ruhr.paD

UA Ruhr Zentrum für
partielle Differentialgleichungen

Contingent derivatives and regularization for noncoercive inverse problems

dedicated to the blessed memory of Jonathan M. Borwein

Ch. Clason, A. Khan, M. Sama and Ch. Tammer

CONTINGENT DERIVATIVES AND REGULARIZATION FOR NONCOERCIVE INVERSE PROBLEMS

Christian Clason[‡]   Akhtar A. Khan[*]   Miguel Sama[‡‡]   Christiane Tammer[**]

February 14, 2018

Abstract   We study the inverse problem of parameter identification in non-coercive variational problems that commonly appear in applied models. We examine the differentiability of the set-valued parameter-to-solution map by using the first-order and the second-order contingent derivatives. We explore the inverse problem by using the output least-squares and the modified output least-squares objectives. By regularizing the non-coercive variational problem, we obtain a single-valued regularized parameter-to-solution map and investigate its smoothness and boundedness. We also consider optimization problems using the output least-squares and the modified output least-squares objectives for the regularized variational problem. We give a complete convergence analysis showing that for the output least-squares and the modified output least-squares, the regularized minimization problems approximate the original optimization problems suitably. We also provide the first-order and the second-order adjoint method for the computation of the first-order and the second-order derivatives of the output least-squares objective. We provide discrete formulas for the gradient and the Hessian calculation and present numerical results.

2000 Mathematics Subject Classification   35R30, 49N45, 65J20, 65J22, 65M30

1 INTRODUCTION

Non-coercive variational problems frequently emerge from applied models (see [21]). However, often less general versions of such models are studied under the coercivity assumption so that useful technical tools can be employed. For instance, in all the available literature on the inverse problem of identifying a variable parameter in elliptic partial differential equations using a variational framework, the bilinear form has always been chosen to be coercive. The coercivity ensures that the variational problem is uniquely solvable and retains stability on the data perturbation. Under coercivity, the parameter-to-solution map is single-valued, well-defined, and infinitely differentiable. Furthermore, coercivity also plays a decisive role in local

[1]Center for Applied and Computational Mathematics, School of Mathematical Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, New York, 14623, USA. (aaksma@rit.edu)

[3]Faculty of Mathematics, University of Duisburg-Essen, 45117 Essen, Germany. (christian.clason@uni-due.de)

[9]Departamento de Matemática Aplicada, Universidad Nacional de Educación a Distancia, Calle Juan del Rosal, 12, 28040 Madrid, Spain. (msama@ind.uned.es)

[7]Institute of Mathematics, Martin-Luther-University of Halle-Wittenberg, Theodor-Lieser-Str. 5, D-06120 Halle-Saale, Germany. (christiane.tammer@mathematik.uni-halle.de)

stability estimates, see [27]. Although the solvability of a noncoercive variational problem can be ensured by other tools (see [3]), the parameter-to-solution map, in this case, is a set-valued map. Therefore, for parameter identification in noncoercive variational problems, to derive optimality conditions, one has to employ a suitable notion of a derivative of set-valued maps. Therefore, the known techniques need to be altered significantly to cope with the involvement of such technical tools. For an overview of the recent developments in the vibrant and expanding field of inverse problems, the reader is referred to [2, 5, 6, 10, 11, 13, 14, 17, 22, 23, 30, 33, 34, 36–38, 40].

A prototypical example of a non-coercive variational problem is the weak formulation of pure Neumann boundary value problem (BVP): Given a bounded open domain Ω and the unit outer unit normal n, consider the problem of finding u such that

$$-\nabla \cdot (a\nabla u) = f \ \text{ in } \Omega, \ \frac{\partial u}{\partial n} = g \ \text{ on } \partial\Omega, \tag{1.1}$$

where $\frac{\partial u}{\partial n}$ is the outer normal derivative of u on the boundary $\partial\Omega$, and f and g are two given functions. It is known that the weak form of the above BVP leads to a noncoercive bilinear form. Moreover, (1.1) is solvable only under the compatibility condition

$$\int_\Omega f + \int_{\partial\Omega} g = 0,$$

whereas, as a consequence of Fredholm alternate, there are infinitely many solutions, with any two solutions only differing by a constant. Furthermore, among these solutions, there is a unique solution under the additional constraint $\int_\Omega u = 0$.

All the research on the inverse problems of parameter identification in pure Neumann BVP, the constraint $\int_\Omega u = 0$ and the compatibility condition have been imposed so that the parameter-to-solution map $a \rightarrow u(a)$ is well-defined and single-valued.

The primary objective of this work is to conduct a thorough study of the inverse problem of parameter identification in noncoercive variational problems. However, before going into the details of the main contribution of this article, we provide a brief review of the existing approaches for parameter identification in partial differential equations (PDEs) and variational problems by focusing on the role of coercivity. Let B be a Banach space and let A be a closed, and convex subset of B with a nonempty interior. Given a Hilbert space V, let $T : B \times V \times V \rightarrow \mathbb{R}$ be a trilinear form with $T(a, u, v)$ symmetric in u, v, and let m be a bounded linear functional on V. Assume there are constants $\alpha > 0$ and $\beta > 0$ such that the following continuity (cf. (1.2)) and coercivity (cf. (1.3)) conditions hold:

$$T(a, u, v) \leq \beta \|a\|_B \|u\|_V \|v\|_V, \ \text{ for all } u, v \in V, \ a \in B, \tag{1.2}$$

$$T(a, u, u) \geq \alpha \|u\|_V^2, \ \text{ for all } u \in V, \ a \in A. \tag{1.3}$$

Consider the following variational problem: Given $a \in A$, find $u = u(a) \in V$ such that

$$T(a, u, v) = m(v), \ \text{ for every } v \in V. \tag{1.4}$$

Due to the conditions imposed on the trilinear map T, the Riesz representation theorem ensures that for any $a \in A$, variational problem (1.4) admits a unique solution $u(a)$ (see [21]). The

inverse problem now seeks the parameter a in (1.4) from a measurement z of u. This inverse problem is often studied in an optimization framework, either formulating the problem as an unconstrained optimization problem or treating it as a constrained optimization problem in which the variational problem itself is the constraint.

Given a Banach space $Z \supseteq V$ equipped with the norm $\| \cdot \|_Z$, the most commonly adopted optimization framework minimizes the following output least-squares (OLS) functional

$$J_1(a) := \|u(a) - z\|_Z^2 \tag{1.5}$$

where $z \in Z$ is the data (the measurement of u) and $u(a)$ solves the variational form (1.4).

For an insight into the abstract framework, consider the boundary value problem (BVP)

$$-\nabla \cdot (a\nabla u) = f \quad \text{in } \Omega, \ u = 0 \ \text{ on } \partial\Omega, \tag{1.6}$$

where Ω is a suitable domain in $\mathbb{R}^2$ or $\mathbb{R}^3$ and $\partial\Omega$ is its boundary. BVP (1.6) models useful real-world problems and has been studied in detail. For example, in (1.6), $u = u(x)$ may represent the steady-state temperature at a point x of a body; then a would be a variable thermal conductivity coefficient, and f the external heat source. BVP (1.6) also models underground steady state aquifers in which the parameter a is the aquifer transmissivity coefficient, u is the hydraulic head, and f is the recharge. The inverse problem in the context of (1.6) is to estimate the parameter a from a measurement z of the solution u.

For (1.6) with $Z = L_2(\Omega)$, optimization problem (1.5) reduces to minimizing

$$\tilde{J}_1(a) := \int_\Omega (u(a) - z)^2, \tag{1.7}$$

where z is the measurement of u and $u(a)$ solves the variational form of (1.6) given by

$$\int_\Omega a\nabla u \cdot \nabla v = \int_\Omega fv, \quad \text{for all } v \in H_0^1(\Omega). \tag{1.8}$$

A variant of (1.5) is the following modified OLS functional (MOLS) introduced in [19]

$$J_2(a) = T(a, u(a) - z, u(a) - z), \tag{1.9}$$

where z is the data and $u(a)$ solves (1.4). In [19], the author established that (1.9) is convex and used it to estimate the Lamé moduli in the equations of isotropic elasticity. Studies related to MOLS functional and its extensions can be found in [15, 20, 26].

The MOLS functional, given in (1.9), was inspired by Knowles [35] who minimized a coefficient-dependent norm

$$\tilde{J}_2(a) := \int_\Omega a\nabla(u(a) - z) \cdot \nabla(u(a) - z), \tag{1.10}$$

where z is the measurement of u, and $u(a)$ solves (1.8).

Besides the OLS functional and the MOLS functional, there are other approaches. The equation error method (cf. [1, 28, 29]), for (1.6), consists of minimizing the functional

$$\tilde{J}_3(a) := \|\nabla(a\nabla z) + f\|_{H^{-1}(\Omega)}^2$$

where $H^{-1}(\Omega)$ is the dual of $H_0^1(\Omega)$ and z is the data. In [16], the equation error approach was explored in an abstract framework.

Finally, we recall the following two results from [19] for the parameter-to-solution map.

Lemma 1.1. *For any $a \in A$, the solution $u(a)$ of the variational problem* (1.4) *satisfies* $\|u(a)\|_V \leq \alpha^{-1}\|m\|_{V^*}$. *Moreover, for any $a, b \in A$, we have*

$$\|u(a) - u(b)\|_V \leq \min\left\{\frac{\beta}{\alpha}\|u(a)\|_V, \frac{\beta}{\alpha}\|u(b)\|_V, \frac{\beta}{\alpha^2}\|m\|_{V^*}\right\}\|b - a\|_B. \tag{1.11}$$

Lemma 1.2. *For each a in the interior of A, the solution $u(a)$ of the variational problem* (1.4) *is infinitely differentiable at a. Given $u = u(a)$, the first derivative $Du(a)\delta a$ of $u(a)$ in the direction δa, is the unique solution of the following variational equation*

$$T(a, Du(a)\delta a, v) = -T(\delta a, u, v), \ \forall\, v \in V. \tag{1.12}$$

Furthermore, the second derivative $D^2 u(a)(\delta a_1, \delta a_2)$ of $u(a)$ in the direction $(\delta a_1, \delta a_2)$, is the unique solution of the following variational equation

$$T(a, D^2 u(a)(\delta a_1, \delta a_2), v) = -T(\delta a_2, Du(a)\delta a_1, v) - T(\delta a_1, Du(a)\delta a_2, v), \ \forall\, v \in V. \tag{1.13}$$

Moreover, the following bounds hold:

$$\|Du(a)\| \leq \frac{\beta}{\alpha}\|u(a)\|_V \leq \frac{\beta}{\alpha^2}\|m\|_{V^*}, \tag{1.14}$$

$$\left\|D^2 u(a)\right\| \leq \frac{2\beta^2}{\alpha^2}\|u(a)\|_V \leq \frac{2\beta^2}{\alpha^3}\|m\|_{V^*}. \tag{1.15}$$

In all of the above results, coercivity condition (1.3) played the most crucial role. It gives the unique solvability of variational problem (1.4), proves bound on the parameter-to-solution map, establishes its Lipschitz continuity and infinite differentiability. As another useful consequences of the coercivity, the first-order and the second-order derivatives of the parameter-to-solution maps are the unique solutions of the variational problems (1.12) and (1.13). Moreover, the useful bounds (1.14) and (1.15) also hold due to the coercivity.

This work aims to study the inverse problem of parameter identification in noncoercive variational problems with perturbed data. Our main contributions are as follows:

(i) Assuming that the noncoercive variational problem is solvable, we give a derivative characterization for the set-valued parameter-to-solution map by using the first-order and the second-order contingent derivatives. To our knowledge, this is the first use of such tools from variational analysis in the study of inverse problems.

(ii) We study the inverse problem by posing optimization problems using the output least-squares and the modified output least-squares functionals for the set-valued parameter to solution map. We regularize the noncoercive variational problem and obtain the single-valued regularized parameter-to-selection map and explore its smoothness. We consider optimization problems using the output least-squares and the modified output least-squares for the regularized variational problem. We prove that the MOLS objective is convex and give a complete convergence analysis showing that the regularized problems approximate the original problem suitably.

4

(iii) To compute the first-order and the second-order derivative of the OLS functional, we give first-order, and second-order adjoint methods in the continuous setting. We provide a discretization scheme and give discrete formulas for the OLS and the MOLS functionals and their gradient and Hessian calculation. As a byproduct of our study, we obtain new insight into the case when the actual trilinear form is coercive, however, for the computations, only its contaminated analog is available which is noncoercive. All the conditions imposed for the convergence analysis are satisfied in this case of practical importance.

We organize the contents of this paper into seven sections. Section 2 introduces the inverse problem and explores the smoothness of the set-valued parameter-to-solution map. Section 3 investigates the inverse problem by using the output-least squares approach and the modified output least squares approach. Section 4 is devoted to the first-order and the second-order adjoint approaches. Section 5 provides a detailed computational framework including the discrete gradient and Hessian formulae. In Section 6, we report the outcome of some preliminary numerical experiments. The paper concludes with some remarks.

2 SET-VALUED SOLUTION MAP FOR A NONCOERCIVE VARIATIONAL PROBLEM

For convenience, we recall the general setting once again. Let B be a Banach space, let $A \subset B$ be a nonempty, closed, and convex set. Let V be a Hilbert space continuously imbedded in a Hilbert space Z. Let $T : B \times V \times V \to \mathbb{R}$ be a trilinear form with $T(a, u, v)$ symmetric in u, v. Let m be a bounded linear functional on V. Assume that T satisfies the continuity assumption (1.2) and the following positivity condition:

$$T(a, u, u) \geq 0, \text{ for all } u \in V, \ a \in A. \tag{2.1}$$

Consider the noncoercive variational problem: Given $a \in A$, find $u = u(a) \in V$ such that

$$T(a, u, v) = m(v), \text{ for every } v \in V. \tag{2.2}$$

Since we do not impose the coercivity condition (see (1.3)) on T, additional conditions are necessary to ensure that (2.2) is solvable. For example, recession analysis can be used to ensure that (2.2) is solvable but such conditions don't guarantee that the solution is unique (see [21]). Therefore, it is natural to study the behavior of the set-valued parameter-to-solution map. For a given parameter $a \in A$, by $\mathcal{U}(a)$ we denote the set of all solutions of variational equation (2.2). In the following, we assume that for each $a \in A$, the set $\mathcal{U}(a)$ is nonempty. The following lemma provides additional information:

Lemma 2.1. *For any $a \in A$, the set of all solutions $\mathcal{U}(a)$ of (2.2) is closed and convex.*

Proof. The proof follows at once from the definition of the set-valued parameter-to-solution map $\mathcal{U} : A \rightrightarrows V$. Indeed, let u and w be two arbitrary elements in $\mathcal{U}(a)$. Then, for every $v \in V$, we have $T(a, u, v) = m(v)$ and $T(a, w, v) = m(v)$. We take $t \in [0, 1]$ and note that for every $v \in V$, we have $T(a, tu, v) = tm(v)$ and $T(a, (1 - t)u, v) = (1 - t)m(v)$. We combine these equations to note that for every $v \in V$, we have $T(a, tu + (1 - t)w, v) = m(v)$. Consequently, $tu + (1 - t)w \in \mathcal{U}(a)$ confirming the convexity of $\mathcal{U}(a)$. The set $\mathcal{U}(a)$ is closed due to the continuity of the trilinear form T. The proof is complete. □

Our goal is to obtain a derivative characterization for the set-valued parameter-to-solution map. In the literature, a variety of derivative concepts have been employed to differentiate set-valued maps (see [32]). We will use first-order and second-order contingent derivatives of the parameter-to-solution set-valued map $\mathcal{U} : A \rightrightarrows V$. These derivatives are defined by using the contingent cone and the second-order contingent set which we recall now.

Definition 2.1. Let X be a normed space, let $S \subset X$ and let $\bar{z} \in \mathrm{cl}(S)$ (closure of S).

(i) The *contingent cone* $C(S, \bar{z})$ of S at $\bar{z}$ is the set of all $z \in X$ such that there are sequences $\{t_n\} \subset \mathbb{P} := \{t \in \mathbb{R} \mid t > 0\}$ and $\{z_n\} \subset X$ with $t_n \downarrow 0$ and $z_n \to z$ such that $\bar{z} + t_n z_n \in S$, for every $n \in \mathbb{N}$.

(ii) The *second-order contingent set* $C^2(S, \bar{z}, w)$ of S at $\bar{z} \in \mathrm{cl}(S)$ in the direction $w \in X$ is the set of all $z \in S$ such that there are a sequence $\{z_n\} \subset X$ with $z_n \to z$ and a sequence $\{t_n\} \subset \mathbb{P}$ with $t_n \downarrow 0$ such that $\bar{z} + t_n w + t_n^2 z_n / 2 \in S$, for every $n \in \mathbb{N}$.

Remark 2.2. It is known that the contingent cone $C(S, \bar{z})$ is a nonempty closed cone. However, $C^2(S, \bar{z}, w)$ is only a closed set (possibly empty), non-connected in general, and it may be nonempty only if $w \in C(S, \bar{z})$. Details of these concepts can be found in [7, 8, 32].

Next we collect some notions for set-valued maps. Given normed spaces X and Y, let $F : X \rightrightarrows Y$ be a set-valued map. The (effective) domain and the graph of F are defined by $\mathrm{dom}(F) := \{x \in X \mid F(x) \neq \emptyset\}$, and $\mathrm{graph}(F) := \{(x, y) \in X \times Y \mid y \in F(x)\}$.

We now introduce first-order and second-order derivatives of set-valued maps.

Definition 2.2. Let X and Y be normed spaces, let $F : X \rightrightarrows Y$ be a set-valued map, and let $(\bar{x}, \bar{y}) \in \mathrm{graph}(F)$. Then the *contingent derivative* of F at $(\bar{x}, \bar{y})$ is the set-valued map $DF(\bar{x}, \bar{y}) : X \rightrightarrows Y$ given by

$$DF(\bar{x}, \bar{y})(x) := \{y \in Y \mid (x, y) \in C(\mathrm{graph}(F), (\bar{x}, \bar{y}))\}.$$

Moreover, the *second-order contingent derivative* of F at $(\bar{x}, \bar{y})$ in the direction $(\bar{u}, \bar{v})$ is the set-valued map $D^2 F(\bar{x}, \bar{y}, \bar{u}, \bar{v}) : X \rightrightarrows Y$ defined by

$$D^2 F(\bar{x}, \bar{y}, \bar{u}, \bar{v})(x) := \left\{y \in Y \mid (x, y) \in C^2(\mathrm{graph}(F), (\bar{x}, \bar{y}), (\bar{u}, \bar{v}))\right\}.$$

The above derivatives have been used extensively in nonsmooth and variational analysis, viability theory, set-valued optimization and numerous other related disciplines, see [32].

We have the following derivative characterization for the parameter-to-solution map:

Theorem 2.3. *For $\bar{a} \in A$, let $\bar{u} \in \mathcal{U}(\bar{a})$ be a given point. Assume that the first-order contingent derivative $D\mathcal{U}(\bar{a}, \bar{u}) : B \rightrightarrows V$ of the set-valued parameter-to-solution map $\mathcal{U} : A \rightrightarrows V$ at the point $(\bar{a}, \bar{u}) \in \mathrm{graph}(\mathcal{U})$ exists. Then for any given direction $\delta a \in B$, any element $\delta u \in D\mathcal{U}(\bar{a}, \bar{u})(\delta a)$ satisfies the following variational problem:*

$$T(\bar{a}, \delta u, v) = -T(\delta a, \bar{u}, v), \quad \text{for every } v \in V. \tag{2.3}$$

Proof. For the given element $(\bar{a}, \bar{u}) \in \mathrm{graph}(\mathcal{U})$ and the given direction δa, for any $\delta u \in DU(\bar{a}, \bar{u})(\delta a)$, we have

$$(\delta a, \delta u) \in \mathrm{graph}(DU(\bar{a}, \bar{u})) = C(\mathrm{graph}(\mathcal{U}), (\bar{a}, \bar{u})),$$

and by the definition of the contingent cone, there are sequences $\{t_n\} \subset \mathbb{P}$ and $\{(a_n, u_n)\}$ with $t_n \to 0$ and $(a_n, u_n) \to (\delta a, \delta u)$ such that $(\bar{a} + t_n a_n, \bar{u} + t_n u_n) \in \mathrm{graph}(\mathcal{U})$, or equivalently $\bar{u} + t_n u_n \in \mathcal{U}(\bar{a} + t_n a_n)$, which, by the definition of the map $\mathcal{U} : A \rightrightarrows V$, implies that

$$T(\bar{a} + t_n a_n, \bar{u} + t_n u_n, v) = m(v), \quad \text{for every } v \in V,$$

and after a rearrangement of this variational problem, we obtain

$$T(\bar{a}, \bar{u}, v) + t_n T(\bar{a}, u_n, v) + t_n T(a_n, \bar{u}, v) + t_n^2 T(a_n, u_n, v) = m(v), \quad \text{for every } v \in V.$$

The condition $(\bar{a}, \bar{u}) \in \mathrm{graph}(\mathcal{U})$ implies that $T(\bar{a}, \bar{u}, v) = m(v)$, for every $v \in V$, and hence

$$T(\bar{a}, u_n, v) = -T(a_n, \bar{u}, v) - t_n T(a_n, u_n, v), \quad \text{for every } v \in V.$$

By passing the above equation to the limit $n \to \infty$, we obtain

$$T(\bar{a}, \delta u, v) = -T(\delta a, \bar{u}, v), \quad \text{for every } v \in V,$$

and the desired identity (2.3) is proved. The proof is complete. $\qquad\square$

Remark 2.4. If the trilinear form T satisfies coercivity condition (1.3), then for every parameter $\bar{a} \in A$, variational problem (2.2) has a unique solution $\bar{u} = u(\bar{a})$, that is, the map $\bar{a} \to u(\bar{a})$ is well-defined and single-valued. Moreover, for any a in the interior of A and any direction δa, the Fréchet derivative $\delta u = D\bar{u}(\bar{a})(\delta a)$ is the unique solution of the variational problem (1.12) which is entirely comparable to the characterization (2.3).

The following is the characterization of the second-order contingent derivative:

Theorem 2.5. *For any $\bar{a} \in A$, let $\bar{u} \in \mathcal{U}(\bar{a})$ be a given element. Assume that second-order contingent derivative of the parameter-to-solution set-valued map $\mathcal{U} : A \rightrightarrows V$ at $(\bar{a}, \bar{u})$ in the direction $(\delta a, \delta u) \in \mathrm{graph}(DU(\bar{a}, \bar{u}))$ exists. Then for any given direction $\delta\tilde{a} \in B$, any element $\delta^2 u \in D^2\mathcal{U}(\bar{a}, \bar{u}, \delta a, \delta u)(\delta\tilde{a})$ satisfies the variational problem:*

$$T(\bar{a}, \delta^2 u, v) = -2T(\delta a, \delta u, v) - T(\delta\tilde{a}, \bar{u}, v), \quad \text{for every } v \in V. \tag{2.4}$$

Proof. For the given $(\bar{a}, \bar{u}) \in \mathrm{graph}(\mathcal{U})$ and the given $(\delta a, \delta u) \in \mathrm{graph}(DU(\bar{a}, \bar{u}))$, let $\delta^2 u \in D^2\mathcal{U}(\bar{a}, \bar{u}, \delta a, \delta u)(\delta\tilde{a})$. Then, we have

$$(\delta\tilde{a}, \delta^2 u) \in \mathrm{graph}(D^2\mathcal{U}(\bar{a}, \bar{u}, \delta a, \delta u)) = C^2(\mathrm{graph}(\mathcal{U}), \bar{a}, \bar{u}, \delta a, \delta u).$$

Therefore, there are sequences $\{t_n\} \subset \mathbb{P}$ and $\{(a_n, u_n)\} \in \mathrm{graph}(\mathcal{U})$ with $t_n \to 0$, and $(a_n, u_n) \to (\delta\tilde{a}, \delta^2 u)$ so that $(\bar{a} + t_n \delta a + \frac{1}{2}t_n^2 a_n, \bar{u} + \frac{1}{2}t_n \delta u + t_n^2 u_n) \in \mathrm{graph}(\mathcal{U})$. By the definition of the parameter-to-solution map, we have

$$T(\bar{a} + t_n \delta a + \frac{1}{2}t_n^2 a_n, \bar{u} + t_n \delta u + \frac{1}{2}t_n^2 u_n, v) = m(v), \quad \text{for every } v \in V.$$

We simplify the above identity as follows

$$T(\bar{a}, \bar{u}, v) + t_n T(\bar{a}, \delta u, v) + \frac{1}{2} t_n^2 T(\bar{a}, u_n, v) + t_n T(\delta a, \bar{u}, v) + t_n^2 T(\delta a, \delta u, v)$$

$$+ \frac{1}{2} t_n^3 T(\delta a, u_n, v) + \frac{1}{2} t_n^2 T(a_n, \bar{u}, v) + \frac{1}{2} t_n^3 T(a_n, \delta u, v) + \frac{1}{4} t_n^4 T(a_n, u_n, v) = m(v), \quad (2.5)$$

which, first by using the fact that $(\bar{a}, \bar{u}) \in \text{graph}(\mathcal{U})$, and then by dividing both sides of the resulting identity by t_n confirms that

$$T(\bar{a}, \delta u, v) + \frac{1}{2} t_n T(\bar{a}, u_n, v) + T(\delta a, \bar{u}, v) + t_n T(\delta a, \delta u, v)$$

$$+ \frac{1}{2} t_n^2 T(\delta a, u_n, v) + \frac{1}{2} t_n T(a_n, \bar{u}, v) + \frac{1}{2} t_n^2 T(a_n, \delta u, v) + \frac{1}{4} t_n^3 T(a_n, u_n, v) = 0,$$

We now first use the fact that $(\delta a, \delta u) \in \text{graph}(D\mathcal{U}(\bar{a}, \bar{u}))$, and then divide both sides of the resulting identity by t_n to obtain

$$\frac{1}{2} T(\bar{a}, u_n, v) + T(\delta a, \delta u, v) + \frac{1}{2} t_n T(\delta a, u_n, v)$$

$$+ \frac{1}{2} T(a_n, \bar{u}, v) + \frac{1}{2} t_n T(a_n, \delta u, v) + \frac{1}{4} t_n^2 T(a_n, u_n, v) = 0,$$

which when passed to the limit $t_n \to 0$, yields

$$T(\bar{a}, \delta^2 u, v) = -2T(\delta a, \delta u, v) - T(\delta \tilde{a}, \bar{u}, v),$$

proving (2.4). The proof is complete. $\qquad \qquad \square$

Note that due to the characterization of the first-order contingent derivative of the set-valued map $\mathcal{U} : A \rightrightarrows V$, variational problem (2.4) is equivalent to

$$T(\bar{a}, \delta^2 u, v) = -2T(\delta a, \delta u, v) + T(\bar{a}, \delta \tilde{u}, v), \qquad (2.6)$$

where $\delta \tilde{u} \in D\mathcal{U}(\bar{a}, \bar{u})(\delta \tilde{a})$ and $v \in V$ is arbitrary.

Clearly, if T satisfies condition (1.3), then the parameter-to-solution map is single-valued and infinitely differentiable in the interior of the domain. Moreover, for any $\bar{a}$ in the interior of A, and suitable directions $(\delta a_1, \delta a_2)$, the second-order derivative $D^2 u(\bar{a})(\delta a_1, \delta a_2)$ is the unique solution of (1.13). In particular, with $\delta a = \delta a_1 = \delta a_2$, we have

$$T(\bar{a}, D^2 u(\bar{a})(\delta a, \delta a), v) = -2T(\delta a, Du(a)\delta a, v), \quad \text{for every } v \in V. \qquad (2.7)$$

We also recall that if a single-valued map $F : X \to Y$ is twice differentiable, then with $DF(x)$ and $D^2 F(x)$ as the first-order and the second-order derivatives, we have (see [41])

$$C^2(\text{graph}(F), (x, F(x), v, DF(x)v)) = \{(y, DF(x)y + D^2 F(x)(v, v)), \ y \in X\}.$$

Consequently, by taking $\delta a = \delta \tilde{a}$, we have

$$\text{graph}(D^2 \mathcal{U}(\bar{a}, \bar{u}, \delta a, \delta u)) = \{(\delta a, D\bar{u}(\bar{a})\delta a + D^2\bar{u}(\bar{a})(\delta a, \delta a)), \ \delta a \in B\}$$

and, as a result, under (1.3), the derivative formula yields

$$T(\bar{a}, D\bar{u}(\bar{a})\delta a + D^2\bar{u}(\bar{a})(\delta a, \delta a), v) = -2T(\delta a, D\bar{u}(\bar{a})(\delta a), v) + T(\bar{a}, D\bar{u}(\bar{a})\delta a, v), \ \text{ for all } v \in V,$$

implying that

$$T(\bar{a}, D^2\bar{u}(\bar{a})(\delta a, \delta a), v) = -2T(\delta a, D\bar{u}(\delta)(\delta a), v), \quad \text{ for every } v \in V,$$

which is in compliance with the second-order formula (2.7).

Remark 2.6. The results given above only offer characterizations of the first-order and the second-order contingent derivatives under the critical assumption that these derivatives exist. This is a natural step as we have not identified conditions under which the variational problem is solvable. A possible extension of these results is singling out conditions ensuring the existence of solutions and then using them to verify the contingent differentiability.

3 RECASTING THE INVERSE PROBLEM IN AN OPTIMIZATION FRAMEWORK

3.1 THE OUTPUT LEAST-SQUARES APPROACH

Let $\mathcal{U} : A \rightrightarrows V$ be the set-valued parameter-to-solution map which assigns to each $a \in A$, the set of all solutions $\mathcal{U}(a)$ of the noncoercive variational problem (2.2). We define the set-valued output least-squares map $\widehat{J} : A \rightrightarrows \mathbb{R}$ which connects to each $a \in A$, the following set

$$\widehat{J}(a) := \left\{ \|u(a) - z\|_Z^2 \ | \ u(a) \in \mathcal{U}(a) \right\},$$

where $z \in Z$ is the measured data.

Using the above set-valued map (and a slight abuse of the notation), we pose the following OLS-based optimization problem

$$\min_{a \in A} \widehat{J}(a). \tag{3.1}$$

The philosophy of the OLS approach is to minimize the gap between the computed solutions $u(a) \in \mathcal{U}(a)$ of (2.2) and the measured data $z \in Z$.

An element $\bar{a} \in A$ is called a minimizer of (3.1), if there exists $u(\bar{a}) \in \mathcal{U}(\bar{a})$ such that

$$\|u(\bar{a}) - z\|_Z^2 \leq \|u(a) - z\|_Z^2, \quad \text{ for every } u(a) \in \mathcal{U}(a), \text{ for every } a \in A. \tag{3.2}$$

To emphasize the role of $u(\bar{a})$, we sometimes say that $(\bar{a}, u(\bar{a})) \in \text{graph}(\mathcal{U})$ is a minimizer.

One of our goals is to approximate (3.1) by a sequence of solutions of optimization problems for which the entire data set of the constraint variational problem is noisy in the sense described below. Let $\{\epsilon_n\}$, $\{\tau_n\}$, $\{\kappa_n\}$, $\{\delta_n\}$, and $\{\nu_n\}$ be sequences of positive reals. Let $\ell \in V^*$ be a given

element. For each $n \in \mathbb{N}$, let $m_{\nu_n} \in V^*$ and $\ell_{\delta_n} \in V^*$ be given elements, and let $z_{\delta_n} \in Z$ be the contaminated data such that the following inequalities hold:

$$\|z_{\delta_n} - z\|_Z \leq \delta_n, \tag{3.3a}$$

$$\|m_{\nu_n} - m\|_{V^*} \leq \nu_n, \tag{3.3b}$$

$$\|\ell_{\delta_n} - \ell\|_{V^*} \leq \delta_n. \tag{3.3c}$$

Furthermore, for each $n \in \mathbb{N}$, let $T_{\tau_n} : B \times V \times V \to \mathbb{R}$ be a trilinear form such that

$$T_{\tau_n}(a, u, u) \geq 0, \text{ for all } u \in V, \ a \in A. \tag{3.4a}$$

$$\left| T_{\tau_n}(a, u, v) - T(a, u, v) \right| \leq \tau_n \|a\|_B \|u\|_V \|v\|_V, \text{ for all } u, v \in V, \ a \in B. \tag{3.4b}$$

Moreover, as $n \to \infty$, the sequences $\{\epsilon_n\}$, $\{\tau_n\}$, $\{\kappa_n\}$, $\{\delta_n\}$, and $\{\nu_n\}$ satisfy

$$\left\{ \tau_n, \epsilon_n, \kappa_n, \nu_n, \delta_n, \frac{\tau_n}{\epsilon_n}, \frac{\delta_n}{\epsilon_n}, \frac{\nu_n}{\epsilon_n} \right\} \to 0. \tag{3.5}$$

Finally, let $S : V \times V \to \mathbb{R}$ be a symmetric bilinear form such that there are constants $\alpha_0 > 0$ and $\beta_0 > 0$ satisfying the following continuity and coercivity conditions

$$S(u, v) \leq \beta_0 \|u\|_V \|v\|_V, \text{ for all } u, v \in V, \tag{3.6a}$$

$$S(u, u) \geq \alpha_0 \|u\|_V^2, \text{ for all } u \in V. \tag{3.6b}$$

With the above preparation, for each $n \in \mathbb{N}$, we now consider the following regularized variational problem: Given $a \in A$, find $u_{\varsigma_n}(a) \in V$ such that

$$T_{\tau_n}(a, u_{\varsigma_n}(a), v) + \epsilon_n S(u_{\varsigma_n}(a), v) = m_{\nu_n}(v) + \epsilon_n \ell_{\delta_n}(v), \text{ for every } v \in V. \tag{3.7}$$

where $\epsilon_n > 0$ is the regularization parameter and for simplicity, we set $\varsigma_n := (\epsilon_n, \tau_n, \nu_n, \delta_n)$.

In view of the above conditions, for a fixed $n \in \mathbb{N}$, and for every $a \in A$, (3.7) has a unique solution $u_{\varsigma_n}(a)$. Therefore, the regularized parameter-to-solution map $a \to u_{\varsigma_n}(a)$ is well-defined and single-valued.

The next result embarks on the smoothness of the regularized parameter-to-solution map:

Theorem 3.1. *For any $n \in \mathbb{N}$ and any parameter a in the interior of A, the regularized parameter-to-solution map $a \to u_{\varsigma_n}(a)$ is infinitely differentiable at a. Moreover, given $u_{\varsigma_n}(a)$, the first-order derivative $Du_{\varsigma_n}(a)\delta a$ in the direction $\delta a \in B$ is the unique solution of the variational equation*

$$T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a, v) + \epsilon_n S(Du_{\varsigma_n}(a)\delta a, v) = -T_{\tau_n}(\delta a, u_{\varsigma_n}, v), \text{ for every } v \in V, \tag{3.8}$$

and the second-order derivative $D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2)$ in the direction $(\delta a_1, \delta a_2) \in B \times B$ is the unique solution of the variational equation

$$T_{\tau_n}(a, D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), v) + \epsilon_n S(D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), v)$$
$$= -T_{\tau_n}(\delta a_2, Du_{\varsigma_n}(a)\delta a_1, v) - T_{\tau_n}(\delta a_1, Du_{\varsigma_n}(a)\delta a_2, v), \text{ for every } v \in V. \tag{3.9}$$

Proof. The proof follows by similar arguments that were used in the proof of Lemma 1.2. The crucial role in the proof is played by the ellipticity of $T_{\tau_n} + \epsilon_n S$. □

Before any further advancement, in the following result we give necessary conditions ensuring that the derivative of the regularized parameter-to-solution map remains bounded.

Theorem 3.2. *For a parameter $\bar{a}$ in the interior of A, let $\bar{u} \in \mathcal{U}(\bar{a})$ be a given point. Assume that the first-order contingent derivative of the set-valued parameter-to-solution map $\mathcal{U} : A \rightrightarrows V$ at the point $(\bar{a}, \bar{u}) \in graph(\mathcal{U})$ exists. If*

$$\|u_{\varsigma_n}(\bar{a}) - \bar{u}\|_V = O(\epsilon_n), \tag{3.10}$$

where $u_{\varsigma_n}(\bar{a})$ is the regularized solution of (3.7) for parameter $\bar{a}$, then the first-order derivative $Du_{\varsigma_n}(\bar{a})\delta a$ of $u_{\varsigma_n}(\bar{a})$ in any direction $\delta a \in B$ is uniformly bounded.

Proof. From Theorem 2.3, for any $\delta a \in B$, and any $\delta u \in D\mathcal{U}(\bar{a}, \bar{u})(\delta a)$, we have

$$T(\bar{a}, \delta u, v) = -T(\delta a, \bar{u}, v), \quad \text{for every } v \in V. \tag{3.11}$$

Furthermore, due to Theorem 3.1, we also have

$$T_{\tau_n}(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a, v) + \epsilon_n S(Du_{\varsigma_n}(\bar{a})\delta a, v) = -T_{\tau_n}(\delta a, u_{\varsigma_n}(\bar{a}), v), \text{ for every } v \in V. \tag{3.12}$$

We subtract (3.11) from (3.12) and rearrange the resulting equation to obtain

$$T(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a - \delta u, v) + \epsilon_n S(Du_{\varsigma_n}(\bar{a})\delta a, v) = T(\delta a, \bar{u} - u_{\varsigma_n}(\bar{a}), v) + T(\delta a, u_{\varsigma_n}(\bar{a}), v)$$
$$- T_{\tau_n}(\delta a, u_{\varsigma_n}(\bar{a}), v) + T(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a, v) - T_{\tau_n}(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a, v).$$

By setting $v = Du_{\varsigma_n}(\bar{a})\delta a - \delta u$ and using the positivity of T, we obtain

$$\epsilon_n \|Du_{\varsigma_n}(\bar{a})\delta a - \delta u\|_V^2$$
$$\leq T(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a - \delta u, Du_{\varsigma_n}(\bar{a})\delta a - \delta u) + \epsilon_n S(Du_{\varsigma_n}(\bar{a})\delta a - \delta u, Du_{\varsigma_n}(\bar{a})\delta a - \delta u)$$
$$= T(\delta a, \bar{u} - u_{\varsigma_n}(\bar{a}), Du_{\varsigma_n}(\bar{a})\delta a - \delta u) + T(\delta a, u_{\varsigma_n}(\bar{a}), Du_{\varsigma_n}(\bar{a})\delta a - \delta u)$$
$$- T_{\tau_n}(\delta a, u_{\varsigma_n}(\bar{a}), Du_{\varsigma_n}(\bar{a})\delta a - \delta u) + T(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a, Du_{\varsigma_n}(\bar{a})\delta a - \delta u)$$
$$- T_{\tau_n}(\bar{a}, Du_{\varsigma_n}(\bar{a})\delta a, Du_{\varsigma_n}(\bar{a})\delta a - \delta u) - \epsilon_n S(\delta u, Du_{\varsigma_n}(\bar{a})\delta a - \delta u),$$

which, due to the properties of T and S, implies that

$$\epsilon_n \|Du_{\varsigma_n}(\bar{a})\delta a - \delta u\|_V \leq \beta \|\delta a\|_B \|\bar{u} - u_{\varsigma_n}(\bar{a})\|_V + \tau_n \|\delta a\|_B \|u_{\varsigma_n}(\bar{a})\|_V$$
$$+ \tau_n \|\bar{a}\|_B \|Du_{\varsigma_n}(\bar{a})\delta a\|_V + \beta_0 \epsilon_n \|\delta u\|,$$

and hence

$$\left(1 - \frac{\tau_n}{\epsilon_n} \|\bar{a}\|_B\right) \|Du_{\varsigma_n}(\bar{a})\delta a - \delta u\|_V \leq [\beta + \tau_n] \|\delta a\|_B \frac{\|u_{\varsigma_n}(\bar{a}) - \bar{u}\|}{\epsilon_n} + \frac{\tau_n}{\epsilon_n} \|\delta a\|_B \|\bar{u}\|$$
$$+ \beta_0 \|\delta u\| + \frac{\tau_n}{\epsilon_n} \|\bar{a}\|_B \|\delta u\|_V.$$

In view of (3.5) and the assumption that $\|u_{\varsigma_n}(\bar{a}) - \bar{u}\| = O(\epsilon_n)$, it follows that there is a constant $c > 0$ such that

$$\left(1 - \frac{\tau_n}{\epsilon_n}\|\bar{a}\|_B\right)\|Du_{\varsigma_n}(\bar{a})\delta a - \delta u\|_V \le c,$$

and since $\frac{\tau_n}{\epsilon_n} \to 0$ as $n \to \infty$, for sufficiently large $n \in \mathbb{N}$, we have $\left(1 - \frac{\tau_n}{\epsilon_n}\|\bar{a}\|_B\right) > 0$ and the boundedness of $\|Du_{\varsigma_n}(\bar{a})\delta a - \delta u\|_V$ follows. The proof is complete. $\qquad\square$

Remark 3.3. The fundamental idea of the elliptic regularization for variational problems is to combine the variational problem with the regularized analog to ensure that the regularized solutions remain bounded (see [31]). Then a subsequence can be extracted and shown to converge weakly to a solution of the variational problem, ensuring its solvability. In the above result, we use this idea to prove the boundedness of the derivatives. In the present context, the role of the original variational problem is played by the derivative characterization involving the first-order contingent derivative. As a consequence, suitable conditions ensuring the boundedness of the derivatives of the regularized parameter-to-solution map can be used to show the contingent differentiability of the set-valued parameter-to-solution map. We also note that if the original trilinear map is elliptic, then (3.10) holds.

To incorporate regularization in the ill-posed inverse problem, we assume:

(i) The Banach space B is continuously embedded in a Banach space L. There is another Banach space $\widehat{B}$ that is compactly embedded in L. The set A is a subset of $B \cap \widehat{B}$, closed and bounded in B and also closed in L.

(ii) $R : \widehat{B} \to \mathbb{R}$ is positive, convex, and lower-semicontinuous in $\|\cdot\|_L$ such that

$$R(a) \ge \tau_1\|a\|_{\widehat{B}} - \tau_2, \quad \text{for every } a \in A, \quad \text{for some } \tau_1 > 0, \ \tau_2 > 0. \tag{3.13}$$

(iii) For any $\{b_k\} \subset B$ with $b_k \to 0$ in L, any bounded $\{u_k\} \subset V$, and fixed $v \in V$, we have

$$T(b_k, u_k, v) \to 0. \tag{3.14}$$

The above framework is inspired by the use of total variation regularization in the identification of discontinuous coefficients. Recall that the total variation of $f \in L^1(\Omega)$ reads

$$\text{TV}(f) = \sup\left\{\int_\Omega f\,(\nabla \cdot g) \,:\, g \in \left(C_0^1(\Omega)\right)^N, \ |g(x)| \le 1 \text{ for all } x \in \Omega\right\}$$

where $|\cdot|$ is the Euclidean norm. Clearly, if $f \in W^{1,1}(\Omega)$, then $\text{TV}(f) = \int_\Omega |\nabla f|$.

If $f \in L^1(\Omega)$ satisfies $\text{TV}(f) < \infty$, then f is said to have bounded variation, and $\text{BV}(\Omega)$ is defined by $\text{BV}(\Omega) = \left\{f \in L^1(\Omega) \,:\, \text{TV}(f) < \infty\right\}$ with norm $\|f\|_{\text{BV}(\Omega)} = \|f\|_{L^1(\Omega)} + \text{TV}(f)$. The functional $\text{TV}(\cdot)$ is a seminorm on $\text{BV}(\Omega)$ and is often called the BV-seminorm, see [39].

We set $B = L^\infty(\Omega)$, $L = L^1(\Omega)$, $\widehat{B} = \text{BV}(\Omega)$, and $R(a) = TV(a)$, and define

$$A = \{a \in L^\infty \,|\, 0 < c_1 \le a(x) \le c_2, \ a.e. \text{ in } \Omega, \ \text{TV}(a) \le c_3 < \infty\}, \tag{3.15}$$

where c_1, c_2, and c_3 are positive constants. Clearly, A is bounded in $\| \cdot \|_{\widehat{B}}$ and compact in L. It is known that $L^\infty(\Omega)$ is continuously embedded in $L^1(\Omega)$, $BV(\Omega)$ is compactly embedded in $L^1(\Omega)$, and $TV(\cdot)$ is convex and lower-semicontinuous in $L^1(\Omega)$-norm. Thus in this setting, assumptions 1 and 2 are satisfied.

Remark 3.4. The regularization framework devised above simplifies if we assume that the set A belongs to a Hilbert space H that is compactly embedded in the space B. An example for this setting is $B = L^\infty(\Omega)$ and $H = H^2(\Omega)$, for a suitable domain Ω.

Our objective is to approximate (3.1) by the following family of regularized optimization problems: For $n \in \mathbb{N}$, find $a_{\varsigma_n} \in A$ by solving

$$\min_{a \in A} \widehat{J}_{\kappa_n}(a) := \frac{1}{2} \| u_{\varsigma_n}(a) - z_{\delta_n} \|_Z^2 + \kappa_n R(a), \tag{3.16}$$

where $u_{\varsigma_n}(a)$ is the unique solution of (3.7), $\kappa_n > 0$, and R is the regularizer defined above.

The following main result of this section shows that (3.16) approximates (3.1):

Theorem 3.5. *Assume that the following conditions hold:*

(i) *The set A is bounded in $\widehat{B}$, the image set $\mathcal{U}(A)$ is a bounded set, and for each $a \in A$, the solution set $\mathcal{U}(a)$ is nonempty.*

(ii) *For $a \in A$, either $\mathcal{U}(a)$ is a singleton, or $Z = V$, $\ell_{\delta_n}(v) = \langle z_{\delta_n}, v \rangle_Z$ and $S(u, v) = \langle u, v \rangle_Z$.*

Then optimization problem (3.1) has a solution, and for each $n \in \mathbb{N}$, optimization problem (3.16) has a solution a_{ς_n}. Moreover, there is a subsequence $\{a_{\varsigma_n}\}$ converging in $\| \cdot \|_L$ to a solution of (3.1). Finally, for any solution a_{ς_n} of (3.1), there is a unique $p_{\varsigma_n} \in V$ such that

$$T_{\tau_n}(a_{\varsigma_n}, p_{\varsigma_n}, v) + \epsilon_n S(p_{\varsigma_n}, v) = \langle z - u_{\varsigma_n}(a_{\varsigma_n}), v \rangle_Z, \quad \text{for every } v \in V, \tag{3.17}$$

$$T_{\tau_n}(a - a_{\varsigma_n}, u_{\varsigma_n}(a_{\varsigma_n}), p_{\varsigma_n}) \geq \kappa_n(R(a_{\varsigma_n}) - R(a)), \quad \text{for every } a \in A. \tag{3.18}$$

Proof. We begin by showing that (3.1) has a solution. Since $\mathcal{U}(a)$ is nonempty for each $a \in A$, optimization problem (3.1) is well-defined. Moreover, since for each $a \in A$, $\widehat{J}(a)$ is bounded from below, there is a minimizing sequence $\{a_n\}$ in A such that $\lim_{n \to \infty} \widehat{J}(a_n) = \inf\{\widehat{J}(a), \ a \in A\}$. Since A is bounded, the sequence $\{a_n\}$ is bounded in $\widehat{B}$, and due to the compact embedding of $\widehat{B}$ into L, it has a subsequence which converges strongly in $\| \cdot \|_L$. Keeping the same notation for subsequences as well, let $\{a_n\}$ be the subsequence converging in $\| \cdot \|_L$ to some $\bar{a} \in A$. Let $u_n \in \mathcal{U}(a_n)$ be arbitrary. Since $\mathcal{U}(A)$ is bounded, $\{u_n\}$ is bounded, and hence contains a weakly convergent subsequence. Let $\{u_n\}$ be the subsequence which converges to some $\bar{u} \in V$. We claim that $\bar{u} \in \mathcal{U}(\bar{a})$. By the definition of (a_n, u_n), we have

$$T(a_n, u_n, v) = m(v), \quad \text{for every } v \in V,$$

which can be rearranged as follows

$$T(a_n - \bar{a}, u_n, v) + T(\bar{a}, u_n - \bar{u}, v) + T(\bar{a}, \bar{u}, v) = m(v), \quad \text{for every } v \in V,$$

and by passing this equation to the limit $n \to \infty$, we obtain

$$T(\bar{a}, \bar{u}, v) = m(v), \quad \text{for every } v \in V,$$

which means that $\bar{u} \in \mathcal{U}(\bar{a})$. The optimality of $\bar{a}$ now is direct consequence of the weak-lower-semicontinuity of $\|\cdot\|_Z$ and the lower-semicontinuity of R.

We now return to (3.16). Evidently, for a fixed $n \in \mathbb{N}$, the existence of a solution a_{ς_n} for (3.16) is a consequence of the arguments just used. Indeed, for any fixed $n \in \mathbb{N}$, (3.7) is uniquely solvable and the solution is bounded because of the ellipticity of $T_{\tau_n} + \epsilon_n S$.

For simplicity, we set $a_n := a_{\varsigma_n}$. Since A is bounded, the sequence of solutions $\{a_n\}$ is uniformly bounded in $\widehat{B}$. As before, let $\{a_n\}$ be a subsequence that converges strongly to some $\bar{a} \in A$ in L. Let $\{u_n\}$, where $u_n := u_{\varsigma_n}(a_n)$, be the corresponding sequence of the solutions of the regularized variational problem (3.7). That is, we have

$$T_{\tau_n}(a_n, u_n, v) + \epsilon_n S(u_n, v) = m_{v_n}(v) + \epsilon_n \ell_{\delta_n}(v), \quad \text{for every } v \in V.$$

We shall prove that $\{u_n\}$ is a bounded sequence. By assumption, for every $a \in A$, the solution set $\mathcal{U}(a)$ is nonempty. Let $\tilde{u}_n \in \mathcal{U}(a_n)$ be chosen arbitrarily. Since $\mathcal{U}(A)$ is bounded by assumption, the sequence $\{\tilde{u}_n\}$ is bounded. Moreover, we have

$$T(a_n, \tilde{u}_n, v) = m(v), \quad \text{for every } v \in V.$$

We combine the above two variational problems and rearrange them to obtain

$$T_{\tau_n}(a_n, \tilde{u}_n, v) - T(a_n, \tilde{u}_n, v) + \epsilon_n S(u_n, v) + m(v) - m_{v_n}(v) - \epsilon_n \ell_{\delta_n}(v) - T_{\tau_n}(a_n, \tilde{u}_n - u_n, v) = 0.$$

Setting $v = \tilde{u}_n - u_n$ and using the fact that $T_{\tau_n}(a_n, \tilde{u}_n - u_n, \tilde{u}_n - u_n) \geq 0$, we obtain

$$\begin{aligned}
\epsilon_n S(u_n, u_n) &\leq \epsilon_n S(u_n, \tilde{u}_n) + T_{\tau_n}(a_n, \tilde{u}_n, \tilde{u}_n - u_n) - T(a_n, \tilde{u}_n, \tilde{u}_n - u_n) - \epsilon_n \ell_{\delta_n}(\tilde{u}_n - u_n). \\
&\quad + m(\tilde{u}_n - u_n) - m_{v_n}(\tilde{u}_n - u_n) \\
&\leq \epsilon_n \beta_0 \|u_n\|_V \|\tilde{u}_n\|_V + \tau_n \|a_n\|_B \|\tilde{u}_n\|_V \|\tilde{u}_n - u_n\|_V + \epsilon_n \|\ell_{\delta_n}\|_{V^*} \|\tilde{u}_n - u_n\|_V \\
&\quad + v_n \|\tilde{u}_n - u_n\|_V,
\end{aligned}$$

implying

$$\|u_n\|_V \leq \frac{\beta_0}{\alpha_0} \|\tilde{u}_n\|_V + \left[\frac{\tau_n}{\alpha_0 \epsilon_n} \|a_n\|_B \|\tilde{u}_n\|_V + \frac{v_n}{\alpha_0 \epsilon_n} + \frac{\delta_n + \|\ell\|_{V^*}}{\alpha_0} \right] \left[\frac{\|\tilde{u}_n\|_V}{\|u_n\|_V} + 1 \right],$$

which confirms the boundedness of $\{u_n\}$.

The reflexivity of V ensures that $\{u_n\}$ has a weakly convergent subsequence. Keeping the same notation for subsequences, let $\{u_n\}$ be a subsequence converging weakly to some $\bar{u}$. We shall show that $\bar{u} \in \mathcal{U}(\bar{a})$. Since a_n is a minimizer of (3.16), we have

$$T_{\tau_n}(a_n, u_n, v) + \epsilon_n S(u_n, v) = m_{v_n}(v) + \epsilon_n \ell_{\delta_n}(v) \quad \text{for every } v \in V,$$

and by using the rearrangement

$$\begin{aligned}
T_{\tau_n}(a_n, u_n, v) &= T(a_n, u_n, v) + T_{\tau_n}(a_n, u_n, v) - T(a_n, u_n, v) \\
&= T(a_n - \bar{a}, u_n, v) + T(\bar{a}, u_n - \bar{u}, v) + T(\bar{a}, \bar{u}, v) \\
&\quad + T_{\tau_n}(a_n, u_n, v) - T(a_n, u_n, v),
\end{aligned}$$

14

we obtain the following equation

$$T(a_n - \bar{a}, u_n, v) + T(\bar{a}, u_n - \bar{u}, v) + T(\bar{a}, \bar{u}, v) + T_{\tau_n}(a_n, u_n, v) - T(a_n, u_n, v)$$
$$+ \epsilon_n S(u_n, v) = m_{\nu_n}(v) + \epsilon_n \ell_{\delta_n}(v),$$

which due to the imposed conditions, when passed to the limit $n \to \infty$, implies that

$$T(\bar{a}, \bar{u}, v) = m(v),$$

and because of the fact that $v \in V$ was chosen arbitrary, confirms that $\bar{u} \in \mathcal{U}(\bar{a})$.

The optimality of $a_n \in A$ for (3.16) means that for $n \in \mathbb{N}$ and each $a \in A$, we have

$$\widehat{J}_{\kappa_n}(a_n) := \frac{1}{2} \|u_{\varsigma_n}(a_n) - z_{\delta_n}\|_Z^2 + \kappa_n R(a_n) \le \frac{1}{2} \|u_{\varsigma_n}(a) - z_{\delta_n}\|_Z^2 + \kappa_n R(a), \tag{3.19}$$

where $u_{\varsigma_n}(a)$ is the solution of regularized optimization problem (3.7).

Let $(\hat{a}, \hat{u})$ be a solution of (3.1). Before any further advancement, we first analyze the behavior of $u_{\varsigma_n}(\hat{a})$. By the definition of $u_{\varsigma_n}(\hat{a})$, we have

$$T_{\tau_n}(\hat{a}, u_{\varsigma_n}(\hat{a}), v) + \epsilon_n S(u_{\varsigma_n}(\hat{a}), v) = m_{\nu_n}(v) + \epsilon_n \ell_{\delta_n}(v), \quad \text{for every } v \in V. \tag{3.20}$$

As in earlier part of this proof, it can be shown that $\{u_{\varsigma_n}(\hat{a})\}$ is uniformly bounded. Therefore, there is a subsequence $\{u_{\varsigma_n}(\hat{a})\}$ converging weakly to some $\bar{u}(\hat{a}) \in \mathcal{U}(\hat{a})$.

Recalling that the set $\mathcal{U}(\hat{a})$ is closed and convex, we consider the following variational inequality: Find $\tilde{u}(\hat{a}) \in \mathcal{U}(\hat{a})$ such that

$$S(\tilde{u}(\hat{a}), w - \tilde{u}(\hat{a})) \ge \ell(w - \tilde{u}(\hat{a})), \text{ for every } w \in \mathcal{U}(\hat{a}). \tag{3.21}$$

Due to the ellipticity of $S(\cdot, \cdot)$, the above variational inequality has a unique solution $\tilde{u}(\hat{a})$. Furthermore, since $\tilde{u}(\hat{a}) \in \mathcal{U}(\hat{a})$, we have

$$T(\hat{a}, \tilde{u}(\hat{a}), v) = m(v), \quad \text{for every } v \in V. \tag{3.22}$$

We combine (3.20) and (3.22) to obtain

$$T(\hat{a}, u_{\varsigma_n}(\hat{a}) - \tilde{u}(\hat{a}), v) + T_{\tau_n}(\hat{a}, u_{\varsigma_n}(\hat{a}), v) - T(\hat{a}, u_{\varsigma_n}(\hat{a}), v) + \epsilon_n S(u_{\varsigma_n}(\hat{a}), v)$$
$$= m_{\nu_n}(v) - m(v) + \epsilon_n \ell_{\delta_n}(v)$$

and by setting $v = \tilde{u}(\hat{a}) - u_{\varsigma_n}(\hat{a})$, and using the positivity of T, we get

$$\left[\frac{\tau_n}{\epsilon_n} \|\hat{a}\|_B \|u_{\varsigma_n}(\hat{a})\|_V + \frac{\nu_n}{\epsilon_n} + \delta_n \right] \|\tilde{u}(\hat{a}) - u_{\varsigma_n}(\hat{a})\|_V - \ell(\tilde{u}(\hat{a}) - u_{\varsigma_n}(\hat{a}))$$
$$\ge S(u_{\varsigma_n}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \tilde{u}(\hat{a})) \ge S(\tilde{u}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \tilde{u}(\hat{a})) \tag{3.23}$$

Since the bilinear form S is positive, we have

$$S(\bar{u}(\hat{a}), \bar{u}(\hat{a})) \le \liminf_{n \to \infty} S(u_{\varsigma_n}(\hat{a}), u_{\varsigma_n}(\hat{a})),$$

15

which, by taking (3.23) into account, implies that

$$S(\bar{u}(\hat{a}), \tilde{u}(\hat{a}) - \bar{u}(\hat{a})) \geq \ell(\tilde{u}(\hat{a}) - \bar{u}(\hat{a})). \tag{3.24}$$

We set $w = \bar{u}(\hat{a})$ in (3.21) to obtain

$$S(\tilde{u}(\hat{a}), \bar{u}(\hat{a}) - \tilde{u}(\hat{a})) \geq \ell(\bar{u}(\hat{a}) - \tilde{u}(\hat{a})),$$

which when combined with (3.24) yields $S(\bar{u}(\hat{a}) - \tilde{u}(\hat{a}), \tilde{u}(\hat{a}) - \bar{u}(\hat{a})) \geq 0$, implying

$$0 \geq S(\bar{u}(\hat{a}) - \tilde{u}(\hat{a}), \bar{u}(\hat{a}) - \tilde{u}(\hat{a})) \geq \alpha_0 \|\bar{u}(\hat{a}) - \tilde{u}(\hat{a})\|_V^2,$$

and hence $\bar{u}(\hat{a}) = \tilde{u}(\hat{a})$. Since $\bar{u}(\hat{a})$ is unique, the whole sequence $u_{\varsigma_n}(\hat{a})$ converges weakly to $\bar{u}(\hat{a})$. The convergence is in fact strong due to (3.23). Indeed, by the coercivity of S, we have

$$\alpha_0 \|u_{\varsigma_n}(\hat{a}) - \bar{u}(\hat{a})\|_V^2 \leq S(u_{\varsigma_n}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \bar{u}(\hat{a})) - S(\bar{u}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \bar{u}(\hat{a})),$$

where $S(u_{\varsigma_n}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \bar{u}(\hat{a})) \to 0$ as $n \to \infty$ by using (3.23) and $S(\bar{u}(\hat{a}), u_{\varsigma_n}(\hat{a}) - \bar{u}(\hat{a})) \to 0$ as $n \to \infty$ by the linearity of S. Hence the strong convergence of $\{u_{\varsigma_n}(\hat{a})\}$ to $\bar{u}(\hat{a})$ follows.

The above observations are valid when $\mathcal{U}$ is a set-valued map. We now prove the final assertion by assuming that $Z = V$, for any $v \in V$, we have $\ell_{\delta_n}(v) = \langle z_{\delta_n}, v \rangle_V$ and $S(u, v) = \langle u, v \rangle_V$. Then, it follows from (3.21) that for an arbitrary $\check{u}(\hat{a}) \in \mathcal{U}(\hat{a})$, we have

$$\langle \bar{u}(\hat{a}) - z, \check{u} - \bar{u}(\hat{a}) \rangle_V \geq 0$$

which implies that

$$\|\bar{u}(\hat{a}) - z\|_Z \leq \|\check{u}(\hat{a}) - z\|_Z,$$

and hence $\bar{u}(\hat{a})$ is the closest element to z among all the elements $\check{u}(\hat{a}) \in \mathcal{U}(\hat{a})$.

Therefore, as before, we have

$$\begin{aligned}
\|u(\bar{a}) - z\|_Z^2 &\leq \liminf_{n \to \infty} \left\{ \|u_{\varsigma_n}(\hat{a}) - z_{\delta_n}\|_Z^2 + \kappa_n R(\hat{a}) \right\}, \\
&\leq \limsup_{n \to \infty} \|u_{\varsigma_n}(\hat{a}) - z\|_Z^2 \\
&= \|\bar{u}(\hat{a}) - z\|_Z^2 \\
&\leq \|\check{u}(\hat{a}) - z\|_Z^2,
\end{aligned}$$

where $\check{u}(\hat{a}) \in \mathcal{U}(\hat{a})$ is arbitrary. In other words, the above inequality confirms the existence of an element $(\bar{a}, u(\bar{a})) \in \mathrm{graph}(\mathcal{U})$ such that for every $(a, u) \in \mathrm{graph}(\mathcal{U})$, we have

$$\|u(\bar{a}) - z\|_Z^2 \leq \|u - z\|_Z^2$$

and hence $\bar{a} \in A$ is a minimizer of (3.1). Evidently, if $\mathcal{U}(a)$ is singleton for each $a \in A$, then the supplied arguments remain valid for any S and ℓ.

Finally, we proceed to prove (3.17) and (3.18). Note that a necessary optimality condition for a_{ς_n} to be a solution of (3.16) is the following variational inequality

$$D\widehat{J}_{\kappa_n}(a_{\varsigma_n})(a - a_{\varsigma_n}) \geq \kappa_n(R(a_{\varsigma_n}) - R(a)), \quad \text{for every } a \in A, \tag{3.25}$$

16

where $\widehat{J}_{\kappa_n}(a_{\varsigma_n}) := \frac{1}{2}\|u_{\varsigma_n}(a_{\varsigma_n}) - z_{\delta_n}\|_Z^2$ and $D\widehat{J}_{\kappa_n}(a)(b) = \langle Du_{\varsigma_n}(a)(b), u_{\varsigma_n}(a) - z_{\delta_n}\rangle_Z$.

For $n \in \mathbb{N}$, we define the adjoint equation: Find $p_{\varsigma_n} \in V$, such that

$$T_{\tau_n}(a_{\varsigma_n}, p_{\varsigma_n}, v) + \epsilon_n S(p_{\varsigma_n}, v) = \langle z_{\delta_n} - u_{\varsigma_n}(a_{\varsigma_n}), v\rangle_Z, \quad \text{for every } v \in V. \tag{3.26}$$

Evidently, (3.26) has a unique solution p_{ς_n}. Taking $v = Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n})$, we get

$$
\begin{aligned}
\langle Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n}), u_{\varsigma_n}(a_{\varsigma_n}) - z_{\delta_n}\rangle_Z &= -T_{\tau_n}(a_{\varsigma_n}, p_{\varsigma_n}, Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n})) \\
&\quad - \epsilon_n S(p_{\varsigma_n}, Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n})) \\
&= -T_{\tau_n}(a_{\varsigma_n}, Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n}), p_{\varsigma_n}) \\
&\quad - \epsilon_n S(Du_{\varsigma_n}(a_{\varsigma_n})(a - a_{\varsigma_n}), p_{\varsigma_n}) \\
&= T_{\tau_n}(a - a_{\varsigma_n}, u_{\varsigma_n}(a_{\varsigma_n}), p_{\varsigma_n}),
\end{aligned}
$$

by (3.8) and (3.18) follows by using the above expression in (3.25). The proof is complete. $\qquad\square$

Remark 3.6. Since $(\hat{a}, \hat{u})$ is a minimizer of (3.1), we have $\|\hat{u}(\hat{a}) - z\|_Z \leq \|u - z\|_Z$, for each $(a, u) \in \text{graph}(\mathcal{U})$. Since $\|\bar{u}(\hat{a}) - z\|_Z \geq \|\hat{u}(\hat{a}) - z\|_Z$ is possible, we can't use $(\hat{a}, \hat{u})$ to show that $(\bar{a}, \bar{u})$ is optimal. We circumvented this difficulty by showing $\|\bar{u}(\hat{a}) - z\|_Z \leq \|\hat{u}(\hat{a}) - z\|_Z$. A practical implication of the condition $Z = V$ is that typically more regular data is required.

Remark 3.7. For $Z = V$, $\ell_{\delta_n}(\cdot) = \langle z_{\delta_n}, \cdot\rangle_V$ and $S(\cdot, \cdot) = \langle \cdot, \cdot\rangle_V$, (3.7) reduces to

$$T_{\tau_n}(a, u_{\varsigma_n}(a), v) + \epsilon_n \langle u_{\varsigma_n}(a) - z_{\delta_n}, v\rangle_V = m_{\nu_n}(v), \quad \text{for every } v \in V,$$

which steers the regularized solutions towards the solution of (2.2) that is closest to z. If $\ell_{\delta_n}(\cdot) = 0$, then the regularized solutions converge to a minimum norm solution of (2.2). We also note that if the sequence of adjoint solutions $\{p_{\varsigma_n}\}$ is bounded, then by passing (3.17) and (3.18) to limit, we shall derive optimality conditions for (3.1). Because an optimality condition for (3.1) would involve the derivative of the set-valued parameter-to-solution map, such convergence result could shed some light on its contingent differentiability.

3.2 THE MODIFIED OLS APPROACH

We shall now focus on the following MOLS-based constrained optimization problem

$$\min_{a \in A} J(a) := \frac{1}{2} T(a, u(a) - z, u(a) - z), \tag{3.27}$$

which aims to minimize the energy associated to underlying noncoercive variational problem (2.2). Here $u(a) \in \mathcal{U}(a)$ and $z \in V$ is the measured data. Studies related to the MOLS functional and its extensions can be found in [15, 18, 20, 26].

We continue to assume that $\{\epsilon_n\}, \{\tau_n\}, \{\kappa_n\}, \{\delta_n\}$, and $\{\nu_n\}$ are sequence of positive reals, $\ell \in V^*$, and for each $n \in \mathbb{N}$, $m_{\nu_n} \in V^*$, $\ell_{\delta_n} \in V^*$, and $z_{\delta_n} \in V$ satisfying (3.3). Furthermore, the trilinear form $T_{\tau_n} : B \times V \times V \to \mathbb{R}$ satisfies (3.4) and the bilinear and symmetric form $S : V \times V \to \mathbb{R}$ satisfies (3.6).

We again consider the regularized problem: Given $a \in A$, find $u_{\varsigma_n}(a) \in V$ such that

$$T_{\tau_n}(a, u_{\varsigma_n}(a), v) + \epsilon_n S(u_{\varsigma_n}(a), v) = m_{v_n}(v) + \epsilon_n \ell_{\delta_n}(v), \quad \text{for every } v \in V, \tag{3.28}$$

where $\epsilon_n > 0$ is a regularization parameter and $\varsigma_n := (\epsilon_n, \tau_n, v_n, \delta_n)$. For a fixed $n \in \mathbb{N}$, let $u_{\varsigma_n}(a)$ be the unique solution of (3.28).

We first consider the following analogue of the MOLS objective with perturbed data:

$$J_{\epsilon_n}(a) := \frac{1}{2} T_{\tau_n}(a, u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}) + \frac{\epsilon_n}{2} S(u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}). \tag{3.29}$$

We have the following result:

Theorem 3.8. *For each $n \in \mathbb{N}$, the modified output least-squares functional (3.29) is convex in A.*

Proof. For each $n \in \mathbb{N}$, the functional J_{ϵ_n} is evidently infinitely differentiable. The first derivative is derived by the using the chain rule:

$$DJ_{\epsilon_n}(a)\delta a = \frac{1}{2} T_{\tau_n}(\delta a, u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}) + T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a, u_{\varsigma_n}(a) - z_{\delta_n})$$
$$+ \epsilon_n S(Du_{\varsigma_n}(a)\delta a, u_{\varsigma_n} - z_{\delta_n}).$$

By using (3.8), we have

$$T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a, u_{\varsigma_n}(a) - z_{\delta_n}) + \epsilon_n S(Du_{\varsigma_n}(a)\delta a, u_{\varsigma_n}(a) - z_{\delta_n}) = -T_{\tau_n}(\delta a, u_{\varsigma_n}, u_{\varsigma_n}(a) - z_{\delta_n}),$$

and hence

$$DJ_{\epsilon_n}(a)\delta a = \frac{1}{2} T_{\tau_n}(\delta a, u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}) - T_{\tau_n}(\delta a, u_{\varsigma_n}(a), u_{\varsigma_n}(a) - z_{\delta_n})$$
$$= -\frac{1}{2} T_{\tau_n}(\delta a, u_{\varsigma_n}(a) + z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}). \tag{3.30}$$

It now follows that

$$D^2 J_{\epsilon_n}(a)(\delta a, \delta a) = -\frac{1}{2} T_{\tau_n}(\delta a, Du_{\varsigma_n}(a)\delta a, u_{\varsigma_n}(a) - z_{\delta_n}) - \frac{1}{2} T_{\tau_n}(\delta a, u_{\varsigma_n}(a) + z_{\delta_n}, Du_{\varsigma_n}(a)\delta a)$$
$$= -T_{\tau_n}(\delta a, u_{\varsigma_n}(a), Du_{\varsigma_n}(a)\delta a)$$
$$= T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a, Du_{\varsigma_n}(a)\delta a) + \epsilon_n S(Du_{\varsigma_n}(a)\delta a, Du_{\varsigma_n}(a)\delta a),$$

where in the last step, we used

$$T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a, Du_{\varsigma_n}(a)\delta a) + \epsilon_n S(Du_{\varsigma_n}(a)\delta a, Du_{\varsigma_n}(a)\delta a) = -T_{\tau_n}(\delta a, u_{\varsigma_n}, Du_{\varsigma_n}(a)\delta a),$$

which follows from (3.8). We notice, in particular, that the following inequality holds for all a in the interior of A:

$$D^2 J_{\epsilon_n}(a)(\delta a, \delta a) \geq \epsilon_n \beta_0 \|Du_{\varsigma_n}(a)\delta a\|_V^2. \tag{3.31}$$

Thus J_{ϵ_n} is a smooth and convex functional. $\qquad \square$

Our objective is to approximate (3.27) by the following family of regularized MOLS based optimization problems: For $n \in \mathbb{N}$, find $a_{\varsigma_n} \in A$ by solving

$$\min_{a \in A} J_{\kappa_n}(a) := \frac{1}{2} T_{\tau_n}(a, u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}) + \frac{\epsilon_n}{2} S(u_{\varsigma_n}(a) - z_{\delta_n}, u_{\varsigma_n}(a) - z_{\delta_n}) + \kappa_n R(a),$$
(3.32)

where $u_{\varsigma_n}(a)$ is the unique solution of the regularized variational problem (3.28).

We have the following result:

Theorem 3.9. *Assume that the following conditions hold:*

(i) *The set A is bounded in $\widehat{B}$, for each $a \in A$, the solution set $\mathcal{U}(a)$ is nonempty, and the image set $\mathcal{U}(A)$ is a bounded set.*

(ii) *For each $a \in A$, either $\mathcal{U}(a)$ is a singleton, or $\ell_{\delta_n}(v) = \langle z_{\delta_n}, v \rangle_V$ and $S(u, v) = \langle u, v \rangle_V$.*

(iii) *For every $a \in A$, and any $u, w \in V$, $\|u - z\|_V \leq \|w - z\|_V$ implies $T(a, u - z, u - z) \leq T(a, w - z, w - z)$.*

Then, the optimization problem (3.27) has a solution, and for each $n \in \mathbb{N}$, optimization problem (3.32) has a solution. Moreover, there is a subsequence $\{a_{\varsigma_n}\} \subset A$ converging in $\| \cdot \|_L$ to a solution of (3.27). Furthermore, a necessary and sufficient optimality condition for any solution a_{ς_n} of (3.27) is the following variational inequality:

$$-\frac{1}{2} T_{\tau_n}(a - a_{\varsigma_n}, u_{\varsigma_n}(a_{\varsigma_n}) + z_{\delta_n}, u_{\varsigma_n}(a_{\varsigma_n}) - z_{\delta_n}) \geq \kappa_n[R(a_{\varsigma_n}) - R(a)], \quad \text{for every } a \in A. \quad (3.33)$$

Finally, (3.33), when passed to the limit $n \to \infty$, results in the following variational inequality

$$-\frac{1}{2} T(a - \bar{a}, u(\bar{a}) + z, u(\bar{a}) - z) \geq 0, \quad \text{for every } a \in A. \quad (3.34)$$

Proof. It follows by standard arguments that for each $n \in \mathbb{N}$, (3.32) has a solution. By assumption $\{a_n\}$ is a bounded sequence, and due to the compact imbedding of $\widehat{B}$ into L, it possesses a strongly convergent subsequence. Let $\{a_n\}$ be the subsequence which converges strongly to some $\bar{a} \in A$. Let $\{u_n\}$ be the corresponding sequence of the solutions of the regularized variational problems. As in the proof of Theorem 3.5, we can show that $\{u_n\}$ is bounded, and there is a subsequence $\{u_n\}$ converging weakly to some $\bar{u} = u(\bar{a})$.

The only new step is to show that for $a_n \to \bar{a}$ and $u_n := u_{\varsigma_n}(a_n) \rightharpoonup u(\bar{a})$, we have

$$T_{\tau_n}(a_n, u_n - z_{\delta_n}, u_n - z_{\delta_n}) \to T(a, u(\bar{a}) - z, u(\bar{a}) - z). \quad (3.35)$$

Indeed, to prove this convergence, we note that for every $v \in V$, we have

$$T_{\tau_n}(a_n, u_n, v) + \epsilon_n S(u_n, v) = m_{v_n}(v) + \epsilon_n \ell_{\delta_n}(v),$$

which, for the choice $v = u_n - z_{\delta_n}$, can be rearranged as follows

$$T(a_n, u_n - z_{\delta_n}, u_n - z_{\delta_n}) = -T(a_n, z_{\delta_n}, u_n - z_{\delta_n}) + m(u_n - z_{\delta_n}) - T_{\tau_n}(a_n, u_n, u_n - z_{\delta_n})$$
$$+ T(a_n, u_n, u_n - z_{\delta_n}) - \epsilon_n S(u_n, u_n - z_{\delta_n}) + \epsilon_n \ell_{\delta_n}(u_n - z_{\delta_n})$$
$$+ m_{v_n}(u_n - z_{\delta_n}) - m(u_n - z_{\delta_n}),$$

19

and since the right-hand side of the above equation converges to $T(\bar{a}, -z, \bar{u} - z) + m(\bar{u} - z)$, which, due to the fact $\bar{u} \in \mathcal{U}(\bar{a})$, equals to $T(\bar{a}, \bar{u} - z, \bar{u} - z)$, the desired convergence follows.

Let $(\hat{a}, \hat{u})$ be a solution of (3.27). Then, by using (3.35), we have

$$
\begin{aligned}
T(\bar{a}, u(\bar{a}) - z, u(\bar{a}) - z) &= \lim_{n \to \infty} T_{\tau_n}(a_n, u_n(a_n) - z_{\delta_n}, u_n(a_n) - z_{\delta_n}) \\
&\leq \liminf_{n \to \infty} \left\{ T_{\tau_n}(a_n, u_n(a_n) - z_{\delta_n}, u_n(a_n) - z_{\delta_n}) + \kappa_n R(a_n) \right\} \\
&\leq \liminf_{n \to \infty} \left\{ T_{\tau_n}(\hat{a}, u_n(\hat{a}) - z_{\delta_n}, u_n(\hat{a}) - z_{\delta_n}) + \kappa_n R(\hat{a}) \right\} \\
&\leq \limsup_{n \to \infty} T(\hat{a}, u_n(\hat{a}) - z, u_n(\hat{a}) - z) \\
&= T(\hat{a}, u(\hat{a}) - z, u(\hat{a}) - z),
\end{aligned}
\tag{3.36}
$$

which, as for the case of the OLS objective ensures that $\bar{a}$ is a solution of (3.36) and the proof is complete. Note that when $\mathcal{U}(a)$, for $a \in A$ is not singleton, we need additionally condition (3) on the trilinear form.

Due to the convexity of the MOLS functional, a necessary and sufficient optimality condition for a_{ς_n} to be a solution of (3.32) is the variational inequality of second-kind

$$
DJ_{\kappa_n}(a_{\varsigma_n})(a - a_{\varsigma_n}) \geq \kappa_n(R(a_{\varsigma_n}) - R(a)), \quad \text{for every } a \in A,
\tag{3.37}
$$

where J_{κ_n} is defined in (3.32). Condition (3.33) is then follows from the derivative characterization (3.30). Variational inequality (3.34) is a consequence of the properties of T and the facts that $a_{\varsigma_n} \to \bar{a}$ in $\| \cdot \|_L$, $u_{\varsigma_n} \to \bar{u} := u(\bar{a})$, and $z_{\delta_n} \to z$. The proof is complete. $\qquad \square$

4 FIRST-ORDER AND SECOND-ORDER ADJOINT APPROACH FOR OLS

We now describe the first-order and the second-order adjoint approaches to compute the first-order and the second-order derivatives of the regularized OLS functional. These formulae can be discretized to derive an efficient scheme for the computation of the gradient and the Hessian of regularized OLS objective. The gradient computation by the adjoint approach avoids a direct computation of the first-order derivative of the regularized parameter-to-solution map whereas the Hessian computation by the second-order adjoint approach avoids a direct computation of the second-order derivative of the regularized parameter-to-solution map. Adjoint methods have been used extensively in the literature and some of the recent developments can be found in [9, 25]) and the cited references therein.

Recall that for a fixed $n \in \mathbb{N}$, the regularized output least-squares functional is given by

$$
\widehat{J}_{\kappa_n}(a) := \frac{1}{2} \| u_{\varsigma_n}(a) - z_{\delta_n} \|_Z^2 + \kappa_n R(u_{\varsigma_n}),
$$

where R is a smooth regularizer and $u_{\varsigma_n}(a)$ is the unique solution of the regularized problem (3.7), that is,

$$
T_{\tau_n}(a, u_{\varsigma_n}(a), v) + \epsilon_n S(u_{\varsigma_n}(a), v) = m_{\nu_n}(v) + \epsilon_n \ell_{\delta_n}(v), \quad \text{for every } v \in V.
\tag{4.1}
$$

Here $\{\epsilon_n\}$, $\{\tau_n\}$, $\{\kappa_n\}$, $\{\delta_n\}$, and $\{\nu_n\}$ are sequence of positive reals, $\ell \in V^*$, and for each $n \in \mathbb{N}$, $m_{\nu_n} \in V^*$, $\ell_{\delta_n} \in V^*$, and $z_{\delta_n} \in Z$ satisfies (3.3). Moreover, $T_{\tau_n} : B \times V \times V \to \mathbb{R}$ satisfies (3.4) and the bilinear and symmetric form $S : V \times V \to \mathbb{R}$ satisfies (3.6).

By using the chain rule, the derivative of $\widehat{J}_{\kappa_n}$ at $a \in A$ in any direction δa is given by

$$D\widehat{J}_{\kappa_n}(a)(\delta a) = \left\langle Du_{\varsigma_n}(a)(\delta a), u_{\varsigma_n}(a) - z_{\delta_n} \right\rangle_Z + \kappa_n DR(a)(\delta a),$$

where $Du_{\varsigma_n}(a)(\delta a)$ is the derivative of the regularized parameter-to-solution map u_{ς_n} and $DR(a)(\delta a)$ is the derivative of the regularizer R, both computed at a in the direction δa.

For an arbitrary $v \in V$, we define the functional $L_{\kappa_n} : B \times V \to \mathbb{R}$ by

$$L_{\kappa_n}(a, v) = \widehat{J}_{\kappa_n}(a) + T_{\tau_n}(a, u_{\varsigma_n}(a), v) + \epsilon_n S(u_{\varsigma_n}(a), v) - m_{\nu_n}(v) - \epsilon_n \ell_{\delta_n}(v).$$

Since $u_{\varsigma_n}(a)$ solves (4.1), for every $v \in V$, we have $L_{\kappa_n}(a, v) = J_{\kappa_n}(a)$, and consequently, for every $v \in V$ and for every direction δa, we have

$$\partial_a L_{\kappa_n}(a, v)(\delta a) = D\widehat{J}_{\kappa_n}(a)(\delta a).$$

The key idea for the first-order adjoint method is to choose v to bypass a direct computation of $Du_{\varsigma_n}(a)(\delta a)$.n To understand a choice of v, we compute

$$\begin{aligned}
\partial_a L_{\kappa_n}(a, v)(\delta a) = {} & \left\langle Du_{\varsigma_n}(a)(\delta a), u_{\varsigma_n} - z_{\delta_n} \right\rangle_Z + \kappa_n DR(a)(\delta a) \\
& + T_{\tau_n}(\delta a, u_{\varsigma_n}, v) + T_{\tau_n}(a, Du_{\varsigma_n}(a)(\delta a), v) + \epsilon_n S(Du_{\varsigma_n}(a)(\delta a), v). \quad (4.2)
\end{aligned}$$

For $a \in A$, and u_{ς_n} satisfying (3.7), let $w_{\varsigma_n}(a)$ be the unique solution of the adjoint problem

$$T_{\tau_n}(a, w_{\varsigma_n}, v) + \epsilon_n S(w_{\varsigma_n}, v) = \left\langle z_{\delta_n} - u_{\varsigma_n}, v \right\rangle_Z, \quad \text{for every } v \in V. \quad (4.3)$$

We set $v = Du_{\varsigma_n}(a)(\delta a)$ in the above equation and use the symmetry of T_{τ_n} and S to obtain

$$T_{\tau_n}(a, Du_{\varsigma_n}(a)(\delta a), w_{\varsigma_n}) + \epsilon_n S(Du_{\varsigma_n}(a)(\delta a), w_{\varsigma_n}) + \left\langle u_{\varsigma_n} - z_{\delta_n}, Du_{\varsigma_n}(a)(\delta a) \right\rangle_Z = 0. \quad (4.4)$$

By plugging $v = w_{\varsigma_n}$ in (4.2), using (4.4), we obtain

$$\begin{aligned}
\partial_a L_{\kappa_n}(a, w_{\varsigma_n})(\delta a) = {} & \left\langle Du_{\varsigma_n}(a)(\delta a), u_{\varsigma_n} - z_{\delta_n} \right\rangle_Z + \kappa_n DR(a)(\delta a) \\
& + T_{\tau_n}(\delta a, u_{\varsigma_n}, w_{\varsigma_n}) + T_{\tau_n}(a, Du_{\varsigma_n}(a)(\delta a), w_{\varsigma_n}) + \epsilon_n S(Du_{\varsigma_n}(a)(\delta a), w_{\varsigma_n}) \\
= {} & \kappa_n DR(a)(\delta a) + T_{\tau_n}(\delta a, u_{\varsigma_n}, w_{\varsigma_n}),
\end{aligned}$$

which gives the following formula for the first-order derivative of J_{κ_n}:

$$D\widehat{J}_{\kappa_n}(a)(\delta a) = \kappa_n DR(a)(\delta a) + T_{\tau_n}(\delta a, u_{\varsigma_n}, w_{\varsigma_n}). \quad (4.5)$$

In summary, the following scheme computes $D\widehat{J}_{\kappa_n}(a)(\delta a)$ for the given direction δa:

1. Compute $u_{\varsigma_n}(a)$ by using (4.1).
2. Compute $w_{\varsigma_n}(a)$ by using (4.3).
3. Compute $D\widehat{J}_{\kappa_n}(a)(\delta a)$ by using (4.5).

We will now derive a second-order adjoint method for the evaluation of the second-order derivative of the regularized OLS functional. The goal is to derive a formula for the second-order derivative that does not require the second-order derivative of the regularized parameter-to-solution map. The central idea is to compute δu_{ς_n} directly by using (3.8) and bypass the computation of $\delta^2 u_{\varsigma_n}$ by an adjoint approach.

Given a fixed direction δa_2 and an arbitrary $v \in V$, we define

$$L_{\kappa_n}(a, v) = D\widehat{J}_{\kappa_n}(a)(\delta a_2) + T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a_2, v) + \epsilon_n S(Du_{\varsigma_n}(a)\delta a_2, v) + T_{\tau_n}(\delta a_2, u_{\varsigma_n}, v)$$
$$= \left\langle Du_{\varsigma_n}(a)(\delta a_2), u_{\varsigma_n} - z_{\delta_n} \right\rangle_Z + \kappa_n DR(a)(\delta a_2) + T_{\tau_n}(a, Du_{\varsigma_n}(a)\delta a_2, v)$$
$$+ \epsilon_n S(Du_{\varsigma_n}(a)\delta a_2, v) + T_{\tau_n}(\delta a_2, u_{\varsigma_n}, v).$$

Since $L_{\kappa_n}(a, v) = D\widehat{J}_{\kappa_n}(a)(\delta a_2)$, for any $v \in V$, and hence for any δa_1, we have

$$\partial_a L_{\kappa_n}(a, v)(\delta a_1) = D^2 \widehat{J}_{\kappa_n}(a)(\delta a_1, \delta a_2).$$

We compute the derivative of L_{κ_n} in the direction δa_1 as follows

$$\partial_a L_{\kappa_n}(a, v)(\delta a_1) = \left\langle D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), u_{\varsigma_n} - z \right\rangle_Z + \left\langle Du_{\varsigma_n}(a)(\delta a_2), Du_{\varsigma_n}(a)(\delta a_1) \right\rangle_Z$$
$$+ \kappa_n D^2 R(a)(\delta a_1, \delta a_2) + T_{\tau_n}(\delta a_1, Du_{\varsigma_n}(a)\delta a_2, v)$$
$$+ T_{\tau_n}(a, D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), v)$$
$$+ \epsilon_n \langle D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), v \rangle_Z + T_{\tau_n}(\delta a_2, Du_{\varsigma_n}(a)\delta a_1, v).$$

Let $w_{\varsigma_n}(a)$ be the solution of the adjoint problem (4.3). We set $v = D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2)$ in (4.3) and use the symmetry of T_{τ_n} and S to obtain

$$T_{\tau_n}(a, D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), w_{\varsigma_n}) + \epsilon_n S(D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), w_{\varsigma_n})$$
$$+ \left\langle u_\varsigma - z_{\delta_n}, D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2) \right\rangle_Z = 0. \quad (4.6)$$

Using (4.6), we have

$$\partial_a L_{\kappa_n}(a, w_{\varsigma_n})(\delta a_1) = \left\langle D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), u_{\varsigma_n} - z \right\rangle_Z + \left\langle Du_{\varsigma_n}(a)(\delta a_2), Du_{\varsigma_n}(a)(\delta a_1) \right\rangle_Z$$
$$+ \kappa_n D^2 R(a)(\delta a_1, \delta a_2) + T_{\tau_n}(\delta a_1, Du_{\varsigma_n}(a)\delta a_2, w_{\varsigma_n})$$
$$+ T_{\tau_n}(a, D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), w_{\varsigma_n})$$
$$+ \epsilon_n S(D^2 u_{\varsigma_n}(a)(\delta a_1, \delta a_2), w_{\varsigma_n}) + T_{\tau_n}(\delta a_2, Du_{\varsigma_n}(a)\delta a_1, w_{\varsigma_n})$$
$$= \kappa_n D^2 R(a)(\delta a_1, \delta a_2) + \left\langle Du_{\varsigma_n}(a)(\delta a_2), Du_{\varsigma_n}(a)(\delta a_1) \right\rangle_Z$$
$$+ T_{\tau_n}(\delta a_1, Du_{\varsigma_n}(a)\delta a_2, w_{\varsigma_n}) + T_{\tau_n}(\delta a_2, Du_{\varsigma_n}(a)\delta a_1, w_{\varsigma_n}),$$

and consequently, we have

$$D^2 \widehat{J}_{\kappa_n}(a)(\delta a_1, \delta a_2) = \kappa_n D^2 R(a)(\delta a_1, \delta a_2) + \left\langle Du_{\varsigma_n}(a)(\delta a_2), Du_{\varsigma_n}(a)(\delta a_1) \right\rangle_Z$$
$$+ T_{\tau_n}(\delta a_1, Du_{\varsigma_n}(a)\delta a_2, w_{\varsigma_n}) + T_{\tau_n}(\delta a_2, Du_{\varsigma_n}(a)\delta a_1, w_{\varsigma_n}).$$

In particular,

$$D^2 \widehat{J}_{\kappa_n}(a)(\delta a, \delta a) = \kappa_n D^2 R(a)(\delta a, \delta a) + \left\langle Du_{\varsigma_n}(a)(\delta a), Du_{\varsigma_n}(a)\delta a \right\rangle_Z$$
$$+ 2T_{\tau_n}(\delta a, Du_{\varsigma_n}(a)\delta a, w_{\varsigma_n}). \quad (4.7)$$

In summary, the following scheme computes $D^2 \widehat{J}_{\kappa_n}(a)(\delta a, \delta a)$ for any direction δa:

1. Compute $u_{\varsigma_n}(a)$ by (4.1).
2. Compute $Du_{\varsigma_n}(a)(\delta a)$ by (3.8).
3. Compute $w_{\varsigma_n}(a)$ by (4.3).
4. Compute $D^2 \widehat{J}_{\kappa_n}(a)(\delta a, \delta a)$ by (4.7).

The above schemes yield efficient formulas for gradient and Hessian computation.

5 COMPUTATIONAL FRAMEWORK

In this section, we develop a finite element method based discretization framework for the direct and the inverse problems. Let $\mathcal{T}_k$ be a triangulation of the domain Ω. We define $\mathcal{A}_k$ to be the space of all continuous piecewise polynomials of degree d_a relative to $\mathcal{T}_k$. Similarly, $\mathcal{U}_k$ will be the space of all continuous piecewise polynomials of degree d_u relative to $\mathcal{T}_k$. Bases for $\mathcal{A}_k$ and $\mathcal{U}_k$ will be represented by $\{\psi_1, \psi_2, \ldots, \psi_m\}$ and $\{\psi_1, \varphi_2, \ldots, \varphi_n\}$, respectively. The space $\mathcal{A}_k$ is then isomorphic to $\mathbb{R}^m$, and for any $a \in \mathcal{A}_k$, we define $A \in \mathbb{R}^m$ by $A_i := a(x_i)$, $i = 1, 2, \ldots, m$, where $\{\psi_1, \psi_2, \ldots, \psi_m\}$ is a nodal basis corresponding to the nodes $\{x_1, x_2, \ldots, x_m\}$. Conversely, each $A \in \mathbb{R}^m$ corresponds to $a \in \mathcal{A}_k$ defined by $a := \sum_{i=1}^{m} A_i \psi_i$. Similarly, $u \in \mathcal{U}_k$ will correspond to $U \in \mathbb{R}^n$, where $U_i := u(y_i)$, $i = 1, 2, \ldots, n$ and $u = \sum_{i=1}^{n} U_i \varphi_i$. Here $y_1, y_2, \ldots, y_n$ are the nodes of the mesh defining $\mathcal{U}_k$. Note that although both $\mathcal{A}_k$ and $\mathcal{U}_k$ are defined relative to the same triangles, the nodes are different.

For a fixed ς_n, we define $\mathbb{F}_{\varsigma_n} : \mathbb{R}^m \to \mathbb{R}^n$ to be the finite element solution operator assigning a coefficient $a \in \mathcal{A}_k$ to the approximate solution $v \in \mathcal{U}_k$. Then $\mathbb{F}_{\varsigma_n}(A) = V_{\varsigma_n}$, where V_{ς_n} is defined by

$$\left[K_{\tau_n}(A) + \epsilon_n W \right] V_{\varsigma_n} = P_{\delta_n} \quad (5.1)$$

and $K_{\tau_n}(A) \in \mathbb{R}^{n \times n}$ is the stiffness matrix, W is the matrix generated by S, and $P_{\delta_n} \in \mathbb{R}^n$ is the load vector:

$$\begin{aligned} K_{\tau_n}(A)_{ij} &= T_{\tau_n}(a, \varphi_i, \varphi_j), & i, j = 1, 2, \ldots, n, \\ W_{ij} &= S(\varphi_i, \varphi_j), & i, j = 1, 2, \ldots, n, \\ P_{\delta_n}^i &= m_{\nu_n}(\varphi_i) + \epsilon_n \ell_{\delta_n}(\varphi_i), & i = 1, 2, \ldots, n. \end{aligned}$$

For future reference, it will be useful to notice that $K_{\tau_n}(A)_{ij} = T_{ijk} A_k$, where the summation convention is used and T is the tensor defined by

$$T_{ijk} = T_{\tau_n}(\psi_k, \varphi_i, \varphi_j), \quad i, j = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, m.$$

The derivative of the regularized parameter-to-solution map is easily computed as

$$D\mathbb{F}_{\varsigma_n}(A)(\delta A) := \delta V_{\varsigma_n} = - \left[K_{\tau_n}(A) + \epsilon_n W \right]^{-1} K_{\tau_n}(\delta A) V_{\varsigma_n}.$$

To write the formula for δV_{ς_n} in a tractable form, we define the matrix $L_{\tau_n}(\mathcal{V})$ by the condition

$$L_{\tau_n}(\mathcal{V})A = K_{\tau_n}(A)\mathcal{V} \text{ for all } A \in \mathbb{R}^m, \mathcal{V} \in \mathbb{R}^n.$$

Using this notation,

$$\delta V_{\varsigma_n} = -\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A.$$

5.1 DISCRETE OLS

In the following, for simplicity, we ignore a discretization of the the regularization term. Using the same notation as above, we can define the L^2-output least-squares objective function by

$$\widehat{J_n}(A) = \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T M(V_{\varsigma_n} - Z_{\delta_n}), \tag{5.2}$$

where V_{ς_n} solves (5.1) and M is the mass matrix.

The gradient of $\widehat{J_{\kappa_n}}$ can be computed as follows

$$\begin{aligned}
D\widehat{J_n}(A)\delta A &= \delta V_{\varsigma_n}^T M(V_{\varsigma_n} - Z_{\delta_n}) \\
&= -\left(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A\right)^T M(V_{\varsigma_n} - Z_{\delta_n}) \\
&= -\delta A^T L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}),
\end{aligned}$$

and hence

$$\nabla\widehat{J_n}(A) = -L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}).$$

We shall now proceed to compute the Hessian. First, we observe that

$$\begin{aligned}
D^2\widehat{J}(A)(\delta A, \delta A) &= \delta A^T(-L_{\tau_n}(\delta V_{\varsigma_n})^T)\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}) \\
&+ \delta A^T L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} K_{\tau_n}(\delta A)\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}) \\
&- \delta A^T L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M\delta V_{\varsigma_n}),
\end{aligned}$$

where we use the formula

$$\begin{aligned}
D\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1}\delta A &= -\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} DK_{\tau_n}(A)\delta A\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} \\
&= -\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} K_{\tau_n}(\delta A)\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1}.
\end{aligned}$$

We will simplify all the the terms involved in the formula for the second-order derivative. First of all, we have

$$\begin{aligned}
&L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} DK_{\tau_n}(A)(\delta A)\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}) \\
&= L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} K_{\tau_n}(\delta A)\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}) \\
&= L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}))\delta A,
\end{aligned}$$

which, by using the fact that

$$L_{\tau_n}(\mathcal{V})^T U = L_{\tau_n}(U)^T \mathcal{V}, \quad \text{for all } U, \mathcal{V} \in \mathbb{R}^n$$

gives

$$-L_{\tau_n}(\delta V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n})$$

$$= L_{\tau_n}(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A)^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n})$$

$$= L_{\tau_n}(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}))^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A.$$

Finally,

$$-L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M\delta V_{\varsigma_n}$$

$$= L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A.$$

Combining the above results, we obtain that

$$\nabla^2 \widehat{J}_{\kappa_n}(A) = L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}))$$

$$+ L_{\tau_n}(\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M(V_{\varsigma_n} - Z_{\delta_n}))^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})$$

$$+ L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} M \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n}).$$

We emphasize that, after a discretization, the first-order and second-order adjoint formulae discussed in the previous section would lead to an alternative scheme for computing the gradient and the Hessian of the OLS objective.

5.2 DISCRETE MOLS

The discrete MOLS function $J_n : \mathbb{R}^m \to \mathbb{R}$ is given by

$$J_n(A) = \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T K_{\tau_n}(A)(V_{\varsigma_n} - Z_{\delta_n}) + \frac{\epsilon_n}{2} \left(V_{\varsigma_n} - Z_{\delta_n}\right)^T W \left(V_{\varsigma_n} - Z_{\delta_n}\right),$$

where V_{ς_n} solves (5.1), W is the symmetric matrix generated by the bilinear form S, and Z_{δ_n} is the discrete data.

We can now compute the gradient as follows

$$DJ_n(A)\delta A := \delta V_{\varsigma_n}^T K_{\tau_n}(A)(V_{\varsigma_n} - Z_{\delta_n}) + \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T DK_{\tau_n}(A)\delta A(V_{\varsigma_n} - Z_{\delta_n})$$

$$+ \epsilon_n \delta V_{\varsigma_n}^T W(V_{\varsigma_n} - Z_{\delta_n})$$

$$= \delta V_{\varsigma_n}^T \left[K_{\tau_n}(A) + \epsilon_n W\right](V_{\varsigma_n} - Z_{\delta_n}) + \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T K_{\tau_n}(\delta A)(V_{\varsigma_n} - Z_{\delta_n})$$

$$= \left[-\left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A\right]^T \left[K_{\tau_n}(A) + \epsilon_n W\right](V_{\varsigma_n} - Z_{\delta_n})$$

$$+ \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T K(\delta A)(V_{\varsigma_n} - Z_{\delta_n})$$

$$= -\delta A^T L_{\tau_n}(V_{\varsigma_n})^T(V_{\varsigma_n} - Z_{\delta_n}) + \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T L_{\tau_n}(V_{\varsigma_n} - Z_{\delta_n})\delta A$$

$$= -\delta A^T L_{\tau_n}(V_{\varsigma_n})^T(V_{\varsigma_n} - Z_{\delta_n}) + \frac{1}{2}\delta A^T L_{\tau_n}(V_{\varsigma_n} - Z_{\delta_n})^T(V_{\varsigma_n} - Z_{\delta_n})$$

$$= -\frac{1}{2}\delta A^T L_{\tau_n}(V_{\varsigma_n} + Z_{\delta_n})^T(V_{\varsigma_n} - Z_{\delta_n}),$$

which yields

$$\nabla J_n(A) = -\frac{1}{2}L_{\tau_n}(V_{\varsigma_n} + Z_{\delta_n})^T(V_{\varsigma_n} - Z_{\delta_n}) = -\frac{1}{2}L_{\tau_n}(V_{\varsigma_n})^T V_{\varsigma_n} + \frac{1}{2}L_{\tau_n}(Z_{\delta_n})^T Z_{\delta_n}.$$

For the second-order derivative, from

$$DJ_n(A)\delta A = -\frac{1}{2}\delta A^T L_{\tau_n}(V_{\varsigma_n})^T V_{\varsigma_n} + \frac{1}{2}\delta A^T L_{\tau_n}(Z_{\delta_n})^T Z_{\delta_n}$$

we evaluate

$$D^2 J_n(A)(\delta A, \delta A) = -\frac{1}{2}\delta A^T L_{\tau_n}(\delta V_{\varsigma_n})^T V_{\varsigma_n} - \frac{1}{2}\delta A^T L_{\tau_n}(V_{\varsigma_n})^T \delta V_{\varsigma_n}.$$

Since $L_{\tau_n}(U)^T \mathcal{V} = L_{\tau_n}(\mathcal{V})^T U$ for all $U, \mathcal{V} \in \mathbb{R}^n$, and hence,

$$L_{\tau_n}(\delta V_{\varsigma_n})^T V_{\varsigma_n} = L_{\tau_n}(V_{\varsigma_n})^T \delta V_{\varsigma_n} = -L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V)\delta A.$$

Consequently,

$$D^2 J_n(A)(\delta A, \delta A) = -\delta A^T L_{\tau_n}(V_{\varsigma_n})^T \delta V_{\varsigma_n}$$
$$= \delta A^T L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n})\delta A,$$

which shows that

$$\nabla^2 J_n(A) = L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n}).$$

Summarizing, the necessary formulas are

$$J_n(A) = \frac{1}{2}(V_{\varsigma_n} - Z_{\delta_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right](V_{\varsigma_n} - Z_{\delta_n}),$$

$$\nabla J_n(A) = -\frac{1}{2}L_{\tau_n}(V_{\varsigma_n})^T V_{\varsigma_n} + \frac{1}{2}L_{\tau_n}(Z_{\delta_n})^T Z_{\delta_n},$$

$$\nabla^2 J_n(A) = L_{\tau_n}(V_{\varsigma_n})^T \left[K_{\tau_n}(A) + \epsilon_n W\right]^{-1} L_{\tau_n}(V_{\varsigma_n}).$$

6 COMPUTATIONAL EXPERIMENTS

We now report preliminary numerical experiments to demonstrate the feasibility of the proposed framework. We will identify the coefficient a in the Neumann boundary value problem (1.1). Our experiments are of synthetic nature, and hence the data vectors are computed, not measured. We solve numerically the regularized problems (3.16) and (3.29) by using the piecewise linear finite elements. We used the finite element library FreeFem++ [24]. For simplicity, we used the $H^1(\Omega)$ norm as the regularizer. The choose the regularization parameters κ_n and ε_n by trial and error. The elliptic regularization of the variational problem was crucial in the identification process. As expected, the reconstruction process failed for $\varepsilon_n = 0$ (see Figure 3). Finding a stable numerical solution of a pure Neumann problem is quite challenging (see [4, 12]) and our computations show that the elliptic regularization does a remarkable job in giving a stable solution.

The (normalized) unique solution is $\bar{u}(x_1, x_2) = \cos(\pi x_1^2)\cos(2\pi x_2)$ and $\bar{a}(x_1, x_2) = 1$. In our experiment, both the OLS objective and the MOLS objective gave quite a satisfactory reconstruction, see Tables 1 and 2 and Figures 1 and 2. To study the influence of noise, we considered contaminated data $z_{\delta_n} = z + \delta_n \eta(t)$, with $\eta(t)$ uniformly distributed in $[0, 1]$. The reconstruction is again quite stable, as seen in Table 3 and Figure 4.

Table 1: Reconstruction error for the OLS for different discretization level and $\kappa = \epsilon = 0.0001$.

h	$\dfrac{\|a^h - I_h \bar{a}_0\|_{L^2(\Omega)}}{\|I_h \bar{a}_0\|_{L^2(\Omega)}}$	$\dfrac{\|u^h - u_0^h\|_{L^2(\Omega)}}{\|u_0^h\|_{L^2(\Omega)}}$	$\dfrac{\|a^h - I_h \bar{a}_0\|_{L^\infty(\Omega)}}{\|I_h \bar{a}_0\|_{L^\infty(\Omega)}}$	$\dfrac{\|u^h - u_0^h\|_{L^\infty(\Omega)}}{\|u_0^h\|_{L^\infty(\Omega)}}$
0.0471405	$1.13 \cdot 10^{-2}$	$2.13 \cdot 10^{-3}$	$3.34 \cdot 10^{-2}$	$9.61 \cdot 10^{-3}$
0.0353553	$6.96 \cdot 10^{-3}$	$1.27 \cdot 10^{-3}$	$1.91 \cdot 10^{-2}$	$6.66 \cdot 10^{-3}$
0.0282843	$5.05 \cdot 10^{-3}$	$8.90 \cdot 10^{-4}$	$1.38 \cdot 10^{-2}$	$4.87 \cdot 10^{-3}$
0.0235702	$4.03 \cdot 10^{-3}$	$7.19 \cdot 10^{-4}$	$9.76 \cdot 10^{-3}$	$3.83 \cdot 10^{-3}$
0.0202031	$3.34 \cdot 10^{-3}$	$6.11 \cdot 10^{-4}$	$8.24 \cdot 10^{-3}$	$2.98 \cdot 10^{-3}$
0.0176777	$3.23 \cdot 10^{-3}$	$6.07 \cdot 10^{-4}$	$8.65 \cdot 10^{-3}$	$2.43 \cdot 10^{-3}$

Table 2: Reconstruction error for the MOLS for different discretization level and $\kappa = 0.01, \epsilon = 0.0001$.

h	$\dfrac{\|a^h - I_h \bar{a}_0\|_{L^2(\Omega)}}{\|I_h \bar{a}_0\|_{L^2(\Omega)}}$	$\dfrac{\|u^h - u_0^h\|_{L^2(\Omega)}}{\|u_0^h\|_{L^2(\Omega)}}$	$\dfrac{\|a^h - I_h \bar{a}_0\|_{L^\infty(\Omega)}}{\|I_h \bar{a}_0\|_{L^\infty(\Omega)}}$	$\dfrac{\|u^h - u_0^h\|_{L^\infty(\Omega)}}{\|u_0^h\|_{L^\infty(\Omega)}}$
0.0471405	$9.54 \cdot 10^{-3}$	$4.37 \cdot 10^{-3}$	$4.32 \cdot 10^{-2}$	$1.16 \cdot 10^{-2}$
0.0353553	$5.83 \cdot 10^{-3}$	$2.50 \cdot 10^{-3}$	$2.50 \cdot 10^{-2}$	$7.50 \cdot 10^{-3}$
0.0282843	$4.24 \cdot 10^{-3}$	$1.66 \cdot 10^{-3}$	$1.70 \cdot 10^{-2}$	$5.49 \cdot 10^{-3}$
0.0235702	$3.34 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$1.23 \cdot 10^{-2}$	$4.16 \cdot 10^{-3}$
0.0202031	$2.82 \cdot 10^{-3}$	$1.04 \cdot 10^{-3}$	$9.77 \cdot 10^{-3}$	$3.54 \cdot 10^{-3}$
0.0176777	$2.36 \cdot 10^{-3}$	$9.21 \cdot 10^{-4}$	$8.18 \cdot 10^{-3}$	$3.24 \cdot 10^{-3}$

Table 3: Reconstruction error for the OLS for different noise levels δ_n for $h = 0.0176777, \kappa = \epsilon = 0.0001$.

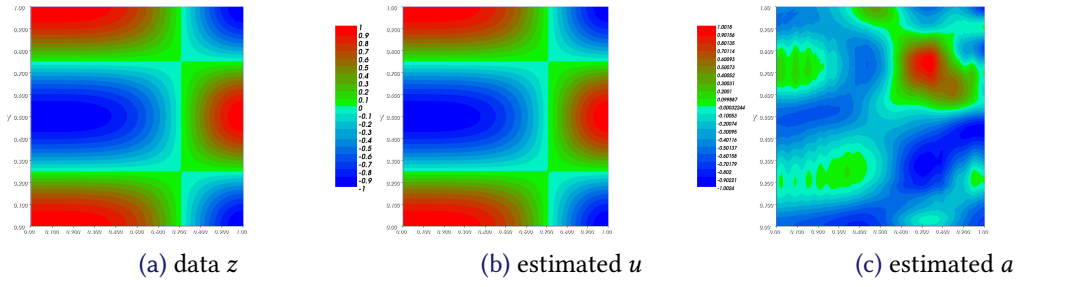| δ_n	$\dfrac{\|a^h - I_h \bar{a}\|_{L^2(\Omega)}}{\|I_h \bar{a}\|_{L^2(\Omega)}}$	$\dfrac{\|u^h - u^h\|_{L^2(\Omega)}}{\|u^h\|_{L^2(\Omega)}}$	$\dfrac{\|a^h - I_h \bar{a}\|_{L^\infty(\Omega)}}{\|I_h \bar{a}\|_{L^\infty(\Omega)}}$	$\dfrac{\|u^h - u^h\|_{L^\infty(\Omega)}}{\|u^h\|_{L^\infty(\Omega)}}$
$1 \cdot 10^{-1}$	$9.22 \cdot 10^{-3}$	$9.01 \cdot 10^{-2}$	$3.53 \cdot 10^{-2}$	$5.69 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	$3.37 \cdot 10^{-3}$	$9.03 \cdot 10^{-3}$	$9.61 \cdot 10^{-3}$	$6.87 \cdot 10^{-3}$
$1 \cdot 10^{-3}$	$3.23 \cdot 10^{-3}$	$1.09 \cdot 10^{-3}$	$8.70 \cdot 10^{-3}$	$2.28 \cdot 10^{-3}$

(a) data z (b) estimated u (c) estimated a

Figure 1: Reconstruction with no noise for $h = 0.0176777$, $\kappa = \varepsilon = 0.0001$ by OLS approach.
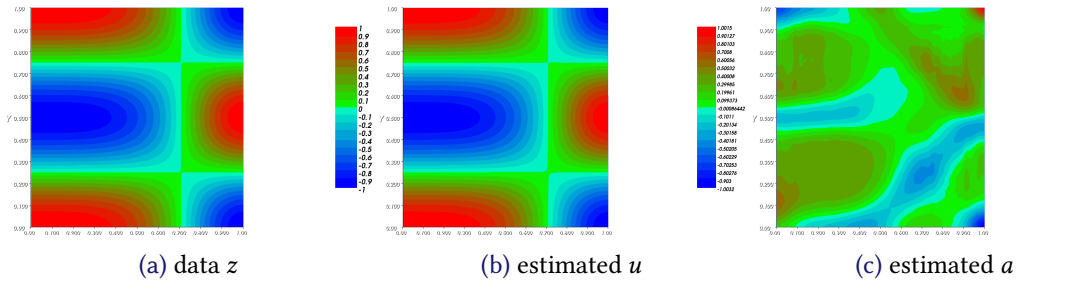

(a) data z (b) estimated u (c) estimated a

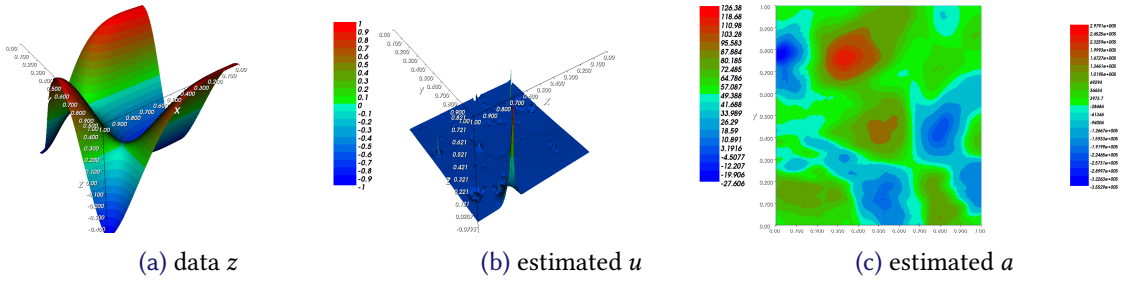Figure 2: Reconstruction with no noise for $h = 0.0176777$, $\kappa = 0.01$, $\varepsilon = 0.0001$ by MOLS approach.

(a) data z (b) estimated u (c) estimated a

Figure 3: Failed reconstruction for $h = 0.0235702$, $\kappa = 0.0001$, $\varepsilon = 0$ by the OLS approach.

(a) data z (b) estimated u (c) estimated a
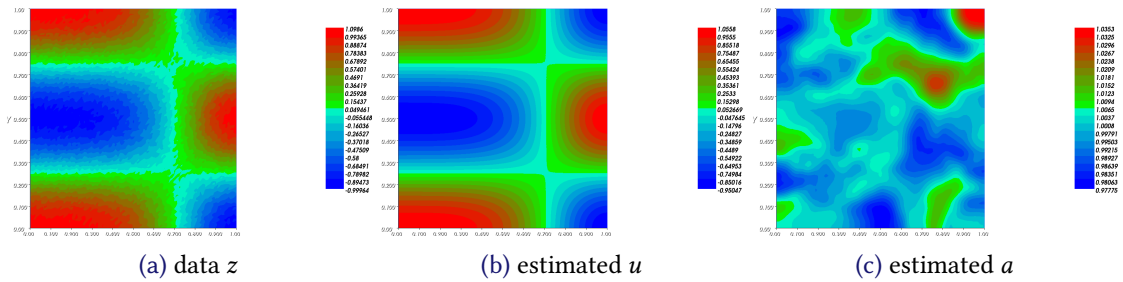
Figure 4: Reconstruction for $h = 0.0176777$, $\kappa = \varepsilon = 0.0001$, and noise level $\delta_n = 0.1$ by the OLS approach.

7 CONCLUDING REMARKS

We explored the inverse problem of parameter identification in non-elliptic variational problems by posing optimization problems using the OLS and the MOLS functionals. We regularized the underlying non-elliptic variational problem and studied the features of the regularized parameter-to-solution map. For the set-valued parameter-to-solution map, we relied on the notion of the first-order and the second-order contingent derivatives. To the best of our knowledge, this is the first work where tools from set-valued optimization have been employed to assist the study of inverse problems of parameter identification. It would be of interest to explore what derivatives of set-valued maps are most convenient for this kind of research. Detailed numerical experimentation, taking into account the data perturbation, is of paramount importance and will be done in future work. An extension of the present approach to inverse problems in noncoercive variational inequalities also seems to be a promising topic to explore.

REFERENCES

[1] ACAR, Identification of the coefficient in elliptic equations, *SIAM J. Control Optim.* 31 (1993), 1221–1244, DOI: 10.1137/0331058.

[2] AMMARI, GARAPON & JOUVE, Separation of scales in elasticity imaging: a numerical study, *J. Comput. Math.* 28 (2010), 354–370, DOI: 10.4208/jcm.2009.12-m1001.

[3] BAIOCCHI, GASTALDI & TOMARELLI, Some existence results on noncoercive variational inequalities, *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* 13 (1986), 617–659, URL: http://www.numdam.org/item?id=ASNSP_1986_4_13_4_617_0.

[4] BOCHEV & LEHOUCQ, On the finite element solution of the pure Neumann problem, *SIAM Rev.* 47 (2005), 50–66, DOI: 10.1137/S0036144503426074.

[5] BOEHM & ULBRICH, A semismooth Newton-CG method for constrained parameter identification in seismic tomography, *SIAM J. Sci. Comput.* 37 (2015), S334–S364, DOI: 10.1137/140968331.

[6] BOIGER & KALTENBACHER, An online parameter identification method for time dependent partial differential equations, *Inverse Problems* 32 (2016), 045006, 28, DOI: 10.1088/0266-5611/32/4/045006.

[7] BORWEIN, Weak tangent cones and optimization in a Banach space, *SIAM J. Control Optimization* 16 (1978), 512–522, DOI: 10.1137/0316034.

[8] Borwein & O'Brien, Tangent cones and convexity, *Canad. Math. Bull.* 19 (1976), 257–261, DOI: 10.4153/CMB-1976-040-x.

[9] Cho, Jadamba, Kahler, Khan & Sama, First-order and second-order adjoint methods for the inverse problem of identifying non-linear parameters in PDEs, in: *Industrial Mathematics and Complex Systems: Emerging Mathematical Models, Methods and Algorithms*, Springer Singapore, 2017, 147–163, DOI: 10.1007/978-981-10-3758-0_9.

[10] Clason, L^∞ fitting for inverse problems with uniform noise, *Inverse Problems* 28 (2012), 104007, DOI: 10.1088/0266-5611/28/10/104007.

[11] Crossen, Gockenbach, Jadamba, Khan & Winkler, An equation error approach for the elasticity imaging inverse problem for predicting tumor location, *Comput. Math. Appl.* 67 (2014), 122–135, DOI: 10.1016/j.camwa.2013.10.006.

[12] Dai, Finite element approximation of the pure Neumann problem using the iterative penalty method, *Appl. Math. Comput.* 186 (2007), 1367–1373, DOI: 10.1016/j.amc.2006.07.148.

[13] Evstigneev, Medvedik & Smirnov, Inverse problem of determining parameters of inhomogeneity of a body from acoustic field measurements, *Comput. Math. Math. Phys.* 56 (2016), 483–490, DOI: 10.1134/S0965542516030040.

[14] Gholami, Mang & Biros, An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas, *J. Math. Biol.* 72 (2016), 409–433, DOI: 10.1007/s00285-015-0888-x.

[15] Gockenbach, Jadamba & Khan, Numerical estimation of discontinuous coefficients by the method of equation error, *Int. J. Math. Comput. Sci.* 1 (2006), 343–359.

[16] Gockenbach, Jadamba & Khan, Equation error approach for elliptic inverse problems with an application to the identification of Lamé parameters, *Inverse Probl. Sci. Eng.* 16 (2008), 349–367, DOI: 10.1080/17415970701602580.

[17] Gockenbach, Jadamba, Khan, Tammer & Winkler, Proximal methods for the elastography inverse problem of tumor identification using an equation error approach, in: *Advances in Variational and Hemivariational Inequalities: Theory, Numerical Analysis, and Applications*, Springer, Cham, 2015, 173–197, DOI: 10.1007/978-3-319-14490-0_7.

[18] Gockenbach & Khan, Identification of Lamé parameters in linear elasticity: a fixed point approach, *J. Ind. Manag. Optim.* 1 (2005), 487–497, DOI: 10.3934/jimo.2005.1.487.

[19] Gockenbach & Khan, An abstract framework for elliptic inverse problems: I. An output least-squares approach, *Math. Mech. Solids* 12 (2007), 259–276, DOI: 10.1177/1081286505055758.

[20] Gockenbach & Khan, An abstract framework for elliptic inverse problems: II. An augmented Lagrangian approach, *Math. Mech. Solids* 14 (2009), 517–539, DOI: 10.1177/1081286507087150.

[21] Goeleven, *Noncoercive Variational Problems and Related Results*, Longman, Harlow, 1996.

[22] Guchhait & Banerjee, Constitutive error based material parameter estimation procedure for hyperelastic material, *Comput. Methods Appl. Mech. Engrg.* 297 (2015), 455–475, DOI: 10.1016/j.cma.2015.09.012.

[23] Hager, Ngo, Yashtini & Zhang, An alternating direction approximate Newton algorithm for ill-conditioned inverse problems with application to parallel MRI, *J. Oper. Res. Soc. China* 3 (2015), 139–162, DOI: 10.1007/s40305-015-0078-y.

[24] Hecht, New development in FreeFem++, *Journal of Numerical Mathematics* 20 (2012), 251–266, DOI: 10.1515/jnum-2012-0013.

[25] Jadamba, Khan, Oberai & Sama, First-order and second-order adjoint methods for parameter identification problems with an application to the elasticity imaging inverse problem, *Inverse Problems in Science and Engineering* (2017), 1–20, DOI: 10.1080/17415977.2017.1289195.

[26] Jadamba, Khan, Rus, Sama & Winkler, A new convex inversion framework for parameter identification in saddle point problems with an application to the elasticity imaging inverse problem of predicting tumor location, *SIAM J. Appl. Math.* 74 (2014), 1486–1510, DOI: 10.1137/130928261.

[27] Jadamba, Khan, Sama & Tammer, On convex modified output least-squares for elliptic inverse problems: stability, regularization, applications, and numerics, *Optimization* 66 (2017), 983–1012, DOI: 10.1080/02331934.2017.1316270.

[28] Al-Jamal & Gockenbach, Stability and error estimates for an equation error method for elliptic equations, *Inverse Problems* 28 (2012), 095006, 15, DOI: 10.1088/0266-5611/28/9/095006.

[29] Kärkkäinen, An equation error method to recover diffusion from the distributed observation, *Inverse Problems* 13 (1997), 1033–1051, DOI: 10.1088/0266-5611/13/4/009.

[30] Khan & Motreanu, Inverse problems for quasi-variational inequalities, *J. Global Optim.* 70 (2018), 401–411, DOI: 10.1007/s10898-017-0597-7.

[31] Khan, Tammer & Zalinescu, Regularization of quasi-variational inequalities, *Optimization* 64 (2015), 1703–1724, DOI: 10.1080/02331934.2015.1028935.

[32] Khan, Tammer & Zalinescu, *Set-Valued Optimization*, Springer, Heidelberg, 2015, DOI: 10.1007/978-3-642-54265-7.

[33] Kindermann, Mutimbu & Resmerita, A numerical study of heuristic parameter choice rules for total variation regularization, *J. Inverse Ill-Posed Probl.* 22 (2014), 63–94, DOI: 10.1515/jip-2012-0074.

[34] Kirisits, Pöschl, Resmerita & Scherzer, Finite-dimensional approximation of convex regularization via hexagonal pixel grids, *Appl. Anal.* 94 (2015), 612–636, DOI: 10.1080/00036811.2014.958998.

[35] Knowles, Parameter identification for elliptic problems, *J. Comput. Appl. Math.* 131 (2001), 175–194, DOI: 10.1016/S0377-0427(00)00275-2.

[36] Kuchment & Steinhauer, Stabilizing inverse problems by internal data. II: Non-local internal data and generic linearized uniqueness, *Anal. Math. Phys.* 5 (2015), 391–425, DOI: 10.1007/s13324-015-0104-6.

[37] Liu, A wavelet multiscale-homotopy method for the parameter identification problem of partial differential equations, *Comput. Math. Appl.* 71 (2016), 1519–1523, DOI: 10.1016/j.camwa.2016.02.036.

[38] Manservisi & Gunzburger, A variational inequality formulation of an inverse elasticity problem, *Appl. Numer. Math.* 34 (2000), 99–126, DOI: 10.1016/S0168-9274(99)00042-2.

[39] Nashed & Scherzer, Least squares and bounded variation regularization with nondifferentiable functionals, *Numer. Funct. Anal. Optim.* 19 (1998), 873–901, DOI: 10.1080/01630569808816863.

[40] Neubauer, Hein, Hofmann, Kindermann & Tautenhahn, Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces, *Appl. Anal.* 89 (2010), 1729–1743, DOI: 10.1080/00036810903517597.

[41] Ward, Calculus for parabolic second-order derivatives, *Set-Valued Anal.* 1 (1993), 213–246, DOI: 10.1007/BF01027635.