assume F' is bounded on U and the regularization parameters satisfy condition (2.38) for some $a > 0$. Then, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 1 in the finite dimensional setting is well-defined and converges super-linearly to x^*.

Proof. As $\mathbb{R}^n$ is a finite dimensional separable Hilbert space for all $n \in \mathbb{N}$, Theorem 2.9 guarantees that the sequence $\{x_k\}_{k\in\mathbb{N}}$ defined by Algorithm 1 in the finite dimensional setting is well-defined and converges linearly to x^*. As $B_0 - J''(x^*)$ being compact is always satisfied in the finite dimensional case [2], the assertion is again a direct consequence of Theorem 2.9. □

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 [18] have already given 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 [18] to improve the robustness by combining regularization to stabilize the solution with globalization to handle poor initial guesses. For this purpose, we will focus on the regularizer $\mathcal{R}^1$.

Algorithm 2 Global regularized structure exploiting QN (GRSE)

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

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 $\nabla J(x_k) = F'(x_k)^T r_k$.

Set $\tilde{B}_k = F'(x_k)^T F'(x_k) + A_k$ and solve for s_k:

$$\left(\tilde{B}_k + \alpha_k I\right) s_k = -\nabla J(x_k)$$

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 $\rho_k \leq c$ 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 $g_{k+1} = \left\|\nabla \tilde{J}(x_{k+1})\right\|$ (successful step)

Compute A_{k+1} according to (3.2) with (3.1) and $v_k = s_k$.

end if

$k = k + 1$

end while

Remark 3.2. Note that $\nabla \tilde{J}(x_k) = \nabla J(x_k)$, as we only deal with the regularizer $\mathcal{R}^1$ and that this globalization is also applicable for the LM method by simply setting $A_k = 0$ for all $k \in \mathbb{N}_0$. Furthermore, we have the relationship $B_k = \tilde{B}_k + \alpha_k I$ between the Hessian approximation matrices of Algorithm 1 and Algorithm 2.

We now proceed with analyzing the convergence of Algorithm 2 both in the PSB case and in the LM case.

Theorem 3.3 (Global Convergence)**.** *Let $V \subset X$ be bounded and let the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 2 be a subset of V, where the termination criterion is neglected, i.e., $\eta = 0$. Let A_0 be regular, J be bounded from below and F, C and Z be Lipschitz continuous on V. Then, for the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 2 it holds that*

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

Furthermore, if ∇J is uniformly continuous on V, then $\lim_{k\to\infty} \|\nabla J(x_k)\| = 0$.

Proof. Preliminarily, we prove that Algorithm 2 performs infinitely many successful steps. Therefore, we assume 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, $\tilde{B}_k = \tilde{B}_{k_0}$ for all $k \geq k_0$. Thus, for sufficiently large $k \geq k_0$ we know that $\tilde{B}_{k_0} + \alpha_k I$ is invertible and that

$$\lim_{k\to\infty} \frac{\left(\tilde{B}_{k_0} + \alpha_k I\right) s}{\left\|\left(\tilde{B}_{k_0} + \alpha_k I\right) s\right\|} = \frac{s}{\|s\|}.$$

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 $(\tilde{B}_{k_0} + \alpha_k I)s_k = -\nabla J(x_{k_0})$ 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 [18] leads to a contradiction. Hence, the set of indices $\mathcal{S} \subset \mathbb{N}$ of successful steps of Algorithm 2 has infinite cardinality. Now, as every step $k \in \mathcal{S}$ is successful, we have for every $k \in \mathcal{S}$ that

$$J(x_k) - J(x_{k+1}) \geq c\text{pred}_k \geq p_{\min} c \|\nabla J(x_k)\| \|s_k\|.$$

We assume for the sake of contradiction that

$$\liminf_{k\to\infty} \|\nabla J(x_k)\| > 0, \tag{3.3}$$

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

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

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 $k \to \infty$. As F, C and S are Lipschitz continuous on V, they are bounded on V. By Lemma 2 in [12] we know that there exist constants $d_1, d_2 \geq 0$ such that

$$A_{k+1} \leq d_1 + d_2 \sum_{j=1}^{k} \|s_j\|. \tag{3.4}$$

Now, for every step $k \in \mathcal{S}$ we have $(\tilde{B}_k+\alpha_k I)s_k = -\nabla J(x_k)$. 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, there exists at least one subsequence $\{\alpha_{k_l}\}_{l\in\mathcal{S}}$ tending to $+\infty$, since otherwise all subsequences would be bounded, which implies that all subsequences $\|\nabla J(x_{k_l})\|$ converge to 0 for $l \to \infty$ as $s_{k_l} \to_{\mathcal{S}} 0$ for $l \to \infty$. This yields that $\|\nabla J(x_k)\|$ converges to 0 for $k \to \infty$, which violates (3.3). Hence one obtains $\alpha_k \to +\infty$, meaning that Algorithm 2 also performs infinitely many unsuccessful steps. With this, arguing now analogously as in the proof of Theorem 1 in [18] yields a contradiction to (3.3). Consequently, $\liminf_{k\to\infty} \|\nabla J(x_k)\| = 0$.
Additionally, if ∇J is uniformly continuous on V, arguing in the same way as in the proof of Theorem 2 in [18], 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 $\{\tilde{B}_k\}_{k\in\mathbb{N}}$. Hence, we can directly apply Lemma 1, Lemma 2, Theorem 1 and Theorem 2 in [18]. □

Remark 3.4. This yields that, given any $\eta > 0$, Algorithm 2 terminates with $\|\nabla J(x_k)\| \leq \eta$ after finitely many iterations.

Corollary 3.5. Let Assumption M1, Assumption 2.2 and the assumptions of Theorem 3.3 be satisfied with $\eta > 0$. Suppose condition (2.38) holds for all $k \in \mathcal{S}$. Then the iterates $\{x_k\}_{k\in\mathbb{N}}$ generated by Algorithm 2 converge super-linearly on U.

Proof. Due to Remark 3.4 we know that Algorithm 2 terminates for $\eta > 0$ after finitely many steps. Therefore, there exists an $N \in \mathbb{N}$ such that for all $k \geq N$ it holds that α_k decreases to 0 geometrically, the condition (2.38) holds and $\{x_k : k \geq N\} \subseteq U$. Then, due to Corollary 3.1, we have that $B_k = \tilde{B}_k + \alpha_k I$ satisfies the Dennis-Moré condition

$$\lim_{k\to\infty} \frac{\left\|\left(\tilde{B}_k + \alpha_k I - J''(x^*)\right) s_k\right\|}{\|s_k\|} = 0,$$

yielding super-linear convergence. □

QN methods depend on the choice of the initial matrix. If $F''(x_0)^T(F(x_0) - y^\delta)$ can be computed at the initial guess x_0, then we suggest choosing A_0 as this matrix. If $F''(x_0)^T(F(x_0) - y^\delta)$ is not accessible, we choose to scale the identity by κ similarly to the initial regularization parameter with $\kappa \geq \alpha_0$.
Additionally, compact representations, in which recursive QN update formulas are given as compact matrix factorizations, are advantageous for large scale inverse problems or limited memory QN methods. The compact representation of structured QN methods was already presented in [5]. In [18] regularized QN methods were discussed with regard to their compact representations. The compact representation of QN methods can be formulated as

$$A_k = A_0 + M_k Q_k^{-1} M_k^T,$$

where A_0 is the initial approximation, M_k is a matrix that captures the updates at iteration k, and Q_k is a matrix related to curvature information.
For the regularized structure exploiting approach with the given regularization parameter $\alpha_k > 0$, the regularized Hessian approximation can be compactly expressed by

$$C(x_k) + A_k + \alpha_k I = C(x_k) + \alpha_k I + A_0 + M_k Q_k^{-1} M_k^T.$$

We define

$$S_k = [s_0, \ldots, s_{k-1}] \in \mathbb{R}^{n\times k} \tag{3.5}$$

$$Y_k = [y_0, \ldots, y_{k-1}] \in \mathbb{R}^{n\times k} \tag{3.6}$$

$$W_k = Y_k - A_0 S_k \in \mathbb{R}^{n\times k} \tag{3.7}$$

$$G_k = \operatorname{diag}(S_k^T W_k) \in \mathbb{R}^{k\times k}. \tag{3.8}$$

Let V_k be the strictly lower triangular part of $S_k^T W_k$ and U_k the non-strictly upper triangular matrix of $S_k^T S_k$. Thus, by Theorem 3 in [18] we obtain

$$M_k := [S_k \quad W_k], \tag{3.9}$$

$$Q_k := \begin{bmatrix} 0 & U_k \\ U_k^T & V_k + G_k + V_k^T \end{bmatrix}. \tag{3.10}$$

This is particularly useful for limited-memory versions of the compact representation. In such cases, only the most recent $k \geq m > 0$ matrices are retained. If we now want to invert $C(x_k) + A_k + \alpha_k I$, we apply the Sherman–Morrison–Woodbury formula. We set $A_0 := \kappa I$, where κ is chosen according to our suggestions above and define $\nu_k := \kappa + \alpha_k$ for all $k \in \mathbb{N}$. Applying the Sherman–Morrison–Woodbury formula yields that the inverse of B_k has the following structure

$$\begin{aligned} B_k^{-1} &= \left(C(x_k) + \nu_k I + M_k Q_k^{-1} M_k^T\right)^{-1} \\ &= \left(C(x_k) + \nu_k I\right)^{-1} \left(I - M_k \left(Q_k + M_k^T \left(C(x_k) + \nu_k I\right)^{-1} M_k\right)^{-1} M_k^T \left(C(x_k) + \nu_k I\right)^{-1}\right). \end{aligned}$$

4. 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. Therefore, the model is dependent on the state u and the parameter p that one wants to identify. As solutions to PDEs are commonly understood in the weak sense, we define the model operator via

$$\langle A(p,u), \zeta \rangle = 0, \tag{4.1}$$

where ζ is the test function. To recover information on the parameter p, we model the observed data

$$O(u) = y \tag{4.2}$$

from observations of the state u, where O denotes the observation operator and y the noiseless data. The reduced approach is common for stating and solving parameter identification problems, where the model is eliminated by introducing a parameter-to-state map S, mapping the parameter to the solution of the underlying PDE model. Hence, S is defined implicitly by satisfying the model for all parameters and test functions, i.e.,

$$\langle A(p, S(p)), \zeta \rangle = 0. \tag{4.3}$$

Then, with noisy measurements y^δ and the forward operator $F(p) = (O \circ S)(p)$, we want to identify p such that

$$F(p) = y^\delta. \tag{4.4}$$

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. We now introduce the specific numerical example. Let $\Omega \subset \mathbb{R}^2$ be a domain with a sufficiently smooth boundary $\partial\Omega$ and define the differential operator

$$\mathcal{B} = \begin{pmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{pmatrix}$$

as well as the parameter matrix

$$P = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix}.$$

We consider the elliptic PDE system

$$-\mathcal{B}^T \mathcal{B} u(x) + P u(x) = f(x), \quad x \in \Omega, \tag{4.5}$$

$$\mathcal{N} \cdot \mathcal{B} u(x) = 0, \qquad x \in \partial\Omega, \tag{4.6}$$

where $\mathcal{N}$ denotes the outward normal vector on the boundary $\partial\Omega$, $u : \Omega \to \mathbb{R}^2$ is the state and $f : \Omega \to \mathbb{R}^2$ is the inhomogeneity. As this system describes the underlying forward problem, it is necessary to ask for the well-posedness of the forward problem. Thus, we need unique existence of solutions to the weak form

$$\int_\Omega (\mathcal{B}u(x))^T \mathcal{B}v(x) + (Pu(x))^T v(x)\,\mathrm{d}x = \int_\Omega (f(x))^T v(x)\,\mathrm{d}x \tag{4.7}$$

for all $v \in H^1(\Omega)$, which is guaranteed by [11]. Then, the inverse parameter identification problem aims to identify the unknown parameters $p_1 \in \mathbb{R}^+$ and $p_2 \in \mathbb{R}^+$ in (4.7) from measurements of the solution u. This system models a variety of stationary physical processes in which diffusion or transport occurs alongside a potential. The term $Pu(x)$ can be interpreted as a potential, whereas the term $-\mathcal{B}^T\mathcal{B}u(x)$ models the diffusion or transport of $u(x)$ through the domain and $f(x)$ describes an external source.
In many industrial and scientific applications, the identification of the parameter matrix P based on observational data is of decisive importance. In our setting, system (4.5)-(4.6) can be used to describe the distribution of mechanical stress and strain in a material, where P can be interpreted as a mechanical damping parameter and $u(x)$ represents displacement. In the analysis of mechanical structures, such as bridges or aircraft, system (4.5)-(4.6) can be used to model the stationary distribution of mechanical stress, particularly when damping or frictional effects need to be considered, see, e.g., [20].
In order to define the observation operator, we utilize measurements taken strictly inside of the domain Ω and do not cover the entire domain, but rather only on disjoint subdomains. Hence, let these measurement areas $\Omega_l \subset \Omega,\ l \in \{1,2,3,4\}$ be pairwise disjoint and let their union be a proper subset of Ω. Then we define the observation operator $O : H^1(\Omega, \mathbb{R}^2) \to \mathbb{R}^4$ via

$$O(u(x)) = \begin{pmatrix} \int_{\Omega_1} u^T u\,\mathrm{d}x \\ \int_{\Omega_2} u^T u\,\mathrm{d}x \\ \int_{\Omega_3} u^T u\,\mathrm{d}x \\ \int_{\Omega_4} u^T u\,\mathrm{d}x \end{pmatrix}.$$

Thus, we are dealing with the forward operator

$$F = O \circ S : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^4,$$

where

$$S : \mathbb{R}^+ \times \mathbb{R}^+ \to H^1(\Omega, \mathbb{R}^2), \qquad S(p_1, p_2) := u(p_1, p_2).$$

Therefore, we want to solve

$$\min_{p_1,p_2} \left\| F(p_1, p_2) - y^\delta \right\|^2 + \mathcal{R}^i(p_1, p_2),$$

where we can choose the regularizer with $i \in \{1, 2\}$.
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), which in practice is realized with the finite element tool FEniCS [1] in dolfin version 2019.2.0.dev0, using AD via the dolfin adjoint [22] library of FEniCS in version 2019.1.0. As domain Ω we consider the unit square and as measurement areas $\Omega_l \subset \Omega,\ l \in \{1,2,3,4\}$ we use squares with the same side length 0.04 within the unit square, where Ω_1 starts at vertex $(0.05, 0.05)$, Ω_2 starts at vertex $(0.55, 0.05)$, Ω_3 starts at vertex $(0.05, 0.55)$ and Ω_4 starts at vertex $(0.55, 0.55)$. The source term f is chosen to be a non-isotropic vector field of the form

$$f(x_1, x_2) = \left(2\exp\left(-\frac{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}{0.02} \right), \exp\left(-\frac{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}{0.02} \right) \right)^T.$$

The generation of the observed data is done with the following set of material parameters

$$p_1 = 1,5, \quad p_2 = 1,$$

where we used a mesh two times finer than the mesh used for the optimization algorithm. These material parameters, consequently, serve as the ground truth. Furthermore, we used the same hyperparameters in all of the following implementations

$$\theta = 0.5, \quad \alpha_0 = 1^{-7}, \quad \kappa = 10^{-6}, \quad A_0 = \kappa I, \quad \sigma = 2, \quad c = 1e-4, \quad p = 10^{-6} \quad \text{and} \quad \eta = 10^{-12}.$$

For the sake of better visualization, the reconstructed parameter values have been normed with the respective true parameter in all figures below, so that convergence to the value 1 is desired. Furthermore, we assumed in the following that the initial guess is an overestimate. In our numerical tests, underestimating the initial guess yielded similar results. We start by focusing on the numerical results of the local methods. Therefore, we choose a sufficiently small deviation of the initial guess from the true parameter of 10 %, which means that the initial guess is 10% larger than the true parameter. We compare the performance during the parameter identification process of the $\mathcal{R}^1$SE-PSB method and the LM method in Figure 1.

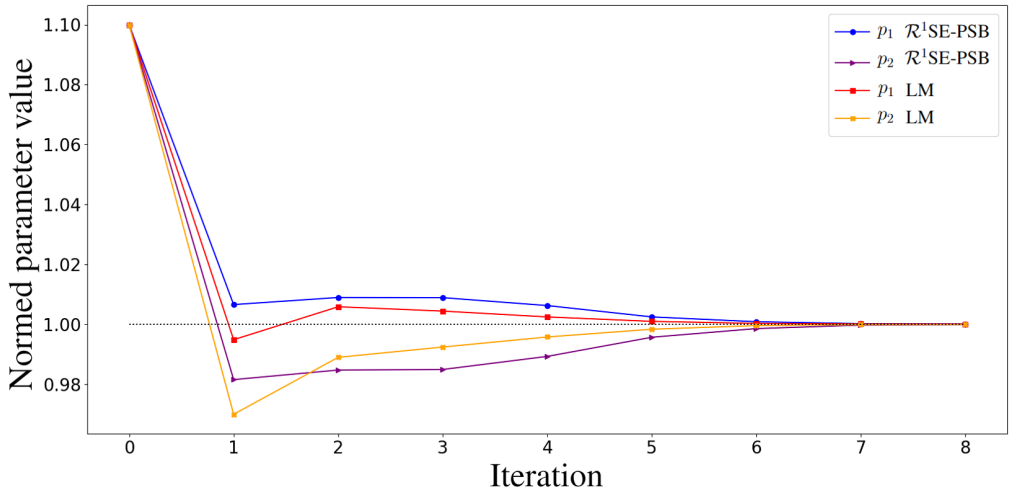

FIGURE 1. Performance of $\mathcal{R}^1$SE-PSB and LM with 10% deviation of the initial guess to the ground truth.

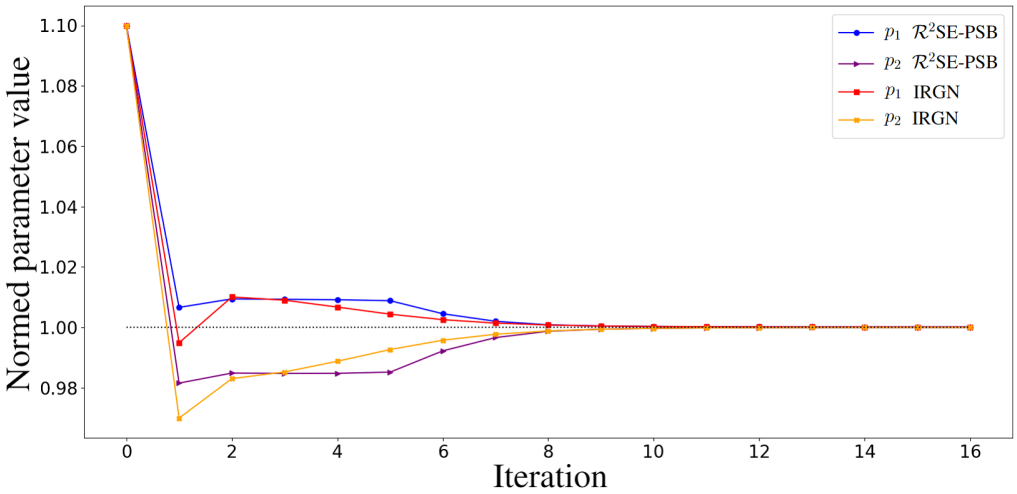

FIGURE 2. Performance of $\mathcal{R}^2$SE-PSB and IRGN with 10% deviation of the initial guess to the ground truth.

These two methods differ in the choice of A_k, as in the $\mathcal{R}^1$SE-PSB method A_k is computed in each step $k \in \mathbb{N}$ using the PSB method under the modified secant condition (2.19) and in the LM method A_k is kept constant 0. The reconstruction results of the $\mathcal{R}^1$SE-PSB method are printed in blue and purple and the reconstruction results of the LM method in red and orange per iteration step. The reconstruction results of the $\mathcal{R}^1$SE-PSB method show clear similarity to the LM method, which is to be expected at this point, since the deviation of the initial guess to the ground truth is small and the forward operator is quadratic. Consequently, the approximation error of the LM method is correspondingly small.
For our second regularization method, we compare the performance of the $\mathcal{R}^2$SE-PSB method and the IRGN method during the parameter identification process, as the key difference between these two methods lies in the choice of A_k. In Figure 2, the reconstruction of the parameters p_1 and p_2 employing the $\mathcal{R}^2$SE-PSB method and the IRGN method, is shown. Analogously, the reconstruction results of the $\mathcal{R}^2$SE-PSB method are shown in blue and purple, and of the IRGN method in red and orange, for each iteration step. Similar to the previous comparison, the convergence results of $\mathcal{R}^2$SE-PSB method are closely resembled to those of the IRGN method for the same reasons as mentioned before. Furthermore, we observe convergence to the true parameter of all four local methods used above.
Finally, the reconstruction results of the $\mathcal{R}^1$SE-PSB and the $\mathcal{R}^2$SE-PSB exhibit super-linear convergence behavior, which can be recognized in Figure 3.

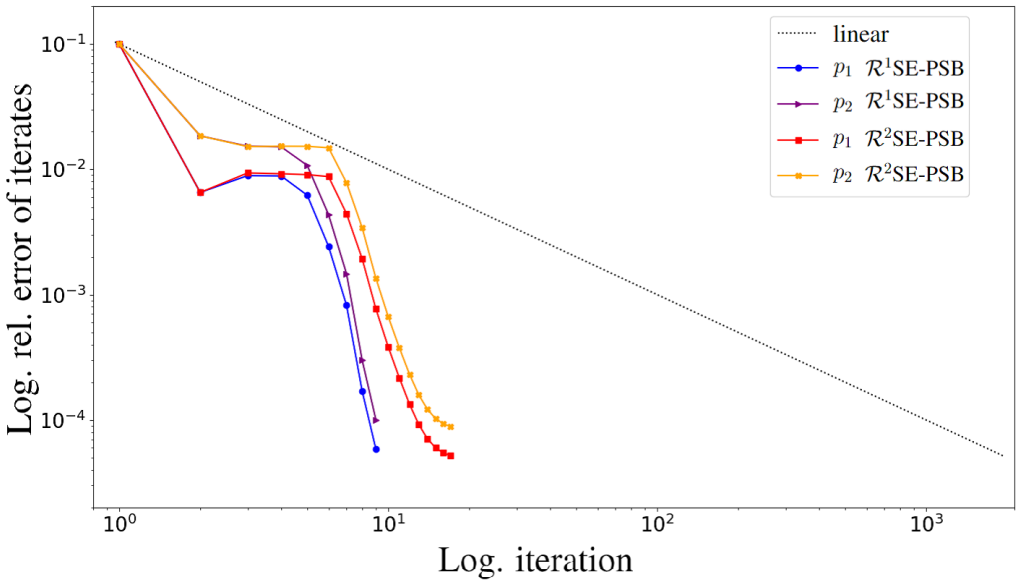

FIGURE 3. Local super-linear convergence behavior of $\mathcal{R}^1$SE-PSB and $\mathcal{R}^2$SE-PSB.

However, since in application the quality of the initial guess may not be sufficient for local methods, we now focus on the numerical results of the global methods. Therefore, we choose a deviation of the initial guess to the true parameter by 100%. Since we have focused on the regularization method defined in (2.5) in the globalization approach, we will now take a closer look at the globalized $\mathcal{R}^1$SE-PSB method (GRSE-PSB) and the globalized LM method (G-LM). Therefore, in Figure 4 we compare the performance of the GRSE-PSB method and the G-LM method during the parameter identification process, due to the fact that these methods differ only in the choice of A_k and show similar local convergence results. For each iteration step, the reconstruction results of the GRSE-PSB method are shown in blue and purple and those of the G-LM method in red and orange. The right figure is a zoomed-in representation of the left figure for improved visualization. It can be clearly seen that the G-LM method requires roughly ten 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.

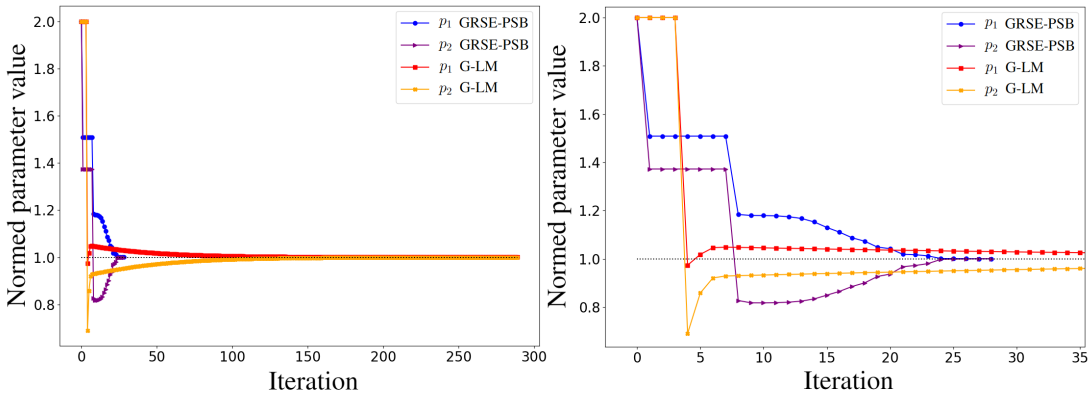

FIGURE 4. Performance of GRSE-PSB and G-LM with 100% deviation of the initial guess to the ground truth.

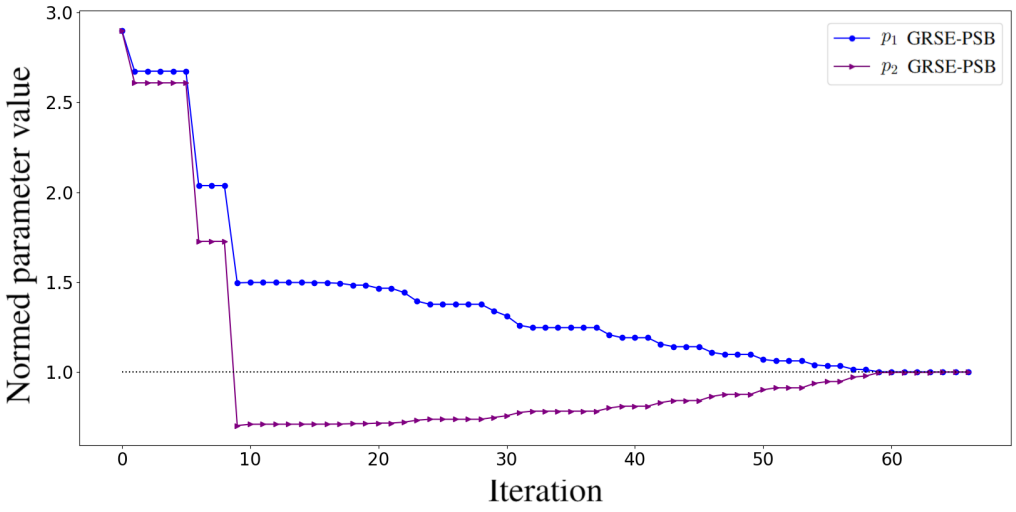

FIGURE 5. Parameter identification performance of GRSE-PSB with 190% deviation of the initial guess to the ground truth.

Note that with all deviations above, every iteration step remains in the feasible set, as both parameters must be positive to maintain well-posedness of the forward problem. Keeping that in mind, the deviation can even be chosen to be 190%, depicted in Figure 5. As the G-LM method does converge, but much slower, we have omitted the G-LM method in the Figure 5 to emphasize the GRSE-PSB method more effectively. Keeping the deviation of 190%, a limited memory implementation of the GRSE-PSB method that exploits the derived compact representation is shown in Figure 6. There we only retained the ten recent matrices. Once more, we observe fast convergence to the true parameter. Furthermore, we see that some segments of the reconstruction graph in Figures 4 - 6 show no changes in the parameter value. This illustrates the control of the regularization parameter and should clarify that the regularization parameter was increased here and the current computed iteration value was rejected.

Lastly, our focus will be on the case of overestimating the true parameter by 100%, where our observed data is contaminated with 2% uniformly distributed random noise, while the mesh on which the data is generated is still twice as fine as the mesh on which the optimization is performed. The performance of the

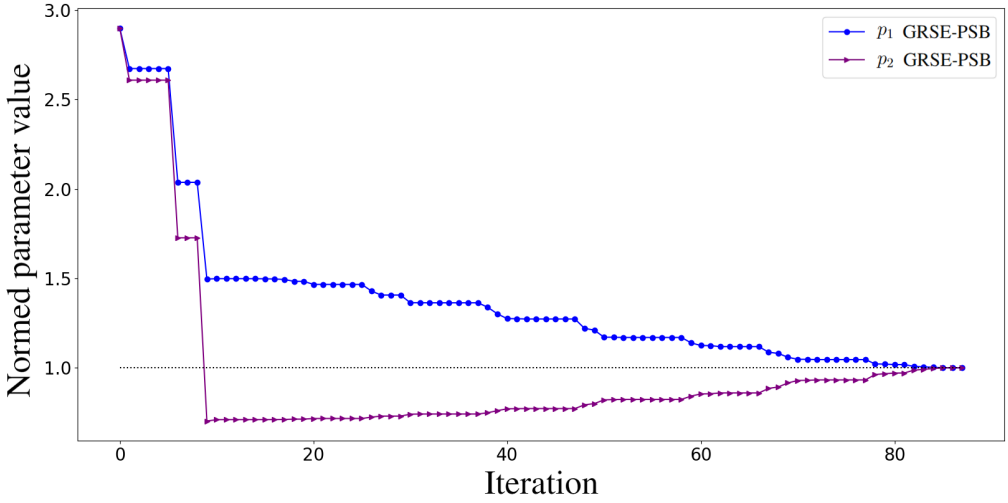

FIGURE 6. Performance of the limited memory GRSE-PSB with 190% deviation of the initial guess to the ground truth.

GRSE-PSB method and the G-LM method during the parameter identification process in the noisy case is compared in Figure 7. Similarly, the reconstruction results of the GRSE-PSB method are shown in blue and purple, whereas those of the G-LM method are shown in red and orange, respectively. The figure on the right is an enlarged representation of the figure on the left for better visualization. As before, it can be seen that the G-LM method requires roughly ten times as many iterations as the GRSE-PSB method to reach the specified threshold value η for the same reasons.

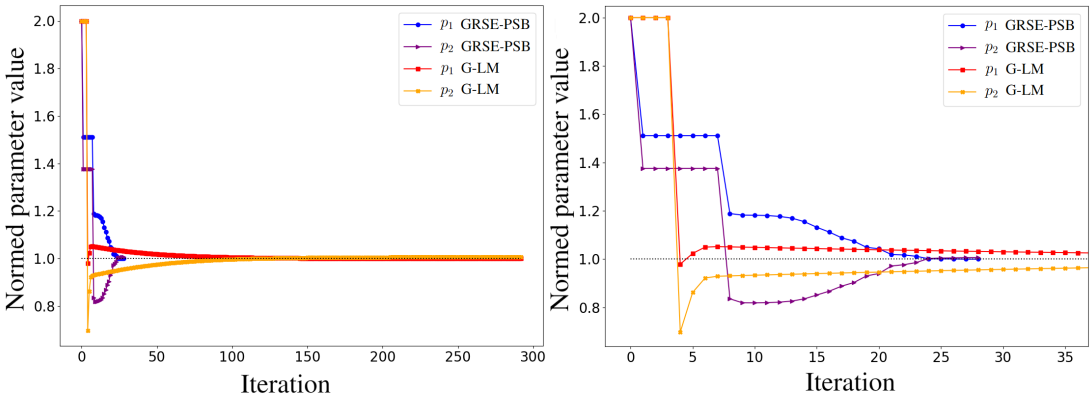

FIGURE 7. Performance of GRSE-PSB and G-LM with 100% deviation of the initial guess to the ground truth and 2% uniformly distributed random noise on the data.

The reconstruction results show convergence for both methods, but the noisy data causes the reconstruction to be less precise, especially for p_2. However, the GRSE-PSB method converges to the same values as the G-LM method.

5. Conclusion

This paper proposed a regularized QN method designed for solving ill-posed inverse problems in both infinite dimensional and finite dimensional settings. Through the derivation of a modified secant equation in an infinite dimensional Hilbert space setting and by directly integrating Tikhonov-type regularization into the QN framework we presented the $\mathcal{R}^1$SE-PSB method and the $\mathcal{R}^2$SE-PSB method, which both overcome challenges inherent in ill-posedness. They leverage the problem structure not only by decomposing in data discrepancy and regularization terms but also by exploiting the structure of the analytical second Fréchet derivative of the respective regularized target functional. The local convergence of the novel regularization methods was demonstrated, and results on its super-linear convergence behavior were established.
To ensure robustness in practical applications, particularly when the initial guess significantly deviates from the true solution, we developed a globalization approach that employs a dynamic control of the regularization parameter, ensuring global convergence and simultaneously stabilizing the ill-posed problem. Therefore, the control approach of the regularization parameter also serves as an a-priori choice.
Towards practical applications, especially in large-scale problems, we have introduced a compact representation of the proposed methods, which enables efficient computation of Hessian approximations, particularly beneficial for limited memory implementations.
In conclusion, this work contributes to the field by connecting regularization techniques and Quasi-Newton methods, offering a comprehensive theoretical and practical framework for solving ill-posed inverse problems.

Acknowledgements

The authors would like to thank the German Research Foundation (DFG) for financial support within the research group 5208 NEPTUN (444955436), support code JU 3254/1-1. 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] 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.
[3] 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.
[4] A. Binder et al. “Weakly closed nonlinear operators and parameter identification in parabolic equations by tikhonov regularization”. In: *Applicable Analysis* 55.3-4 (1994), pp. 215–234.
[5] J. J. Brust et al. “Compact representations of structured BFGS matrices”. In: *Computational Optimization and Applications* 80.1 (2021), pp. 55–88.
[6] J. E. Dennis. *On some methods based on Broyden’s secant approximation to the Hessian*. Tech. rep. Cornell University, 1971.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] L. C. Evans. *Partial differential equations*. Providence, R.I.: American Mathematical Society, 2010.
[12] R. Fletcher. “An algorithm for solving linearly constrained optimization problems”. In: *Mathematical Programming* 2 (1972), pp. 133–165.

[13] 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.

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] A. I. Lurie. *Theory of elasticity*. Springer Science & Business Media, 2010.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.