ADAPTIVITY WITH RELAXATION FOR ILL-POSED PROBLEMS AND GLOBAL CONVERGENCE FOR A COEFFICIENT INVERSE PROBLEM *

LARISA BEILINA†, MICHAEL V. KLIBANOV ‡, AND MIKHAIL YU. KOKURIN §

Abstract. A new framework of the Functional Analysis is developed for the adaptive FEM (adaptivity) for the Tikhonov regularization functional for ill-posed problems. As a result, the relaxation property for adaptive mesh refinements is established. An application to a multidimensional Coefficient Inverse Problem for a hyperbolic equation is discussed. This problem arises in the inverse scattering of acoustic and electromagnetic waves. First, a globally convergent numerical method provides a good approximation for the correct solution of this problem. Next, this approximation is enhanced via the subsequent application of the adaptivity. Analytical results are computationally verified

Key words. ill-posed problems, globally convergent numerical method for a coefficient inverse problem, two-stage numerical procedure, adaptivity for the Tikhonov functional, relaxation property, orthogonal projection operators

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction.

We develop a new framework of the Functional Analysis for the Finite Element Adaptive technique (adaptivity for brevity) for the Tikhonov functional for ill-posed problems. For the *first time* the so-called relaxation property for the adaptive mesh refinements is proved (see below in this section). We use the adaptivity as a complementary tool to a globally convergent numerical method, which was recently developed in [8] for a Coefficient Inverse Problem (CIP) for a hyperbolic PDE (section 5). This CIP has applications in acoustics and electromagnetics. CIPs for PDEs are both ill-posed and nonlinear, which causes serious difficulties for their numerical solutions. In particular, least squares residual functionals for CIPs suffer from the problem of multiple local minima and ravines, see, e.g. [19] for some examples. Because of the phenomenon of local minima, conventional numerical methods for CIPs are locally convergent ones. The numerical method of [8] relies on the structure of the PDE operator and thus, is not using least squares. The convergence estimate in the global convergence theorem of [8] depends on a small parameter $\eta > 0$. This parameter incorporates the level of the error in the boundary data as well as some approximation errors of the technique of [8].

This paper is motivated by our recent numerical experience. Namely, although η is small, we have observed that it cannot be made infinitely small in practical computations, because of above approximation errors of the method of [8]. On the other hand, locally convergent numerical methods for CIPs are independent on these approximation errors. This led us to the idea of enhancing images resulting from the globally convergent method via a subsequent application of a locally convergent

* This preprint is available on-line at http://www.ma.utexas.edu/mp_arc (posting date November 20, 2009) and at http://www.math.chalmers.se/Math/Research/Preprints, Chalmers Preprint Series, ISSN 1652-9715, 2009. This work was supported by US Army Research Laboratory and US Army Research Office by the grant number W911NF-08-1-0470 as well as by the grant RFBR N09-01-00273a from Russian Foundation for Basic Research. Computations were performed in the Center for Scientific and Technical Computing (C3SE) at Chalmers University of Technology, Gothenburg, Sweden.

† Department of Mathematical Sciences, Chalmers University of Technology and Gothenburg University, SE-42196 Gothenburg, Sweden, (larisa@chalmers.se)

‡ Department of Mathematics and Statistics University of North Carolina at Charlotte, Charlotte, NC 28223, USA, (mklibanv@uncc.edu)

§ Department of Mathematics, Mari State University, 424000, Yoshkar Ola, Russia, (kokurin@marsu.ru).

one. On the other hand, it is well known that a good first approximation for the correct solution is one of the key inputs for any locally convergent method. Therefore, our natural conclusion was that one should have a two stage numerical procedure. On the first stage one should get a good first approximation for the solution by the globally convergent method of [8]. And on the second stage one should use this approximation as a first guess for a further enhancement via an appropriate locally convergent numerical method. An *important* point here is that since η is small, then there is a *rigorous guarantee* that the globally convergent part indeed provides the above input. This idea is carried out in numerical experiments of section 8.

The next question to ask was about the choice of a proper locally convergent numerical method. We have observed numerically (section 8) that a straightforward application of the quasi-Newton method on the same mesh where the globally convergent part worked does not improve the solution provided by the first stage. Thus, based on the previous numerical experience of the first author for the same CIP [5-7], we have concluded that a sequence of adaptive local mesh refinements should be used. It is shown numerically here that the adaptivity indeed refines images obtained on the globally convergent stage. Therefore, we study here the problem of successive approximations of the regularized solution via a sequence of adaptive mesh refinements for a given value of the regularization parameter α. In our computations α is chosen experimentally. The question of an optimal choice of α is outside of the scope of this publication. We refer to [17] for a detailed study of this question for the adaptivity technique.

In this paper the Tikhonov functional J_α is constructed for a general nonlinear operator F, and J_α is linked with the FEM. Our functional analytical framework for the adaptivity is independent on a specific procedure of the minimization of J_α. One of the *key assumptions* below is that a first good approximation for the exact solution is available, which is in conjunction with the above idea about the two stage procedure. Since the adaptivity is a locally convergent numerical method, then our analysis is inevitably an "asymptotic" one, as it is always the case in such scenarios. In other words, we assume that the error in the data is sufficiently small.

In addition to the above framework, the following five (5) new results are presented in this paper: **(1)** We prove the strict convexity of J_α in a small neighborhood of the regularized solution, provided that the originating nonlinear operator F has the first Lipschitz continuous Frechet derivative. A similar result was proven earlier in [24,25] under the condition that the nonlinear operator F has the second continuous Frechet derivative. Note that such a result for the case of a bounded linear operator is trivial. **(2)** We prove the relaxation property of the Tikhonov functional with respect to adaptive mesh refinements, see (1.1) below, which is our *main* result. **(3)** We derive the Frechet derivative of the Tikhonov functional for our CIP and prove that it equals to the so-called "all-at-once" Frechet derivative of the Lagrangian used in [5-7]. The connection between these two derivatives was not clarified in [5-7]. **(4)** Results of items 1, 2 are specified for our CIP. We prove a posteriori error estimate for the computed regularized unknown coefficient of our CIP, which, in particular, also approximately estimates the accuracy of the exact coefficient (Lemma 2.1). In previous publications on the adaptivity for CIPs only the accuracy of Lagrangians was estimated, see, e.g. [5-7]. Our estimate uses the local strict convexity of the Tikhonov functional instead of the traditional apparatus of the Galerkin orthogonality. **(5)** In our numerical tests for the above two stage procedure the medium consists of small inclusions embedded in a slowly varying background, whereas the background function was constant in [8]. The relaxation property (1.1) is numerically verified.

The adaptivity is about adaptive mesh refinements in the FEM to improve the accuracy of the solution. This is a classic tool for forward problems [1], and it is also applied both to CIPs and parameter identification problems, see, e.g. [5-7,17]. Mesh refinements can be either local, i.e., in some subdomains of the original domain, or global, i.e. in the whole domain. Local refinements are preferable, because a globally fine mesh imposes extra demands on the computer's capacity. The following two questions are of an interest in the adaptivity technique: **(A)** Where to refine the mesh? **(B)** Is it possible to estimate the distance between the solution obtained on the refined mesh and the regularized one **via** that distance obtained on the previous coarser mesh? Let x_α be the regularized solution and V_ρ be a certain neighborhood of x_α of the radius $\rho \in (0,\alpha)$. Let $x_n \in V_\rho$ and $x_{n+1} \in V_\rho$ be minimizers of the above Tikhonov functional after n and $n+1$ adaptive mesh refinements respectively. So, x_{n+1} is obtained on a finer mesh than x_n. Although the intuition seems to be saying that x_{n+1} should be closer to x_α than x_n, the authors are unaware about published estimates of the ratio $\|x_{n+1} - x_\alpha\| / \|x_n - x_\alpha\|$ for a general Tikhonov functional. In fact, because of the ill-posedness of CIPs, previously known a posteriori estimates of the accuracy of Lagrangians do not imply such estimates for regularized coefficients. Hence, that intuitive feeling was not rigorously justified so far. So, we prove the following relaxation property (under certain conditions)

$$\|x_{n+1} - x_\alpha\| \leq r \|x_n - x_\alpha\|, \text{ where } r \in (0,1). \tag{1.1}$$

In the case of forward problems the above question (A) is addressed via a posteriori error analysis, which estimates the difference between computed and exact solutions [1]. It is important that instead of the knowledge of the exact solution, this analysis assumes only the knowledge of an upper estimate for this solution. The latter is usually obtained on the basis of classic a priori estimates for solutions of these problems. In addition, the well posed nature of forward problems enables one to obtain a posteriori error estimates for computed solutions. Unlike this, the ill-posedness of CIPs *radically* changes the situation. As a result, only the accuracy of Lagrange functionals is estimated instead of that of the unknown coefficient [5-7,17]. In those estimates for CIPs a priori upper bounds of solutions are imposed rather than proved. The latter is going along well with the Tikhonov concept for ill-posed problems, which states that some a priori bounds can be imposed on solutions of such problems [3,13,26].

In section 2 a new framework of the Functional Analysis for the adaptivity is introduced. In section 3 the local strict convexity of the Tikhonov functional is proved, the main problem of the interest of this paper is formulated and the existence of local minimizers on subspaces is established. The relaxation property (1.1) is established in section 4. In section 5 we state our CIP and outline the globally numerical convergent numerical method of [8] for it. In section 6 Frechet derivatives with respect to the unknown coefficient of solutions of state and adjoint problems are derived. In section 7 the Tikhonov functional for the CIP is constructed, its Frechet derivative is derived and results of section 4 are specified for this case. In section 8 numerical tests are presented.

2. The Framework Of the Functional Analysis. We work only with piecewise linear finite elements, because they are used in our numerical studies. An extension of our analysis on other finite elements is outside of the scope of this publication. Let $\Omega \subset \mathbb{R}^m, m = 2,3$ be a bounded domain. Consider a triangulation T_0 of this domain with a rather coarse mesh. We obtain a polygonal domain $\sigma \subseteq \Omega$. All subsequent mesh refinement via other triangulations will be done via embedding (in a certain well known manner) smaller triangles/tetrahedra in triangles/tetrahedra forming T_0. Hence, all those triangles/tetrahedra will be located inside of the domain σ. Let T be one of those triangulations. Then we have associated piecewise linear functions $\{e_j(x,T)\}_{j=1}^{\widetilde{p}}$. We now construct a linear space of these functions similarly with the subsection 7.4 of the book

[15]. The function $e_j(x,T)$ is a first order polynomial within the triangle/tetrahedra number j, which we denote as $(Tr)_j$. This function equals 1 at one vertex $(Vs)_j$ of $(Tr)_j$ and it equals zero at all other vertices of $(Tr)_j$. We extend the function $e_j(x,T)$ outside of $(Tr)_j$ for all $x \in \overline{\sigma}$ as follows. Let $(Tr)_k$ be another triangle/tetrahedra of T. Assume first that $(Vs)_j \in \left(\overline{(Tr)_j} \cap \overline{(Tr)_k}\right)$. Then we extend $e_j(x,T)$ in $(Tr)_k$ as $e_j(x,T) := e_k(x,T), x \in (Tr)_k$. Suppose now that $(Vs)_j \notin \left(\overline{(Tr)_j} \cap \overline{(Tr)_k}\right)$. Then we set for the extension $e_j(x,T) := 0, x \in (Tr)_k$. It is clear that if $(Vs)_j = (Vs)_k \in \left(\overline{(Tr)_j} \cap \overline{(Tr)_k}\right)$, then so obtained functions $e_j(x,T)$ and $e_k(x,T)$ are equal to each other, $e_j(x,T) = e_k(x,T), \forall x \in \sigma$. So, we do not differentiate between these equal functions. Hence, each so obtained function $e_j(x,T) = 1$ at the vertex $(Tr)_j$, it equals zero at all other vertices and has a localized support in σ. In addition, each so obtained function $e_j(x,T)$ is piecewise linear in σ. Since these functions are linearly independent ones, we take them as the basis $B(T) := \{e_j(x,T)\}$ for the linear space $Span(e_j(x,T))$.

Let h' be the minimal diameter of triangles/tetrahedra which form T and ϖ' be the radius of the maximal circle/sphere contained in that triangle/tetrahedra. We assume that for all possible triangulations T which we consider below

$$a_1 \leq h' \leq \varpi' a_2; \quad a_1, a_2 = const. > 0, \forall T. \tag{2.1}$$

Thus, the first inequality (2.1) means that we do not decrease the size of triangles/tetrahedra indefinitely. The second inequality (2.1) means that all our triangulations are regular ones, see [12]. It follows from this construction that there exists only a finite number $\widetilde{N}$ of possible triangulations satisfying (2.1). Denote $H = \cup_T Span(B(T))$. Then H is a subspace of $L_2(\sigma)$ and $\dim H := d_H := d_H\left(\widetilde{N}\right) < \infty$. Furthermore,

$$H \subset \left(H^1(\sigma) \cap C(\overline{\sigma})\right) \text{ as a set, } \partial_{x_i} f \in L_\infty(\sigma), \forall f \in H. \tag{2.2}$$

We set the scalar product in H to be the same as one in $L_2(\sigma)$ and denote $(,)$ and $\|\cdot\|$ the scalar product and the corresponding norm in H respectively. The space H can be viewed as an "ideal" space of very fine finite elements, which is never reached in practical computations.

We now construct subspaces $M_n \subset H$ associated with our triangulations T_n. We need to construct these subspaces in such a way that

$$M_n \subset M_{n+1}. \tag{2.3}$$

First, we define the subspace $M_0 := Span(B(T_0)) \subset H$. Next, given the pair (T_n, M_n), the pair (T_{n+1}, M_{n+1}) is constructed as follows. First, we refine the mesh and obtain T_{n+1} and $B(T_{n+1})$. Let $\left\{e_j^n(x)\right\}_{j=1}^{p_n}$ be the basis in M_n. To form the basis of M_{n+1}, we first take functions from $B(T_{n+1})$. Next, we add to $B(T_{n+1})$ such functions from the set $\left\{e_j^n(x)\right\}_{j=1}^{p_n}$ that $e_j^n(x) \notin Span(B(T_{n+1}))$, provided of course that such functions $e_j^n(x)$ exist (alternatively $B(T_{n+1})$ is the basis in M_{n+1}). Thus, we obtain the basis $\left\{e_j^{n+1}(x)\right\}_{j=1}^{p_{n+1}}$ of the subspace $M_{n+1} \subseteq H$. Since $\left\{e_j^n(x)\right\}_{j=1}^{p_n} \subset Span\left(\left\{e_j^{n+1}(x)\right\}_{j=1}^{p_{n+1}}\right)$, then (2.3) holds.

For any subspace $M \subset H$ let $P_M : H \to M$ be the operator of the orthogonal projection of H onto M. Since we use the subspace M_n many times below, we denote for brevity $P_n := P_{M_n}, P_{n+1} := P_{M_{n+1}}$. Below I is the identity operator on H. Let the function $f \in H^1(\sigma) \cap C(\overline{\sigma})$

and its $\partial_{x_i} f_{x_i} \in L_\infty(\sigma)$. Let h_n be the maximal diameter of the above triangles/tetrahedra which are involved in T_n. By the construction of above subspaces $h_{n+1} \leq h_n$. For any function $f \in H$, let f_n be its standard interpolant [15] on triangles/tetrahedra involved in T_n. Then by one of properties of orthogonal projection operators $\|f - P_n f\|_{L_2(\sigma)} \leq \|f - f_n\|_{L_2(\sigma)}$. Hence, it follows from (2.2) and formula 76.3 of the book [15] that with a positive constant $K = K(\sigma)$ depending only on the domain σ

$$\|f - P_n f\|_{L_2(\sigma)} \leq K \|\nabla f\|_{L_\infty(\sigma)} h_n, \forall f \in H. \tag{2.4}$$

Let H_1 be another real valued Hilbert space, whose norm is denoted as $\|\cdot\|_1$. Let $\widetilde{F} : H \to H_1$ be a continuous operator, which does not necessary has a "good" continuous inverse. In general, even if an ill-posed problem in an infinitely dimensional space is "turned" into a well-posed one via a finite dimensional approximation, still the corresponding operator often does not have a "good" continuous inverse, because of that "heritage" from the ill-posed case. Thus, one should apply regularization. Consider the equation $\widetilde{F}(x) = y$. By the Tikhonov concept for ill-posed problems [26], we assume that there exists an "ideal" exact solution $x^* \in H$ of this equation with the "ideal" exact right hand side $y = y^*$, where y^* is given without an error, i.e. $\widetilde{F}(x^*) = y^*$. However, in practice the right hand side y is always given with a small error of the level $\delta \in (0,1)$, $\|y - y^*\|_1 \leq \delta$. Denote $F(x) = \widetilde{F}(x) - y$. Hence, in a small neighborhood of x^* we should find an approximate solution of the following equation

$$F(x) = 0, \ x \in H. \tag{2.5}$$

So, we assume throughout the paper that

$$\|F(x^*)\|_1 \leq \delta, \ \delta \in (0,1). \tag{2.6}$$

For any $d > 0$ denote $V_d(x^*) = \{x \in H : \|x - x^*\| < d\}$. We also assume throughout the paper that the operator F has the Frechet derivative $F'(x)$ for $x \in V_1(x^*) = \{\|x - x^*\| < 1\}$, and this derivative is Lipschitz continuous, i.e. for certain positive constants N_1, N_2

$$\|F'(x)\| \leq N_1, \|F'(x) - F'(y)\| \leq N_2 \|x - y\|, \forall x, y \in V_1(x^*). \tag{2.7}$$

Let x_{glob} be a good first guess for the exact solution x^*. For example, for our CIP of section 5 a good first guess can be obtained by a globally convergent numerical method of [8]. Consider the Tikhonov functional J_α with the regularization parameter $\alpha \in (0,1)$,

$$J_\alpha(x) = \frac{1}{2}\|F(x)\|_1^2 + \frac{\alpha}{2}\|x - x_{glob}\|^2. \tag{2.8}$$

Remark 2.1. In principle, by the Tikhonov theory [26], one should use a stronger norm in the second term of the right hand side of (2.8) to ensure the existence of a minimizer of J_α. However, since all norms in the finite dimensional space H are equivalent, we use a simpler $L_2(\sigma)$ norm here. By our numerical experience with the adaptivity both in section 8 and in previous publications [5-7], this norm is sufficient for our CIP.

Let $J'_\alpha(x)$ be the gradient (i.e. the Frechet derivative) of the functional $J_\alpha(x)$. Then by (2.4)

$$J'_\alpha(x) = (F'(x))^* F(x) + \alpha(x - x_0). \tag{2.9}$$

Let $N_3 = N_3(N_1, N_2) = const. > 0$ be such that

$$\|J_\alpha'(x) - J_\alpha'(y)\| \le N_3 \|x - y\|, \forall x, y \in W_1. \tag{2.10}$$

Below $C = C(N_1, N_2) > 0$ denotes a finite number of different constants depending only on N_1, N_2. We now assume that

$$\|x_{glob} - x^*\| \le \delta^{\mu_1}, \mu_1 = const. \in (0,1), \tag{2.11}$$

$$\alpha = \delta^{\mu_2}, \mu_2 = const. \in (0, \min(\mu_1, 2(1-\mu_1))) \tag{2.12}$$

We impose these assumptions on parameters μ_1, μ_2 to ensure that the distance between the first approximation x_{glob} and the exact solution x^* as well as the regularization parameter α far exceed the error in the data δ for sufficiently small δ, since one cannot perform better than the level of the error in the data. In addition, (2.11b) ensures that points x^*, x_{glob} belong to an appropriate neighborhood of the regularized solution, see Lemmata 2.1 and 3.2.

Lemma 2.1. *A minimizer x_α of the functional $J_\alpha(x)$ on the space H exists for any value of the regularization parameter α. For any $r > 0$ denote $V_r(x_\alpha) = \{x \in H : \|x - x_\alpha\| < r\}$. Assume that conditions (2.11), (2.12) hold. Then $x_{glob} \in V_{\sqrt{2}\delta^{\mu_1}}(x_\alpha)$ and $x^* \in V_{(1+\sqrt{2})\delta^{\mu_1}}(x_\alpha)$. Let $\beta_1 \in (0,1)$ be any number. Then there exists a sufficiently small number $\delta_0 = \delta_0(\mu_1, \mu_2, \beta_1) \in (0,1)$ such that if $\delta \in (0, \delta_0)$, then $x^*, x_{glob} \in V_{\beta_1\alpha}(x_\alpha)$.*

Proof. Since $\dim H < \infty$, then $\lim_{\|x\|\to\infty} J_\alpha(x) = \infty$ implies the existence of a minimizer x_α. Since $J_\alpha(x_\alpha) \le J_\alpha(x^*)$ and by (2.6), (2.8) and (2.11), (2.12) $J_\alpha(x^*) \le \left(\delta^2 + \alpha\delta^{2\mu_1}\right)/2 < \alpha\delta^{2\mu_1}$, then $J_\alpha(x_\alpha) < \alpha\delta^{2\mu_1}$. Hence, by (2.8) $\|x_\alpha - x_{glob}\| \le \sqrt{2}\delta^{\mu_1}$. Hence, $\|x_\alpha - x^*\| \le \|x_\alpha - x_{glob}\| + \|x_{glob} - x^*\| \le \left(1 + \sqrt{2}\right)\delta^{\mu_1}$. To finish the proof, note that by (2.12) $\left(1+\sqrt{2}\right)\delta^{\mu_1} < \beta_1\alpha = \beta_1\delta^{\mu_2}$ for sufficiently small δ. $\square$

The point x_α is called the *regularized solution* of equation (2.3) [3,13,26]. In general, the classic Theorem 2 of Tikhonov on page 65 of [26] states that one can often choose the regularization parameter as $\alpha(\delta) = \delta^\varrho, \varrho \in (0,1)$, which implies $\alpha(\delta) >> \delta$ for sufficiently small δ. Hence, (2.12) is in a good agreement with this result. The proof of the following lemma is rather standard and is therefore omitted.

Lemma 2.2. *Let $M \subset H$ be a subspace and $x_M \in M$ be a point of a local minimum of the functional J_α on M. Then $(J_\alpha'(x_M), z) = 0, \forall z \in M$. Hence,*

$$J_\alpha'(x_\alpha) = 0, \tag{2.13}$$

$$P_M J_\alpha'(x_M) = 0. \tag{2.14}$$

3. Local Strict Convexity of J_α, Problem Statement and Minimizers on Subspaces.

3.1. Convexity. Lemma 3.1 ([22], chapter 10). *Let $U \subset H$ be a convex set and $G : U \to R$ be a continuous functional. Let $(G'(u), z), \forall z \in H$ be its Frechet derivative at the point $u \in U$. Assume that $G'(u)$ is continuous for $u \in U$. Then each of conditions (3.1) and (3.2) is both necessary and sufficient for the strict convexity of the functional G on U with the strict convexity parameter $\kappa = const. > 0$*

$$G(u) - G(v) \ge (G'(v), u - v) + \kappa \|u - v\|^2, \forall u, v \in U, \tag{3.1}$$

$$(G'(u) - G'(v), u - v) \ge 2\kappa \|u - v\|^2, \forall u, v \in U. \tag{3.2}$$

Theorem 3.1. *Assume that conditions (2.11), (2.12) hold. Then there exists numbers* $\beta_1 = \beta_1(N_1, N_2) \in (0,1)$ *and* $\delta_1 = \delta_1(\mu_1, \mu_2, N_2, \beta_1) \in (0,1)$ *depending only on listed parameters such that if* $\rho = \beta_1 \alpha$*, then for any* $\delta \in (0, \delta_1)$ *the functional* J_α *is strictly convex in the neighborhood* $V_\rho(x_\alpha)$ *of the point* x_α *with the strict convexity parameter* $\kappa = \alpha/4$*. Furthermore, by Lemma 2.1 points* $x_{glob}, x^* \in V_\rho(x_\alpha)$*.*

Proof. Let $\beta_1 \in (0,1)$ be the number which we will choose below in this proof, $\rho = \beta_1 \alpha$ and $x, y \in V_\rho(x_\alpha)$ be two arbitrary points. By (2.9)

$$\begin{aligned}
\left(J'_\alpha(x) - J'_\alpha(y), x - y\right) &= \alpha \|x-y\|^2 + \left(F'^*(x)F(x) - F'^*(y)F(y), x-y\right) \\
&= \alpha \|x-y\|^2 + \left(F'^*(x)F(x) - F'^*(x)F(y), x-y\right) \\
&\quad + \left(F'^*(x)F(y) - F'^*(y)F(y), x-y\right).
\end{aligned} \tag{3.3}$$

Denote $A_1 = \left(F'^*(x)F(x) - F'^*(x)F(y), x-y\right), A_2 = \left(F'^*(x)F(y) - F'^*(y)F(y), x-y\right)$ and estimate A_1, A_2 from the below.

Since $A_1 = A_1 - \left(F'^*(x)F'(x)(x-y), x-y\right) + \left(F'^*(x)F'(x)(x-y), x-y\right)$, then

$$\begin{aligned}
A_1 = F'^*(x) &\int_0^1 \left(F'(y+\theta(x-y)) - F'(x)\right)(x-y)\,d\theta, x-y \\
&+ \left(F'^*(x)F'(x)(x-y), x-y\right).
\end{aligned}$$

Using (2.7), we obtain

$$\begin{aligned}
&\left| F'^*(x) \int_0^1 \left[F'(y+\theta(x-y)) - F'(x)\right](x-y)\,d\theta, x-y \right| \\
&\leq \|F'(x)\| \int_0^1 \left\|\left[F'(y+\theta(x-y)) - F'(x)\right](x-y)\right\| d\theta \cdot \|x-y\| \leq \frac{1}{2} N_1 N_2 \|x-y\|^3.
\end{aligned}$$

Also,

$$\left(F'^*(x)F'(x)(x-y), x-y\right) = \left(F'(x)(x-y), F'(x)(x-y)\right)_2 = \|F'(x)(x-y)\|_2^2 \geq 0.$$

Hence, $A_1 \geq N_1 N_2 \|x-y\|^3/2$. Now we estimate A_2,

$$|A_2| \leq \|F(y)\|_2 \|F'(x) - F'(y)\| \|x-y\| \leq N_2 \|x-y\|^2 \|F(y)\|_2.$$

Since $\|F(x_\alpha)\|_2 \leq \|F(y) - F(x_\alpha)\|_2 + \|F(x_\alpha)\|_2 \leq N_1 \|y - x_\alpha\| + \|F(x_\alpha)\|_2$, then

$$|A_2| \leq N_2 \|x-y\|^2 \left(N_1 \|y - x_\alpha\| + \|F(x_\alpha)\|_2\right). \tag{3.4}$$

By (2.6), (2.7) and Lemma 2.1 $\|F(x_\alpha)\|_2 \leq \|F(x_\alpha) - F(x^*)\|_2 + \delta \leq \alpha\beta_1 N_2 + \delta$. Hence, by (3.4)

$$A_2 \geq -N_2 \|x-y\|^2 \left(N_1 \|y - x_\alpha\| + \alpha\beta_1 N_2 + \delta\right).$$

Combining this with (3.3) and the above estimate for A_1, we obtain

$$\left(J_\alpha'(x) - J_\alpha'(y), x-y\right) \geq \tag{3.5}$$

$$\|x-y\|^2 \left[\alpha - \frac{N_1 N_2}{2}\|x-y\| - N_1 N_2 \|y - x_\alpha\| - N_2\left(N_2 \alpha \beta_1 + \delta\right)\right].$$

We have

$$N_1 N_2 \left(\frac{\|x-y\|}{2} + \|y-x_\alpha\|\right) + N_2\left(N_2 \alpha\beta_1 + \delta\right) \leq 2\alpha\beta_1 N_2 \left(N_1 + N_2\right) + N_2 \delta. \tag{3.6}$$

Choose $\beta_1 = \beta_1(N_1, N_2) \in (0,1)$ such that $2\beta_1 N_2 (N_1 + N_2) \leq 1/4$. Given this β_1, choose $\delta_1 = \delta_1(\mu_1, \mu_2, N_1, N_2) \in (0,1)$ so small that $N_2 \delta < \delta^{\mu_2}/4 = \alpha/4$ and $2\delta^{\mu_1} < \beta_1 \delta^{\mu_2} = \beta_1 \alpha, \forall \delta \in (0, \delta_1)$. Then (3.5), (3.6) and (3.2) imply that Theorem 3.1 is proven. □

Lemma 3.2. *Assume that conditions of Theorem 3.1 hold. Then in the neighborhood $V_{\left(1+\sqrt{2}\right)\delta^{\mu_1}}(x^*)$ of x^* there exists unique minimizer x_α of the functional $J_\alpha(x)$. Furthermore, $V_{\left(1+\sqrt{2}\right)\delta^{\mu_1}}(x^*) \subset V_\rho(x_\alpha)$. If the operator F is one-to-one, then x^* is unique and therefore x_α is unique also.*

Note that, unless the operator F is one-to-one, there is no guarantee that the exact solution of equation (2.5) is unique. The proof of Lemma 3.2 follows immediately from Lemma 2.1 and Theorem 3.1. Hence, even though there might exist several exact solutions of equation (2.5), still as long as a good first guess x_{glob} about one of these solutions is available and conditions (2.11), (2.12) are satisfied, one can guarantee uniqueness of the regularized solution in a small neighborhood of that exact solution. Hence, below we work only with such an exact solution x^* that satisfies (2.11), assuming of course that x^* exists for the given vector x_{glob}. As to x_α, all what we know about this vector is it exists, is unique and by Lemma 2.1 $x_\alpha \in V_{\left(1+\sqrt{2}\right)\delta^{\mu_1}}(x^*)$. Thus, we denote below for brevity $V_\rho(x_\alpha) := V_\rho$. Therefore the statement of the Problem 3.1 has no ambiguity now in terms of x_α. The following problem is the main interest of our study below.

Problem 3.1. *Suppose that conditions of Theorem 3.1 are satisfied and $\delta \in (0, \delta_1)$. For a fixed value of the regularization parameter α, approximate the regularized solution x_α in the norm of $L_2(\sigma)$ via a finite number of above described mesh refinements.*

3.2. Local minimizers on subspaces. In this subsection we establish the existence and uniqueness of a minimizer $x_n \in M_n \cap \left(\overline{V}_\rho \diagdown \partial \overline{V}_\rho\right)$ of the functional J_α. To do so, we first reformulate Proposition 6.3.4 of [23], which is derived there from the Leray-Schauder theorem.

Proposition 3.1. *Let $D \subset \mathbb{R}^k$ be an open domain, $\Phi : \overline{D} \to \mathbb{R}^k$ be a continuous mapping and $x^0 \in \overline{D} \diagdown \partial D$ be an arbitrary point. Assume that $\left[\Phi(x), x - x^0\right] \geq 0, \forall x \in \partial D$, where $[,]$ is the scalar product in $\mathbb{R}^k$. Then there exists a point $\widetilde{x} \in \overline{D}$ such that $\Phi(\widetilde{x}) = 0$.*

Theorem 3.2. *Assume that conditions of Theorem 3.1 hold. Suppose that there exists an integer $\overline{n} \geq 1$ such that with the constant K from (2.4)*

$$K \|\nabla x_\alpha\|_{L_\infty(\sigma)} h_{\overline{n}} := \Delta_{\overline{n}}' < \frac{\beta_1 \alpha^2}{\sqrt{4N_3^2 + \alpha^2}} = \frac{\alpha\rho}{\sqrt{4N_3^2 + \alpha^2}}. \tag{3.7}$$

Let $M' \subseteq H$ be any subspace such that $M_{\overline{n}} \subseteq M'$. Then $V_\rho \cap M' \neq \varnothing$. Furthermore, there exists a unique point $x_{M'} \in \left(\overline{V}_\rho \diagdown \partial V_\rho\right) \cap M'$ at which the functional $J_\alpha(x)$ attains its minimal value on the set $V_\rho \cap M'$.

Proof of Theorem 3.2. We first prove this theorem for $M' = M_{\overline{n}}$. Denote

$$\Delta_{\overline{n}} = \|x_\alpha - P_{M_{\overline{n}}} x_\alpha\|, R_{M_{\overline{n}}} = \sqrt{\rho^2 - \Delta_{\overline{n}}^2}, S_{\overline{n}} = \{x \in M_{\overline{n}} : \|x - P_{M_{\overline{n}}} x_\alpha\| < R_{M_{\overline{n}}}\}. \tag{3.8}$$

By (2.4) and (3.7)

$$\Delta_n \leq \Delta'_{\overline{n}} < \rho. \tag{3.9}$$

Let $x \in M_{\overline{n}}$ be an arbitrary point. Since $(x - P_{M_{\overline{n}}} x_\alpha) \in M_{\overline{n}}$, then vectors $(x - P_{M_{\overline{n}}} x_\alpha)$ and $(x_\alpha - P_{M_{\overline{n}}} x_\alpha)$ are orthogonal. Hence,

$$\begin{aligned} \|x - x_\alpha\|^2 &= \|x - P_{M_{\overline{n}}} x_\alpha + P_{M_{\overline{n}}} x_\alpha - x_\alpha\|^2 = \|x - P_{M_{\overline{n}}} x_\alpha\|^2 + \|P_{M_{\overline{n}}} x_\alpha - x_\alpha\|^2 \\ &< \|x - P_{M_{\overline{n}}} x_\alpha\|^2 + \Delta_{\overline{n}}^2 \leq \rho^2 - \Delta_{\overline{n}}^2 + \Delta_{\overline{n}}^2 = \rho^2, \forall x \in S_{\overline{n}}. \end{aligned}$$

Hence,

$$S_{\overline{n}} \subseteq V_\rho \cap M_{\overline{n}} \text{ implying that } V_\rho \cap M_{\overline{n}} \neq \varnothing. \tag{3.10}$$

Define the functional $J_{\alpha, M_{\overline{n}}} : M_{\overline{n}} \to \mathbb{R}$ as $J_{\alpha, M_{\overline{n}}}(x) := J_\alpha(x), \forall x \in M_{\overline{n}}$. Then the gradient of $J_{\alpha, M_{\overline{n}}}(x)$ is $P_{M_{\overline{n}}} J'_\alpha(x), \forall x \in M_{\overline{n}}$. Hence, it follows from (3.2), (3.10) and Theorem 3.1 that the functional $J_{\alpha, M_n}(x)$ is strictly convex on $V_\rho \cap M_{\overline{n}}$. Hence, (3.10) and (2.14) imply that it is sufficient to prove the existence of a point $x_{M_{\overline{n}}} \in S_n$ such that $J'_{\alpha, M_{\overline{n}}}(x_{M_{\overline{n}}}) = 0$. To make sure that the point $x_{M_{\overline{n}}} \in \overline{S}_{\overline{n}} \diagdown \partial \overline{S}_{\overline{n}}$, consider a small number $\varepsilon \in (0,1)$ which will be chosen later. Let $S_{\overline{n}}(\varepsilon) = \{x \in M_{\overline{n}} : \|x - P_{M_{\overline{n}}} x_\alpha\| = (1-\varepsilon) R_{M_{\overline{n}}}\}$. Hence, $S_{\overline{n}}(\varepsilon) \subset S_{\overline{n}}$. Using (2.10), (2.13), (3.2), Theorem 3.1 and (3.8), we obtain for $x \in S_{\overline{n}}(\varepsilon)$

$$\begin{aligned} \left(J'_{\alpha, M_n}(x), x - P_{M_{\overline{n}}} x_\alpha\right) &= \left(P_{M_{\overline{n}}} J'_\alpha(x) - J'_\alpha(x_\alpha), x - P_{M_{\overline{n}}} x_\alpha\right) \\ &= \left(J'_\alpha(x) - J'_\alpha(P_{M_{\overline{n}}} x_\alpha), x - P_{M_{\overline{n}}} x_\alpha\right) \\ &\quad + \left(J'_\alpha(P_{M_{\overline{n}}} x_\alpha) - J'_\alpha(x_\alpha), x - P_{M_{\overline{n}}} x_\alpha\right) \\ &\geq \frac{\alpha}{2} \|x - P_{M_{\overline{n}}} x_\alpha\|^2 + \left(J'_\alpha(P_{M_{\overline{n}}} x_\alpha) - J'_\alpha(x_\alpha), x - P_{M_{\overline{n}}} x_\alpha\right) \\ &\geq \frac{\alpha (1-\varepsilon)^2 R_{M_{\overline{n}}}^2}{2} - N_3 (1-\varepsilon) R_{M_{\overline{n}}} \Delta_{\overline{n}}. \end{aligned}$$

Hence, (3.7), (3.9) and elementary calculations show that one can choose a sufficiently small ε such that $\left(J'_{\alpha, M_n}(x), x - P_{M_{\overline{n}}} x_\alpha\right) > 0, \forall x \in S_{\overline{n}}(\varepsilon)$. Hence, Proposition 3.1 implies the existence of the above point $x_{M_{\overline{n}}}$. By Theorem 3.1 this point is unique. Finally, if $M_{\overline{n}} \subseteq M'$, then $\|x_\alpha - P_{M'} x_\alpha\| \leq \|x_\alpha - P_{M_{\overline{n}}} x_\alpha\|$, which means that the above proof is applicable to M' as well. $\square$

4. Relaxation. In this section we use without restating various properties of orthogonal projection operators in Hilbert spaces, which are well known from the standard Functional Analysis course. In particular, we use the following three properties

$$P_M^2 = P_M; \; P_M^* = P_M; \; P_M(z) = z, \forall z \in M; \; (x - P_M x.y) = 0, \; \forall x \in H, \forall y \in M.$$

In sections 4 and 7 we assume without restating that the following Assumption 4.1 is valid.

Assumption 4.1. We assume that conditions of Theorem 3.2 hold, which implies that conditions of Theorem 3.1 and Lemma 2.1 are also in place. In particular, we impose a priori upper bound on the regularized solution x_α. The latter is going along well with the above mentioned (section 1) Tikhonov concept for ill-posed problems, by which a priori bounds should be imposed [3,13,26]. Namely, we assume that $\|\nabla x_\alpha\|_{L_\infty(\sigma)} \leq A$, where A is a given constant. Hence, we assume below that $n \geq \overline{n}$ and impose a little bit stronger condition than (3.7),

$$h_n < \frac{\beta_1 \alpha^2}{AK\sqrt{4N_3^2+\alpha^2}}. \tag{4.1}$$

By Theorem 3.2, there exists unique point $x_n \in \left(\overline{V}_\rho \diagdown \partial V_\rho\right) \cap M_n$ at which the functional $J_\alpha(x)$ attains its minimal value on this set. Hence, by (2.14)

$$P_n J'_\alpha(x_n) = 0. \tag{4.2}$$

For any two vectors $a, b \in H$ let $An(a,b) \in [0,\pi]$ be the angle between them, provided that at least one of them is non zero. If one of them is zero, then $An(a,b) := 0$. The number $\cos[An(a,b)]$ is defined via the scalar product. Lemma 4.1 is elementary.

Lemma 4.1. *Let $u, v \in H$ be two orthogonal vectors, $u + v \neq 0$ and $\varphi = An(u, u+v)$. Then $\varphi \in [0, \pi/2], \|u\| = \|u+v\| \cos\varphi$ and $\|v\| = \|u+v\| \sin\varphi$.*

Consider the functional $J_\alpha(x)$ for $x \in V_\rho \cap M_n$. It is reasonable to assume that

$$J'_\alpha(x_n) \neq 0. \tag{4.3}$$

Indeed, if (4.3) is not true, then by (2.13) $J'_\alpha(x_n) = J'_\alpha(x_\alpha) = 0$ and Theorem 3.1 implies that $x_n = x_\alpha$ and the Problem 3.1 is solved in this case. Assume that the subspace M_{n+1} is also chosen. Recall that by (2.3) $M_n \subset M_{n+1}$. Since by (4.2) the gradient $J'_\alpha(x_n)$ is orthogonal to the subspace M_n, then one can consider two *auxiliary* subspaces,

$$G_{n+1} = M_n \oplus J'_\alpha(x_n), \tag{4.4}$$

$$\widetilde{G}_{n+1} = P_{n+1} G_{n+1}, \tag{4.5}$$

where "$\oplus$" denotes the orthogonal sum. By (4.4) $M_n \subset G_{n+1}$. Also, since $M_n \subset M_{n+1}$, then by (4.4) $P_{n+1}x = P_n x + \lambda(x) P_{n+1} J'_\alpha(x_n), \forall x \in G_{n+1}$, where $\lambda(x)$ is a certain number depending on x. Since, $P_n x \in M_n$ and $P_n M_n = M_n$, then by (4.5) $M_n \subset \widetilde{G}_{n+1}$. Therefore, Theorem 3.2 and Assumption 4.1 imply that there exists two auxiliary minimizers $x_{n+1}^g \in G_{n+1}, \widetilde{x}_{n+1}^g \in \widetilde{G}_{n+1}$ of the functional J_α (each one of them is unique) such that

$$J_\alpha\left(x_{n+1}^g\right) = \min_{\overline{V}_\rho \cap G_{n+1}} J_\alpha(x), \; x_{n+1}^g \in \left(\overline{V}_\rho \diagdown \partial \overline{V}_\rho\right) \cap G_{n+1}, \tag{4.6}$$

$$J_\alpha\left(\widetilde{x}_{n+1}^g\right) = \min_{\overline{V}_\rho \cap \widetilde{G}_{n+1}} J_\alpha(x), \; \widetilde{x}_{n+1}^g \in \left(\overline{V}_\rho \diagdown \partial \overline{V}_\rho\right) \cap \widetilde{G}_{n+1}. \tag{4.7}$$

Hence, by (4.6) and (4.7) there exist numbers $\lambda_{n+1}, \widetilde{\lambda}_{n+1} \in \mathbb{R}$ such that vectors $x_{n+1}^g, \widetilde{x}_{n+1}^g$ can be represented as

$$x_{n+1}^g = y_{n+1} + \lambda_{n+1} J'_\alpha(x_n), \; \widetilde{x}_{n+1}^g = \widetilde{y}_{n+1} + \widetilde{\lambda}_{n+1} P_{n+1} J'_\alpha(x_n); \; y_{n+1}, \widetilde{y}_{n+1} \in M_n. \tag{4.8}$$

Lemma 4.2. *Let condition (4.3) holds. Then the following estimate is valid*

$$\left\|x_{n+1}^g - x_\alpha\right\| \leq \widetilde{r} \|x_n - x_\alpha\| + \widetilde{r}\Delta_n + \Delta_{n+1}^g, \tag{4.9}$$

$$\Delta_n = \|x_n - P_n x_\alpha\|, \; \Delta_{n+1}^g = \left\|x_{n+1}^g - P_{G_{n+1}} x_\alpha\right\|, \; \widetilde{r} = \sqrt{1 - \frac{\alpha^2}{4N_3^2}}. \tag{4.10}$$

Proof. Consider the unit vector $p_n = J'_\alpha(x_n)/\|J'_\alpha(x_n)\|$. Then by (4.4) $P_{G_{n+1}}x = P_n x + (x - P_n x, p_n) p_n, \forall x \in H$. Consider vectors $u = P_n x_\alpha - P_{G_{n+1}} x_\alpha, v = P_{G_{n+1}} x_\alpha - x_\alpha$. Since $u = (P_n x_\alpha - x_\alpha, p_n) p_n \in G_{n+1}$ and v is orthogonal to G_{n+1}, then $(u, v) = 0$. Since by (4.3) and (2.13) $x_\alpha \notin M_n$, then $u + v = P_n x_\alpha - x_\alpha \neq 0$. Hence, by Lemma 4.1

$$\left\|x_\alpha - P_{G_{n+1}} x_\alpha\right\| = g_n \left\|x_\alpha - P_n x_\alpha\right\|, g_n = \sin\varphi_n, \tag{4.11}$$

where $\varphi_n = An(u, x_\alpha - P_n x_\alpha)$. Using (4.11), we now estimate the norm $\left\|x_{n+1}^g - x_\alpha\right\|$,

$$\begin{aligned}\left\|x_{n+1}^g - x_\alpha\right\| &\leq \left\|x_{n+1}^g - P_{G_{n+1}} x_\alpha\right\| + \left\|x_\alpha - P_{G_{n+1}} x_\alpha\right\| \\ &= \left\|x_{n+1}^g - P_{G_{n+1}} x_\alpha\right\| + g_n \left\|x_\alpha - P_n x_\alpha\right\| \\ &\leq g_n \left\|x_n - P_n x_\alpha\right\| + g_n \left\|x_n - x_\alpha\right\| + \left\|x_{n+1}^g - P_{G_{n+1}} x_\alpha\right\|.\end{aligned}$$

Hence, taking into account notations (4.10), we obtain (4.9) in which $\widetilde{r}$ is replaced with g_n.

We now estimate g_n from the above. By (4.2) $(J'_\alpha(x_n), P_n x_\alpha - x_n) = 0$. Hence,

$$(J'_\alpha(x_n), P_n x_\alpha - x_\alpha) = (J'_\alpha(x_n), P_n x_\alpha - x_n) + (J'_\alpha(x_n), x_n - x_\alpha) = (J'_\alpha(x_n), x_n - x_\alpha).$$

Comparing this with (4.11), we obtain $\|J'_\alpha(x_n)\| \|P_n x_\alpha - x_\alpha\| \cos\varphi_n = (J'_\alpha(x_n), x_n - x_\alpha)$. By (2.13) and Theorem 3.1

$$(J'_\alpha(x_n), x_n - x_\alpha) = (J'_\alpha(x_n) - J'_\alpha(x_\alpha), x_n - x_\alpha) \geq \frac{\alpha}{2}\|x_\alpha - x_n\|^2.$$

Hence,

$$\|J'_\alpha(x_n)\| \|x_\alpha - P_n x_\alpha\| \cos\varphi_n \geq \frac{\alpha}{2}\|x_\alpha - x_n\|^2. \tag{4.12}$$

By (2.10) and (2.13) $\|J'_\alpha(x_n)\| = \|J'_\alpha(x_n) - J'_\alpha(x_\alpha)\| \leq N_3 \|x_n - x_\alpha\|$. Combining this with (4.12) and using the fact that by one of the properties of orthogonal projection operators $\|x_\alpha - x_n\| \geq \|x_\alpha - P_n x_\alpha\|$, we obtain

$$\cos\varphi_n \geq \frac{\alpha}{2N_3} \cdot \frac{\|x_\alpha - x_n\|}{\|x_\alpha - P_n x_\alpha\|} \geq \frac{\alpha}{2N_3}.$$

Hence, by (4.10) and (4.11) $g_n = \sin\varphi_n \leq \sqrt{1 - \alpha^2 (2N_3)^{-2}} = \widetilde{r}$. $\square$

Numbers Δ_n and Δ_{n+1}^g in (4.10) characterize approximating properties of subspaces M_n and G_{n+1} with respect to the regularized solution x_α. In the proof of Lemma 4.2 we have not used the fact that x_{n+1}^g is the minimizer of $J_\alpha(x)$ on $V_\rho \cap G_{n+1}$, see (4.6). We use (4.6) in Theorem 4.1. In the proof of this theorem we first obtain an upper estimate of $\|x_{n+1} - x_\alpha\|$ via numbers $\|x_n - x_\alpha\|, \Delta_n, \Delta_{n+1}^g$ and $\left\|x_{n+1}^g - \widetilde{x}_{n+1}^g\right\|$. Next, we estimate $\left\|x_{n+1}^g - \widetilde{x}_{n+1}^g\right\|$ from the above via $\|(I - P_{n+1}) J'_\alpha(x_n)\| / \|J'_\alpha(x_n)\|$, which is the most technical part of the proof. Finally, we estimate numbers Δ_n, Δ_{n+1} and Δ_{n+1}^g from the above via $\|(I - P_n) x_\alpha\|$. Next, in the proof of Theorem 4.2 we estimate from the above numbers $\|(I - P_{n+1}) J'_\alpha(x_n)\| / \|J'_\alpha(x_n)\|$ and $\|(I - P_n) x_\alpha\|$ via $\|x_n - x_\alpha\|$, thus ending up with the target estimate (1.1). By (4.3) there exists such a subspace $M_{n+1} \subset H$, $M_n \subset M_{n+1}$ that

$$P_{n+1} J'_\alpha(x_n) \neq 0, \text{ which is equivalent with } (P_{n+1} J'_\alpha(x_n), J'_\alpha(x_n)) \neq 0. \tag{4.13}$$

Theorem 4.1. *Assume that condition (4.3) holds. Then with the constant* $\widetilde{r}\in(0,1)$ *of (4.10) the following estimate is valid*

$$\|x_{n+1}-x_\alpha\|\leq\widetilde{r}\,\|x_n-x_\alpha\|+C\frac{\|(I-P_{n+1})\,x_\alpha\|}{\sqrt{\alpha}}+C\sqrt{\alpha}\frac{\|(I-P_{n+1})\,J'_\alpha\,(x_n)\|^{1/2}}{\|J'_\alpha\,(x_n)\|^{1/2}}. \tag{4.14}$$

Proof. We have

$$\left\|x^g_{n+1}-x_\alpha\right\|\geq\left\|\widetilde{x}^g_{n+1}-x_\alpha\right\|-\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|. \tag{4.15}$$

Since $\widetilde{x}^g_{n+1}\in\widetilde{G}_{n+1}\subset M_{n+1}$, then, using (4.10), we obtain

$$\left\|\widetilde{x}^g_{n+1}-x_\alpha\right\|\geq\|P_{n+1}x_\alpha-x_\alpha\|\geq\|x_{n+1}-x_\alpha\|-\|x_{n+1}-P_{n+1}x_\alpha\|=\|x_{n+1}-x_\alpha\|-\Delta_{n+1}.$$

Hence, it follows from (4.15) that $\left\|x^g_{n+1}-x_\alpha\right\|\geq\|x_{n+1}-x_\alpha\|-\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|-\Delta_{n+1}$. Substituting this inequality in (4.9), we obtain

$$\|x_{n+1}-x_\alpha\|-\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|-\Delta_{n+1}\leq\left\|x^g_{n+1}-x_\alpha\right\|\leq\widetilde{r}\,\|x_n-x_\alpha\|+\widetilde{r}\Delta_n+\Delta^g_{n+1},$$

which implies that

$$\|x_{n+1}-x_\alpha\|\leq\widetilde{r}\,\|x_n-x_\alpha\|+\widetilde{r}\Delta_n+\Delta_{n+1}+\Delta^g_{n+1}+\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|. \tag{4.16}$$

We now estimate the norm $\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|$ of the last term of (4.16) from the above. Using (4.6), (4.9) and (2.14), we obtain

$$\begin{aligned}
&\left(J'_\alpha\left(x^g_{n+1}\right),x^g_{n+1}-\widetilde{x}^g_{n+1}\right)=\left(J'_\alpha\left(x^g_{n+1}\right),x^g_{n+1}-P_{G_{n+1}}\widetilde{x}^g_{n+1}\right)\\
&+\left(J'_\alpha\left(x^g_{n+1}\right),P_{G_{n+1}}\widetilde{x}^g_{n+1}-\widetilde{x}^g_{n+1}\right)=\left(J'_\alpha\left(x^g_{n+1}\right),\left(P_{G_{n+1}}-I\right)\widetilde{x}^g_{n+1}\right),\\
&-\left(J'_\alpha\left(\widetilde{x}^g_{n+1}\right),x^g_{n+1}-\widetilde{x}^g_{n+1}\right)=-\left(J'_\alpha\left(\widetilde{x}^g_{n+1}\right),x^g_{n+1}-P_{\widetilde{G}_{n+1}}x^g_{n+1}\right)\\
&-\left(J'_\alpha\left(\widetilde{x}^g_{n+1}\right),P_{\widetilde{G}_{n+1}}x^g_{n+1}-\widetilde{x}^g_{n+1}\right)=-\left(J'_\alpha\left(\widetilde{x}^g_{n+1}\right),\left(P_{\widetilde{G}_{n+1}}-I\right)x^g_{n+1}\right).
\end{aligned}$$

Hence, (3.2) and Theorem 3.1 imply that

$$\begin{aligned}
\frac{\alpha}{2}\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|^2&\leq\left(J'_\alpha\left(x^g_{n+1}\right)-J'_\alpha\left(\widetilde{x}^g_{n+1}\right),x^g_{n+1}-\widetilde{x}^g_{n+1}\right)\\
&=\left(J'_\alpha\left(x^g_{n+1}\right),\left(P_{G_{n+1}}-I\right)\widetilde{x}^g_{n+1}\right)-\left(J'_\alpha\left(\widetilde{x}^g_{n+1}\right),\left(P_{\widetilde{G}_{n+1}}-I\right)x^g_{n+1}\right)\\
&\leq\left\|J'_\alpha\left(x^g_{n+1}\right)\right\|\left\|\left(I-P_{G_{n+1}}\right)\widetilde{x}^g_{n+1}\right\|+\left\|J'_\alpha\left(\widetilde{x}^g_{n+1}\right)\right\|\left\|\left(I-P_{\widetilde{G}_{n+1}}\right)x^g_{n+1}\right\|.
\end{aligned} \tag{4.17}$$

Since $x^g_{n+1},\widetilde{x}^g_{n+1}\in V_\rho,\rho=\beta_1\alpha$ and the constant β_1 depends only on constants N_1,N_2, we can temporary set $\beta_1:=C$. Hence (2.10) and (2.13) imply that with another constant C, $\left\|J'_\alpha\left(x^g_{n+1}\right)\right\|=\left\|J'_\alpha\left(x^g_{n+1}\right)-J'_\alpha\left(x_\alpha\right)\right\|\leq N_3\left\|x^g_{n+1}-x_\alpha\right\|\leq C\alpha$. Similarly $\left\|J'_\alpha\left(\widetilde{x}^g_{n+1}\right)\right\|\leq C\alpha$. Hence, (4.17) implies that

$$\left\|x^g_{n+1}-\widetilde{x}^g_{n+1}\right\|^2\leq C\left(\left\|\left(I-P_{G_{n+1}}\right)\widetilde{x}^g_{n+1}\right\|+\left\|\left(I-P_{\widetilde{G}_{n+1}}\right)x^g_{n+1}\right\|\right). \tag{4.18}$$

By (4.3) and (4.13) the following angle is properly defined $\psi_n = An\left(J'_\alpha(x_n), P_{n+1}J'_\alpha(x_n)\right)$. We now prove that

$$\left\|\left(I - P_{G_{n+1}}\right)\widetilde{x}^g_{n+1}\right\| = \left\|\left(I - P_n\right)\widetilde{x}^g_{n+1}\right\| \sin\psi_n, \tag{4.19}$$

$$\left\|\left(I - P_{\widetilde{G}_{n+1}}\right)x^g_{n+1}\right\| = \left\|\left(I - P_n\right)x^g_{n+1}\right\| \sin\psi_n. \tag{4.20}$$

First, we figure out the form of the vector $P_{G_{n+1}}\widetilde{x}^g_{n+1}$. By (4.4) and (4.8)

$$P_{G_{n+1}}\widetilde{x}^g_{n+1} = \widetilde{y}_{n+1} + \widetilde{\lambda}_{n+1}P_{G_{n+1}}P_{n+1}J'_\alpha(x_n), \widetilde{y}_{n+1} \in M_n. \tag{4.21}$$

By (2.3) and (4.2) $(y, P_{n+1}J'_\alpha(x_n)) = (P_{n+1}y, J'_\alpha(x_n)) = (y, J'_\alpha(x_n)) = 0, \forall y \in M_n$. Hence, by (4.4) $P_{G_{n+1}}P_{n+1}J'_\alpha(x_n) = a \cdot J'_\alpha(x_n)$, where $a \in \mathbb{R}$. Compute the number a using (4.4),

$$\begin{aligned} a\left\|J'_\alpha(x_n)\right\|^2 &= \left(P_{G_{n+1}}P_{n+1}J'_\alpha(x_n), J'_\alpha(x_n)\right) = \left(P_{n+1}J'_\alpha(x_n), P_{G_{n+1}}J'_\alpha(x_n)\right) \\ &= \left(P_{n+1}J'_\alpha(x_n), J'_\alpha(x_n)\right) = \left(P_{n+1}J'_\alpha(x_n), P_{n+1}J'_\alpha(x_n)\right) = \left\|P_{n+1}J'_\alpha(x_n)\right\|^2. \end{aligned}$$

Hence, $a = \left\|P_{n+1}J'_\alpha(x_n)\right\|^2 \left\|J'_\alpha(x_n)\right\|^{-2}$. Hence, (4.21) leads to

$$P_{G_{n+1}}\widetilde{x}^g_{n+1} = \widetilde{y}_{n+1} + \widetilde{\lambda}_{n+1}\frac{\left\|P_{n+1}J'_\alpha(x_n)\right\|^2}{\left\|J'_\alpha(x_n)\right\|^2} \cdot J'_\alpha(x_n), \widetilde{y}_{n+1} \in M_n. \tag{4.22}$$

Let $u_1 = P_{G_{n+1}}\widetilde{x}^g_{n+1} - P_n\widetilde{x}^g_{n+1}, v_1 = \left(I - P_{G_{n+1}}\right)\widetilde{x}^g_{n+1}$. First, we show that $(u_1, v_1) = 0$. Indeed, by (4.22) and (4.8)

$$u_1 = \widetilde{\lambda}_{n+1}\frac{\left\|P_{n+1}J'_\alpha(x_n)\right\|^2}{\left\|J'_\alpha(x_n)\right\|^2}J'_\alpha(x_n), \tag{4.23}$$

$$v_1 = \widetilde{\lambda}_{n+1}\left[P_{n+1}J'_\alpha(x_n) - \frac{\left\|P_{n+1}J'_\alpha(x_n)\right\|^2}{\left\|J'_\alpha(x_n)\right\|^2}J'_\alpha(x_n)\right].$$

Hence,

$$(u_1, v_1) = \widetilde{\lambda}^2_{n+1}\frac{\left\|P_{n+1}J'_\alpha(x_n)\right\|^2}{\left\|J'_\alpha(x_n)\right\|^2}\left[\left\|P_{n+1}J'_\alpha(x_n)\right\|^2 - \left\|P_{n+1}J'_\alpha(x_n)\right\|^2\right] = 0. \tag{4.24}$$

Next, $u_1 + v_1 = \widetilde{\lambda}_{n+1}P_{n+1}J'_\alpha(x_n)$. Thus, (4.23) and Lemma 4.1 imply (4.19) if $\widetilde{\lambda}_{n+1} \neq 0$. If, however, $\widetilde{\lambda}_{n+1} = 0$, then it follows from (4.8) and (4.22) that in (4.19) $\left(I - P_{G_{n+1}}\right)\widetilde{x}^g_{n+1} = (I - P_n)\widetilde{x}^g_{n+1} = 0$, which again implies (4.19).

To prove (4.20), denote

$$u_2 = P_{\widetilde{G}_{n+1}}x^g_{n+1} - P_nx^g_{n+1}, v_2 = \left(I - P_{\widetilde{G}_{n+1}}\right)x^g_{n+1}.$$

By (4.4), (4.5) and (4.8)

$$P_{\widetilde{G}_{n+1}}x^g_{n+1} = P_{n+1}x^g_{n+1} = y_{n+1} + \lambda_{n+1}P_{n+1}J'_\alpha(x_n).$$

Hence, $u_2 = \lambda_{n+1} P_{n+1} J'_\alpha (x_n)$ and $v_2 = \lambda_{n+1} (I - P_{n+1}) J'_\alpha (x_n)$. Thus, $(u_2, v_2) = 0$. Next, by (4.4) and (4.8) $u_2 + v_2 = (I - P_n) x^g_{n+1} = \lambda_{n+1} J'_\alpha (x_n)$. Hence, if $\lambda_{n+1} \neq 0$, then the angle between vectors u_2 and $u_2 + v_2$ is the same as the angle between vectors $J'_\alpha (x_n)$ and $P_{n+1} J'_\alpha (x_n)$, i.e., this is the angle ψ_n introduced above. Hence, using Lemma 4.1, we obtain (4.20) for $\lambda_{n+1} \neq 0$. In the case $\lambda_{n+1} = 0$ we have $\left(I - P_{\widetilde{G}_{n+1}}\right) x^g_{n+1} = (I - P_n) x^g_{n+1} = 0$, which implies (4.20).

Since by (4.13) $\|(I - P_{n+1}) J'_\alpha (x_n)\| < \|J'_\alpha (x_n)\|$, then by (4.20)

$$\sin \psi_n = \frac{\|v_2\|}{\|u_2 + v_2\|} = \frac{\|(I - P_{n+1}) J'_\alpha (x_n)\|}{\|J'_\alpha (x_n)\|} < 1. \tag{4.25}$$

Thus, (4.18)-(4.20) and (4.25) imply that

$$\left\|x^g_{n+1} - \widetilde{x}^g_{n+1}\right\|^2 \leq C \left(\left\|(I - P_n) x^g_{n+1}\right\| + \left\|(I - P_n) \widetilde{x}^g_{n+1}\right\|\right) \cdot \frac{\|(I - P_{n+1}) J'_\alpha (x_n)\|}{\|J'_\alpha (x_n)\|}. \tag{4.26}$$

Estimate the term in the parenthesis in the right hand side of (4.26). By (4.6) and (4.7)

$$\left\|x^g_{n+1} - x_n\right\|, \left\|\widetilde{x}^g_{n+1} - x_n\right\| \leq C\alpha.$$

Also, since $x_n \in M_n$, then $(I - P_n)(x_n) = 0$. In addition, $\|I - P_n\| \leq \|I\| + \|P_n\| \leq 2$. Hence,

$$\begin{aligned} &\left\|(I - P_n) x^g_{n+1}\right\| + \left\|(I - P_n) \widetilde{x}^g_{n+1}\right\| = \left\|(I - P_n) \left(x^g_{n+1} - x_n\right)\right\| \\ &+ \left\|(I - P_n) \left(\widetilde{x}^g_{n+1} - x_n\right)\right\| \leq \left\|x^g_{n+1} - x_n\right\| + \left\|\widetilde{x}^g_{n+1} - x_n\right\| \leq C\alpha. \end{aligned} \tag{4.27}$$

Hence, (4.26) and (4.27) lead to

$$\left\|x^g_{n+1} - \widetilde{x}^g_{n+1}\right\| \leq C\sqrt{\alpha} \frac{\|(I - P_{n+1}) J'_\alpha (x_n)\|^{1/2}}{\|J'_\alpha (x_n)\|^{1/2}}. \tag{4.28}$$

Therefore, it follows from (4.16) and (4.28) that

$$\|x_{n+1} - x_\alpha\| \leq \widetilde{r} \|x_n - x_\alpha\| + \widetilde{r}\Delta_n + \Delta_{n+1} + \Delta^g_{n+1} + C\sqrt{\alpha} \frac{\|(I - P_{n+1}) J'_\alpha (x_n)\|^{1/2}}{\|P_{n+1} J'_\alpha (x_n)\|^{1/2}}. \tag{4.29}$$

We now estimate from the above terms Δ_n, Δ_{n+1} and Δ^g_{n+1} in (4.29). We have

$$J_\alpha (x) - J_\alpha (y) - (J'_\alpha (y), x - y) = \int_0^1 (J'_\alpha (y + \theta (x - y)) - J'_\alpha (y), x - y) \, d\theta.$$

Hence, by (2.10)

$$|J_\alpha (x) - J_\alpha (y) - (J'_\alpha (y), x - y)| \leq C \|x - y\|^2, \forall x, y \in V_\rho. \tag{4.30}$$

Substituting in (4.30) $x := P_n x_\alpha, y := x_\alpha$ and using (2.13), we obtain

$$J_\alpha (P_n x_\alpha) - J_\alpha (x_\alpha) \leq C \|P_n x_\alpha - x_\alpha\|^2 = C \|(I - P_n) x_\alpha\|^2. \tag{4.31}$$

On the other hand, since $J_\alpha(x_n) \geq J_\alpha(x_\alpha)$, then using (3.2), Theorem 3.1, (2.14) and (4.31), we obtain

$$C\|(I-P_n)x_\alpha\|^2 \geq J_\alpha(P_n x_\alpha) - J_\alpha(x_\alpha) \geq J_\alpha(P_n x_\alpha) - J_\alpha(x_n) \geq (J'_\alpha(x_n), P_n x_\alpha - x_n) + \frac{\alpha}{2}\|x_n - P_n x_\alpha\|^2 = \frac{\alpha}{2}\|x_n - P_n x_\alpha\|^2. \quad (4.32)$$

Since by (2.3) and (4.4) $M_n \subset M_{n+1}$ and $M_n \subset G_{n+1}$, then $\|(I-P_{n+1})x_\alpha\| \leq \|(I-P_n)x_\alpha\|$ and $\|(I-G_{n+1})x_\alpha\| \leq \|(I-P_n)x_\alpha\|$. On the other hand, two inequalities, similar with (4.32), can be proven similarly via replacing the pair $(\|x_n - P_n x_\alpha\|, \|(I-P_n)x_\alpha\|)$ first with the pair

$$(\|x_{n+1} - P_{n+1}x_\alpha\|, \|(I-P_{n+1})x_\alpha\|)$$

and then with the pair

$$\left(\left\|x^g_{n+1} - P_{G_{n+1}}x_\alpha\right\|, \|(I-G_{n+1})x_\alpha\|\right).$$

Hence,

$$\frac{\alpha}{2}\left(\|x_{n+1} - P_{n+1}x_\alpha\|^2 + \left\|x^g_{n+1} - P_{G_{n+1}}x_\alpha\right\|^2\right) \leq C\|(I-P_n)x_\alpha\|^2. \quad (4.33)$$

Thus, (4.10), (4.32) and (4.33) imply the following three inequalities

$$\Delta_n, \Delta_{n+1}, \Delta^g_{n+1} \leq C\alpha^{-1/2}\|(I-P_n)x_\alpha\|.$$

Substitution of these three in (4.29) leads to (4.14). □

It is assumed in Theorem 4.1 that the vector $J'_\alpha(x_n)$ can be calculated exactly. In the computational practice, however, this vector is calculated with an error and the minimization process on $V_\rho \cap M_n$ is usually stopped at such a point $\overline{x}_n$ for which the norm $\|P_n J'_\alpha(\overline{x}_n)\|$ is sufficiently small, although still non-zero. These considerations are reflected in Theorem 4.2, which establishes (1.1).

Theorem 4.2. *Assume that the Frechet derivative $J'_\alpha(x), x \in V_\rho$ is calculated with a small error $\tau \in [0,1)$. In other words, for any point $x \in V_\rho$ one actually calculates the vector $S_\alpha(x) \in H$ and $\|J'_\alpha(x) - S_\alpha(x)\| \leq \tau$, $\forall x \in V_\rho$. Let $\overline{n}$ be the integer of Assumption 4.1. Suppose that for any subspace M_k with $k \geq \overline{n}$ the minimization process of the functional $J_\alpha(x)$ on the set $V_\rho \cap M_k$ is stopped at such a point $x_{k,\tau}$ that $\|P_{M_k}S_\alpha(x_{k,\tau})\| \leq \tau$. Let a_1 be the number from (2.1). Consider the function of spatial variables $S_{n,\tau}(y) := S_\alpha(x_{n,\tau}), y \in \sigma$. Assume that there exists a number $r_n \in (\widetilde{r}, 1)$ such that*

$$C\left(AK\frac{h_n}{\sqrt{\alpha}} + \sqrt{\alpha\tau} + \frac{\tau}{\alpha}\right) \leq (r_n - \widetilde{r})\|S_\alpha(x_{n,\tau})\|, \quad (4.34)$$

$$a_1 CK\sqrt{\alpha}\left\|\nabla\widetilde{S}(y)\right\|_{L_\infty(\widetilde{\sigma})} < \frac{r_n - \widetilde{r}}{8N_3}\|S_\alpha(x_{n,\tau})\|^{3/2}. \quad (4.35)$$

Let $\delta_1 = \delta_1(\mu_1, \mu_2, N_1, N_2)$ be the number defined in Theorem 3.1. Then there exists a sufficiently small number $\delta_2 \in (0, \delta_1]$ and a subspace $M_{n+1} \subseteq H, M_n \subset M_{n+1}$ such that if $\delta \in (0, \delta_2]$, then the following relaxation property holds

$$\|x_{n+1,\tau} - x_\alpha\| \leq r_n\|x_{n,\tau} - x_\alpha\|. \quad (4.36)$$

If at least one of inequalities (4.34), (4.35) is invalid, then the mesh refinement process should be stopped. If $\tau = 0$, *then the above holds with the replacement of the pair* $\{S_\alpha(x_{n,\tau}), x_{n,\tau}\}$ *by the pair* $\{J'_\alpha(x_n), x_n\}$. *Let* $r \in (\widetilde{r}, 1)$ *be the maximal value of corresponding numbers* r_n *for a certain finite number of such mesh refinements. Then (4.36) is valid with the replacement of* r_n *with* r, *which turns (4.36) into (1.1).*

Remark 4.1. Although both constants C and N_3 depend on numbers N_1, N_2 introduced in (2.7), the inequality (4.35) makes sense, since these constants can be explicitly estimated via N_1, N_2. The latter would turn both inequalities (4.34), (4.35) in more explicit forms. Following a common tradition of the PDE theory, we are not providing such explicit estimates for brevity only. The same is true for Theorem 7.4 (section 7) with respect to numbers C_2 and N_4.

Proof of Theorem 4.2. Since $\alpha = \delta^{\mu_2}$, then by (2.12) one can choose

$\delta_2 = \delta_2(\mu_1, \mu_2, N_1, N_2) \in (0, \delta_1]$ so small that $C/\alpha \geq 2, \forall \delta \in (0, \delta_2]$. Hence, by (4.34) we can assume that

$$\frac{\|S_\alpha(x_{n,\tau})\|}{2} \geq \tau. \tag{4.37}$$

Using (4.2), we obtain

$$\tau\|x_{n,\tau} - x_n\| \geq (S_\alpha(x_{n,\tau}), x_{n,\tau} - x_n) = (S_\alpha(x_{n,\tau}) - J'_\alpha(x_{n,\tau}), x_{n,\tau} - x_n)$$
$$+ (J'_\alpha(x_{n,\tau}) - J'_\alpha(x_n), x_{n,\tau} - x_n) \geq \frac{\alpha}{2}\|x_{n,\tau} - x_n\|^2 - \tau\|x_{n,\tau} - x_n\|.$$

Hence, $\|x_{n,\tau} - x_n\| \leq 4\tau/\alpha$. Using (2.10) and (4.34), we obtain

$$\|J'_\alpha(x_n)\| = \|J'_\alpha(x_{n,\tau}) - (J'_\alpha(x_{n,\tau}) - J'_\alpha(x_n))\| \geq \|J'_\alpha(x_{n,\tau})\| - \|J'_\alpha(x_{n,\tau}) - J'_\alpha(x_n)\|$$
$$\geq \frac{\|S_\alpha(x_{n,\tau})\|}{2} - N_3\frac{4\tau}{\alpha} \geq C\|S_\alpha(x_{n,\tau})\| > 0. \tag{4.38}$$

Similarly (4.34) and (4.37) lead to

$$\|J'_\alpha(x_n)\| \leq \|J'_\alpha(x_{n,\tau})\| + \|J'_\alpha(x_{n,\tau}) - J'_\alpha(x_n)\| \tag{4.39}$$
$$\leq \|S_\alpha(x_{n,\tau})\| + \tau + N_3\frac{4\tau}{\alpha} \leq C\|S_\alpha(x_{n,\tau})\|,$$

where the constant C is different from one in (4.38).

It follows from (4.38) that (4.3) holds, which implies in turn the existence of such a subspace M_{n+1} that (4.13) is valid. Hence, the point $x_{n+1,\tau}$ exists and $\|x_{n+1,\tau} - x_n\| \leq 4\tau/\alpha$. Hence, by (4.14)

$$\|x_{n+1,\tau} - x_\alpha\| \leq \widetilde{r}\|x_{n,\tau} - x_\alpha\| + \frac{8\tau}{\alpha} + C\frac{\|(I - P_{n+1})x_\alpha\|}{\sqrt{\alpha}} \tag{4.40}$$
$$+C\sqrt{\alpha}\frac{\|(I - P_{n+1})J'_\alpha(x_n)\|^{1/2}}{\|J'_\alpha(x_n)\|^{1/2}}$$

Since $h_{n+1} \leq h_n$, then by (2.4) $\|(I - P_{n+1})x_\alpha\| \leq K\|\nabla x_\alpha\|_{L_2(\sigma)} h_n \leq AKh_n$. Hence, by (4.39) and (4.40)

$$\|x_{n+1,\tau} - x_\alpha\| \leq \widetilde{r}\|x_{n,\tau} - x_\alpha\| + CAK\frac{h_n}{\sqrt{\alpha}} + \frac{8\tau}{\alpha} \tag{4.41}$$
$$\frac{C\sqrt{\alpha\tau}}{\|S_\alpha(x_{n,\tau})\|^{1/2}} + C\sqrt{\alpha}\frac{\|(I - P_{n+1})S_\alpha(x_{n,\tau})\|^{1/2}}{\|S(x_{n,\tau})\|^{1/2}}.$$

By (2.10), (2.13) and (4.37)

$$\|x_{n,\tau}-x_\alpha\| \geq \frac{\|J'_\alpha(x_{n,\tau})\|}{N_3} \geq \frac{\|S(x_{n,\tau})\|-\tau}{N_3} \geq \frac{\|S(x_{n,\tau})\|}{2N_3}.$$

By (4.34) we can assume that

$$CAK\frac{h_n}{\sqrt{\alpha}} + C\sqrt{\alpha\tau} + \frac{8\tau}{\alpha} \leq \frac{(r_n-\widetilde{r})}{4N_3}\|S(x_{n,\tau})\|. \tag{4.42}$$

Suppose that

$$C\sqrt{\alpha}\,\|(I-P_{n+1})\,S_\alpha(x_{n,\tau})\|^{1/2} \leq \frac{r_n-\widetilde{r}}{4N_3}\|S_\alpha(x_{n,\tau})\|^{3/2}. \tag{4.43}$$

Then (4.41) and (4.42) imply that (4.36) holds for M_{n+1}. So, we now construct the subspace M_{n+1}. Let $\widetilde{\sigma} \subseteq \sigma$ be a subdomain in which one wants to refine the mesh and suppose that the mesh is not refined in $\sigma\diagdown\widetilde{\sigma}$. By refining the mesh in $\widetilde{\sigma}$ and not refining it in $\sigma\diagdown\widetilde{\sigma}$, one obtains the target subspace M_{n+1}. If $meas\,(\sigma\diagdown\widetilde{\sigma})$ is not too small, then one obtains a local mesh refinement. We have $P_{n+1}S_\alpha(x_{n,\tau}) = S_{n,\tau}(y)$ for $y \in \sigma\diagdown\widetilde{\sigma}$. Hence,

$$\|(I-P_{n+1})\,S_\alpha(x_{n,\tau})\|^{1/2} \leq \|(I-P_{n+1})\,S_{n,\tau}(y)\|^{1/2}_{L_2(\widetilde{\sigma})} + \|S_{n,\tau}(y)\|^{1/2}_{L_2(\sigma\diagdown\widetilde{\sigma})}. \tag{4.44}$$

Since the limiting case of $\widetilde{\sigma}$ is simply $\widetilde{\sigma}=\sigma$, then one can always choose $\widetilde{\sigma}$ such that

$$C\sqrt{\alpha}\,\|S_{n,\tau}(y)\|^{1/2}_{L_2(\sigma\diagdown\widetilde{\sigma})} \leq \frac{r_n-\widetilde{r}}{8N_3}\|S_\alpha(x_{n,\tau})\|^{3/2}. \tag{4.45}$$

Since $S_\alpha(x_{n,\tau}) \in H$, then $S_{n,\tau}(y) \in H^1(\sigma)$ and $\partial_{y_i}S_{n,\tau}(y) \in L_\infty(\sigma)$. Let $\widetilde{h}_{n+1}$ be the maximal mesh size for the new mesh in $\widetilde{\sigma}$. Then by (2.4)

$$\|(I-P_{n+1})\,S_{n,\tau}(y)\|_{L_2(\widetilde{\sigma})} \leq K\,\|\nabla S_{n,\tau}(y)\|_{L_\infty(\widetilde{\sigma})}\widetilde{h}_{n+1}. \tag{4.46}$$

By (4.35) we can choose $\widetilde{h}_{n+1} \in (a_1,1)$ such that

$$CK\widetilde{h}_{n+1}\sqrt{\alpha}\,\|\nabla S_{n,\tau}(y)\|_{L_\infty(\widetilde{\sigma})} \leq \frac{r_n-\widetilde{r}}{8N_3}\|S_\alpha(x_{n,\tau})\|^{3/2}. \tag{4.47}$$

Estimates (4.44)-(4.47) imply (4.43), which in turn leads to (4.36). □

There is no point to have errors or parameters in calculations less than the level of error δ in the data. Hence, assuming that conditions of Theorem 4.2 hold, we now show the existence of an interval for the number μ_2 in (2.11b), which guarantees that one indeed can choose parameters h_n,τ satisfying above conditions and such that $h_n,\tau >> \delta$ for $\delta \in (0,\delta_2]$. Estimating the right hand side of (4.34) from the above and assuming that $\tau<\alpha$, we obtain, $(r_n-\widetilde{r})\,\|S_\alpha(x_{n,\tau})\| \leq C\alpha^3$. Hence, (4.34), (4.35) and (2.12) imply that one should have $CAKh_n < \alpha^{3.5} = \delta^{3.5\mu_2}, \tau \leq C\alpha^5 = C\delta^{5\mu_2}$. The first of these inequalities is stronger than (4.1). Hence, if $\mu_2 \in (0,1/5)$, then one can always choose numbers h_n,τ such that $h_n,\tau >> \delta$ for $\delta \in (0,\delta_2]$ and (4.34) holds. The same is true for $\widetilde{h}_{n+1}$ in (4.46), provided that $a_1 >> \delta$.

Recommendation for the mesh refinement. By refining the mesh in $\widetilde{\sigma}$, one actually decreases the value of $\|(I-P_{n+1})\, S_{n,\tau}(y)\|_{L_2(\widetilde{\sigma})}$ and therefore "paves the way" for the validity of the relaxation estimate (4.36). Hence, estimates (4.44)-(4.46) indicate that the mesh should be refined in such a subdomain $\widetilde{\sigma}$ of σ in which values of $|S_{n,\tau}(y)|$ are close to $\max_{\overline{\sigma}} |S_{n,\tau}(y)|$, and it should not be refined in subdomains where these values are rather low. This is exactly what is done in section 8 as well as in the past publications [5-7,17].

Remark 4.2. While Theorem 4.2 establishes the existence of such a subspace M_{n+1} that relaxation property (4.36) is valid, one can pose the question on how to computationally decide whether this subspace exists. A simple recipe for this follows from (4.36) and we actually use this approach in our computations in section 8, also see [5-7]. Namely, having found the point $x_{n,\tau}$, one should refine the mesh and minimize the functional J_α on the refined mesh. It follows from (4.36) that if the change in the resulting solution is significant compared with the previous mesh, then the subspace M_{n+1} exists, it is represented by this new mesh and the mesh refinement process should be continued. Otherwise it should be stopped.

5. The Coefficient Inverse Problem. In this section we state our Coefficient Inverse Problem and outline the globally convergent numerical method of [8] for it. We refer to [8] for more details about this method. In addition, we outline in subsection 5.2 some discrepancies between our theory and numerical implementation. Consider the Cauchy problem for the hyperbolic equation

$$c(x)u_{tt} = \Delta u \text{ in } \mathbb{R}^m \times (0,\infty)\,, m = 2,3, \tag{5.1}$$

$$u\,(x,0) = 0, u_t\,(x,0) = \delta\,(x-x_0)\,. \tag{5.2}$$

Equation (5.1) governs a wide range of applications, including, e.g. propagation of acoustic and electromagnetic waves. In the acoustical case $1/\sqrt{c(x)}$ is the sound speed. In the 2-D case of EM waves propagation in a non-magnetic medium, the dimensionless coefficient is $c(x) = \varepsilon_r(x)$, where $\varepsilon_r(x)$ is the spatially distributed dielectric constant of the medium, see [11], where this equation was derived from Maxwell's equations in the 2-D case. Let d_1 and d_2, be two positive numbers, $d_1 < d_2$. We assume that the coefficient $c\,(x)$ of equation (5.1) is such that

$$c\,(x) \in [d_1, d_2]\,, c\,(x) = d_1 \text{ for } x \in \mathbb{R}^m \diagdown \Omega, \tag{5.3}$$

$$c \in C^2\left(\mathbb{R}^3\right) \tag{5.4}$$

Coefficient Inverse Problem (CIP). Let $\Omega \subset \mathbb{R}^m, m = 2,3$ *be a convex bounded domain with the boundary* $\partial\Omega \in C^3$. *Suppose that the coefficient* $c\,(x)$ *satisfies conditions (5.3) and (5.4), where the numbers* d_1 *and* d_2 *are given. Assume that the function* $c\,(x)$ *is unknown in* Ω. *Determine the function* $c\,(x)$ *for* $x \in \Omega$, *assuming that the following function* $g\,(x,t)$ *is known for a single source position* $x_0 \notin \overline{\Omega}$

$$u\,(x,t) = g\,(x,t)\,, \forall\,(x,t) \in \partial\Omega \times (0,\infty)\,. \tag{5.5}$$

The reason why we assume here that the source $x_0 \notin \overline{\Omega}$ is that we do not want to deal with singularities near the source location, see an applied scenario for this in, e.g. [2]. In applications the assumption $c\,(x) = d_1$ for $x \in \mathbb{R}^3 \diagdown \Omega$ means that the target coefficient $c\,(x)$ has a known constant value outside of the domain of interest Ω. Since we do not impose any "smallness" conditions on numbers d_1 and d_2, the numerical method is not a locally convergent one. The function $g\,(x,t)$ models time dependent measurements of the wave field at the boundary of the domain of interest. Practical measurements are performed at a number of detectors, of course. In this case the function

$g(x,t)$ can be obtained via one of standard interpolation procedures, which is outside the scope of this publication. Uniqueness theorem for this inverse problem is a long standing and well known open question, which is addressed positively only in the case when the $\delta-$ function in (5.2) is replaced with a function, which is non vanishing in the entire domain $\overline{\Omega}$ [18,19]. It is an opinion of the authors that it is still worthy to develop numerical methods for this CIP because of applications.

5.1. Outline of the globally convergent numerical method of [8]. Let the function $w(x,s)$ be the Laplace transform of the function $u(x,t)$ with respect to t with the parameter $s > \underline{s} = const. > 0$. We call s "pseudo frequency". One can prove that $w(x,s) > 0$. Let $q(x,s) = \partial_s\left[s^{-2}\ln w(x,s)\right]$. The function q solves the following boundary value problem for a nonlinear integral differential equation in which the unknown coefficient is not present

$$\Delta q - 2s^2\nabla q\cdot\int\limits_s^{\overline{s}}\nabla q(x,\tau)\,d\tau + 2s\left[\int\limits_s^{\overline{s}}\nabla q(x,\tau)\,d\tau\right]^2 \tag{5.6}$$

$$+2s^2\nabla q\nabla V - 2s\nabla V\cdot\int\limits_s^{\overline{s}}\nabla q(x,\tau)\,d\tau + 2s(\nabla V)^2 = 0, q\mid_{\Omega} = \psi(x,s), (x,s)\in\partial\Omega\times[\underline{s},\overline{s}].$$

where the function ψ is generated by the function g in (5.5). Here $\overline{s}$ is the truncation pseudo frequency, it is one of regularization parameters here and it is assumed to be large. Numbers $\underline{s}$ and $\overline{s}$ should be chosen in numerical experiments. The truncation of integrals at a large value of the pseudo frequency $\overline{s}$ is similar to a routine truncation of high frequencies in science and engineering, and so our truncation is natural in this sense. In (5.6) $V(x,\overline{s}) = \overline{s}^{-2}\ln w(x,\overline{s})$ is the so-called "tail" function, and it is unknown. The presence of s-integrals as well as of the tail function implies the nonlinearity and thus, leads to the main difficulty of the globally convergent stage of our method.

One can prove that, under certain conditions,

$$|V(x,\overline{s})|_{2+\gamma} = O\left(\frac{1}{\overline{s}}\right), \overline{s}\to\infty, \tag{5.7}$$

Here $|\cdot|_{k+\gamma}$ is the norm in the Hōlder space $C^{k+\gamma}\left(\overline{\Omega}\right)$. Although (5.6) implies that the tail is small for large $\overline{s}$, it was found in numerical experiments in section 8 that resulting solutions have a better quality if we approximate the tail via the procedure described below, rather than simply neglect it. Equation (5.6) has two unknown functions q and V. The reason why we can accurately approximate both these functions is that we treat them differently, see below.

We consider a layer stripping procedure with respect to s partitioning the interval $[\underline{s},\overline{s}]$ into N small subintervals with the step size $\kappa = s_{n-1} - s_n, \underline{s} = s_N < s_{N-1} < ... < s_0 = \overline{s}$. Approximate the function $q(x,s)$ as a piecewise constant function with respect to s, $q(x,s) = q_n(x)$ for $s\in[s_n, s_{n-1})$. Let $\mathbb{C}_{n,\lambda}(s) = \exp[\nu(s - s_{n-1})]$ be the s-dependent Carleman Weight Function (CWF), where $\nu > 1$ is a large parameter, which is chosen in numerical experiments. Multiplying both sides of equation (5.6) by $\mathbb{C}_{n,\lambda}(s)$ and integrating over $[s_n, s_{n-1})$, we obtain the following finite sequence of nonlinear second order elliptic equations for functions $q_n(x)$ with Dirichlet boundary conditions $\psi_n(x)$, which are derived from the function $\psi(x,s)$,

$$L_n(q_n) \; : \; = \Delta q_n - A_{1,n}\left(h\sum_{i=1}^{n-1}\nabla q_i\right)\nabla q_n + A_{1n}\nabla q_n \nabla V_n - \varkappa q_n \tag{5.8}$$
$$= B_n(\nabla q_n)^2 - A_{2,n}h^2\left(\sum_{i=1}^{n-1}\nabla q_i(x)\right)^2 + 2A_{2,n}\nabla V_n\left(h\sum_{i=1}^{n-1}\nabla q_i\right) - A_{2,n}(\nabla V_n)^2,$$
$$q_n \mid_{\partial\Omega} = \psi_n(x), n = 1, ..., N.$$

Here $A_{1,n}, A_{2,n}, B_n$ are certain numbers depending on ν, κ, n and $\varkappa > 0$ is a small parameter of ones choice. We use in (5.8) V_n instead of V for convenience of notations, see below in this paragraph. It is important that $\lim_{\nu\to\infty} B_n = 0$ uniformly for all n due to the presence of the CWF. Hence, the presence of the CWF with $\nu >> 1$ mitigates the influence of the nonlinear term $(\nabla q_n)^2$, which enables us to solve the boundary value problem for each q_n iteratively via solving a linear elliptic problem on each step. Still, the computational experience shows that we cannot take ν exceedingly large, which would effectively turn equations (5.8) into linear ones. Starting from $n = 1$, we solve problems (5.8) sequentially with respect to n. For each n we have inner iterations with respect to the tail function and calculate functions $q_{n,i}$ until convergence occurs. We set $q_0 := 0$. The first approximation $V_{1,1}$ for the tail was $V_{1,1} \equiv 0$ in [8], and in section 8 we use $V_{1,1}(x,\overline{s}) = \overline{s}^{-2}\ln w_0(x,\overline{s})$, where $w_{d_1}(x,\overline{s})$ is the Laplace transform of the solution of the problem (5.1), (5.2) for the case $c_0(x) \equiv d_1$. Substituting $V_{n,1} \in C^{2+\gamma}(\mathbb{R}^m)$ in (5.8) for V_n, we find the first approximation $q_{n,1} \in C^{2+\gamma}(\overline{\Omega})$ for q_n via solving the boundary value problem (5.8). This is our inner iteration, in which we set $(\nabla q_n)^2 := (\nabla q_{n-1})^2$. To find the next approximation for the tail via the outer iteration, we first find the new approximation $c_{n,1} \in C^{\gamma}(\mathbb{R}^m), c_{n,1}(x) = d_1$ in $\mathbb{R}^m \diagdown \Omega$ via a simple backwards calculation. Next, we solve the problem (5.1), (5.2) with $c := c_{n,1}$, calculate the Laplace transform $w_{n,1}$ and set $V_{n,2}(x,\overline{s}) = \overline{s}^{-2}[\ln w_{n,1}(x,\overline{s})]$. Then we find a new approximation $q_{n,2}$ for q_n, etc.. Suppose that convergence of inner iterations occurs at q_{n,m_n}. Then we set $(q_{n,m_n}, c_{n,m_n}, V_{n,m_n}) := (q_n, c_n, V_{n+1,1}) \in C^{2+\gamma}(\overline{\Omega}) \times C^{\gamma}(\mathbb{R}^m) \times C^{2+\gamma}(\mathbb{R}^m)$, where $c_n(x) = d_1$ in $\mathbb{R}^m \diagdown \Omega$, and repeat the above process for $n := n+1$. The convergence for both $q_{n,i}$ (with respect to i) and q_n is evaluated via evaluating the residuals at a part of the boundary, see section 8. We have added the term $-\varkappa q_n$ to the left hand side of equation (5.8) to improve the stability property of the Dirichlet value problem (5.8) because of the maximum principle [20] (Chapter 3).

Now we briefly outline the global convergence theorem of [8]. Because of (5.7), we assume that $|V_n(x,\overline{s})|_{2+\gamma} \leq \xi$, $\forall n$, where ξ is a small number. Let δ be the level of the error in the data g. Denote $\eta = 2(\kappa + \delta + \varkappa + \xi)$. Hence, η is a small parameter, which, in particular, depends on two regularization parameters of our method, $\varkappa$ and $\overline{s}$. It is important that the second stage of our two stage procedure, the adaptivity, is independent on parameters $\kappa, \varkappa, \xi$, also see the second paragraph of section 1. Let $c^*(x)$ be the exact solution of our CIP. Let $\overline{N} \in [1, N]$ be the total number of functions q_n we have calculated, and $\beta_2 = \kappa\overline{N}$ be the length of the interval $s \in [\overline{s} - \beta_2, \overline{s}]$ covered this way. We assume that the number β_2 is small. Indeed, equations (5.8) are generated by equation (5.6), which contains Volterra integrals in nonlinear terms. It is well known from, e.g. the classic ODE course that one can guarantee a "good" behavior of solutions of such equations only on a small interval. Hence, for a given thickness of the s-layer κ, the number $\overline{N}$ of computed functions c_n is another regularization parameter here, and we set $c_{\overline{N}} := c_{glob}$. This is going along well with one of main ideas of the theory of Ill-Posed Problems, by which the iteration number can serve as a regularization parameter, see pages 156 and 157 in [13]. The following global convergence estimate was proven in [8]

$$|c_n - c^*|_\gamma \leq B_1\eta, \ \forall n \in \left[1, \overline{N}\right], \tag{5.9}$$

with a certain positive constant B_1. Since η is small, then (5.9) guarantees that one obtains a good approximation for the solution for each n. On the other hand, although η is small, we see in our numerical experiments that it is impossible to make it infinitely small in practical computations. The latter two factors pave the way for a subsequent application of the adaptivity technique, which enhances the solution c_{glob}. This technique uses the function c_{glob} as its starting point for a subsequent enhancement.

5.2. Some discrepancies between our theory and computational experiments. Since the above CIP is a quite complex problem with many yet unknown factors, it is hard to anticipate that practical computations would not have any deviations from the theory and also that the theories of two stages of our numerical method would exactly match each other. So, as it is often the case when numerical methods for some complicated nonlinear ill-posed problems are backed up analytically, some discrepancies of this sort take place in this paper. We list them in this subsection. Still, the main point is that, regardless on these discrepancies, the above theory of the globally convergent numerical method still works, including the convergence estimate (5.9).

The 1^{st} discrepancy is that, because of some conveniences of our past computational practice [5-8] and because the main focus of this paper is analytical rather than numerical, we use a generating plane wave instead of the point source in (5.2). We launch this plane wave outside of the target domain Ω. Note that we have used the point source only to justify the asymptotic behavior (5.6), see Lemma 2.1 in [8]. We verify this asymptotic behavior computationally, see subsection 7.2 of [8]. The 2^{nd} discrepancy is that we solve boundary value problems (5.8) in a square, whose boundary is non-smooth. In principle, this might result in singularities near the corners. However, we have not observed such singularities in our computations. Although the boundary of this square is not smooth, as required in subsection 5.1, a modification of the convergence estimate (5.9) can be proven in this case if considering solutions of FEM analogs of (5.8) with a step size bounded from below and applying the Lax-Milgram theorem instead of the Schauder theorem, also see subsection 7.2 of [8].

The 3^{rd} discrepancy is that in order to figure out the Frechet derivative of the Tikhonov functional for the above CIP for the adaptivity, we need to assume that solutions of certain hyperbolic initial boundary value problems are sufficiently smooth. These conditions cannot be guaranteed for the fundamental solution of the hyperbolic equation (5.1). Still, they can be guaranteed if the function $\delta\left(x-x_0\right)$ in (5.2) is replaced with

$$\delta_\theta\left(x-x_0\right) = \left\{ \begin{array}{c} C_\theta \exp\left(\frac{1}{|x-x_0|^2-\theta^2}\right), |x-x_0| < \theta \\ 0, |x-x_0| > \theta \end{array} \right\}, \int\limits_{\mathbb{R}^m} \delta_\theta\left(x-x_0\right) dx = 1,$$

for a sufficiently small $\theta > 0$. Hence, since $x_0 \notin \overline{\Omega}$, then $\delta_\theta\left(x-x_0\right) = 0$ for in $\overline{\Omega}$ as well as in a small neighborhood of $\partial\Omega$ outside of Ω. Here the constant $C_\theta > 0$ is such that the above integral equals unity. We stress that we have introduced this function only to show that the required smoothness of sections 6 and 7 can indeed be ensured for an initial condition, which is close to (5.2) in the distribution sense. The theory of the globally convergent numerical method works for this case, including (5.9).

To consider the Frechet derivative in sections 6,7, we need to vary the coefficient c. To do this, it is convenient to introduce the set of functions $Z = Z(d_1, d_2, \omega, H)$,

$$Z = \left\{ \begin{array}{c} c : c(x) \in H \text{ for } x \in \sigma, c(x) \in (d_1 - \omega, d_2 + \omega) \text{ for } x \in \overline{\Omega}, \\ c \in C(\mathbb{R}^m), c - d_1 \in H^1(\mathbb{R}^m), c(x) = d_1 \text{ in } \mathbb{R}^m \diagdown \sigma \end{array} \right\}, \tag{5.10}$$

where $\omega \in (0, d_1)$ is a small positive number. Because of (5.10), denote Z' the set of all functions $b \in H^1(\mathbb{R}^m) \cap C(\mathbb{R}^m)$ such that

$$b(x) \in H \text{ for } x \in \sigma, \partial_{x_i} b \in L_\infty(\mathbb{R}^m), b(x) = 0 \text{ for } x \in \mathbb{R}^m \diagdown \sigma. \tag{5.11}$$

By (5.10) and (5.11) $c_1 - c_2 \in Z'$, $\forall c_1, c_2 \in Z$. Since H is a finite dimensional space, then we can estimate $C(\overline{\sigma})$ norms via $L_2(\sigma)$ norms, which is important for our derivations in section 6,

$$\|c_1 - c_2\|_{C(\overline{\sigma})} \leq \widetilde{C}_1 \|c_1 - c_2\|_{L_2(\sigma)}, \ \forall c_1, c_2 \in Z; \ \|b\|_{C(\overline{\sigma})} \leq \widetilde{C}_1 \|b\|_{L_2(\sigma)}, \ \forall b \in Z', \tag{5.12}$$

for a positive constant $\widetilde{C}_1 = \widetilde{C}_1(Z)$. Hence, Z can be considered as an open subset of the space $L_2\left(\widetilde{\Omega}\right)$ for any bounded domain $\widetilde{\Omega}$ such that $\sigma \subset \widetilde{\Omega}$. While conditions (5.10), (5.11) are suitable for our theory of sections 2-4, condition (5.4) is violated for functions $c \in Z$, and this is our 4$^{\text{th}}$ discrepancy. Still, we need the adaptivity only on the second stage of our procedure, and also in actual computations of the first stage we obtain the function $c_{glob} \in Z$.

6. Frechet Derivatives. In this section we derive Frechet derivatives of solutions of certain hyperbolic initial boundary value problems for equation (5.1) with respect to the coefficient $c \in Z$. Let $T = const > 0$. Let Ω_1 be a convex bounded domain such that $\Omega \subset \Omega_1, \partial\Omega \cap \partial\Omega_1 = \varnothing, \partial\Omega_1 \in C^\infty$. Denote $Q_T = \Omega_1 \times (0,T), S_T = \partial\Omega_1 \times (0,T)$. We replace in sections 6,7 the $\delta(x - x_0)$ function in (5.2) with the function $\delta_\theta(x - x_0)$ defined in section 5 and assume that $x_0 \notin \overline{\Omega}_1$ and θ is so small that $\delta_\theta(x - x_0) = 0$ in Ω_1. Using results of Chapter 4 of [21], one can prove that the function $u \in C^\infty(\mathbb{R}^m \times [0,T])$. We also assume that there exists a function $a(x) \in C^\infty\left(\overline{\Omega}_1\right)$ such that $\partial_n a \mid_{\partial\Omega_1} = 1, a \mid_{\partial\Omega_1} = 0, a(x) = 0$ in Ω. For example, if $\Omega_1 = \{x : |x| < R\}$, then one can choose $a(x) = \chi(x)\left(|x|^2 - R^2\right)/(2R)$, where the function $\chi \in C^\infty\left(\overline{\Omega}_1\right), \chi \mid_{\partial\Omega_1} = 1$ and $\chi(x) = 0$ in Ω. Although the existence of such functions $a(x)$ might also be established for more general domains, we are not doing this here for brevity.

Since the function $c(x) = d_1$ in $\mathbb{R}^m \diagdown \Omega$ and the constant d_1 is known, we can uniquely solve the resulting initial boundary value problem (5.1), (5.2), (5.5) in the domain $(\mathbb{R}^m \diagdown \Omega) \times (0,T)$. Hence, the following two functions $\widetilde{g}, p$ can be uniquely determined: $\widetilde{g}(x,t) = u \mid_{S_T}$, $p(x,t) = \partial_n u \mid_{S_T}$. We assume that there exist two functions P, G such that

$$P, G \in H^{m+2}(Q_T), \tag{6.1}$$

$$\partial_n P \mid_{S_T} = p(x,t), \partial_n G \mid_{S_T} = \widetilde{g}(x,t) \tag{6.2}$$

$$P(x,t) = G(x,t) = 0 \text{ for } x \in \Omega, \tag{6.3}$$

$$\partial_t^j P(x,0) = 0 \text{ in } \Omega_1, j = 0, ..., 3. \tag{6.4}$$

We impose these assumptions because the function g in (5.5) might be given with an error, meaning that the solution of the initial boundary value problem (5.1), (5.2), (5.5) in $(\mathbb{R}^m \diagdown \Omega) \times (0,T)$ would not necessarily belong to C^∞ then. Next, we consider solutions u and λ of the following initial boundary value problems (6.5) and (6.6) (we do not use a new notation for u for brevity),

$$\begin{aligned} c(x)\, u_{tt} &= \Delta u \text{ in } Q_T, \\ u(x,0) &= u_t(x,0) = 0, \\ \partial_n u \mid_{S_T} &= p(x,t)\,; \end{aligned} \tag{6.5}$$

$$\begin{aligned} c(x)\, \lambda_{tt} &= \Delta \lambda \text{ in } Q_T, \\ \lambda(x,T) &= \lambda_t(x,T) = 0, \\ \partial_n \lambda \mid_{S_T} &= (\widetilde{g} - u\mid_{S_T})\, \zeta_{\varepsilon_2}(t)\,. \end{aligned} \tag{6.6}$$

We call problems (6.5) and (6.6) the "state problem" and the "adjoint problem" respectively. So, (6.6) is the problem with the reversed time, and the boundary condition for it is known only if the function $u\mid_{S_T}$ is known. Hence, for a given coefficient $c(x)$, one should first solve the state problem and next solve the adjoint problem. In (6.6) $\zeta_{\varepsilon_2}(t)$ is a cut-off function, which is introduced to ensure that the compatibility condition is satisfied at $\overline{S}_T \cap \{t = T\}$, where $\varepsilon_2 > 0$ is a small number. So, we choose such a function ζ_{ε_2} that $\zeta_{\varepsilon_2} \in C^\infty[0,T]$, $\zeta_{\varepsilon_2}(t) = 1$ for $\in [0, T-\varepsilon_2]$, $\zeta_{\varepsilon_2}(t) = 0$ for $t \in (T - \varepsilon_2/2, T]$ and $\zeta_{\varepsilon_2}(t) \in [0,1]$ for $t \in (T-\varepsilon_2, T-\varepsilon_2/2]$.

We now remind a result from the classic theory of hyperbolic PDEs with the Neumann boundary condition, see Theorems 5 and 6 in section 7.2 of [16]. We formulate it for our specific needs rather than providing a more general formulation of [16]. Although those Theorems 5 and 6 are proven for the Dirichlet boundary data, extensions of those proofs to the case of Neumann boundary data are rather straightforward, see, e.g. Theorem 5.1 of Chapter 4 in [21]. Consider the following initial boundary value problem

$$\begin{aligned} c(x)\, v_{tt} &= \Delta v + f \text{ in } Q_T, \\ v(x,0) &= v_t(x,0) = 0, \\ \partial_n v \mid_{S_T} &= v^{(n)}(x,t) \in L_2(S_T)\,, \end{aligned} \tag{6.7}$$

where the function $f \in H^k(Q_T)$. By the definition, the weak solution $v \in H^1(Q_T)$ of the problem (6.7) should satisfy the following integral identity (see an analogue for $y = 0$ in §5 of Chapter 4 in [21]) for all functions $z \in H^1(Q_T)$ such that $z(x,T) = 0$

$$\int\limits_{Q_T} (-c(x)\, v_t z_t + \nabla v \nabla z)\, dxdt - \int\limits_{S_T} v^{(n)} z dS - \int\limits_{Q_T} f z dxdt = 0. \tag{6.8}$$

Assume that there exists such an extension $W(x,t)$ of the function $v^{(n)}$ from the boundary S_T in the domain Q_T that $\partial_n W \mid_{S_T} = y(x,t)\,, W \in H^{k+2}(Q_T)\,, W(x,t) = 0$ for $x \in \Omega, \partial_t^j W(x,0) = 0, j = 0,...,k$. In the case $k \geq 2$ we also assume that $\partial_t^i f(x,0) = 0, i = 0..,k-2$. Consider the function $v - W$. Let the function $c \in Z$. Dividing both sides of equation (6.7) by $c(x)$ and using $c^{-1}\Delta v = \nabla \cdot \left(c^{-1} \nabla v\right) - \nabla\left(c^{-1}\right) \nabla v$, we obtain that $v \in H^{k+1}(Q_T)$ and the following estimate holds

$$\|v\|_{H^{k+1}(Q_T)} \leq C_1 \left[\|W\|_{H^{k+2}(Q_T)} + \|f\|_{H^k(Q_T)} \right]. \tag{6.9}$$

Here and below $C_1 = C_1(Z, Q_T, a(x))$ and

$C_2 = C_2 \left(Z, Q_T, \zeta_{\varepsilon_2}, a(x), \|P\|_{H^{m+2}(Q_T)}, \|G\|_{H^{m+2}(Q_T)} \right)$ denote different positive constants depending on listed parameters. Consider functions $\widehat{u} = u - P, \widehat{\lambda} = \lambda - (G - a(x)u)$ and substitute them in (6.5), (6.6). Then, using (6.1)-(6.4), (6.7) and (6.9), we obtain that $u, \lambda \in H^{m+1}(Q_T)$ and

$$\|u\|_{H^{m+1}(Q_T)} \leq C_1 \|P\|_{H^{m+2}(Q_T)}, \ \ \|\lambda\|_{H^{m+1}(Q_T)} \leq C_1 \left(\|P\|_{H^{m+2}(Q_T)} + \|G\|_{H^{m+2}(Q_T)} \right). \quad (6.10)$$

Theorem 6.1. *Let domains Ω, Ω_1 be those specified above. Assume functions P, G satisfying conditions (6.1)-(6.4) exist. Consider the set Z as an open set in the space $L_2(\Omega_1)$ (see (5.12)). Let operators $A_1 : Z \to H^2(Q_T)$ and $A_2 : Z \to H^2(Q_T)$ map every function $c \in Z$ in the solution $u(x,t,c) \in H^2(Q_T)$ of the problem (6.5) and the solution $\lambda(x,t,c) \in H^2(Q_T)$ of the problem (6.6) respectively, where in (6.7) $u \mid_{S_T} := u(x,t,c) \mid_{S_T}$. Let $\varepsilon_3 \in (0,1)$ be a number and the function $c \in Z$ be such that $d_1 - \omega(1-\varepsilon_3) \leq c(x) \leq d_2 + \omega(1-\varepsilon_3)$. Then each of the operators A_1, A_2 has the Frechet derivative at this point c, $A_i'(c)(b) = \widetilde{u}(x,t,c,b), A_2'(c)(b) = \widetilde{\lambda}(x,t,c,b)$, where $b(x) \in Z' \subset L_2(\Omega_1)$ is an arbitrary function. Functions $\widetilde{u}, \widetilde{\lambda} \in H^2(Q_T)$ and they are solutions of the following initial boundary value problems*

$$\begin{aligned} &c(x)\widetilde{u}_{tt} = \Delta \widetilde{u} - b(x) u_{tt}(x,t,c), \text{ in } Q_T, \\ &\widetilde{u}(x,0) = \widetilde{u}_t(x,0) = 0, \partial_n \widetilde{u} \mid_{S_T} = 0; \end{aligned} \quad (6.11)$$

$$\begin{aligned} &c(x)\widetilde{\lambda}_{tt} = \Delta \widetilde{\lambda} - b(x) \lambda_{tt}(x,t,c), \text{ in } Q_T, \\ &\widetilde{\lambda}(x,T) = \widetilde{\lambda}_t(x,T) = 0, \partial_n \widetilde{\lambda} \mid_{S_T} = -\zeta_{\varepsilon_2} \widetilde{u} \mid_{S_T}. \end{aligned} \quad (6.12)$$

Proof. Since $m = 2, 3$, then by the embedding theorem $H^{m+1}(Q_T) \subset C^1(\overline{Q}_T)$ and $\|f\|_{C^1(\overline{Q}_T)} \leq B_2 \|f\|_{H^{m+1}(Q_T)}, \forall f \in H^{m+1}(Q_T)$, where the positive constant $B_2 = B_2(Q_T)$ depends only on the domain Q_T. Let the function $b \in Z'$ be such that $\|b\|_{C(\overline{\Omega}_1)} < \varepsilon_3 \omega$. Then $c + b \in Z$. It follows from (6.10) that

$$\|u\|_{C^1(\overline{Q}_T)} \leq B_2 \|u\|_{H^{m+1}(Q_T)} \leq C_1 \|P\|_{H^{m+2}(Q_T)}. \quad (6.13)$$

By (5.12), (6.9)-(6.11) and (6.13) $\widetilde{u} \in H^2(Q_T)$ and

$$\|\widetilde{u}\|_{H^2(Q_T)} \leq C_1 \|P\|_{H^{m+2}(Q_T)} \cdot \|b\|_{L_2(\sigma)}. \quad (6.14)$$

Consider now the function $w_1(x,t,c,b) = u(x,t,c+b) - u(x,t,c) - \widetilde{u}(x,t,c,b)$. Then $w_1 \in H^2(Q_T)$. Using (6.5), we obtain

$$(c+b) w_{1tt} = \Delta w_1 - b\widetilde{u}_{tt}; \quad w_1(x,0) = w_{1t}(x,0) = 0, \quad \partial_n w_1 \mid_{S_T} = 0.$$

Hence, by (5.12), (6.9) and (6.14) $\|w_1\|_{H^2(Q_T)} \leq C_2 \|b\|_{L_2(\sigma)}^2$. Since the function $\widetilde{u}(x,t,c,b)$ depends linearly on b, then the latter inequality proves that the function $\widetilde{u}$ is indeed the Frechet derivative of the operator A_1 at the point c. Hence, we now can consider $\widetilde{u}(x,t,c,b)$ for all functions $b \in Z'$. The proof for the operator A_2 is similar. □

Theorem 6.2. *Let conditions of Theorem 6.1 be satisfied. Consider the operator $A_3 : Z \to L_2(\sigma)$ defined as*

$$A_3(c)(x) = \int_0^T (u_t \lambda_t)(x,t,c)\,dt, x \in \sigma, \forall c \in Z,$$

where functions $u,\lambda \in H^{m+1}(Q_T)$ *are solutions of initial boundary value problems (6.5), (6.6). Then the function* $A_3(c)(x) \in C(\overline{\Omega})$ *and the operator* A *is Lipschitz continuous,*

$$\|A_3(c_1) - A_3(c_2)\|_{L_2(\sigma)} \leq C_2 \|c_1 - c_2\|_{L_2(\sigma)}, \forall c_1, c_2 \in Z.$$

Proof. Since by (6.10) and the embedding theorem functions $u,\lambda \in C^1\left(\overline{Q}_T\right)$, then $A_3(c) \in C(\overline{\sigma})$. For $i = 1,2$ let $u_i = u(x,t,c_i), \lambda_i = \lambda(x,t,c_i)$. Denote $U = u_1 - u_2, \Lambda = \lambda_1 - \lambda_2$. Then

$$c_1 U_{tt} = \Delta U - (c_1 - c_2) u_{2tt}, \quad U(x,0) = U_t(x,0) = 0, \quad \partial_n U \mid_{S_T} = 0, \tag{6.15}$$

$$c_1 \Lambda_{tt} = \Delta \Lambda - (c_1 - c_2) \lambda_{2tt}, \quad \Lambda(x,T) = \Lambda_t(x,T) = 0, \quad \partial_n \Lambda \mid_{S_T} = -\zeta_{\varepsilon_2} U \mid_{S_T}. \tag{6.16}$$

Hence, using (5.12) and (6.10), we obtain from (6.15) and (6.16)

$$\begin{aligned}\|A_3(c_1) - A_3(c_2)\|_{L_2(\sigma)} &\leq T \|\lambda_1\|_{C^1(\overline{Q}_T)} \|U\|_{H^2(Q_T)} + T \|u_2\|_{C^1(\overline{Q}_T)} \|\Lambda\|_{H^2(Q_T)} \\ &\leq C_2 \|c_1 - c_2\|_{L_2(\sigma)}. \ \square\end{aligned}$$

7. The Tikhonov Functional for the CIP. To apply results of sections 2-4 to our CIP, we specify in this section the Tikhonov functional for this CIP and derive the Frechet derivative for it. We assume in this section that conditions of Theorem 6.1 hold and consider now the set Z as an open subset of the space H (see the paragraph after (5.12)). Recall that the norm in H is $L_2(\sigma)$ and the set $Z \subset H$. Let $c \in Z$ be an arbitrary function and $u = u(x,t,c) \in H^{m+1}(Q_T)$ be the solution of the problem (6.5). Denote $H_1 := L_2(S_T)$. Consider the operator $F: Z \to H_1$ defined as

$$F(c)(x,t) := (\widetilde{g} - u(x,t,c) \mid_{S_T}) \zeta_{\varepsilon_2}(t). \tag{7.1}$$

Since the function $\widetilde{g}(x,t), (x,t) \in S_T$ is actually generated by the data $g(x,t)$ in (5.5) for our CIP, we assume that $\widetilde{g}(x,t) = \widetilde{g}^*(x,t) + \widetilde{g}_\delta(x,t)$, where $\widetilde{g}^*$ corresponds to the exact solution c^* (section 5) and $\widetilde{g}_\delta$ corresponds to the error in the data with a sufficiently small level of error $\delta \in (0,1)$. Hence, $\widetilde{g}^*(x,t) - u(x,t,c^*) \mid_{S_T} \equiv 0$ and by (7.1) $F(c^*) = \widetilde{g}_\delta(x,t)$. Following (6.1)-(6.4), we assume that there exist functions G^*, G_δ, such that

$$\begin{gathered} G^*, G_\delta \in H^{m+2}(Q_T), G = G^* + G_\delta, G^*(x,t) = G_\delta(x,t) = 0, \text{for } x \in \Omega, \\ \partial_n G^* \mid_{S_T} = g^*, \partial_n G_\delta \mid_{S_T} = g_\delta, \ \|G_\delta\|_{H^{m+2}(Q_T)} \leq \delta. \end{gathered} \tag{7.2}$$

Obviously one can take, e.g. $G^* = a(x) u(x,t,c^*)$. Hence, we assume that

$$\|F(c^*)\|_{L_2(S_T)} \leq \delta, \tag{7.3}$$

which is required by (2.6). In addition, by Theorem 6.1 and the trace theorem the operator F has the Frechet derivative $F'(c)(b)$ at every point $c \in Z$,

$$F'(c)(b) = -\zeta_{\varepsilon_2}(t) \widetilde{u}(x,t,c,b) \mid_{S_T}, \forall b \in Z'. \tag{7.4}$$

Lemma 7.1. *Assume that conditions of Theorem 6.1 and consider* Z *is a subset of* H*. Then the Frechet derivative* $F'(c)$ *satisfies the Lipschitz condition*

$$\|F'(c_1) - F'(c_2)\| \leq C_2 \|c_1 - c_2\|_{L_2(\sigma)}, \forall c_1, c_2 \in Z.$$

Proof. For $i=1,2$ denote $u_i=u_i(x,t,c_i)$ and $\widetilde{u}_i=\widetilde{u}_i(x,t,c_i,b)$ solutions of problem (6.13) and (6.15) respectively with $c=c_i$. Similarly with the proof of Theorem 6.2 let $U=u_1-u_2,\widetilde{U}=\widetilde{u}_1-\widetilde{u}_2$. Hence, $U\in H^{m+1}(Q_T),\widetilde{U}\in H^2(Q_T)$. By (6.11)

$$c_1\widetilde{U}_{tt}=\Delta\widetilde{U}-bU_{tt}-(c_1-c_2)\widetilde{u}_{2tt},\quad \widetilde{U}(x,0)=\widetilde{U}_t(x,0)=0,\quad \partial_n\widetilde{U}\mid_{S_T}=0. \tag{7.5}$$

It follows (5.12), (6.9), (6.11) and (6.15) that

$$\|bU_{tt}\|_{L_2(Q_T)}+\|(c_1-c_2)\widetilde{u}_{2tt}\|_{L_2(Q_T)}\le C_2\|c_1-c_2\|_{L_2(\sigma)}\|b\|_{L_2(\sigma)}.$$

Hence, by (6.9), (7.4), (7.5) and the trace theorem

$$\|F'(c_1)(b)-F'(c_2)(b)\|_{H_1}\le C_2\|c_1-c_2\|_{L_2(\sigma)}\|b\|_{L_2(\sigma)}. \quad \square$$

Recall that the function $c_{glob}\in Z$ (subsection 5.2) and consider the Tikhonov functional $Y_\alpha: Z\to\mathbb{R}$ for the operator $F(c)$ in (7.1) (also, see Remark 2.1),

$$Y_\alpha(c)=\frac{1}{2}\|F(c)\|_{H_1}^2+\frac{\alpha}{2}\|c-c_{glob}\|_{L_2(\sigma)}^2, \tag{7.6}$$

In order to find the Frechet derivative $Y_\alpha'(c)$, consider the Lagrange functional $L(c)$,

$$L(c)=Y_\alpha(c)+\int\limits_{Q_T}(-c(x)u_t\lambda_t+\nabla u\nabla\lambda)\,dxdt-\int\limits_{S_T}p\lambda dSdt, \tag{7.7}$$

where functions $u(x,t,c),\lambda(x,t,c)\in H^{m+1}(Q_T)$ are solutions of initial boundary value problems (6.5), (6.6). By (6.5), (6.6) and (6.8) the integral term in (7.7) equals zero. Hence, $L(c)=Y_\alpha(c),\forall c\in Z$. However, it is not straightforward to figure out the analytic expression for $(F'(c))^*F(c)$ for the operator F in (7.1). The latter is required by (2.9) for the calculation of the Frechet derivative $Y_\alpha'(c)$. The reason why $L(c)$ is introduced is that it is easier to calculate its Frechet derivative $L'(c)$ compared with the one of $Y_\alpha(c)$. To obtain the explicit expression for $L'(c)$, we need, similarly with section 6, to vary the function c via considering $c+b\in Z$ for $b\in Z'$ and then to single out the term, which is linear with respect to b. When varying c, we also need to consider respective variations of functions u and λ in (7.7), since these functions depend on c as solutions of state and adjoint problems. And linear, with respect to c, parts of these variations will be functions $\widetilde{u}(x,t,c,b)$, $\widetilde{\lambda}(x,t,c,b)$. Unlike this, the "all-at-once" approach of [5-7], assumes that in (7.7) $L:=\widetilde{L}(c,u,\lambda)$, where functions c,u,λ are treated as mutually independent ones with variations $(b,\overline{u},\overline{\lambda})$ of (c,u,λ) satisfying

$$\overline{u},\overline{\lambda}\in H^1(Q_T),\overline{u}(x,0)=\overline{\lambda}(x,T)=0. \tag{7.8}$$

The resulting expression $\widetilde{L}'(c,u,\lambda)(b,\overline{u},\overline{\lambda})$ is considered as the "all-at-once" Frechet derivative of the Lagrangian $\widetilde{L}(c,u,\lambda)$ rather than the one of the Tikhonov functional $Y_\alpha(c)$. One of assertions of Theorem 7.1 is that these two derivatives are equal to each other.

Theorem 7.1. *Assume that conditions of Theorem 6.1 hold. Then*

$$Y_\alpha'(c)(b)=L'(c)(b)=\int\limits_\Omega\left[\alpha(c-c_{glob})-\int\limits_0^T u_t\lambda_t dt\right]b(x)\,dx,\ \forall c\in Z,\ \forall b\in Z'. \tag{7.9}$$

In particular, since by (5.11) $b(x)=0$ *for* $x\in\mathbb{R}^m\diagdown\sigma$, *then*

$$Y'_\alpha(c)(x)=\alpha(c-c_{glob})(x)-\int_0^T(u_t\lambda_t)(x,t,c)\,dt,\ \ x\in\sigma,\ \forall c\in Z, \tag{7.10}$$

and by Theorem 6.2 $Y'_\alpha(c)\in C(\overline{\sigma})$. *The same expression (7.9) holds for the all-at-once Frechet derivative of the Lagrangian,* $\widetilde{L}'(c,u,\lambda)\left(b,\overline{u},\overline{\lambda}\right)=Y'_\alpha(c)(b),\forall c\in Z,\forall b\in Z'$, *i.e. the all-at-once Frechet derivative of the Lagrangian equals the Frechet derivative of the Tikhonov functional.*

Proof. Considering in (7.7) $L(c+b)-L(c)$, singling out the term, which is linear with respect to b and using (7.4), (7.6) and Theorem 6.1, we obtain

$$\begin{aligned}L'(c)(b)=Y'_\alpha(c)(b)&=\int_\Omega\left[\alpha(c-c_{glob})-\int_0^T u_t\lambda_t dt\right]b(x)\,dx\\&+\int_{Q_T}\left(-cu_t\widetilde{\lambda}_t+\nabla u\nabla\widetilde{\lambda}\right)dxdt-\int_{S_T}p\widetilde{\lambda}dSdt\end{aligned} \tag{7.11}$$

$$+\int_{Q_T}\left(-c\lambda_t\widetilde{u}_t+\nabla\lambda\nabla\widetilde{u}\right)dxdt-\int_{S_T}\left(g-u\mid_{S_T}\right)\zeta_{\varepsilon_2}(t)\,\widetilde{u}dSdt,\forall c\in Z,\forall b\in Z',$$

where $\widetilde{u}$ and $\widetilde{\lambda}$ are solutions of problems (6.11) and (6.12) respectively. Since $\widetilde{u}(x,0)=\widetilde{\lambda}(x,T)=0$, then (6.8), (6.11) and (6.12) imply that second and third lines in (7.11) equal zero, which proves (7.9). Consider now the all-at-once Frechet derivative via considering $\widetilde{L}\left(c+b,u+\overline{u},\lambda+\overline{\lambda}\right)-\widetilde{L}(c,u,\lambda)$ and singling out in this expression the term, which is linear with respect to $\left(b,\overline{u},\overline{\lambda}\right)$. Then we obtain the same expression as in (7.11) where functions $\widetilde{u},\widetilde{\lambda}$ are replaced of with $\overline{u},\overline{\lambda}$. Hence, (6.5)-(6.8) and (7.8) imply that second and third lines in the latter expression equal zero. □

Remark 7.1. We refer to the earlier work [10] where the Frechet derivative for the Tikhonov functional for the parameter identification problem (which is different from a CIP) was derived, although the proof was not presented: by the rules of that journal. The forward problem in [10] is the Cauchy problem for a hyperbolic equation. A private communication with the author of [10] has revealed that the complete proof was presented in his Ph.D. thesis (1971). Since the Lagrangian was not introduced in [10], then the above equality of two derivatives was not proved in [10].

Now we are ready to reformulate theorems of sections 2-4 for our CIP. To do this, it is convenient to consider another set $Z_1\subset H$, which is the set of restrictions of all functions $c\in Z$ on the polygonal domain σ. Hence, when considering solutions u and λ of state and adjoint problems in the functional $Y_\alpha(c)$, we assume that the coefficient $c\in Z$. However, when subsequently applying the theory of sections 2-4 to this functional, we assume that $c\in Z_1$. Since we always work with the gradient $Y'_\alpha(c)$ in that theory, then (7.9) and (7.10) imply that this theory is not affected this way.

There is no guarantee that the function $Y'_\alpha(c)\in H$, because of the integral term in (7.10). Hence, in order to apply the theory of section 4, we should consider the function $PY'_\alpha(c)$ instead of $Y'_\alpha(c)$, where $P:L_2(\sigma)\to H$ is the operator of the orthogonal projection of $L_2(\sigma)$ onto H. In practical computations we actually compute the interpolant of $Y'_\alpha(c)$ on the corresponding mesh instead of $PY_{\alpha}^{\prime}(c)$, and this is one of sources of error, see Theorems 7.3 and 7.4. Increasing the smoothness of functions P,G by 1 in (6.1), one can prove that in Theorem 6.2 $A(c) \in C^{1}(\overline{\sigma})$, which leads to an estimate of this error via (2.4). We keep in mind below that $(PY_{\alpha}^{\prime}(c), f) = (Y_{\alpha}^{\prime}(c), f), \forall c \in Z_1, \forall f \in H$. It follows from Lemma 7.1 that there exists a number $N_4 = N_4(C_2) > 0$ such that $\|Y_{\alpha}^{\prime}(c_1) - Y_{\alpha}^{\prime}(c_1)\|_{L_2(\sigma)} \leq N_4 \|c_1 - c_2\|_{L_2(\sigma)}, \forall c_1, c_2 \in Z_1$. Obviously, lemmata and theorems of sections 2 and 3 are applicable now with the natural replacement of the vector $(x^*, x_{glob}, x_{\alpha}, N_3)$ with the vector $(c^*, c_{glob}, c_{\alpha}, N_4)$ and assuming that conditions (2.11), (2.12) hold. Hence, below we still regard, without restating, Assumption 4.1 as a standing one. Also, in Lemma 3.2 and Theorems 3.1, 3.2 we now have $\beta_1 = \beta_1(C_2) \in (0,1), \rho = \beta_1\alpha = \beta_1\delta^{\mu_2}, \delta_1 = \delta_1(\mu_1,\mu_2,C_2), \delta \in (0,\delta_1)$ and

$$V_{(1+\sqrt{2})\delta^{\mu_1}}(c^*) := \left\{ f \in H : \|f - c^*\|_{L_2(\sigma)} < \left(1+\sqrt{2}\right)\delta^{\mu_1} \right\},$$
$$V_{\rho} := \left\{ f \in H : \|f - c_{\alpha}\|_{L_2(\sigma)} < \rho \right\}.$$

In addition, (4.1) holds where A is a given positive constant and $\|\nabla c_{\alpha}\|_{L_{\infty}(\sigma)} \leq A$. The proof of Theorem 7.2 follows immediately from (5.10)-(5.12), (6.1)-(6.4), Lemmata 2.1, 3.2 and Theorems 3.1, 3.2.

Theorem 7.2. *Assume that conditions of Theorem 6.1 hold, functions $c^*, c_{glob} \in Z_1$ and in particular $\|c_{glob} - c^*\|_{L_2(\sigma)} \leq \delta^{\mu_1}$. Then one can choose the number $\delta_2 = \delta_2(\mu_1,\mu_2,C_2) \in (0,\delta_1]$ so small that for $\delta \in (0,\delta_2)$ we have: $V_{(1+\sqrt{2})\delta^{\mu_1}}(c^*) \subset V_{\rho} \subset Z_1$, the functional $Y_{\alpha}(c)$ is strictly convex on V_{ρ} with the strict convexity parameter $\kappa = \alpha/4$ and there exists the unique minimizer c_{α} of $Y_{\alpha}(c)$ on the set $V_{(1+\sqrt{2})\delta^{\mu_1}}(c^*)$ as well as the unique minimizer c_n on the set $\left(\partial\overline{V}_{\rho} \diagdown V_{\rho}\right) \cap M_n$.*

Theorem 7.3. *Let conditions of Theorem 7.2 hold. Suppose that the function $PY_{\alpha}^{\prime}(c) \in H$ is calculated with an error. That is, one calculates the function $S_{\alpha}(c) \in H$ instead of $PY_{\alpha}^{\prime}(c)$ and $\|PY_{\alpha}^{\prime}(c) - S_{\alpha}(c)\|_{L_2(\sigma)} \leq \tau, \forall c \in V_{\rho}$, where the number $\tau \in [0,1)$ is sufficiently small. Suppose that the minimization process of the functional $Y_{\alpha}(c)$ on the set $M_n \cap V_{\rho}$ with $n \geq \overline{n}$ is stopped at such a point $c_{n,\tau}$ that $\tau \leq \|P_n S_{\alpha}(c_{n,\tau})\|_{L_2(\sigma)}/2$. Then the following a posteriori error estimate of the reconstruction of the regularized coefficient holds*

$$\|c_{n,\tau} - c_{\alpha}\|_{L_2(\sigma)} \leq \frac{3}{\alpha}\|S_{\alpha}(c_{n,\tau})\|_{L_2(\sigma)}.$$

In particular, if $\tau = 0$, then

$$\|c_n - c_{\alpha}\|_{L_2(\sigma)} \leq \frac{3}{\alpha}\|PY_{\alpha}^{\prime}(c_n)\|_{L_2(\sigma)} \leq \frac{3}{\alpha}\|Y_{\alpha}^{\prime}(c_n)\|_{L_2(\sigma)}.$$

Proof. Since $Y_{\alpha}^{\prime}(c_{\alpha}) = 0$, then by (3.2) and Theorem 7.2

$$\alpha\|c_{n,\tau} - c_{\alpha}\|_{L_2(\sigma)} \leq 2\|Y_{\alpha}^{\prime}(c_{n,\tau}) - Y_{\alpha}^{\prime}(c_{\alpha})\|_{L_2(\sigma)} \leq 2\|S(c_{n,\tau})\|_{L_2(\sigma)} + 2\tau \leq 3\|S(c_{n,\tau})\|_{L_2(\sigma)}. \square$$

We now reformulate the relaxation Theorem 4.2 for our CIP. This is Theorem 7.4, which immediately follows from Theorems 4.2, 7.2 and 7.3.

Theorem 7.4. Denote $\widehat{r} = \sqrt{1 - \alpha^2(2N_4)^{-2}}$. *Assume that conditions of Theorem 7.3 hold and $\overline{n}$ is the integer of Assumption 4.1. Suppose that for any subspace M_k with $k \geq \overline{n}$ the minimization process of the functional $Y_{\alpha}(c)$ on the set $V_{\rho} \cap M_n$ is stopped at such a point $c_{k,\tau}$ that*

$\|P_{M_k} S(c_{k,\tau})\|_{L_2(\sigma)} \leq \tau$. *Let* $n \geq \overline{n}$ *and* a_1 *be the number from (2.1). Assume that there exists a number* $r_n \in (\widetilde{r}, 1)$ *such that*

$$C_2\left(AK\frac{h_n}{\sqrt{\alpha}} + \sqrt{\alpha\tau} + \frac{\tau}{\alpha}\right) \leq (r_n - \widehat{r})\|S(c_{n,\tau})\|_{L_2(\sigma)}, \tag{7.12}$$

$$a_1 C_2 K\sqrt{\alpha}\|\nabla S(c_{n,\tau})\|_{L_\infty(\widetilde{\sigma})} \leq \frac{r_n - \widehat{r}}{8N_4}\|S(c_{n,\tau})\|^{3/2}_{L_2(\sigma)}. \tag{7.13}$$

Let $\delta_2 \in (0, \delta_1]$ *be the number of Theorem 7.2. Then there exists such a subspace* M_{n+1} *of the space* H *that* $M_n \subset M_{n+1}$ *and for* $\delta \in (0, \delta_2)$ *the following relaxation property holds*

$$\|c_{n+1,\tau} - c_\alpha\|_{L_2(\sigma)} \leq r_n\|c_{n,\tau} - c_\alpha\|_{L_2(\sigma)}. \tag{7.14}$$

If at least one of inequalities (7.12), (7.13) is invalid, then the mesh refinement process should be stopped. If $\tau = 0$, *then the above holds with the replacement of the pair* $\{S(c_{n,\tau}), c_{n,\tau}\}$ *by the pair* $\{PY'_\alpha(c_n), c_n\}$. *Let* $r \in (\widehat{r}, 1)$ *be the maximal value of corresponding numbers* r_n *for a certain finite number of such mesh refinements. Then (7.14) is valid with the replacement of* r_n *with* r, *which turns (7.14) into (1.1).*

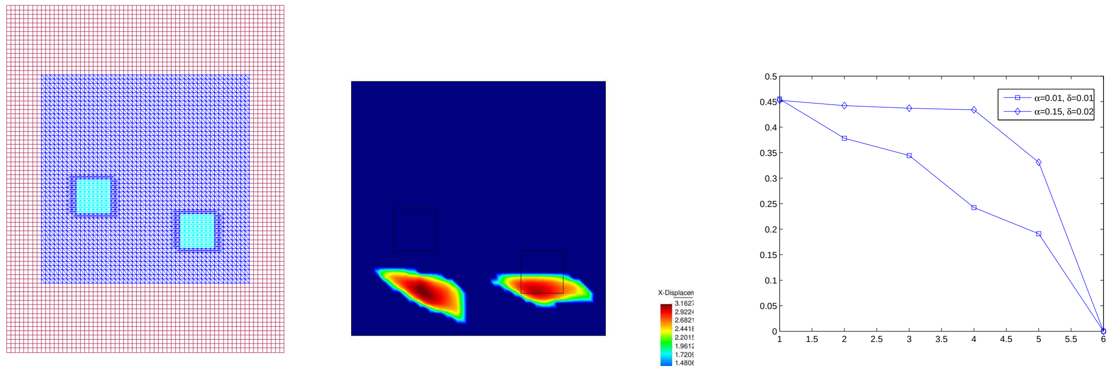
Remark 7.2. In reference to numbers C_2 and N_4 in (7.13), see Remark 4.1. Since the local strict convexity of the functional Y_α on the set V_ρ in combination with Assumption 4.1 implies convergence of gradient-like methods of minimization of Y_α on sets $V_\rho \cap M_n$, a corresponding global convergence theorem for the entire two-stage procedure for the above CIP to the above defined regularized solution c_α can be proven, unlike the current first stage only of [8]. This can be done, provided that the globally convergent stage would be modified for the smooth initial condition $\delta_\theta(x - x_0)$ (see section 5).

8. Numerical Studies.

8.1. Computation of the forward problem. In this paper we work with the computationally simulated data. That is, the data for the CIP are generated by computing the forward problem with the given function $c(x) := c^*(x)$. To solve the forward problem, it is convenient to use the hybrid FEM/FDM method described in [4]. The computational domain for the forward problem is $G = [-4.0, 4.0] \times [-5.0, 5.0]$. This domain is split into a finite element domain $G_{FEM} := \Omega = [-3.0, 3.0] \times [-3.0, 3.0]$ and a surrounding domain G_{FDM} with a structured mesh, $G = G_{FEM} \cup G_{FDM}$, see 8.1-a). The reason of the convenience of the hybrid method is that there is no need to have the unstructured mesh in the domain $G \diagdown \Omega$, since $c(x) = 1$ in this domain. The space mesh in Ω consists of triangles and in G_{FDM} - of squares with the mesh size $\widetilde{h} = 0.125$ both in the overlapping regions and in $G \diagdown \Omega$. At the top and bottom boundaries of G we use first-order absorbing boundary conditions [14]. At the lateral boundaries, mirror boundary conditions allow us to assume an infinite space domain in the lateral direction. The coefficient $c(x)$ is unknown only in the square $\Omega \subset G$,

$$c(x) = \left\{\begin{array}{c} 1 := d_1 \text{ in } G\diagdown\Omega \\ 1 + k(x) \text{ in } \Omega, \\ \widetilde{c} = 4 \text{ in small squares} \end{array}\right\}, \tag{8.1}$$

$$k(x) = \left\{\begin{array}{c} 0.5\sin^2\left(\frac{\pi x_1}{2.875}\right)\sin^2\left(\frac{\pi x_2}{2.875}\right), \text{ for } |x_1|, |x_2| < 2.875 \\ 0 \text{ otherwise, including small squares} \end{array}\right\}.$$



(a) $G = G_{FEM} \cup G_{FDM}$ (b) $c_{7,9}$ (c) $||c_n - c_\alpha||_{L_2}$

FIG. 8.1. *a) The computational domain for the forward problem is the rectangle. The dark blue square is the domain Ω. b) The spatial distribution of the function $c_{7,9}(x) := c_7(x) := c_{glob}(x)$ resulting from the globally convergent stage of our method. The maximal value of this function within imaged inclusions is 3.1 (correct maximal value is 4). Also, $c_{glob}(x) = 1$ outside of imaged inclusions. Hence, it is desirable to enhance the image in two ways: (1) it would be good to increase the value of the function c within imaged inclusions from 3.1 to 4, and (2) it is desirable to move up the location of the left imaged inclusion. This paves the way for the subsequent application of the adaptivity technique. c) Computed norms $\|c_n - c_\alpha\|_{L_2(\sigma)}$ on five (5) adaptively refined meshes including the initial coarse mesh. Two cases are presented: (1) $\varsigma = 0.02 \approx \delta, \alpha = 0.15 \approx \varsigma^{0.48}$ and (2) $\varsigma = \alpha = 0.01$, see explanations in the text. The relaxation property (7.14) is evident from this figure. In the first case the relaxation is more pronounced on the 4 th mesh refinement with $r_4 \approx 0.79$, although it is also clear that $0.95 < r_1, r_2, r_3 < 1$. In the second case $r_1 \approx 0.82, r_2 \approx 0.89, r_3 \approx r_4 \approx 0.71$.*

Hence, (8.1) means that $c(x) = 1$ both near the boundary of the square Ω and outside of this square and $c(x) \geq 1 := d_1$ everywhere. The constant $\widetilde{c}$ characterizes the inclusion/background contrast in small squares. The number 0.5 is the maximal amplitude of the slowly changing background function. We point out that our goal is to image small squares of Figure 8.1 (c) rather than to image the slowly changing background function. Another approach to imaging of small inclusions can be found in, e.g. [2]. The trace of the solution of the forward problem is recorded at the boundary $\partial\Omega$ as the function $g(x,t)$, see (5.5). Next, the coefficient $c(x)$ is "forgotten", and our goal is to reconstruct this coefficient for $x \in \Omega$ from the data $g(x,t)$. The boundary of the domain G is $\partial G = \partial G_1 \cup \partial G_2 \cup \partial G_3$. Here, ∂G_1 and ∂G_2 are respectively top and bottom sides of the largest domain of 8.1-a) and ∂G_3 is the union of left and right sides of this domain. Let $t_1 := 2\pi/\overline{s}, T = 17.8t_1$. We initialize the plane wave $f(t)$ on the top boundary ∂G_1 (also, see subsection 5.2), where $f(t) = 0.1\left(\sin\left(\overline{s}t - \pi/2\right) + 1\right),\ 0 \leq t \leq t_1, f(t) = 0, t \in (t_1, T)$. Hence, it is initialized for $t \in (0, t_1]$ and propagates into G. In the integral of the Laplace transform (subsection 5.1) we integrate for

$t \in (0,T)$. Thus, the forward problem in our numerical test is

$$
\begin{aligned}
c(x)\, u_{tt} - \triangle u &= 0, \quad \text{in } G \times (0,T),\\
u(x,0) &= u_t(x,0) = 0, \text{ in } G,\\
\partial_n u\big|_{\partial G_1} &= f(t), \text{ on } \partial G_1 \times (0,t_1],\\
\partial_n u\big|_{\partial G_1} &= -\partial_t u, \text{ on } \partial G_1 \times (t_1,T),\\
\partial_n u\big|_{\partial G_2} &= -\partial_t u, \text{ on } \partial G_2 \times (0,T),\\
\partial_n u\big|_{\partial G_3} &= 0, \text{ on } \partial G_3 \times (0,T).
\end{aligned}
\tag{8.2}
$$

To see how our algorithm works with the noisy date, we introduce the multiplicative random noise in the data g. thus considering the following function g_ς

$$
g_\varsigma\left(x^i,t^j\right) = g\left(x^i,t^j\right)\left[1+\varsigma\alpha_j\left(g_{\max}-g_{\min}\right)\right], \tag{8.3}
$$

where $x^i \in \partial\Omega, t^j \in [0,T]$ are mesh points at the boundary of the square Ω and in the time interval $[0,T]\,, \alpha_j \in [-1,1]$ is the random number taken from the uniform distribution, $\varsigma \approx \delta$ is the noise level, where $g_{\max}$ and $g_{\min}$ are maximal and minimal values of the function g. However, we have differentiated the Laplace transform $w(x,s)$ with respect to s using the finite difference, because the Laplace transforms smooths out the noise.

8.2. Reconstruction result on the globally convergent stage. In our numerical studies we have used the interval $s \in [\underline{s},\overline{s}] = [6.95, 7.45]$, which is a part of the interval $[6.7, 7.45]$ used in [8]. We have taken its partition step size $\kappa = 0.05$, which means that $N = 10$. We have taken the following values of parameters: $\nu = 20$, $\varkappa_n = 0$ for $n = 1,2$ and $\varkappa_n = 0.0001$ for $n \in [3,10]\,, \varsigma = 0.05$. Thus, the noise level on the first stage of our two stage procedure was 5%. We have solved Dirichlet boundary value problems (5.8) by the FEM. If in our computations we saw that $c_{n,i}(x') \leq 0.5$ for any point $x' \in \overline{\Omega}$, then we have set a new value as $c_{n,i}(x') := 1 = d_1$ in order to ensure that the operator $c_{n,i}(x')\,\partial_t^2 - \Delta$ is a hyperbolic one when solving the forward problem (8.2) with the function $c_{n,i}$, which we need for iterations with respect to tails (subsection 5.1). The latter cut-off procedure prevents us from imaging the slowly changing background, which is not our goal anyway (see subsection 8.1).

To monitor the convergence of our method, we have evaluated norms

$$
F_{n,i} = \frac{\left\| q_{n,i}\,|_{\Gamma_h} - \psi_n \right\|_{L_2(\partial\Omega)}}{\left\|\psi_n\right\|_{L_2(\partial\Omega)}}\,. \tag{8.4}
$$

In (8.4) values of calculated functions $q_{n,i}^k$ are taken at the points h-inside from the lower boundary, where $h = 0.125$. We stop inner iterations with respect to tails (i.e., with respect to i), when either $F_{n,i+1} \geq F_{n,i}$ or $|F_{n,i+1} - F_{n,i}| \leq 0.0005$. One can see from Table 8.1 that norms $F_{n,i}$ decay first with the grow of $n \in [1,7]$. Next, they start to grow at $n = 8$ and grow sharply at $n = 9$. Hence, we stop the globally convergent part at $c_{7,9} := c_7 := c_{glob}$, see label for Table 8.1. Another reason of the growth at $n = 8$ might be that the s-interval covered at $n = 7$ has the length of 0.35, and this might be the limit for the number β_2 of subsection 5.1. Figure 8.1-b) displays the resulting image and its legend explains details.

i	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$
1	1.07519	0.979843	0.957188	0.960068	0.840414	0.799041	0.188793	0.380556	0.563063
2	0.98301	0.978974	0.955977	0.934431	0.763071	0.826884	0.197357	0.397676	1.563063
3	0.98301	0.978574	0.957078	0.931403	0.753745	0.826884	0.203472	0.399297	
4			0.956932	0.931034	0.768198		0.21208	0.400714	
5			0.956501	0.931012	0.768198		0.214845	0.414013	
6			0.955725				0.1983	0.435663	
7			0.955006				0.201933	0.426121	
8			0.954221				0.19723	0.420526	
9			0.953986				0.195233	0.420526	
10			0.953287				0.199145		
11			0.952856				0.199145		
12			0.952856						

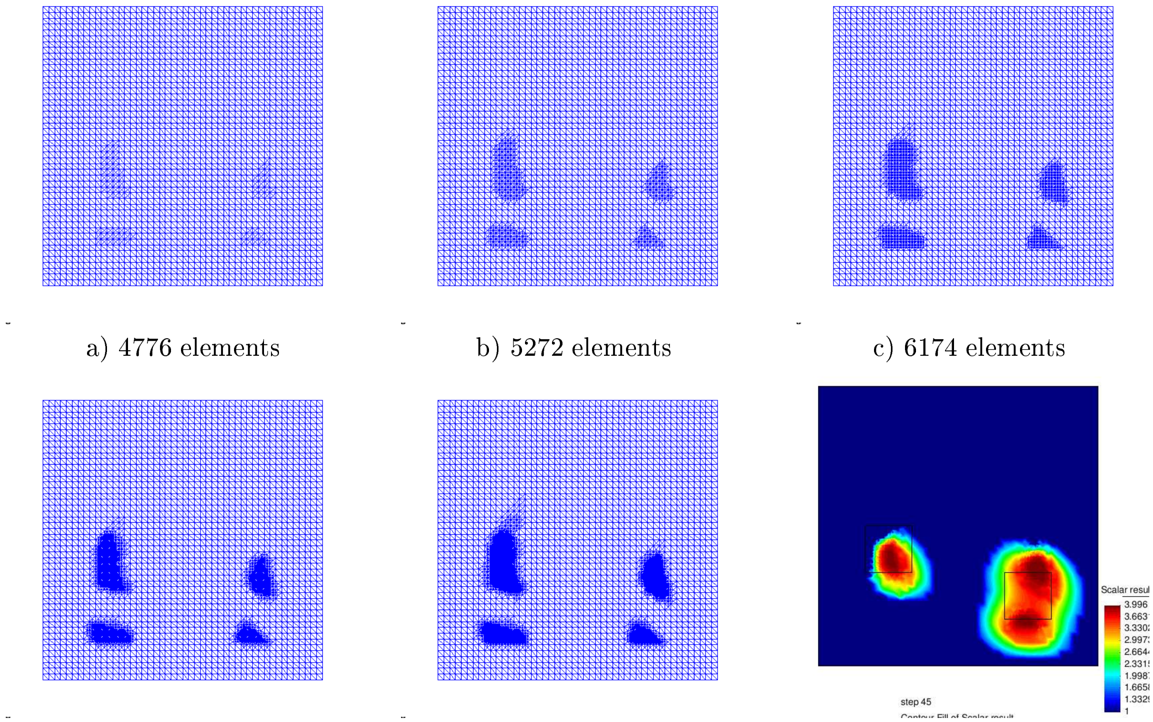
TABLE 8.1

Values of numbers $F_{n,i}$ in (8.4). One can see that they generally decrease until $n = 7$. And stabilization with respect to i is also observed. Next, they start to increase at $n = 8$ and grow sharply at $n = 9$. Therefore, we stop the globally convergent part at $n := \overline{N} = 7$ and set $c_{7,9} := c_7 := c_{glob}$. This is going along well with one of basic ideas of the theory of Ill-Posed Problems by which the iteration number can be one of regularization parameters, see pages 156, 157 of [13].

8.3. The second stage of the two-stage procedure. In this second stage of our two stage procedure we use the adaptivity technique, which is the main focus of the analytical study of this publication. We take the above function c_{glob} (Figure 8.1-b)) as the first guess for our method. On each mesh we use the quasi-Newton method to find an approximate solution of the equation $(Y'_\alpha)^I(c) = 0$, where the function $Y'_\alpha(c)$ is given in (7.10), see [7] for details of our implementation of the quasi-Newton method. Here the superscript "I" stands for the standard interpolant of the function $Y'_\alpha(c)$ on this mesh (see section 7 for some details). On each mesh we stop iterations of the quasi-Newton method on such a function $c^{(n)}$ that either $\left\|(Y'_\alpha)^I\left(c^{(n)}\right)\right\|_{L_2(\sigma)} \leq 10^{-5}$ or these norms are stabilized. Usually norms are stabilized and the resulting norm $\left\|(Y'_\alpha)^I\left(c^{(n)}\right)\right\|_{L_2(\sigma)} \neq 0$. Hence, we refine the mesh in such subdomains of σ that

$$\left|(Y'_\alpha)^I\left(c^{(n)}\right)(x)\right| \geq \upsilon \max_{\overline{\sigma}} \left|(Y'_\alpha)^I\left(c^{(n)}\right)(x)\right|,$$

where $\upsilon = 0.6$ was chosen in numerical experiments. This corresponds to the mesh refinement recommendation presented after the proof of Theorem 4.2. In our case is the domain $\Omega_1 = \{x_2 > -3\} \cap G$ (section 6). We do not use the cut-off function $\zeta_{\varepsilon_2}(t)$ in (6.6) and (7.1), since we have observed computationally that $u(x,T) \approx 0$. Since the convergence estimate (5.9) guarantees that the correct solution is not far from c_{glob}, then we use some constrains for the reconstructed coefficient. We impose these constraints using the solution obtained on the globally convergent stage. The idea is that the solution obtained on the second stage should not be too far from the function c_{glob}. Thus, in all adaptive meshes we enforce that the coefficient $c(x)$ belongs to the following set of admissible parameters, $c(x) \in \{c \in H : 1 \leq c(x) \leq 4\}$. The solution computed on the mesh, which was obtained after n refinements, is denoted here as $c^{(n)}$, for convenience. In



a) 4776 elements b) 5272 elements c) 6174 elements

d) 7682 elements e) 11068 elements f) $c_{h\alpha} \approx 3.99$

FIG. 8.2. *Computational results for the second stage of our two stage numerical procedure. We have taken on this stage the noise level in (8.3)* $\varsigma = 0.02 \approx \delta$ *and the regularization parameter* $\alpha = 0.15 \approx \varsigma^{0.48}$. *Hence,* $\mu_2 \approx 0.48$, *where the number* μ_2 *is defined in (2.12). Adaptively refined meshes on five consequtive mesh refinements are shown on a)-e). Fig. 8.2 f) displays the resulting image of the coefficient* $c^{(5)}(x)$ *after five (5) mesh refinements, see details in the text. Locations of both inclusions are accurately imaged (compare with Fig. 8.1 a)). The maximal value of the function* $c^{(5)}(x) = 4$ *inside of each imaged inclusion, which means that the inclusion/background contrast is also accurately imaged. In addition,* $c^{(5)}(x) = 1$ *outside of imaged inclusions. We set* $c_\alpha(x) := c^{(5)}(x)$.

addition, we use a cut-off parameter C_{cut} for the reconstructed coefficient $c^{(n)}$,

$$c^{(n)}(x) = \begin{cases} c^{(n)}(x), & \text{if } c^{(n)}(x) \geq C_{cut} \max_{\overline{\sigma}} c^{(n)}(x) \\ c_{glob}, & \text{elsewhere.} \end{cases}.$$

In our numerical experiments we have taken $C_{cut} = 0.75$ and in the adaptivity technique we have taken in (8.3) $\varsigma = 0.02 \approx \delta$ which corresponds to 2% of the noise level, and we have taken $\alpha = 0.15 \approx \delta^{0.48}$, which means that in (2.12) $\mu_2 \approx 0.48$. First, we use the quasi-Newton method on the same coarse mesh where the globally convergent method worked and have obtained the same image quality (not shown) as on Figure 8.1b. Next, we have performed our testing on 5 times refined meshes. As a result, the image was stabilized. This stabilization basically means that the norm $\left\| (Y'_\alpha)^I \left(c^{(5)}\right) \right\|_{L_2(\sigma)}$ became too small, indicating that at least one of conditions (7.12), (7.13) is likely invalid at $n = 5$ and thus, the mesh refinement process should be stopped (Theorem 7.4). Figure 8.2 displays those mesh refinements as well as the resulting image on the finally refined mesh. One can see that the image quality is significantly enhanced compared with Figure 8.1-b). Namely, the maximal value of the imaged coefficient within both inclusions is now 4, which is the correct value, and locations of both imaged inclusions are also imaged accurately.

An important additional point is to computationally verify the relaxation property (7.14). As c_α we have taken the function obtained on the finally refined mesh (see Figure 8.1-c)). Next, we compute norms $\left\| c^{(n)} - c_\alpha \right\|_{L_2(\sigma)}$, where $c^{(n)}$ is the approximation for the function c obtained after n mesh refinements. Each function $c^{(n)}$ is linearly interpolated on the finally refined mesh. Since $c^{(n)}$ is a piecewise linear function, this interpolation does not change it. Figure 8.1-d) displays computed values of norms $\left\| c^{(n)} - c_\alpha \right\|_{L_2(\sigma)}$ for all six meshes on which computations have been performed. The relaxation property (7.14) is evident. Note that this figure also displays result of another test with $\varsigma = \alpha = 0.01$, which we have performed. We observe that computed norms for this second test are slightly lower than those of the first, so as relaxation numbers r_n. The resulting image for the second test (not shown) was of about the same quality as the one for the first.

Acknowledgment

The authors are profoundly grateful to Professor Anatoly B. Bakushinskii for his constant and very friendly consulting during the work on this paper.

References

1. M. Ainsworth and J.T. Oden, *A Posteriori Error Estimation in Finite Element Analysis*, Wiley, New York, 2000

2. H. Ammari, E. Iakovleva, and D. Lesselier, Music-type electromagnetic imaging of a collection of small three dimensional inclusions. *SIAM J. Sci.Comp.*, 29, 674-709, 2007.

3. A.B. Bakushinsky and M.Yu. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*, Springer, 2004.

4. L. Beilina, K. Samuelsson and K. Åhlander, Efficiency of a hybrid method for the wave equation. In *International Conference on Finite Element Methods*, Gakuto International Series Mathematical Sciences and Applications. Gakkotosho CO., LTD, 2001.

5. L. Beilina and C. Johnson. A hybrid FEM/FDM method for an inverse scattering problem. In *Numerical Mathematics and Advanced Applications - ENUMATH 2001*. Springer-Verlag, 2001.

6. L. Beilina, Adaptive finite element/difference method for inverse elastic scattering waves, *Appl. Comput. Math.*, 1, 158-174, 2002.

7. L. Beilina and C. Clason, An adaptive hybrid FEM/FDM method for an inverse scattering problem in scanning acoustic microscopy. *SIAM J. Sci. Comp.*, 28, 382-402, 2006.

8. L. Beilina and M. V. Klibanov, A globally convergent numerical method for a coefficient inverse problem, *SIAM J. Sci. Comp.*, 31, 478-509, 2008.

9. L. Beilina, M. V. Klibanov and M.Yu Kokurin, Adaptivity with relaxation for ill-posed problems and global convergence for a coefficient inverse problem, preprint, available on-line at http://www.ma.utexas.edu/mp_arc (posting date November 20, 2009) and at

http://www.math.chalmers.se/Math/Research/Preprints, *Chalmers Preprint Series*, ISSN 1652-9715, 2009.

10. G. Chavent, Deux résultats sur le probléme inverse dans les équations aux dérivées partielles du deuxième ordre an t et sur l'unicité de la solution du problème inverse de la diffusion, *C.R. Acad. Sc. Paris*, 270, 25-28, 1970.

11. M. Cheney and D. Isaacson, Inverse problems for a perturbed dissipative half-space, *Inverse Problems*, 11, 865-888, 1995.

12. P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, SIAM Publications, 2002.

13. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Boston, 2000.

14. B. Engquist and A. Majda, Absorbing boundary conditions for the numerical simulation of waves, *Math. Comp.*, 31, 629-651, 1977.

15. K. Eriksson, D. Estep and C. Johnson, *Calculus in Several Dimensions*, Springer, Berlin, 2004.

16. L. C. Evans, *Partial Differential Equations,* AMS, Providence, RI, 1998.

17. A. Grisbaum, B. Kaltenbacher and B. Vexler, Efficient computation of the Tikhonov regularization parameter by goal-oriented adaptive discretization, *Inverse Problems*, 24, 025025, 2008.

18. M. V. Klibanov, Inverse problems and Carleman estimates, *Inverse Problems*, 8, 575-596, 1992.

19. M. V. Klibanov and A. Timonov, *Carleman Estimates for Coefficient Inverse Problems and Numerical Applications*, VSP, Utrecht, The Netherlands, 2004

20. O. A. Ladyzhenskaya and N. N. Uralceva, *Linear and Quasilinear Elliptic Equations*, Academic Press, New York, 1969

21. O. A. Ladyzhenskaya, *Boundary Value Problems of Mathematical Physics*, Springer Verlag, Berlin, 1985

22. M. Minoux, *Programmation Mathematique*, Bordas, Paris, 1989.

23. J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

24. R. Ramlau, A steepest descent algorithm for the global minimization of the Tikhonov functional, *Inverse Problems*, 18, 381-405, 2002.

25. R. Ramlau, TIGRA- an iterative algorithm for regularizing nonlinear ill-posed problems, *Inverse Problems*, 19, 433-465, 2003.

26. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems,* Winston and Sons, Washington, DC, 1977.