CHALMERS | UNIVERSITY OF GOTHENBURG

PREPRINT 2014:3

Globally strongly convex cost functional for a coefficient inverse problem

LARISA BEILINA
MICHAEL V. KLIBANOV

Department of Mathematical Sciences
Division of Mathematics
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg Sweden 2014

Preprint 2014:3

Globally strongly convex cost functional for a coefficient inverse problem

Larisa Beilina and Michael V. Klibanov

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg, Sweden
Gothenburg, January 2014

Preprint 2014:3
ISSN 1652-9715

Matematiska vetenskaper
Göteborg 2014

Globally strongly convex cost functional for a coefficient inverse problem

Larisa Beilina* and Michael V. Klibanov◇

Abstract

A Carleman Weight Function (CWF) is used to construct a new cost functional for a Coefficient Inverse Problems for a hyperbolic PDE. Given a bounded set of an arbitrary size in a certain Sobolev space, one can choose the parameter of the CWF in such a way that the constructed cost functional will be strongly convex on that set. Next, convergence of the gradient method, which starts from an arbitrary point of that set, is established. Since restrictions on the size of that set are not imposed, then this is the global convergence.

[1]

Key Words: Coefficient inverse problem, wave-like equation, Carleman weight function, strongly convex cost functional.

AMS subject classifications. 35R30, 35L05.

1 Introduction

We consider a Coefficient Inverse Problem (CIP) for the equation $c(x)\,u_{tt} = \Delta u, (x,t) \in \mathbb{R}^3 \times (0,T)$ with initial conditions $u(x,0) = f(x), u_t(x,0) = 0$ and the unknown coefficient $c(x)$. Hence, the function $u = u(x,t,c)$ depends on the coefficient $c(x)$ nonlinearly. First, we derive an initial-boundary value problem for a nonlinear integral differential equation with Volterra integrals and with both Dirichlet and Neumann boundary conditions. The coefficient $c(x)$ is not present in this equation. As soon as the solution of this initial-boundary value problem is found, the function $c(x)$ can be easily calculated. To solve this problem, we construct a weighted least squares Tikhonov-like functional for the latter problem. The weight is the Carleman Weight Function (CWF), which is involved in the Carleman estimate for the operator $c(x)\,\partial_t^2 - \Delta$. The main new result is Theorem 2 (subsection 2.3), which ensures that, given a certain convex set of an arbitrary finite size in a Hilbert space, one can

[1]*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

choose the large parameter $\lambda > 1$ of this CWF in such a way that the above functional is strongly convex on this set. Since restrictions on the size of that set are not imposed, we call this "global strong convexity".

To prove the strong convexity, we first prove a new Carleman estimate in Theorem 1 (subsection 2.2). This estimate is derived for case when the conventional term $(c(x) u_{tt} - \Delta u)^2$ is summed up with the nonlinear term $g(x,t) u(x,0) u_{tt}(x,t)$ with a certain function $g(x,t)$. The difficulty here is in the presence of the second derivative $u_{tt}(x,t)$ in this term, since it is also present in the principal part of the hyperbolic operator $c(x) \partial_t^2 - \Delta$. As far as the authors are aware of, all currently known Carleman estimates are obtained only for the case when the square of the linear principal part of a PDE operator is involved.

The idea of obtaining a nonlinear integral differential equation without the unknown coefficient present goes back to the method of proofs of global uniqueness theorems for CIPs using Carleman estimates. This method was originally proposed in [7], also see, e.g. sections 1.10, 1.11 in [4] as well as surveys [14, 19] and references cited there.

We prove the global convergence of the gradient method of the minimization of our functional. More precisely, we prove that this method converges to the unique minimizer on the above set if starting from an arbitrary point of that set. In addition, the distance between the minimizer and the exact solution of the original problem is estimated in the case when the data are given with an error. Also, keeping in mind future numerical studies, we prove similar results for some finite dimensional approximations of that integral differential equation. Finally, we outline an algorithm of solving that minimization problem using the FEM. The latter might be useful in computations. Numerical testing of these ideas would require a substantial additional effort, which is outside of the scope of the current paper. We plan to do this in the future.

The assumption that we work on a priori given bounded set is going along well with the Tikhonov concept for ill-posed problems, see, e.g. section 1.4 of [4]. By this concept, one should seek for the solution of an ill-posed on an a priori given bounded set, which can also be considered as the set of admissible parameters.

A cost functional with the CWF in it, which is strongly convex on a bounded convex set of an arbitrary size, was constructed for a similar CIP in the earlier work of the second author [13]. However, the strong convexity was established only in the case when a series resulting from the separation of x and t variables was truncated. It is unclear how the result of [13] would change if the number of terms in that truncated series would tend to infinity. As a result of the separation of variables, the CWF for the Laplace operator was used in [13]. Compared with [13], the main new element of this paper is that we do not truncate any series in Theorem 2. In addition, we do not consider the separation of x and t variables and use the CWF for the above hyperbolic operator.

In the recent interesting work [3] a different CWF was used to construct a globally convergent numerical method for an analog of our CIP for the case of the equation $v_{tt} = \Delta v - q(x)v$ with the unknown coefficient $q(x)$. The method of [3] requires the minimization on each iterative step of a certain quadratic functional with the CWF in it. As a result, a good approximation for $q(x)$ is obtained, if starting from any point of a bounded set of an arbitrary radius (in a certain space). This approach is different from ours in the sense that a nonlinear integral differential equation was not obtained and a globally strongly convex cost functional was not constructed in [3].

Conventional numerical methods for CIPs, such as, e.g. gradient method and Newton method, are based on the minimization of conventional least squares cost functionals for CIPs. These functionals suffer from the phenomenon of local minima and ravines. Thus, these methods converge locally, i.e. their convergence is guaranteed only if the starting point is located in a small neighborhood of the solution. Recently a globally convergent method for our CIP was developed analytically and verified computationally in a series of publications of the authors, which were summarized in their book [4]. This method is not using an optimization procedure and is significantly different from the one of the current paper.

In some theorems below we analyze the case when an error in the data is present. We assume that its level is sufficiently small. The smallness assumption about the error is a natural one for an ill-posed problem. Indeed, the intuition says that a numerical method cannot work well if the data contain a large error. Another argument here is that it is well known that the theory is usually more pessimistic than computations. For example, the global convergence theorem 2.9.4 of [4] also works with a small error in the data. Nevertheless, that method works well with experimental data, which are very noisy: see Chapter 5 and section 6.9 of [4] as well as [6] and references cited in these publications.

In section 2 we formulate our inverse problem and theorems. In sections 3-5 we prove these theorems. In section 6 we outline an algorithm which works with finite elements.

2 The Coefficient Inverse Problem and Main Results

Let $\Omega \subset \mathbb{R}^3$ be a convex bounded domain with a piecewise-smooth boundary $\partial\Omega$. For any $T > 0$ denote $Q_T = \Omega \times (0,T), S_T = \partial\Omega \times (0,T)$. Let the function $c(x)$ satisfies the following conditions

$$1 \leq c(x) \leq 1 + b, \forall x \in \mathbb{R}^3, \tag{1}$$

$$c(x) = \overline{b} = const. \in [1, 1+b], x \in \mathbb{R}^3 \diagdown \Omega, \tag{2}$$

$$c \in C^1\left(\mathbb{R}^3\right), \tag{3}$$

where numbers $b, \overline{b} > 0$ are given. We use "1" here in (2) for the normalization only. In addition, we assume that there exists a point $x_0 \in \mathbb{R}^3 \diagdown \overline{\Omega}$ such that

$$(\nabla c, x - x_0) \geq 0, \forall x \in \overline{\Omega}, \tag{4}$$

where $(,)$ denotes the scalar product in $\mathbb{R}^3$. Condition (4) is imposed to guarantee the validity of the Carleman estimate for the operator $c(x)\partial_t^2 - \Delta$, see Theorem 1.10.2 and Corollary 1.10.2 in [4]. Let $\Omega' \subset \mathbb{R}^3$ be another bounded domain such that $\Omega \subset \Omega', \partial\Omega \cap \partial\Omega' = \varnothing$. We assume that the function $f(x)$ satisfies the following conditions

$$f \in H^7\left(\mathbb{R}^3\right), \tag{5}$$

$$f(x) = 0 \text{ for } x \in \mathbb{R}^3 \diagdown \Omega'. \tag{6}$$

The embedding theorem, (5) and (6) imply that $f \in C^5(\mathbb{R}^3)$. In addition, we assume below that

$$\Delta f \geq \xi = const. > 0 \text{ for } x \in \overline{\Omega}. \tag{7}$$

A discussion of condition (7) can be found in subsection 2.4.

2.1 Coefficient Inverse Problem

Consider the following Cauchy problem

$$c(x) u_{tt} = \Delta u, (x,t) \in \mathbb{R}^3 \times (0,T), \tag{8}$$

$$u(x,0) = f(x), u_t(x,0) = 0. \tag{9}$$

It follows from corollary 4.2 of Chapter 4 of the book [15] that conditions (1)-(3), (5) and (6) guarantee uniqueness and existence of the solution $u \in H^2(\mathbb{R}^3 \times (0,T))$ of the problem (8), (9). To apply our technique, we need $u_{tt} \in C^3(\mathbb{R}^3 \times [0,T])$. Since by the embedding theorem $H^6(\mathbb{R}^3 \times (0,T)) \subset C^3(\mathbb{R}^3 \times [0,T])$, then we assume that

$$u \in H^8\left(\mathbb{R}^3 \times (0,T)\right). \tag{10}$$

Hence, the trace theorem justifies the smoothness (5) for $f(x) = u(x,0)$. Using results of Chapter 4 of [15], one can show that some smoothness conditions being imposed on functions c, f guarantee (10). However, we leave aside these conditions for brevity and just assume (10) below. Note that the issue of the minimal smoothness of coefficients is rarely a concern in the theory of CIPs, since these problems are quite difficult ones even for sufficiently smooth coefficients, see, e.g. [16, 17].

Equation (8) is the acoustics equation, where $c^{-1/2}(x)$ is the speed of sound and $u(x,t)$ is the pressure of the acoustic wave at the point x at the moment of time t. In the 2-d case (8) can be derived from the Maxwell's equations. In this case $c(x)$ is the spatially distributed dielectric constant and $u(x,t)$ is the amplitude of one of components of the electric wave. In addition, (8) was successfully used in [4, 6] to model the propagation in $\mathbb{R}^3$ of one of components of the electric field in the case of experimental data with the same interpretation of functions $c(x)$ and $u(x,t)$ as above. A numerical explanation of the latter can be found in [5].

Coefficient Inverse Problem (CIP). Suppose that conditions (1)-(10) are satisfied. Assume that the coefficient $c(x)$ is unknown inside of the domain Ω. Determine the function $c(x)$ for $x \in \Omega$, assuming that the following function $s(x,t)$ is known

$$u|_{S_T} = s(x,t). \tag{11}$$

The function $s(x,t)$ in (11) models the boundary measurement. Having $s(x,t)$, one can uniquely solve the initial boundary value problem for equation (8) with the initial condition (9) outside of the domain Ω. Hence, the normal derivative is known,

$$\partial_\nu u \mid_{S_T} = p(x,t). \tag{12}$$

Remark 1. It would be sufficient to know the function $p(x,t)$ in (12) only on a part of the boundary $\partial\Omega$. To extend our method to this case, one should specify a part of the Carleman estimate of Theorem 1, see, e.g. [3] for such a specification. We do not follow this root for brevity. Uniqueness of this CIP under conditions (4), (7) was established via the method of [7], see, e.g. theorem 1.10.5.1 in [4] and theorem 3.1 in [14].

2.2 The Carleman estimate with a nonlinear term

Choose a point $x_0 \in \mathbb{R}^3 \diagdown \overline{\Omega}$. Let the number $\eta \in (0,1)$ and let $\lambda > 1$ be a large parameter. Consider functions ψ, φ_λ,

$$\psi(x,t) = |x-x_0|^2 - \eta t^2, \varphi_\lambda(x,t) = \exp(\lambda\psi(x,t)).$$

For any number $\eta \in (0,1)$ one can choose a sufficiently large T such that

$$N = N(\Omega, x_0, \eta, T) = \eta T^2 - \max_{x\in\overline{\Omega}} |x-x_0|^2 > 0. \tag{13}$$

Choose a number $d \in \left(0, \min_{x\in\overline{\Omega}} |x-x_0|^2\right)$. Hence, $\Omega \subset \left\{|x-x_0|^2 > d\right\}$. Consider the domain P_d and the number $M > 0$,

$$P_d = \{(x,t) \in Q_T : \psi(x,t) > d\}, M = \max_{\overline{Q}_T} \psi(x,t) = \max_{x\in\overline{\Omega}} |x-x_0|^2.$$

By (13) $\overline{P}_d \cap \{t = T\} = \varnothing$. In fact, P_d is a part of the hyperboloid $\left\{(x,t) : |x-x_0|^2 - \eta t^2 > d\right\}$. Denote $\Omega_T = \{(x,t) : x \in \Omega, t = T\}$. Let the function $g(x,t)$ be such that $g, \partial_t g \in C\left(\overline{Q}_T\right)$.

Theorem 1. *Assume that the function $c(x)$ satisfies conditions (1), (3) and (4). Then there exist a sufficiently small number $\eta = \eta\left(\Omega, x_0, \|c\|_{C^1(\overline{\Omega})}\right) \in (0,1)$, a positive number $C_1 = C_1\left(\eta, b, \|g\|_{C(\overline{Q}_T)}, \|\partial_t g\|_{C(\overline{Q}_T)}\right)$ and a sufficiently large number $\overline{\lambda} = \overline{\lambda}\left(\eta, T, b, \|g\|_{C(\overline{Q}_T)}, \|\partial_t g\|_{C(\overline{Q}_T)}\right) > 1$, all three numbers depending only on listed parameters, such that if $T = T(\eta, x_0, \Omega) > 0$ is so large that (13) holds, then for all $\lambda \geq \overline{\lambda}$ the following Carleman estimate is valid*

$$\int_{Q_T} (c(x) u_{tt} - \Delta u)^2 \varphi_\lambda^2 dxdt + \int_{Q_T} g(x,t) u(x,0) u_{tt}(x,t) \varphi_\lambda^2 dxdt$$

$$+C_1 \lambda^3 \exp(2\lambda M) \left(\|\partial_\nu u \mid_{S_T}\|^2_{L_2(S_T)} + \|u\|^2_{H^1(S_T)}\right)$$

$$+C_1 \lambda^3 \exp(-2\lambda N) \left(\|u_t\|^2_{L_2(\Omega_T)} + \|u\|^2_{H^1(\Omega_T)}\right) + C_1 \exp(-\lambda N) \|u_t\|^2_{L_2(Q_T)} \tag{14}$$

$$\geq C_1 \int_{Q_T} \left(\lambda u_t^2 + \lambda (\nabla u)^2 + \lambda^3 u^2\right) \varphi_\lambda^2 dxdt,$$

$$\forall u\in\left\{U\in H^{2}\left(Q_{T}\right):U_{t}\left(x,0\right)=0\right\}.$$

In the case $c\left(x\right)\equiv 1$ *one can choose any number* $\eta\in\left(0,1\right)$.

Remarks 2:

1. The main difficulty of the proof of this theorem (section 3) is due to the presence of the second derivative u_{tt} in the nonlinear term $g\left(x,t\right)u_{tt}\left(x,t\right)u\left(x,0\right)$, since this derivative is also a part of the operator $c\left(x\right)\partial_{t}^{2}-\Delta$. On the other hand, $\left(c\left(x\right)u_{tt}-\Delta u\right)^{2}$ in (14) is the standard term in the Carleman estimate for this operator, see, e.g. Theorem 1.10.2 and Corollary 1.10.2 of [4]. The only non-standard element here is the absence of the integral over $\overline{Q}_{T}\cap\left\{t=0\right\}$. This absence is due to the fact that formulae (1.86) and (1.87) in [4] imply that the corresponding integral over $\overline{Q}_{T}\cap\left\{t=0\right\}$ equals zero, since $u_{t}\left(x,0\right)=0$.

2. If the nonlinear term $g\left(x,t\right)u\left(x,0\right)u_{tt}\left(x,t\right)$ would be absent, then we would not need terms in the third line of (14), and the fourth line would be replaced simply with $\left\Vert u\right\Vert_{H^{1}\left(Q_{T}\right)}^{2}$. This follows from a slight modification of the technique, which was proposed for the first time in [12]. This technique can also be found in, e.g. theorem 5.1 of the more recent publication [14], also, see references in section 5.5 of [14].

2.3 Strong convexity

In this subsection we formulate our main result. Denote

$$\overline{s}\left(x,t\right)=\partial_{t}^{2}s\left(x,t\right),\overline{p}\left(x,t\right)=\partial_{t}^{2}p\left(x,t\right). \tag{15}$$

Let $\widetilde{w}=u_{tt}$. Then by (10) $\widetilde{w}\in H_{0}^{6}\left(Q_{T}\right)$. Hence, (8), (9) and (15) imply that

$$c\left(x\right)=\frac{\left(\Delta f\right)\left(x\right)}{\widetilde{w}\left(x,0\right)},x\in\Omega, \tag{16}$$

$$\frac{\left(\Delta f\right)\left(x\right)}{\widetilde{w}\left(x,0\right)}\widetilde{w}_{tt}-\Delta\widetilde{w}=0,\left(x,t\right)\in Q_{T}, \tag{17}$$

$$\widetilde{w}_{t}\left(x,0\right)=0,\widetilde{w}\mid_{S_{T}}=\overline{s}\left(x,t\right),\partial_{\nu}\widetilde{w}\mid_{S_{T}}=\overline{p}\left(x,t\right). \tag{18}$$

We now want to obtain zero boundary conditions at S_{T}. To do this, assume that there exists a function $F\left(x,t\right)$ such that

$$F\in H^{6}\left(Q_{T}\right),F_{t}\left(x,0\right)=0,F\mid_{S_{T}}=\overline{s}\left(x,t\right),\partial_{\nu}F\mid_{S_{T}}=\overline{p}\left(x,t\right). \tag{19}$$

Denote $w=\widetilde{w}-F$. Hence, by (16)

$$c\left(x\right)=\frac{\left(\Delta f\right)\left(x\right)}{\left(w+F\right)\left(x,0\right)}. \tag{20}$$

For any integer $s\geq 2$ denote

$$H_{0}^{s}\left(Q_{T}\right)=\left\{u\in H^{s}\left(Q_{T}\right):u_{t}\left(x,0\right)=0,u\mid_{S_{T}}=0,\partial_{\nu}u\mid_{S_{T}}=0\right\}. \tag{21}$$

Then (17)-(21) imply that

$$\frac{(\Delta f)(x)}{[w(x,0)+F(x,0)]}(w+F)_{tt}-\Delta(w+F)=0,(x,t)\in Q_T, \tag{22}$$

$$w_t(x,0)=0, w\mid_{S_T}=0, \partial_\nu w\mid_{S_T}=0, \text{ i.e. } w\in H_0^6(Q_T). \tag{23}$$

We focus below on the solution of the problem (22), (23). Indeed, if an approximate solution of this problem is found, then the corresponding approximation for the target coefficient $c(x)$ can be found via (16) where $\widetilde{w}=w+F$. Note that equation (22) is nonlinear with respect to the function w.

We need the smoothness $w\in H_0^6(Q_T)$ because the proof of Theorem 2 uses the fact that three derivatives of the function w are bounded. This can be ensured if, e.g. $w\in C^3\left(\overline{Q}_T\right)$. Indeed, by the embedding theorem

$$H^6(Q_T)\subset C^3\left(\overline{Q}_T\right), \|y\|_{C^3\left(\overline{Q}_T\right)}\leq C_2\|y\|_{H^6(Q_T)}, \forall y\in H^6(Q_T), \tag{24}$$

where the number $C_2=C_2(Q_T)>0$ depends only on the domain Q_T. Also, by the trace theorem

$$\|y(x,0)\|_{H^5(\Omega)}\leq C_2\|y\|_{H^6(Q_T)}, \forall y\in H^6(Q_T). \tag{25}$$

Let $R>0$ be an arbitrary number such that

$$\|\Delta f\|_{H^5(\Omega)}\leq C_2 R. \tag{26}$$

For every function $v\in H_0^6(Q_T)$ such that $(v+F)(x,0)>0$ in $\overline{\Omega}$ denote

$$A(v)=\frac{(\Delta f)(x)}{v(x,0)+F(x,0)}, x\in\Omega. \tag{27}$$

Consider the set of functions $G(Q_T,b,R,f,F)$ defined as

$$G=G(Q_T,b,R,f,F)=\left\{\begin{array}{c} v\in H_0^6(Q_T), \\ \|v\|_{H^6(Q_T)}\leq R, \\ (1+b)^{-1}(\Delta f)(x)\leq v(x,0)+F(x,0)\leq(\Delta f)(x) \text{ in } \overline{\Omega}, \\ (\nabla A(v)(x), x-x_0)\geq 0 \text{ in } \overline{\Omega}. \end{array}\right. \tag{28}$$

Denote $Int(G)$ the open set of interior points of G. Inequality (26) guarantees that the function $(\Delta f)(x)$ is in the proper range. Indeed, $\|v(x,0)+F(x,0)\|_{H^5(\Omega)}\leq C_2R$ in this case: by (25) and (28). If $c=c(x,v)=A(v), v\in G$, then by (7), (25) and (26) the function c satisfies conditions (1), (3), (4).

We need Proposition 1, since convex functionals are typically defined on convex sets.

Proposition 1. *G is a convex set.*

Proof. Let $\beta \in [0,1]$ be an arbitrary number and let $v_1, v_2 \in G$ be two arbitrary functions. We should prove that the function $\overline{v} = \beta v_1 + (1-\beta) v_2 \in G$. Obviously this function satisfies conditions of first three lines of (28). Consider now the fourth line. By (27)

$$(\nabla A(\overline{v})(x), x - x_0) = \frac{\beta}{(\overline{v}+F)^2(x,0)} ((\nabla(\Delta f)(v_1+F)(x,0) - \Delta f \nabla (v_1+F)(x,0)), x - x_0)$$

$$+\frac{(1-\beta)}{(\overline{v}+F)^2(x,0)} ((\nabla(\Delta f)(v_2+F)(x,0) - \Delta f \nabla (v_2+F)(x,0)), x - x_0).$$

Since $(\nabla A(v_j), x - x_0) \geq 0$ for $j = 1,2$, then

$$((\nabla(\Delta f)(v_j+F)(x,0) - \Delta f \nabla (v_j+F)(x,0)), x - x_0) \geq 0.$$

Hence, $(\nabla A(\overline{v})(x), x - x_0) \geq 0$. □

We now reformulate the problem (22), (23) as: *Find a function $w(x,t)$ such that*

$$Y(w) := A(w)(w+F)_{tt} - \Delta(w+F) = 0, (x,t) \in Q_T, w \in G. \tag{29}$$

Let $\alpha \in (0,1)$ be the regularization parameter. To solve the problem (29), we construct the following weighted Tikhonov regularization functional $J_{\lambda,\alpha}(w) : G \to \mathbb{R}$,

$$J_{\lambda,\alpha}(w) = \int\limits_{Q_T} [A(w)(w+F)_{tt} - \Delta(w+F)]^2 \varphi_\lambda^2 dxdt + \alpha \|w\|^2_{H^6(Q_T)}, w \in G, \tag{30}$$

Theorem 2. *Let η and T be numbers of Theorem 1 and let $Int(G) \neq \varnothing$. Let $J'_{\lambda,\alpha}(w)(h), \forall h \in H_0^6(Q_T)$ be the Fréchet derivative of the functional $J_{\lambda,\alpha}$ at the point $w \in Int(G)$. Then there exists a constant $C = C(\eta, T, b, R, G) > 0$ and a sufficiently large number $\lambda_0 = \lambda_0(\eta, T, b, R, G) > 1$, both depending only on listed parameters, such that for all $\lambda \geq \lambda_0$ and for all $\alpha \geq 2C\exp(-\lambda N)$ the functional $J_{\lambda,\alpha}(w)$ is strongly convex on the set G. More precisely*

$$J_{\lambda,\alpha}(w_2) - J_{\lambda,\alpha}(w_1) - J'_{\lambda,\alpha}(w_1)(w_2 - w_1)$$

$$\geq C \int\limits_{Q_T} \left[\lambda(\partial_t(w_2-w_1))^2 + \lambda(\nabla(w_2-w_1))^2 + \lambda^3(w_2-w_1)^2\right] \varphi_\lambda^2 dxdt \tag{31}$$

$$+\frac{\alpha}{2}\|(w_2-w_1)\|^2_{H^6(Q_T)} \geq \frac{\alpha}{2}\|w_2-w_1\|^2_{H^6(Q_T)}, \forall w_1 \in Int(G), \forall w_2 \in G.$$

In particular, (31) implies that

$$J_{\lambda,\alpha}(w_2) - J_{\lambda,\alpha}(w_1) - J'_{\lambda,\alpha}(w_1)(w_2 - w_1)$$

$$\geq C\exp(2\lambda d)\|w_2-w_1\|^2_{H^1(P_d)} + \frac{\alpha}{2}\|w_2-w_1\|^2_{H^6(Q_T)}, \forall w_1 \in Int(G), \forall w_2 \in G. \tag{32}$$

Corollary 1. *Let*

$$\widetilde{J}_{\lambda}(w)=\int\limits_{Q_T}\left[A(w)(w+F)_{tt}-\Delta(w+F)\right]^2\varphi_{\lambda}^2 x\,dx\,dt. \tag{33}$$

Then

$$\widetilde{J}_{\lambda}(w_2)-\widetilde{J}_{\lambda}(w_1)-\widetilde{J}_{\lambda}'(w_1)(w_2-w_1)\geq C\exp(2\lambda d)\|w_2-w_1\|_{H^1(P_d)}^2$$
$$-C\lambda^3\exp(-2\lambda N)\left(\|\partial_t(w_2-w_1)\|_{L_2(\Omega_T)}^2+\|w_2-w_1\|_{H^1(\Omega_T)}^2\right) \tag{34}$$
$$-C\exp(-\lambda N)\|w_2-w_1\|_{L_2(Q_T)}^2,\ \forall w_1\in Int(G),\forall w_2\in G.$$

Below $C>0$ denotes different constants depending on the same parameters as ones in Theorem 2.

2.4 Discussion of Theorem 2

Even though numerical studies are outside of the scope of the current publications, we briefly discuss in this subsection some computational aspects of Theorem 2. We point out that it is well known that computations often work well under conditions which are less restrictive than the theory is. Still, the theory, such as, e.g. the one of this paper, usually provides an important guidance for computations.

The maximal value M of the function $\psi(x,t)$ is achieved at the point $(\overline{x}(x_0),0)$, such that $\overline{x}(x_0)\in\partial\Omega$ and $|\overline{x}(x_0)-x_0|^2=\max_{x\in\overline{\Omega}}|x-x_0|^2$. On the other hand, since the function $\varphi_{\lambda}^2(x,t)$ changes rapidly with respect to (x,t), then it seems to be that the coefficient $c(x)$ will be accurately reconstructed in practical computations only in a small neighborhood of the point $\overline{x}(x_0)$. Changing x_0, one can cover a neighborhood of a part $\partial'\Omega$ of the boundary $\partial\Omega$.

We point out that our *ultimate goal* is to apply the technique of this paper to our experimental data, which are described in [6]. Next, we will compare its performance on these data with the performance of the globally convergent algorithm of [4], which was used in [6]. A potential application of the work [6] is in imaging and identification of explosive devices. Those devices typically have small sizes. The data of [6] were collected on a part of a plane in the backscattering case. The distance between this plane and the front surface of any target of interest was about 80 centimeters (cm). Using the arrival time of the backscattering signal, these distances were accurately estimated for all targets.

The raw data of [6] are very far from the range of the operator of the forward problem. Hence, it was necessary to use a heuristic data pre-processing procedure as a preliminary step before applying any reconstruction algorithm. The pre-processed data were used as the input for the algorithm of [4]. One of steps of data pre-processing was data propagation, which provides an approximation of the data at a part Γ' of a plane Γ. The distances between Γ' and the front surface of a target can vary from 0 cm to 4 cm. Since targets of interest have rather small sizes of a few centimeters, then imaging of a small neighborhood of Γ' might be sufficient for the experimental data of [6]. Besides, using a layer stripping procedure, one might cover a somewhat larger neighborhood of Γ'.

Next, the domain Ω was chosen, where the solution of the inverse problem was computed. The surface $\Gamma' \subset \partial\Omega$ was a part of the boundary of this domain. Another step of the data pre-processing procedure of [6] was complementing the data on Γ' by the data on the rest of the boundary $\partial\Omega$. Those additional data resulted from the solution of the forward problem for equation (8) for the case $c(x) \equiv 1$. Accurate reconstructions were obtained in [6] even for the most difficult cases of completely blind data. We believe, therefore, that the technique of the current paper might be applicable to the case of backscattering data, if complementing those data as in [6].

In Theorems 3-7 below we assume that the point of minimum of our functional is an interior point of either the set G or its finite dimensional analog. For such a point, the condition of the fourth line of (28) becomes $(\nabla A(v)(x), x - x_0) = (\nabla c(x), x - x_0) > 0$. Thus, $c(x) \equiv const.$ does not satisfy this condition. However, the case $c(x) \equiv const.$ as the solution of our CIP is of no interest to us, since the experimental data of [6] are quite different for this scenario from the case when a target of interest is present.

As to the condition (7), in principle it would be better to assume instead that $f(x) = \delta(x - x')$ for a certain point $x' \in \Omega' \diagdown \Omega$. However, it is well known that the technique of [7] does not work in this case. On the other hand, a narrow Gaussian centered at the point $\{x'\}$ approximates $\delta(x - x')$ in the sense of distributions [18]. Thus, it was pointed out on pages 47, 48 of [4] and on pages 480, 481 of [14] that if the function $\delta(x - x')$ would be replaced with that Gaussian, then this would be equivalent to $\delta(x - x')$ from the Physics standpoint and would provide only an insignificant difference in the data $s(x,t), p(x,t)$. We believe that, for the data of [6], such a replacement can be handled well computationally by the technique of the current data since this method is stable (Theorems 4,7 below). On the other hand, both the method of [7] and the technique of this paper work in the case of this replacement. To ensure (6), one should multiply that Gaussian by a function $\chi \in C^{\infty}(\mathbb{R}^3)$ such that $\chi(x) = 1$ in Ω, $\chi(x) = 0$ in $\mathbb{R}^3 \diagdown \Omega'$ and the integral of the resulting product over Ω' would be equal to unity.

Although the topic of differentiation of noisy data is outside of the scope of this paper, we now briefly comment on it. Functions $\overline{s}(x,t), \overline{p}(x,t)$ amount to the second $t-$derivative of measured data $s(x,t), p(x,t)$ in (11), (12), which naturally contain noise. By (19) the function F depends on functions $\overline{s}, \overline{p}$ and needs to have more derivatives. Hence, in Theorems 4 and 7 below we consider the case when the function F is given with an error. It is well known that the ill-posed problem of stable differentiation of noisy data can be addressed via a variety of regularization algorithms, see, e.g. [1]. We only mention that our experience of working with experimental data tells us that such a question can usually be addressed via either one of methods of [1] or, again, a proper heuristic data pre-preprocessing procedure. In fact, differentiation of noisy data is one of elements of the technique of [4], and this was successfully done for experimental data in Chapter 5 of [4] and in [6].

Here is an additional consideration about the differentiation. Although we obtain zero Dirichlet and Neumann boundary conditions in (23) for the function w via the introduction of the function F with properties (19), we actually can prove an analog of the main Theorem 2 in the case of non-zero boundary conditions $\overline{s}, \overline{p}$ in (18) for the function $\widetilde{w}$. However, this is inconvenient for the practical implementation of the gradient method (35), since in this case the term $\gamma J'_{\lambda,\alpha}(w_n)$ should have zero boundary conditions. It is not immediately clear how to arrange the latter. On the other hand, if working with the finite difference approximation of the problem (29), then it is clear how to arrange zero boundary conditions for this term: one would need to arrange this term only for interior points of the domain Q_T. Thus, in this case one would need to stably calculate only the second $t-$derivatives of noisy data in (15). It is well known that regularization techniques can handle well the first and second derivatives of noisy data. An extension of results of this paper to the case of finite differences amounts to a significant effort, which is outside of our scope here.

2.5 Global convergence of the gradient method

We now formulate global convergence theorems of the gradient method of the minimization of the functional $J_{\lambda,\alpha}$. Consider an arbitrary function $w_1 \in Int(G)$. Let $\gamma > 0$ be a number. For brevity we do not indicate the dependence of functions w_n on parameters λ, α, γ. Consider the sequence $\{w_n\}_{n=1}^{\infty}$ of the gradient method,

$$w_{n+1} = w_n - \gamma J'_{\lambda,\alpha}(w_n), n = 1, 2, ... \tag{35}$$

Theorem 3. *Let η and T be numbers of Theorem 1, λ_0 be the number of Theorem 2 and let $Int(G) \neq \varnothing$. Choose parameters $\lambda \geq \lambda_0$ and $\alpha \geq 2C \exp(-\lambda N)$. Assume that the functional $J_{\lambda,\alpha}$ achieves its minimal value on the set G at a point $w_{\min} \in Int(G)$. Then such a point $w_{\min}$ is unique. Consider the sequence (35), where $w_1 \in Int(G)$ is an arbitrary point. Assume that $\{w_n\}_{n=1}^{\infty} \subset Int(G)$. Then there exists a sufficiently small number $\gamma = \gamma(\lambda, \alpha, \eta, T, b, R, G) \in (0,1)$ and a number $q = q(\gamma) \in (0,1)$, both dependent only on listed parameters, such that the sequence (35) converges to the point $w_{\min}$,*

$$\|w_{n+1} - w_{\min}\|_{H^6(Q_T)} \leq q^n \|w_1 - w_{\min}\|_{H^6(Q_T)}. \tag{36}$$

Following subsection 2.4, assume now that the function F is given with an error. Then the natural question is on how far are points $w_{\min}$ and w_n from the exact solution of the problem (29) with errorless data. Theorem 4 addresses this question in terms of the norm of the space $H^1(P_d)$. The latter might be sufficient for computations. We use the Tikhonov concept for ill-posed problems mentioned in Introduction. Namely, we assume that there exists the exact solution w^* of the problem (29) with errorless data $F^* \in H^6(Q_T)$,

$$w^* \in G(Q_T, b, R, f, F^*) := G^* \neq \varnothing. \tag{37}$$

Theorem 4. *Let η and T be numbers of Theorem 1, λ_0 be the number of Theorem 2 and let $Int(G) \neq \varnothing$. Consider the problem (29). Assume that conditions of Theorem 3 about functions $w_1, w_{\min}$ and the sequence (35) hold and that (37) is valid. In addition, assume that the function F is given with an error of the level $\delta > 0$, i.e. $\|F - F^*\|_{C^3(\overline{Q}_T)} \leq \delta$, where*

$\delta \in (0, \delta_0)$, where the number $\delta_0 \in (0,1)$ is so small that $\delta_0 \leq \min\left(C_2 R, (1+b)^{-1}\xi/2\right)$ and also $\lambda_1 = \ln\left(\delta_0^{-1/(2M)}\right) \geq \lambda_0$. Choose $\lambda = \lambda(\delta) = \ln\left(\delta^{-1/(2M)}\right)$ and $\alpha = \alpha(\delta) = 2C\delta^{N/(2M)}$. Then for the same numbers γ, q as in Theorem 3

$$\|w^* - w_{\min}\|_{H^1(P_d)} \leq C\delta^\rho, \rho = \min\left(\frac{1}{2}, \frac{N}{4M}\right), \tag{38}$$

$$\|w_{n+1} - w^*\|_{H^1(P_d)} \leq q^n \|w_1 - w^*\|_{H^6(Q_T)} + C\delta^\rho, n = 1, 2, ... \tag{39}$$

Estimates (38) and (39) remain true in the case of errorless data with $\delta = 0$ if setting $\alpha = \alpha(\lambda) = 2C\exp(-\lambda N)$ and replacing $C\delta^\rho$ with $C\exp(-\lambda N/2)$ for $\lambda \geq \lambda_0$.

2.6 The finite dimensional case

Consider the weighted space $L_2^\lambda(Q_T)$,

$$L_2^\lambda(Q_T) = \left\{ w : \|w\|_{L_2^\lambda(Q_T)} = \left(\int\limits_{Q_T} w^2 \varphi_\lambda dxdt \right)^{1/2} < \infty \right\}.$$

Recall that $Y(w)$ is the nonlinear operator in the left hand side of (29) supplied with initial and boundary conditions (23). Hence, we can consider the operator Y as $Y : G \to L_2^\lambda(Q_T)$. Let $Y'(w)$ be the Fréchet derivative of Y at the point w. Then $J'_{\lambda,\alpha}(w) = Y'^*(w)(Y(w))$, where the linear bounded operator $Y'^*(w) : L_2^\lambda(Q_T) \to H_0^6(Q_T)$ is adjoint to the operator $Y'(w)$, see section 8.1 of [2]. Since $L_2^\lambda(Q_T) \neq H_0^6(Q_T)$, then it is both not easy and time consuming to calculate the Fréchet derivative $J'_{\lambda,\alpha}(w_n)$ in (35) for each n. On the other hand, in computations one always works with a finite dimensional space. In this case one deals with vectors of parameters in an Euclidian space. Hence, the gradient of a finite dimensional analog of the functional $J_{\lambda,\alpha}$ can be easily computed. Thus, analogs of above theorems for the finite dimensional case might be useful for computations. These analogs are formulated in the current subsection.

In this subsection we work with finite dimensional subspaces of two spaces: $H_0^6(Q_T)$ and $H_0^3(Q_T)$. In the case of $H_0^6(Q_T)$ constants in the convergence Theorem 6 for the gradient method are independent on the dimension of the subspace. Furthermore, Theorem 7 estimates distances between points calculated by the gradient method and the exact solution w^*. The space $H_0^3(Q_T)$ is simpler to work with. On the other hand, the constant $C_3 > 0$ in the corresponding Theorem 8 depends on the dimension of the subspace and those distances are not estimated.

For $m = 3, 6$ let

$$H_0^m(\Omega) = \{u \in H^m(\Omega) : u\mid_{\partial\Omega} = \partial_\nu u\mid_{\partial\Omega} = 0\}, H_0^m(0,T) = \{\psi(t) \in H^m(0,T) : \psi'(0) = 0\}.$$

Let $\{\phi_{i,m}(x)\}_{i=1}^\infty \subset H_0^m(\Omega)$ be an orthonormal basis in $H_0^m(\Omega)$ and $\{\psi_{j,m}(t)\}_{j=1}^\infty \subset H_0^m(0,T)$ be an orthonormal basis in $H_0^m(0,T)$. For an integer $k \geq 1$, let $H_0^{m,k}(Q_T)$ be the subspace of

$H_0^m(Q_T)$ with the orthonormal basis $\{\phi_{i,m}(x)\psi_{j,m}(t)\}_{i,j=1}^k$. Hence, $\dim\left(H_0^{m,k}(Q_T)\right)=k^2$ and

$$v(x,t)=\sum_{i,j=1}^{k} b_{i,j}\phi_{i,m}(x)\psi_{j,m}(t),(x,t)\in Q_T,\forall v\in H_0^{m,k}(Q_T), \tag{40}$$

where $b_{i,j}=b_{i,j}(v)$ are numbers. Denote $B=(b_{1,1},...,b_{k,k})^T\in\mathbb{R}^{k^2}$. For each vector $B\in\mathbb{R}^{k^2}$, let $v_B(x,t)$ be the function $v(x,t)$ represented via (40).

Let $Z_k^m:H_0^m(Q_T)\to H_0^{m,k}(Q_T)$ be the operator of the orthogonal projection of the space $H_0^m(Q_T)$ onto its subspace $H_0^{m,k}(Q_T)$. Based on (28), we define the set $G_{m,k}(Q_T,b,R,f,F):=G_{m,k}\subset\mathbb{R}^{k^2}$ as

$$G_{m,k}=\{B\}\in\mathbb{R}^{k^2}:\left\{\begin{array}{c}|B|\le R,\\ v_B\in H_0^{m,k}(Q_T),\\ (1+b)^{-1}(\Delta f)(x)\le v_B(x,0)+F(x,0)\le(\Delta f)(x),\forall x\in\overline{\Omega},\\ (\nabla A(v_B)(x),x-x_0)\ge 0,\forall x\in\overline{\Omega}.\end{array}\right.$$

Denote $Int(G_{m,k})$ the open set of interior points of the set $G_{m,k}$. The convexity of the set $G_{m,k}$ can be proven similarly with Proposition 1.

2.6.1 The case of $H_0^{6,k}(Q_T)$

As in (37), let $w^*\in G^*$ be the exact solution of the problem (29) with $F:=F^*$. Let $w_k^*=Z_k^6(w^*)$. Then there exists a function $\theta(k)>0$ such that

$$\|w^*-w_k^*\|_{H^6(Q_T)}\le\theta(k),\lim_{k\to\infty}\theta(k)=0. \tag{41}$$

Consider the following analog $J_{\lambda,\alpha,k}:G_{6,k}\to R$ of the functional $J_{\lambda,\alpha}$

$$J_{\lambda,\alpha,k}(B)=\int\limits_{Q_T}\left[A(w_B)(w_B+F)_{tt}-\Delta(w_B+F)\right]^2\varphi_\lambda^2dxdt+\alpha\|w\|_{H^6(Q_T)}^2,B\in G_{6,k}.$$

Theorem 5. *Let η and T be numbers of Theorem 1, λ_0 be the number of Theorem 2 and let $Int(G_{6.k})\neq\varnothing$. Then for all $\lambda\ge\lambda_0,\alpha\ge 2C\exp(-\lambda N)$ the functional $J_{\lambda,\alpha,k}(B)$ is strongly convex on the set $G_{6,k}$ and estimate (31) holds for $J_{\lambda,\alpha,k}(B)$, where $w_1:=w_{B_1},w_2:=w_{B_2},\forall B_1\in Int(G_{6,k}),\forall B_2\in G_{6,k}$.*

Let $\sigma>0$ be a number which will be chosen in Theorem 6. Consider and arbitrary point $B_1\in Int(G_{6.k})$ and define the gradient method for the functional $J_{\lambda,\alpha,k}$ as

$$B_{n+1}=B_n-\sigma\nabla J_{\lambda,\alpha,k}(B_n),n=1,2,... \tag{42}$$

Theorem 6 assures the global convergence of the gradient method (42).

Theorem 6. *Let η and T be numbers of Theorem 1, λ_0 be the number of Theorem 2 and let $Int(G_{6.k})\neq\varnothing$. Let $\lambda\ge\lambda_0,\alpha\ge 2C\exp(-\lambda N)$. Assume that the functional $J_{\lambda,\alpha,k}$*

achieves its minimal value on the set $G_{6,k}$ at a point $B_{\min} \in Int\left(G_{6.k}\right)$. Then the point $B_{\min}$ is unique. Assume that in the sequence (42) $\{B_n\}_{n=1}^{\infty} \subset Int\left(G_{6.k}\right)$. Then there exists a sufficiently small number $\sigma = \sigma\left(\lambda, \alpha, T, b, R, G_{6,k}\right) \in (0,1)$ and a number $q = q\left(\sigma\right) \in (0,1)$, both dependent on only listed parameters but independent on k, such that the sequence (42) converges to the point $B_{\min}$,

$$\left|B_{n+1} - B_{\min}\right| \leq q^n \left|B_1 - B_{\min}\right|. \tag{43}$$

We now consider the case of an error in the data (subsection 2.4) and estimate distances between functions $w_n, w_{k,\min}$ and the exact solution w^*.

Theorem 7. *Let η and T be numbers of Theorem 1, λ_0 be the number of Theorem 2, let $Int\left(G_{6.k}\right) \neq \varnothing$ and be valid. Assume that the function F is given with an error of the level $\delta > 0$, i.e. $\left\|F - F^*\right\|_{C^3\left(\overline{Q}_T\right)} \leq \delta$, $\delta \in (0, \delta_0)$, where the number $\delta_0 \in (0,1)$ is so small that $\delta_0 \leq \min\left(C_2 R, (1+b)^{-1}\xi/2\right)$ and also $\lambda_1 = \ln\left(\delta_0^{-1/(2M)}\right) \geq \lambda_0$. Choose $\lambda = \lambda\left(\delta\right) = \ln\left(\delta^{-1/(2M)}\right)$ and $\alpha = \alpha\left(\delta\right) = 2C\delta^{N/(2M)}$. Also, let the dimension $k^2 = k^2\left(\delta\right)$ of the space $H_0^{6,k}\left(Q_T\right)$ be so large that $\theta\left(k\left(\delta\right)\right) \leq \delta$, where the function $\theta\left(k\right)$ is defined in (41). Assume that conditions of Theorem 6 about $B_1, B_{\min}$ and the sequence (42) hold. Denote $w_{k,\min} = w_{B_{\min}}$. Then for the same numbers σ and q as in Theorem 6*

$$\left\|w_{k,\min} - w^*\right\|_{H^1(P_d)} \leq C\delta^{\rho}, \rho = \min\left(\frac{1}{2}, \frac{N}{4M}\right), \tag{44}$$

$$\left\|w_{B_{n+1}} - w^*\right\|_{H^1(P_d)} \leq q^n \left|B_1 - B_{\min}\right| + C\delta^{\rho}, n = 1, 2, ... \tag{45}$$

In the case of errorless data with $\delta = 0$ we set $\lambda = \lambda\left(k\right) = \ln\left(\theta\left(k\right)^{-1/(2M)}\right) \geq \lambda_0$, $\alpha = \alpha\left(k\right) = 2C\theta^{\rho}\left(k\right)$ for sufficiently large k. Then estimates (44) and (45) are valid, where δ^{ρ} is replaced with $\theta^{\rho}\left(k\right)$.

2.6.2 The case of $H_0^{3,k}\left(Q_T\right)$

Consider the orthonormal basis $\{\phi_{i,3}\left(x\right)\psi_{j,3}\left(t\right)\}_{i,j=1}^{k}$ in $H_0^{3,k}\left(Q_T\right)$, see (40). Define the functional $\overline{J}_{\lambda} : G_{3,k} \to R$,

$$\overline{J}_{\lambda}\left(B\right) = \int\limits_{Q_T} \left[A\left(w_B\right)\left(w_B + F\right)_{tt} - \Delta\left(w_B + F\right)\right]^2 \varphi_{\lambda}^2 dxdt, B \in G_{3,k}. \tag{46}$$

Theorem 8 for the functional $\overline{J}_{\lambda}$ is an analog of Theorem 2. Unlike Theorems 5-7, the constant C_3 of Theorem 8 depends on k. On the other hand, allowing this dependence, enables us not to include the regularization term with α in (46).

Theorem 8. *Let η and T be numbers of Theorem 1 and let $Int\left(G_{3,k}\right) \neq \varnothing$. Assume that for $i, j = 1, ..., k$ functions $D^{\beta}\phi_{i,3}\left(x\right) \in L_{\infty}\left(\Omega\right), \psi_{j,3}^{(s)}\left(t\right) \in L_{\infty}\left(0, T\right)$, where $|\beta| \leq 3$ and $s =$*

0, 1, 2, 3. *Then there exists a sufficiently large number* $\lambda_2 = \lambda_2(\eta, T, b, R, G_{3,k}, k) > 1$ *depending only on listed parameters, such that for all* $\lambda \geq \lambda_2$ *the functional* $\overline{J}_\lambda(B)$ *is strongly convex on the set* $G_{3,k}$. *More precisely, there exists a constant* $C_3 = C_3(\eta, T, b, R, G_{3,k}, P_d, k) > 0$ *such that*

$$\overline{J}_\lambda(B_2) - \overline{J}_\lambda(B_1) - \left(\nabla \overline{J}'_\lambda(B_1), B_2 - B_1\right)_k \geq C_3 \exp(\lambda d) |B_2 - B_1|^2, \tag{47}$$

$\forall B_1 \in Int(G_{3,k}), \forall B_2 \in G_{3,k}$, *where* $(,)_k$ *is the scalar product in* $\mathbb{R}^{k^2}$. *In addition, the analog of Theorems 6 is valid, in which* λ_0 *and* $G_{6,k}$ *are replaced with* λ_2 *and* $G_{3,k}$ *respectively and the parameter* α *is not counted.*

3 Proof of Theorem 1

In this proof $C_1 > 0$ denotes different positive constants depending on the same parameters as in the formulation of this theorem. First, we estimate from the below the term with the function g in (14). We have

$$u(x,0) = u(x,t) - \int_0^t u_t(x,\tau)\, d\tau. \tag{48}$$

Hence,

$$g(x,t)\, u(x,0)\, u_{tt}(x,t)\, \varphi_\lambda^2 = \left((g u_{tt} u)(x,t) - (g u_{tt})(x,t) \int_0^t u_t(x,\tau)\, d\tau \right) \varphi_\lambda^2(x,t). \tag{49}$$

Since $\varphi_\lambda^2(x,t) = \exp\left[2\lambda\left(|x - x_0|^2 - \eta t^2\right)\right]$, then, applying the Cauchy inequality, we obtain

$$\begin{aligned}(g u_{tt} u)(x,t)\, \varphi_\lambda^2(x,t) &= \partial_t\left[g u_t u \varphi_\lambda^2\right] - g u_t^2 \varphi_\lambda^2 - g_t u_t u \varphi_\lambda^2 + 4\lambda\eta t g u_t u \varphi_\lambda^2 \\ &\geq \partial_t\left[g u_t u \varphi_\lambda^2\right] - C_1 u_t^2 \varphi_\lambda^2 - C_1 \lambda^2 u^2 \varphi_\lambda^2. \end{aligned} \tag{50}$$

Now we estimate the term with the integral in (49),

$$-\varphi_\lambda^2 (g u_{tt})(x,t) \int_0^t u_t(x,\tau)\, d\tau$$

$$= \partial_t \left(-g u_t \varphi_\lambda^2 \int_0^t u_t(x,\tau)\, d\tau \right) + g u_t^2 \varphi_\lambda^2 + g_t u_t \varphi_\lambda^2 \int_0^t u_t(x,\tau)\, d\tau - 4\lambda\eta t g u_t \varphi_\lambda^2 \int_0^t u_t(x,\tau)\, d\tau$$

$$\geq \partial_t \left[-g u_t \varphi_\lambda^2 \int_0^t u_t(x,\tau)\, d\tau \right] - C_1 u_t^2 \varphi_\lambda^2 - C_1 \left(\int_0^t u_t(x,\tau)\, d\tau \right)^2 \varphi_\lambda^2 \tag{51}$$

$$-C_1\lambda\beta u_t^2\varphi_\lambda^2 - \frac{C_1}{\beta}\lambda\left(\int\limits_0^t u_t\left(x,\tau\right)d\tau\right)^2\varphi_\lambda^2$$

$$\geq \partial_t\left[-gu_t\varphi_\lambda^2\int\limits_0^t u_t\left(x,\tau\right)d\tau\right] - C_1\lambda\beta\left(1+\frac{1}{\lambda\beta}\right)u_t^2\varphi_\lambda^2$$

$$-\frac{C_1}{\beta}\lambda\left(1+\frac{\beta}{\lambda}\right)\left(\int\limits_0^t u_t\left(x,\tau\right)d\tau\right)^2\varphi_\lambda^2,$$

where $\beta > 0$ is a number which will be chosen later, and it is independent on λ. Here we have used the so-called "Cauchy inequality with the parameter β", i.e. $2ab \geq -\beta a^2 - b^2/\beta, \forall a,b\in\mathbb{R}, \forall\beta>0$. Summing up (50) and (51), comparing this sum with (49) and taking $\lambda > \max\left(\beta, 1/\beta\right)$, we obtain

$$gu\left(x,0\right)u_{tt}\varphi_\lambda^2 \geq \partial_t\left(gu_t u\varphi_\lambda^2 - gu_t\varphi_\lambda^2\int\limits_0^t u_t\left(x,\tau\right)d\tau\right) \tag{52}$$

$$-C_1\lambda\beta u_t^2\varphi_\lambda^2 - C_1\lambda^2u^2\varphi_\lambda^2 - \frac{C_1}{\beta}\lambda\left(\int\limits_0^t u_t\left(x,\tau\right)d\tau\right)^2\varphi_\lambda^2.$$

Integrating (52) over Q_T and recalling that $u_t\left(x,0\right)=0$, we obtain

$$\int\limits_{Q_T} gu\left(x,0\right)u_{tt}\varphi_\lambda^2dxdt \geq \int\limits_{\Omega_T} gu_tu\varphi_\lambda^2dx - \int\limits_{\Omega_T}\left(gu_t\int\limits_0^T u_t\left(x,\tau\right)d\tau\right)\varphi_\lambda^2dx$$

$$-C_1\int\limits_{Q_T}\left(\lambda\beta u_t^2+\lambda^2u^2\right)\varphi_\lambda^2dxdt - \frac{C_1}{\beta}\lambda\int\limits_{Q_T}\left(\int\limits_0^t u_t\left(x,\tau\right)d\tau\right)^2\varphi_\lambda^2dxdt. \tag{53}$$

Since $\varphi_\lambda^2\mid_{\Omega_T}\leq\exp\left(-2\lambda N\right)$, then, applying the Cauchy inequality, we obtain

$$\int\limits_{\Omega_T} gu_tu\varphi_\lambda^2dx - \int\limits_{\Omega_T}\left(gu_t\int\limits_0^T u_t\left(x,\tau\right)d\tau\right)\varphi_\lambda^2dx$$

$$\geq -C_1\exp\left(-2\lambda N\right)\int\limits_{\Omega_T}\left(u_t^2+u^2\right)dx - C_1\exp\left(-2\lambda N\right)\int\limits_{\Omega_T}\left(\int\limits_0^T u_t\left(x,\tau\right)d\tau\right)^2dx \tag{54}$$

$$\geq -C_1 \exp\left(-2\lambda N\right) \int\limits_{\Omega_T} \left(u_t^2 + u^2\right) dx - C_1 \exp\left(-2\lambda N\right) T \int\limits_{Q_T} u_t^2 dxdt.$$

Next, since $\varphi_\lambda^2 \left(x,t\right) \geq \exp\left(2\lambda d\right)$ for $\left(x,t\right) \in P_d$,

$$\exp\left(2\lambda d\right) > \exp\left(-2\lambda N\right) T \text{ and } \exp\left(-2\lambda N\right) T < \exp\left(-\lambda N\right) \text{ for } \lambda \geq \overline{\lambda},$$

then

$$-C_1 \exp\left(-2\lambda N\right) T \int\limits_{Q_T} u_t^2 dxdt \geq -C_1 \int\limits_{P_d} u_t^2 \varphi_\lambda^2 dxdt - C_1 \exp\left(-\lambda N\right) \int\limits_{Q_T \diagdown P_d} u_t^2 dxdt \tag{55}$$

$$\geq -C_1 \int\limits_{Q_T} u_t^2 \varphi_\lambda^2 dxdt - C_1 \exp\left(-\lambda N\right) \int\limits_{Q_T} u_t^2 dxdt.$$

The inequality of the second line of (55) follows from the inequality of the first line, since $P_d \subset Q_T$ and $Q_T \diagdown P_d \subset Q_T$. By lemma 1.10.3 of [4] the following estimate holds

$$\int\limits_{Q_T} \left(\int\limits_0^t u_t \left(x,\tau\right) d\tau \right)^2 \varphi_\lambda^2 dxdt \leq \frac{C_1}{\lambda} \int\limits_{Q_T} u_t^2 \varphi_\lambda^2 dxdt. \tag{56}$$

Combining (53), (54), (55) and (56) and taking $\lambda > 1/\beta^2$, we obtain

$$\int\limits_{Q_T} gu\left(x,0\right) u_{tt} \varphi_\lambda^2 dxdt \geq -C_1 \int\limits_{Q_T} \left(\lambda \beta u_t^2 + \lambda^2 u^2\right) \varphi_\lambda^2 dxdt \tag{57}$$

$$-C_1 \exp\left(-2\lambda N\right) \int\limits_{\Omega_T} \left(u_t^2 + u^2\right) dx - C_1 \exp\left(-\lambda N\right) \int\limits_{Q_T} u_t^2 dxdt.$$

The Carleman estimate of Theorem 1.10.2 of [4] for the hyperbolic operator $c\left(x\right) \partial_t^2 - \Delta$ leads to

$$\int\limits_{Q_T} \left(c\left(x\right) u_{tt} - \Delta u\right)^2 \varphi_\lambda^2 dxdt + C_1 \lambda^3 \exp\left(2\lambda M\right) \left(\left\| \partial_\nu u \mid_{S_T} \right\|_{L_2(S_T)}^2 + \left\| u \right\|_{H^1(S_T)}^2 \right) \tag{58}$$

$$+C_1 \lambda^3 \exp\left(-2\lambda N\right) \left(\left\| u_t \right\|_{L_2(\Omega_T)}^2 + \left\| u \right\|_{H^1(\Omega_T)}^2 \right) \geq C_1 \int\limits_{Q_T} \left(\lambda \left(\nabla u\right)^2 + \lambda u_t^2 + \lambda^3 u^2\right) \varphi_\lambda^2 dxdt.$$

Sum up (57) with (58) and choose $\beta = 1/2$. Then we obtain (14). The assertion of this theorem about $c\left(x\right) \equiv 1$ follows from the above and Corollary 1.10.2 of [4]. $\square$

4 Proofs of Theorem 2 and Corollary 1

Proof of Theorem 2. Consider two arbitrary functions $w_1 \in Int(G), w_2 \in G$. Let $h = w_2 - w_1$. Since $h \in H_0^6(Q_T)$, then

$$h_t(x,0) = 0, h\mid_{S_T} = \partial_\nu h\mid_{S_T} = 0. \tag{59}$$

Next, the triangle inequality combined with the second line of (28) implies that $\|h\|_{H^6(Q_T)} \leq 2R$. Hence, (26) leads to

$$\|h\|_{C^3(\overline{Q}_T)} \leq C. \tag{60}$$

First, we evaluate the difference $J_{\lambda,\alpha}(w_1 + h) - J_{\lambda,\alpha}(w_1)$ and single out the linear term with respect to h, since this term is $J'_{\lambda,\alpha}(w_1)(h)$. Denote

$$I_1 = \left[A(w_1 + h)(w_1 + h + F)_{tt} - \Delta(w_1 + h + F)\right]^2. \tag{61}$$

Then

$$J_{\lambda,\alpha}(w_2) = J_{\lambda,\alpha}(w_1 + h) = \int\limits_{Q_T} I_1 \varphi_\lambda^2 dxdt + \alpha\|w_1 + h\|_{H^6(Q_T)}^2. \tag{62}$$

By (27)

$$A(w_1 + h)$$
$$= (\Delta f)(x)\left(\frac{1}{(w_1 + F)(x,0)} - \frac{h(x,0)}{(w_1 + F)^2(x,0)} + \frac{h^2(x,0)}{(w_1 + F)^2(x,0)(w_1 + F + h)(x,0)}\right).$$

Hence,

$$A(w_1 + h) = A(w_1) - \frac{A(w_1)}{(w_1 + F)(x,0)}h(x,0) + \frac{A(w_1 + h)}{(w_1 + F)^2(x,0)}h^2(x,0).$$

Hence,

$$A(w_1 + h)(w_1 + h + F)_{tt} = A(w_1 + h)(w_1 + F)_{tt} + A(w_1 + h)h_{tt} = Q_1 + Q_2.$$

$$Q_1 = A(w_1 + h)(w_1 + h + F)_{tt} = A(w_1)(w_1 + F)_{tt}$$
$$-\frac{A(w_1)}{(w_1 + F)(x,0)}(w_1 + F)_{tt}h(x,0) + \frac{A(w_1 + h)}{(w_1 + F)^2(x,0)}(w_1 + F)_{tt}h^2(x,0).$$

$$Q_2 = A(w_1 + h)h_{tt} = A(w_1)h_{tt} - \frac{A(w_1)}{(w_1 + F)(x,0)}h(x,0)h_{tt} + \frac{A(w_1 + h)}{(w_1 + F)^2(x,0)}h_{tt}h^2(x,0).$$

Summing up Q_1 and Q_2, we obtain

$$A(w_1 + h)(w_1 + h + F)_{tt} = Q_1 + Q_2$$

$$= A(w_1)(w_1 + F)_{tt} + A(w_1)h_{tt} - \frac{A(w_1)}{(w_1 + F)(x,0)}(w_1 + F)_{tt}h(x,0)$$

$$-\frac{A(w_1)}{(w_1+F)(x,0)} h(x,0)\, h_{tt} + \frac{A(w_1+h)}{(w_1+F)^2(x,0)} \left[(w_1+F)_{tt} + h_{tt}\right] h^2(x,0).$$

Hence, by (61)

$$I_1 =$$

$$\left\{ \left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right] + \left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right) + Z \right\}^2,$$

$$Z = -\frac{A(w_1)}{(w_1+F)(x,0)} h(x,0)\, h_{tt} + \frac{A(w_1+h)}{(w_1+F)^2(x,0)} \left[(w_1+F)_{tt} + h_{tt}\right] h^2(x,0). \tag{63}$$

Hence,

$$I_1 = \left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right]^2$$

$$+2\left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right] \left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right)$$

$$+2\left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right] Z$$

$$+\left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right)^2$$

$$+2\left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right) Z + Z^2.$$

Let $I_{1,linear}$ be the part of I_1, which is linear with respect to h,

$$I_{1,linear} = 2\left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right] \tag{64}$$

$$\times \left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right).$$

Hence, $I_{1,linear}$ generates the Fréchet derivative,

$$J'_{\lambda,\alpha}(w_1)(h) = \int\limits_{Q_T} I_{1,linear} \varphi_\lambda^2 dxdt + 2\alpha\left[w_1, h\right], \tag{65}$$

where $[,]$ is the scalar product in $H^6(P_d)$. Denote

$$I_2 = I_1 - \left[A(w_1)(w_1+F)_{tt} - \Delta(w_1+F)\right]^2 - I_{1,linear}. \tag{66}$$

Ignoring the term with Z^2 in I_2 and using the Cauchy inequality, we obtain

$$I_2 \geq \left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right)^2$$

$$+2\left(A(w_1) h_{tt} - \Delta h - \frac{A(w_1)}{(w_1+F)(x,0)} (w_1+F)_{tt}\, h(x,0) \right) Z$$

$$+2\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]Z$$

$$\geq\frac{1}{2}\left(A\left(w_{1}\right)h_{tt}-\Delta h\right)^{2}-Ch^{2}\left(x,0\right) \tag{67}$$

$$+2\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)Z$$

$$+2\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]Z.$$

We now use the expression (63) for Z,

$$2\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)Z$$

$$=-\left[\frac{2A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)\right]h\left(x,0\right)h_{tt}$$

$$+2\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)\frac{A\left(w_{1}+h\right)}{\left(w_{1}+F\right)^{2}\left(x,0\right)}\left[\left(w_{1}+F\right)_{tt}+h_{tt}\right]$$

$$\times h^{2}\left(x,0\right).$$

Hence,

$$2\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)Z \tag{68}$$

$$\geq-\left\{\frac{2A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)\right\}h\left(x,0\right)h_{tt}$$

$$-Ch^{2}\left(x,0\right),$$

Similarly

$$2\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]Z$$

$$\geq-\left\{\frac{2A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]\right\}h\left(x,0\right)h_{tt}-Ch^{2}\left(x,0\right). \tag{69}$$

Denote

$$g\left(x,t\right)=-\left\{\frac{2A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(A\left(w_{1}\right)h_{tt}-\Delta h-\frac{A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left(w_{1}+F\right)_{tt}h\left(x,0\right)\right)\right\}$$

$$-\left\{\frac{2A\left(w_{1}\right)}{\left(w_{1}+F\right)\left(x,0\right)}\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]\right\}. \tag{70}$$

It follows from (24), (28) and (70) that functions $g,g_{t}\in C\left(\overline{Q}_{T}\right)$ and $\|g\|_{C\left(\overline{Q}_{T}\right)},\|g_{t}\|_{C\left(\overline{Q}_{T}\right)}\leq C$. Combining (48) with (66)-(69), we obtain

$$I_{1}-\left[A\left(w_{1}\right)\left(w_{1}+F\right)_{tt}-\Delta\left(w_{1}+F\right)\right]^{2}-I_{1,linear}$$

$$\geq \frac{1}{2}\left(A\left(w_1\right)h_{tt}-\Delta h\right)^2 - Ch^2\left(x,t\right) - C\left(\int\limits_0^t h_t\left(x,\tau\right)d\tau\right)^2 + g\left(x,t\right)h\left(x,0\right)h_{tt}.$$

Hence, (30) and (65) imply that

$$J_{\lambda,\alpha}\left(w_1+h\right) - J_{\lambda,\alpha}\left(w_1\right) - J'_{\lambda,\alpha}\left(w_1\right)\left(h\right)$$

$$\geq \frac{1}{2}\int\limits_{Q_T}\left(A\left(w_1\right)h_{tt}-\Delta h\right)^2\varphi_\lambda^2 dxdt + \int\limits_{Q_T} g\left(x,t\right)h\left(x,0\right)h_{tt}\varphi_\lambda^2 dxdt$$

$$-C\int\limits_{Q_T} h^2\varphi_\lambda^2 dxdt - C\int\limits_{Q_T}\left(\int\limits_0^t h_t\left(x,\tau\right)d\tau\right)^2\varphi_\lambda^2 dxdt + \alpha\left\|h\right\|_{H^6(Q_T)}^2.$$

Applying Theorem 1, (56) and (59), we obtain

$$J_{\lambda,\alpha}\left(w_1+h\right) - J_{\lambda,\alpha}\left(w_1\right) - J'_{\lambda,\alpha}\left(w_1\right)\left(h\right) \geq C\int\limits_{Q_T}\left(\lambda h_t^2 + \lambda\left(\nabla h\right)^2 + \lambda^3 h^2\right)\varphi_\lambda^2 dxdt \tag{71}$$

$$-C\exp\left(-2\lambda N\right)\left(\left\|h_t\right\|_{L_2(\Omega_T)}^2 + \left\|h\right\|_{H^1(\Omega_T)}^2\right) - C\exp\left(-\lambda N\right)\left\|h_t\right\|_{L_2(Q_T)}^2 + \alpha\left\|h\right\|_{H^6(Q_T)}^2$$

Since

$$\left\|h_t\right\|_{L_2(\Omega_T)}^2 + \left\|h\right\|_{H^1(\Omega_T)}^2 + \left\|h_t\right\|_{L_2(Q_T)}^2 \leq C\left\|h\right\|_{H^6(Q_T)}^2 \text{ and } \alpha \geq 2C\exp\left(-\lambda N\right),$$

then, using (71), we obtain

$$J_{\lambda,\alpha}\left(w_1+h\right) - J_{\lambda,\alpha}\left(w_1\right) - J'_{\lambda,\alpha}\left(w_1\right)\left(h\right)$$

$$\geq C\int\limits_{Q_T}\left(\lambda h_t^2 + \lambda\left(\nabla h\right)^2 + \lambda^3 h^2\right)\varphi_\lambda^2 dxdt + \frac{\alpha}{2}\left\|h\right\|_{H^6(Q_T)}^2, \tag{72}$$

$\forall w_1 \in Int\left(G\right), \forall w_2 = w_1 + h \in G$, which is (31). Next, since $P_d \subset Q_T$ and $\varphi_\lambda^2 \mid_{P_d} \geq \exp\left(2\lambda d\right)$, then (72) implies (32). $\square$

Proof of Corollary 1. Estimate (34) follows immediately from (33) and (71). $\square$

5 Proofs of Theorems 3-8

Proof of Theorem 3. Recall that $Y\left(w\right) = A\left(w\right)\left(w+F\right)_{tt} - \Delta\left(w+F\right)$ is the operator in (29) and $Y: G \to L_2^\lambda\left(Q_T\right)$ (subsection 2.6). The bounded linear operator of the Fréchet derivative $Y'\left(w\right): H_0^6\left(Q_T\right) \to L_2^\lambda\left(Q_T\right)$. Let $L\left(H_0^6\left(Q_T\right), L_2^\lambda\left(Q_T\right)\right)$ be the space of bounded linear operators mapping $H_0^6\left(Q_T\right)$ in $L_2^\lambda\left(Q_T\right)$. It follows from results of section 8.2 of [2] that if we would prove that the norm of this operator $\|Y'(w)\|_{L\left(H_0^6(Q_T),L_2^\lambda(Q_T)\right)}$ is uniformly bounded for all $w \in Int(G)$ and that the map $Y'(w) : Int(G) \to L\left(H_0^6(Q_T), L_2^\lambda(Q_T)\right)$ is Lipschitz continuous on the set $Int(G)$, then Theorem 2 combined with conditions of Theorem 3 and the convexity of the set G (Proposition 1) would imply that the assertion of Theorem 3 is true. It follows from the above expression (64) for $I_{1,linear}$ that

$$Y'(w)(h) = A(w)h_{tt} - \Delta h - \frac{A(w)}{(w+F)(x,0)}(w+F)_{tt}h(x,0), \forall w \in Int(G), \forall h \in H_0^6(Q_T).$$

Hence,

$$\|Y'(w)(h)\|_{L_2^\lambda(Q_T)} \le C\exp(\lambda M)\|h\|_{H^6(Q_T)}, \forall h \in H_0^6(Q_T), \forall w \in Int(G).$$

Hence,

$$\|Y'(w)\|_{L\left(H_0^6(Q_T),L_2^\lambda(Q_T)\right)} \le C\exp(\lambda M), \forall w \in Int(G). \tag{73}$$

To prove the Lipschitz continuity of the map $Y'(w) : Int(G) \to L\left(H_0^6(Q_T), L_2^\lambda(Q_T)\right)$, estimate the norm

$$\|Y'(w_1)(h) - Y'(w_2)(h)\|_{L_2^\lambda(Q_T)}, \forall w_1, w_2 \in Int(G), \forall h \in H_0^6(Q_T).$$

We have for $(x,t) \in Q_T$

$$|Y'(w_1)(h) - Y'(w_2)(h)|$$

$$= \left|(A(w_1) - A(w_2))h_{tt} + \left(\frac{A(w_2)}{(w_2+F)(x,0)}(w_2+F)_{tt} - \frac{A(w_1)}{(w_1+F)(x,0)}(w_2+F)_{tt}\right)h(x,0)\right|$$

$$\le C(|w_1 - w_2| + |w_{1tt} - w_{2tt}|)(|h_{tt}| + |h(x,0)|) \le C\|w_1 - w_2\|_{H^6(Q_T)}\|h\|_{H^6(Q_T)}.$$

Hence,

$$\|Y'(w_1) - Y'(w_2)\|_{L\left(H_0^6(Q_T),L_2^\lambda(Q_T)\right)} \le C\exp(\lambda M)\|w_1 - w_2\|_{H^6(Q_T)}, \forall w_1, w_2 \in Int(G). \tag{74}$$

Thus, (73) and (74) ensure that the operator $Y'(w)$ of the Fréchet derivative is uniformly bounded and the map $w \to Y'(w) \in L\left(H_0^6(Q_T), L_2^\lambda(Q_T)\right)$ is Lipschitz continuous on the set $Int(G)$. $\square$

Proof of Theorem 4. Let the function $w \in Int(G)$. Since by (37) $w^* \in G^*$, then estimate (14) and a slight modification of the proof of Theorem 2 lead to the following analog of (32)

$$J_{\lambda,\alpha}(w^*) - J_{\lambda,\alpha}(w) - J'_{\lambda,\alpha}(w)(w^* - w) \ge C\exp(2\lambda d)\|w^* - w\|_{H^1(P_d)}^2$$

$$-C\lambda^3\exp(-2\lambda N)\left(\|\partial_t(w^* - w)\|_{L_2(\Omega_T)}^2 + \|w^* - w\|_{H^1(\Omega_T)}^2\right) \tag{75}$$

$$-C\exp(-\lambda N)\|w^* - w\|_{L_2(Q_T)}^2 + \alpha\|w^* - w\|_{H^6(Q_T)}^2.$$

Here the function F, rather than F^*, is involved in $J_{\lambda,\alpha}(w^*)$. Also, we use the fact that by (7), (26) and (28) for all $x \in \overline{\Omega}$

$$w^*(x,0) + F(x,0) = w^*(x,0) + F^*(x,0) + (F - F^*)(x,0) \leq (\Delta f)(x) + \delta \leq 2C_2 R,$$

$$w^*(x,0) + F(x,0) = w^*(x,0) + F^*(x,0) + (F - F^*)(x,0)$$
$$\geq (1+b)^{-1}(\Delta f)(x) - \delta \geq (1+b)^{-1}\xi - \delta \geq (1+b)^{-1}\xi/2.$$

Using the same arguments as ones in the end of the proof of Theorem 2, we conclude that terms of (75) with $\exp(-2\lambda N)$ and $\exp(-\lambda N)$ are absorbed by the term $\alpha \|w^* - \overline{w}\|^2_{H^6(Q_T)}/2$. Hence, we obtain

$$J_{\lambda,\alpha}(w^*) - J_{\lambda,\alpha}(w) - J'_{\lambda,\alpha}(w)(w^* - w) \geq C \exp(2\lambda d) \|w^* - w\|^2_{H^1(P_d)}. \tag{76}$$

We now estimate $J_{\lambda,\alpha}(w^*)$ from the above. Let the functional $J^*_{\lambda,\alpha}(w^*)$ be obtained from $J_{\lambda,\alpha}(w^*)$ via replacement in (30) F with F^*. Since the function w^* is the solution of the problem (29) with $F := F^*$, then $J^*_{\lambda,\alpha}(w^*) = \alpha \|w^*\|^2_{H^6(Q_T)}$. Hence, representing F in $J_{\lambda,\alpha}(w^*)$ as $F = F^* + (F - F^*)$ and using $\|F - F^*\|_{C^3(\overline{Q}_T)} \leq \delta$, we obtain $J_{\lambda,\alpha}(w^*) \leq C\exp(2\lambda M)\delta^2 + \alpha R^2$. Hence, using $J'_{\lambda,\alpha}(w_{\min}) = 0$, we obtain from (76)

$$\|w^* - w_{\min}\|^2_{H^1(P_d)} \leq C\delta^2 \exp(2\lambda M) + \alpha R^2. \tag{77}$$

Choose $\lambda = \lambda(\delta)$ and $\alpha = \alpha(\delta)$ as in the formulation of this theorem. Then $C\delta^2\exp(2\lambda M) + \alpha R^2 \leq C\left(\delta + \delta^{N/(2M)}\right) \leq C\delta^{2\rho}$. Hence, (77) implies that $\|w^* - w_{\min}\|_{H^1(P_d)} \leq C\delta^{\rho}$, which is (38). By the triangle inequality $\|w_{n+1} - w_{\min}\|_{H^1(P_d)} \geq \|w_{n+1} - w^*\|_{H^1(P_d)} - \|w^* - w_{\min}\|_{H^1(P_d)}$. Hence, combining (36) with (38), we obtain $\|w_{n+1} - w^*\|_{H^1(P_d)} \leq q^n \|w_1 - w_{\min}\|_{H^6(Q_T)} + C\delta^{\rho}$, which is (39). Considerations in the case of errorless data are similar. □

Proofs of Theorems 5,6. Theorems 5 and 6 follow immediately from Theorems 2 and 3 respectively. □

Proof of Theorem 7. Denote $w_{k,\min} = w_{B_{\min}}$. As in Theorem 3,

$$\left\|w_{B_{n+1}} - w_{k,\min}\right\|_{H^1(P_d)} \leq \left\|w_{B_{n+1}} - w_{k,\min}\right\|_{H^6(Q_T)} = |B_{n+1} - B_{\min}| \leq q^n |B_1 - B_{\min}|. \tag{78}$$

We now prove an analog of estimate (38). Recall that $w^*_k = Z^6_k(w^*)$, i.e. the function w^*_k is the orthogonal projection of the function w^* on the k^2-dimensional subspace $H^{6,k}_0(Q_T)$ of the space $H^6_0(Q_T)$. Similarly with (76) we obtain

$$J_{\lambda,\alpha}(w^*_k) \geq \|w^* - w_{k,\min}\|^2_{H^1(P_d)}. \tag{79}$$

We now estimate $J_{\lambda,\alpha}(w^*_k)$ from the above. We have

$$J_{\lambda,\alpha,k}(w^*_k) = \widetilde{J}_{\lambda,k}(w^*_k) + \alpha \|w^*_k\|^2_{H^6(Q_T)}, \tag{80}$$

$$\widetilde{J}_{\lambda,k}(w^*_k) = \int\limits_{Q_T} \left[A(w^*_k)(w^*_k + F)_{tt} - \Delta(w^*_k + F)\right]^2 \varphi^2_{\lambda} dxdt. \tag{81}$$

It follows from (24) and (41) that

$$\left|\left[A\left(w_k^*\right)\left(w_k^*+F\right)_{tt}-\Delta\left(w_k^*+F\right)\right]-\left[A\left(w^*\right)\left(w^*+F^*\right)_{tt}-\Delta\left(w^*+F^*\right)\right]\right|\leq C\left(\delta+\theta\left(k\right)\right)\leq C\delta.$$

Since $A\left(w^*\right)\left(w^*+F^*\right)_{tt}-\Delta\left(w^*+F^*\right)=0$, then (81) implies that $\widetilde{J}_{\lambda,k}\left(w_k^*\right)\leq C\delta^2\exp\left(2\lambda M\right).$ Hence, by (79) and (80)

$$\left\|w^*-w_{k,\min}\right\|_{H^1(P_d)}^2\leq C\delta^2\exp\left(2\lambda M\right)+C\alpha. \tag{82}$$

Choose $\lambda=\lambda\left(\delta\right),\alpha=\alpha\left(\delta\right)$ as in conditions of this theorem. Then (82) implies that $\left\|w^*-w_{k,\min}\right\|_{H^1(P_d)}\leq C\delta^\rho$, which is (44). Next, by the triangle inequality

$$\left\|w_{B_{n+1}}-w_{k,\min}\right\|_{H^1(P_d)}\geq\left\|w_{B_{n+1}}-w^*\right\|_{H^1(P_d)}-\left\|w^*-w_{k,\min}\right\|_{H^1(P_d)}$$

$$\geq\left\|w_{B_{n+1}}-w^*\right\|_{H^1(P_d)}-C\delta^\rho.$$

Hence, $\left\|w_{B_{n+1}}-w^*\right\|_{H^1(P_d)}\leq\left\|w_{B_{n+1}}-w_{k,\min}\right\|_{H^1(P_d)}+C\delta^\rho$. Combining this with (78), we obtain the desired estimate (45). Considerations in the case of errorless data are similar. $\square$

Proof of Theorem 8. In this proof $C_3=C_3\left(\eta,b,\xi,G_{3,k},P_d,k\right)>0$ denotes different constants depending on listed parameters. For functions $w_B\in G_{3,k}$ the functional $\overline{J}_\lambda$ in (46) is the same as the functional $\widetilde{J}_\lambda$ in (33) with the only difference that $\overline{J}_\lambda$ depends on the vector B, whereas $\widetilde{J}_\lambda$ depends on the function w. Orthogonalize functions $\left\{\phi_{i,m}\left(x\right)\psi_{j,m}\left(t\right)\right\}_{i,j=1}^k$ in the space $H^1\left(P_d\right).$ Then, using (34), we obtain (47) via

$$\overline{J}_\lambda\left(B_2\right)-\overline{J}_\lambda\left(B_1\right)-\left(\nabla\overline{J}_\lambda\left(B_1\right),B_2-B_1\right)$$

$$\geq C_3\exp\left(2\lambda d\right)\left|B_2-B_1\right|^2-C_3\exp\left(-\lambda N\right)\left|B_2-B_1\right|^2$$

$$\geq C_3\exp\left(\lambda d\right)\left|B_2-B_1\right|^2,\forall B_1\in Int\left(G_{3,k}\right),\forall B_2\in G_{3,k}.\ \square$$

6 A Finite Element Method for the Reconstruction of the Coefficient $c(x)$

In this section we explain how to compute the coefficient $c(x)$ in (20) explicitly via finite elements, as soon as the solution w of the problem (29) is computed via the minimization of the functional (46) by the gradient method. In addition, we outline an algorithm of computing the minimizer of this functional using finite elements. This might be useful for computations. The space $H_0^{3,k}\left(Q_T\right)$ is used since it is easier to work with finite elements of the third order rather than with those of the sixth order of $H_0^{6,k}\left(Q_T\right).$ Thus, below $w=w_B,B\in G_{3,k}$, and we rely in Theorem 8. As basis functions $\left\{\phi_{i,3}\left(x\right)\right\}_{i=1}^k\subset H^3\left(\Omega\right),\left\{\psi_{j,3}\left(t\right)\right\}_{j=1}^k\subset H_0^3\left(0,T\right)$ in (40), we use standard finite elements of the third order. They can be orthogonalized in terms of the space $H^1\left(P_d\right)$, since the proof of Theorem 8 requires this: recall that the constant C_3 in this theorem depends on k.

Assume that a minimizer $B_{\min}$ of the functional $\overline{J}_{\lambda}(B)$ is an interior point of the set $G_{3,k}$. Consider an arbitrary point $B_1 \in Int(G_{3,k})$ as the initial guess and assume that all points obtained by the gradient method for the functional $\overline{J}_{\lambda}(B)$ belong to $Int(G_{3,k})$. Then, applying an analog of Theorem 6, we conclude that (43) is true, i.e. the gradient method results in computing the minimizer $B_{\min}$ as well as of the corresponding function $w_{B_{\min}}(x,t)$.

6.1 Spaces of finite elements

When minimizing the functional $\overline{J}_{\lambda}(B)$ in (46), we search for its stationary point satisfying

$$\nabla \overline{J}_{\lambda}(B) = 0, B \in Int(G_{3,k}). \tag{83}$$

Consider a triangulation of the domain Ω by non-overlapping tetrahedral elements $K_j \subset \Omega$. These elements form the mesh $Ms = \{K_j\}_{j=1}^{k_1}$, where k_1 is the total number of elements in Ω, and $\Omega = \cup_{j=1}^{k_1} K_j$. We also introduce the time discretization D_j of the time domain $(0,T)$ into subintervals $D_j = (t_{j-1}, t_j]$ of the uniform length $\tau = t_j - t_{j-1}, j = 1, ...k_2$. For each element $K_j \subset \Omega$ let $P_3(K_j)$ be the set of polynomials of the third degree defined on K_j. Similarly with section 76.4 of [8] and sections 3.1, 3.2 of [10], we define the finite element space $V_{x,h,0}$ as

$$V_{h,x,0} = \left\{ v \in H^3(\Omega) : v\mid_{K_j} \in P_3(K_j), j = 1, ...k_1; v\mid_{\partial\Omega} = \partial_{\nu} v \mid_{\partial\Omega} = 0 \right\}. \tag{84}$$

Similarly for each time subinterval D_j let $P_3(D_j)$ be the set of polynomials of the third degree defined on D_j. We introduce the finite element space $V_{h,t,0}$ as

$$V_{h,t,0} = \left\{ v \in H^3(0,T) : v\mid_{D_j} \in P_3(D_j), j = 1, ...k_2; v_t \mid_{t=0} = 0 \right\}. \tag{85}$$

To formulate the finite element method for (83) we introduce the finite element spaces $W_{h,0}$ as

$$W_{h,0} = V_{h,x,0} \times V_{h,t,0}. \tag{86}$$

It follows from (84), (85) and (86) that for a certain integer $k = k(k_1, k_2) > 0$

$$W_{h,0} \subset H_0^{3,k}(Q_T), \text{ i.e. } w_t(x,0) = 0, w\mid_{S_T} = 0, \partial_{\nu} w \mid_{S_T} = 0, \forall w \in W_{h,0}.$$

For brevity and without loss of generality assume that $\dim V_{h,x,0} = \dim V_{h,t,0} := k = k_1 = k_2$.Consider linear bases $\{\phi_i(x)\}_{i=1}^{k}$and $\{\psi_j(t)\}_{j=1}^{k}$ in spaces $V_{h,x,0}$ and $V_{h,t,0}$ respectively. Unlike section 2.6, products $\phi_i(x)\psi_j(t)$ are not orthonormal in $H^3(Q_T)$. Still, they are linearly independent. Thus, similarly with (40)

$$w(x,t) = w_B(x,t) = \sum_{i,j=1}^{k} b_{i,j}\phi_i(x)\psi_j(t), \forall (x,t) \in Q_T, \forall w \in W_{h,0}, \tag{87}$$

$$B = \{b_{1,1}, ..., b_{k,k}\}^T \in \mathbb{R}^{k^2}. \tag{88}$$

To approximate functions $c(x)$, we use the space of piecewise constant functions C_h,

$$C_h := \{v \in L_2(\Omega) : v|_{K_j} = v_j = const., j = 1, ..., k_1\}.$$

6.2 A finite element method for the coefficient $c(x)$

Let $[,]_2$ be the scalar product in $L_2(\Omega)$. To compute the function $c(x)$ we formulate the finite element method for (20) as: *Suppose that the function* $w(x,t) = w_{B_{\min}}(x,t) \in W_{h,0}$ *is computed as the solution of the problem (83). Assuming that the function* $f(x) \in H^7(\mathbb{R}^3)$ *is known, approximate the function* $c(x) \in C_h$ *such that*

$$[c(x)\, w(x,0), v]_2 = [\Delta f, v]_2\,, \forall v \in V_h. \tag{89}$$

We express the function $w(x,0)$ as

$$w(x,0) = \sum_{i=1}^{k} w_i \phi_i(x), \tag{90}$$

where w_i are numbers. Consider an auxiliary vector $\widetilde{c} = (\widetilde{c}_1, ..., \widetilde{c}_k)^T$ and assume for a moment that in (89)

$$c(x)\, w(x,0) = \sum_{i=1}^{k} \widetilde{c}_i w_i \phi_i(x). \tag{91}$$

Substituting (90) into (89) and choosing $v(x) = \phi_j(x)$, we obtain the following system of linear algebraic equations

$$\sum_{i=1}^{k} \widetilde{c}_i \left[w_i \phi_i, \phi_j\right]_2 = \sum_{i=1}^{k} \left[\Delta f, \phi_j\right]_2, j = 1, ..., k. \tag{92}$$

The system (92) can be rewritten in the matrix form for the unknown vector $\widetilde{c}$ and the known vector $w = (w_1, ..., w_k)^T$ as

$$Q\widetilde{c} = Z. \tag{93}$$

In (93) the matrix Q is the block mass matrix in space and Z is the load vector.

To obtain an explicit scheme for the computation of the vector $\widetilde{c}$, we approximate the matrix Q by the lumped mass matrix Q^L in space, i.e., the diagonal approximation is obtained by taking the row sum of Q [11]. Thus, we obtain following explicit formula for the computation of the vector $\widetilde{c}$:

$$\widetilde{c} = (Q^L)^{-1} Z. \tag{94}$$

Given $\widetilde{c}$ from (94), we approximate values of $c(x)$ on every tetrahedron K_j as

$$c_j := c|_{K_j} \approx \frac{1}{\widetilde{k}_j} \sum_{i=1}^{\widetilde{k}_j} \widetilde{c}_i, j = 1, ...k_1,$$

where $\widetilde{k}_j$ is the number of tetrahedra $K_i \in Ms$ which have at least one point of intersection with the boundary of the tetrahedron K_j, and k_1 is common number of elements in the mesh. Hence, K_j is among them. As to the numbers $\widetilde{c}_i$, we define them as components of the vector $\widetilde{c}$, which correspond to such functions $\phi_i(x)$ in (91), which are third order polynomials in those tetrahedra K_i. Thus, so defined vector $\overline{c} = (c_1, ..., c_k)^T$ represents a piecewise constant function $\overline{c}\,(x) \in C_h$, which we consider as an approximation for our target unknown coefficient $c\,(x)$. The lumping procedure does not include approximation errors in the case of linear Lagrange elements. For the case of higher order finite elements elements we refer to [11] for an approximation of the mass matrix by lumped mass matrix.

6.3 An outline of the algorithm

We now outline the algorithm of the minimization of the functional $\overline{J}_\lambda\,(B)$ using the gradient method. The condition (83) for the minimizer is

$$\nabla \overline{J}_\lambda(B_{\min}) = \int\limits_{Q_T} \nabla \left[A\left(w_{B_{\min}}\right)\left(w_{B_{\min}} + F\right)_{tt} - \Delta\left(w_{B_{\min}} + F\right)\right]^2 \varphi_\lambda^2 dxdt = 0, B_{\min} \in Int\,(G_{3,k}) \cap W_{h,0}. \tag{95}$$

Consider the vector $g^n \in \mathbb{R}^{k^2}$ defined as

$$g^n = \int\limits_{Q_T} \nabla \left[A\left(w_{B_n}\right)\left(w_{B_n} + F\right)_{tt} - \Delta\left(w_{B_n} + F\right)\right]^2 \varphi_\lambda^2 dxdt, w_{B_n} \in W_{h,0}, \tag{96}$$

where B_n is the vector B obtained at n iterations of the gradient method.

Algorithm

Step 0. Choose a point $B_1 \in Int\,(G_{3,k}) \cap W_{h,0}$ and the corresponding function w_1, which is computed via (87).

Step 1. Compute the vector g^n via (96).

Step 2. Update the vector B_n in the gradient method similarly with (42) as $B_{n+1} = B_n - \sigma g^n$ where σ is step-size in the gradient method. Also, obtain the corresponding function $w_{B_{n+1}}$ via (87).

Step 3. Stop computing vectors B_n if $|g^n| \leq \theta$. Otherwise set $n := n + 1$ and go to step 1. Here θ is the tolerance in the gradient method.

Step 4. Compute the function $c_h \in C_h$ via the finite element discretization as in (94).

Acknowledgments

This research was supported by US Army Research Laboratory and US Army Research Office grant number W911NF-11-1-0399, the Swedish Research Council and the Swedish Foundation for Strategic Research (SSF) through the Gothenburg Mathematical Modelling Centre (GMMC).

References

[1] V.V. Arestov, Approximation of unbounded operators by bounded operators and related extremal problems, *Russian Mathematical Surveys*, 51, 1093-1126, 1996.

[2] A. Bakushinsky, M. Y. Kokurin and A. Smirnova, *Iterative Methods for Ill-Posed Problems*, De Gruyter, 2011.

[3] L. Baudouin, M. de Buhan and S. Ervedoza, Global Carleman estimates for waves and applications, *Communications in Partial Differential Equations*, 38, 823-859, 2013.

[4] L. Beilina and M.V. Klibanov, *Approximate Global Convergence and Adaptivity for Coefficient Inverse Problems*, Springer, 2012.

[5] L. Beilina, Energy estimates and numerical verification of the stabilized domain decomposition finite element/finite difference approach for the Maxwell's system in time domain, *Central European Journal of Mathematics*, 11, 702-733, 2013.

[6] L. Beilina, N.T. Thánh, M.V. Klibanov and M.A. Fiddy, Reconstruction from blind experimental data for an inverse problem for a hyperbolic equation, *Inverse Problems*, accepted for publication, 2013.

[7] A.L. Bukhgeim and M.V. Klibanov, Uniqueness in the large of a class of multidimensional inverse problems, *Soviet Math. Doklady*, 17, 244-247, 1981.

[8] K. Eriksson, D. Estep amd C. Johnson, *Calculus in Several Dimensions*, Springer, Berlin, 2004.

[9] V. Isakov, *Inverse Problems for Partial Differential Equations*, Second Edition, Springer, New York, 2006.

[10] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Dover Publications, New York, 2009.

[11] P. Joly, *Variational Methods for Time Dependent Wave Propagation Problems*, Lecture Notes in Computational Science and Engineering, Springer, 2003.

[12] M.V. Klibanov and J. Malinsky, Newton-Kantorovich method for 3-dimensional potential inverse scattering problem and stability for the hyperbolic Cauchy problem with time dependent data, *Inverse Problems*, 7, 577-596, 1991.

[13] M.V. Klibanov, Global convexity in a three-dimensional inverse acoustic problem, *SIAM J. Math. Analysis*, 28, 1371-1388, 1997.

[14] M.V. Klibanov, Carleman estimates for global uniqueness, stability and numerical methods for coefficient inverse problems, survey, *J. Inverse and Ill-Posed Problems*, 21, 477-560, 2013.

[15] O.A. Ladyzhenskaya, *Boundary Value Problems of Mathematical Physics*, Springer, New York, 1985.

[16] R.G. Novikov, The inverse scattering problem on a fixed energy level for the two-dimensional Schrödinger operator, *J. Functional Analysis*, 103, 409-463, 1992.

[17] R.G. Novikov, $\partial-$bar approach to approximate inverse scattering at fixed energy in three dimensions, *International Math. Research Peports*, 6, 287-349, 2005.

[18] V.S. Vladimirov, *Equations of Mathematical Physics*, M. Dekker, New York, 1971.

[19] M. Yamamoto, Carleman estimates for parabolic equations. Topical Review. *Inverse Problems*, 25, 123013, 2009.