7-29-2014

Iterative Methods for the Elasticity Imaging Inverse Problem

Brian C. Winkler

Iterative Methods for the Elasticity Imaging Inverse Problem

by

Brian C. Winkler

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Applied Mathematics
School of Mathematical Sciences, College of Science

Rochester Institute of Technology
Rochester, NY

Date of Approval: July 29, 2014

Committee Approval:

R·I·T
College of
Science

Akhtar A. Khan — Date

School of Mathematical Sciences

Thesis Advisor

Baasansuren Jadamba — Date

School of Mathematical Sciences

Thesis Co-advisor

Patricia Clark — Date

School of Mathematical Sciences

Committee Member

Tamas Wiandt — Date

School of Mathematical Sciences

Committee Member

Nathan Cahill, D.Phil. — Date

School of Mathematical Sciences

Director of Graduate Programs

Abstract

Cancers of the soft tissue reign among the deadliest diseases throughout the world and effective treatments for such cancers rely on early and accurate detection of tumors within the interior of the body. One such diagnostic tool, known as elasticity imaging or elastography, uses measurements of tissue displacement to reconstruct the variable elasticity between healthy and unhealthy tissue inside the body. This gives rise to a challenging parameter identification inverse problem, that of identifying the Lamé parameter μ in a system of partial differential equations in linear elasticity. Due to the near incompressibility of human tissue, however, common techniques for solving the direct and inverse problems are rendered ineffective due to a phenomenon known as the "locking effect". Alternative methods, such as mixed finite element methods, must be applied to overcome this complication. Using these methods, this work reposes the problem as a generalized saddle point problem along with a presentation of several optimization formulations, including the modified output least squares (MOLS), energy output least squares (EOLS), and equation error (EE) frameworks, for solving the elasticity imaging inverse problem. Subsequently, numerous iterative optimization methods, including gradient, extragradient, and proximal point methods, are explored and applied to solve the related optimization problem. Implementations of all of the iterative techniques under consideration are applied to all of the developed optimization frameworks using a representative numerical example in elasticity imaging. A thorough analysis and comparison of the methods is subsequently presented.

Contents

I. Introduction

I.1 Problem Background

Combined, deaths from soft-tissue cancers make up a majority of cancer deaths worldwide [*Cancer* (Fact Sheet No. 297), WHO, 2013]. Even the most effective treatments for these types of cancer benefit greatly from as early a detection as possible, but finding soft-tissue tumors is a notoriously difficult task. Palpation is a standard medical practice for finding soft-tissue tumors where a doctor manually applies a force to the patient's body and feels for harder"lumps" in the surrounding softer tissue. Palpation uses the fact that changes in the macroscopic/microscopic structure of tissue correlate with changes in tissue health (see [45]). Unfortunately, the actual practice of palpation is subjective and usually limited to finding exceptionally hard nodules near the skin's surface. Other techniques, like ultrasound, can also take advantage of these same differences in tissue stiffness and be used to diagnose tumors further within the body, but even hard growths or other lesions can lack the particular acoustical properties for detection. Researchers have recently sought to use these elastic differences to develop a more quantitative approach to tumor detection, giving rise to the field of elasticity imaging methods for tumor detection (see [3] and the cited references therein).

Elasticity imaging seeks to extend methods like palpation and ultrasound to identify likely tumors. A relatively small external quasistatic compression force is applied to the tissue under examination and the tissue's axial displacement field is determined either through direct measurement, or alternatively through an indirect method such as the comparison of an undeformed and deformed image.

The measured data along with a method for solving a challenging inverse problem in linear elasticity (outlined in the sections below) then allow for the recovery of the tissue's variable elasticity and the subsequent determination of likely tumor sites.

I.2 Model of Linear Elasticity

We consider a model of elasticity in the human body based on the following system of partial differential equations:

$$-\nabla \cdot \sigma = f \quad \text{in } \Omega, \tag{I.1a}$$

$$\sigma = 2\mu\varepsilon(u) + \lambda \operatorname{div} u\, I, \tag{I.1b}$$

$$u = g \quad \text{on } \Gamma_1, \tag{I.1c}$$

$$\sigma n = h \quad \text{on } \Gamma_2. \tag{I.1d}$$

In the above system, the domain Ω is taken as a subset of $\mathbb{R}^2$ (or $\mathbb{R}^3$) with the boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2$. f is the external force applied to the body and $u = u(x) = \begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix}$ is the vector-valued displacement that satisfies (I.1). (I.1a) arises from Newton's laws of motion and taken along with (I.1b), indicates that u is the solution that minimizes the potential energy stored in the internal stresses of the object [51]. (I.1b), known as the stress-strain law, holds under the assumption that the material is isotropic and that the displacement remains small enough to ensure a linear relationship. In (I.1b), $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^{\mathrm{T}})$ is the linearized strain tensor where ∇u is the so-called displacement gradient [40]. In two dimensions, the displacement gradient takes the form

$$\nabla u = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} \end{bmatrix},$$

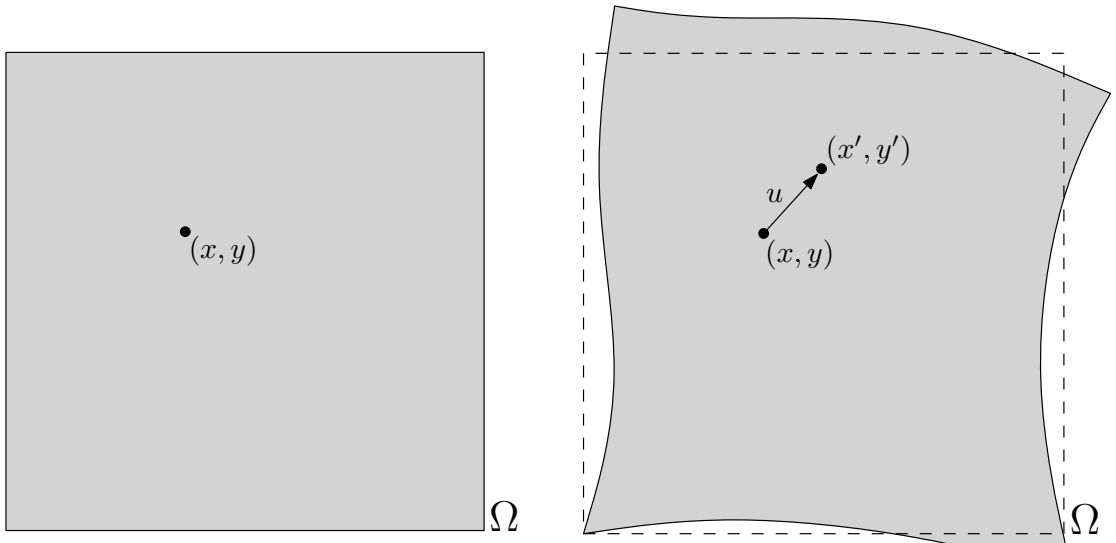
thus giving

$$\varepsilon(u) = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2}(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}) \\ \frac{1}{2}(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}) & \frac{\partial u_2}{\partial x_2} \end{bmatrix}.$$

g and h are the boundary conditions that u must meet on the sub-boundaries Γ_1 and Γ_2, respectively, with n being the unit outward normal. The coefficients μ and λ are the Lamé parameters which completely describe the object's physical elastic properties.

To illustrate the problem further, we examine a simple two-dimensional square elastic membrane in figure 1. In figure 1a, prior to the application of f to the object, we consider some "speck", (x, y), on the membrane. After f is applied in 1b, (x, y) has been displaced to a new point (x', y'). $u(x, y)$ is then the vector in $\mathbb{R}^2$ taking (x, y) to (x', y').

The direct problem is thus to find the displacement u that satisfies I.1 when the functions g, h, the variable coefficients μ and λ, and the force f are all given. The related parameter identification



(a) Before f is applied

(b) After f is applied

Figure 1: $u(x, y)$ as a displacement.

inverse problem would be to identify μ and λ given a measurement z of u. However, in the specific case of the elasticity imaging problem, we consider only the identification of μ.

Soft tissue's underlying structure consists of cells, collagen, and other constituents all immersed in a fluid consisting primarily of water. The properties of the structure are described by the shear modulus μ and the fluid's properties are instead dominated by the compressibility modulus $M = \lambda + 2\mu$. Thus when $\lambda \gg \mu$, it follows that $M \approx \lambda$. Materials having this condition, such as soft tissue within the human body (see [146] for details), are known as *nearly incompressible* and as will be detailed below, render certain classical methods of solving both the direct and inverse problem ineffective. See [57, 58, 58, 74, 77] for a thorough development of the more general inverse problem of identifying both Lamé parameters.

In Figure 2, we see a scenario in which the model I.1 can be applied to the identification of tumors inside the body.

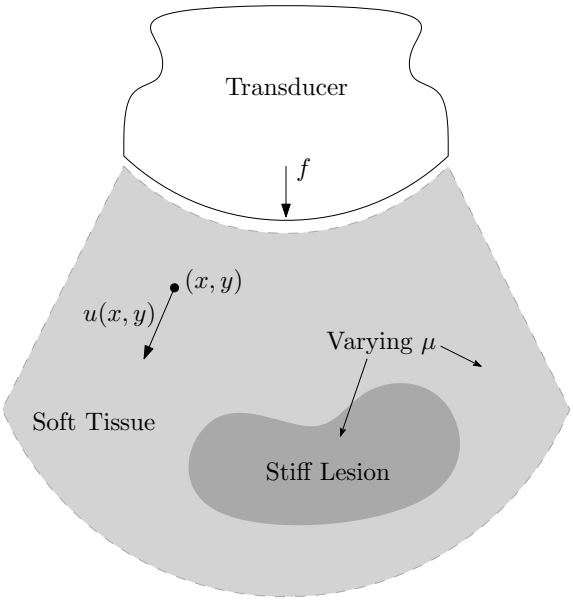


Figure 2: Diagram of possible elasticity imaging scenario

I.3 Mathematical Challenges in the Tumor Location Problem

Apart from the general complexity of the isotropic linear elasticity model in I.1 [1], the near incompressibility of the material in question and several other factors give rise to additional significant complications.

Before continuing with a discussion of these complexities, though, we first review some necessary nomenclature and notation.

First, we denote the tensor product of two 2-tensors A and B by $A : B$ with

$$A : B = A_{11}B_{11} + A_{12}B_{12} + A_{21}B_{21} + A_{22}B_{22}$$

Given a smooth domain $\Omega \subset \mathbb{R}^2$, the L_2-norm of a tensor-valued function $A = A(x)$ is given by

$$\|A\|_{L^2}^2 = \|A\|_{L^2(\Omega)}^2 = \int_\Omega A : A = \int_\Omega \left(A_{11}^2 + A_{12}^2 + A_{21}^2 + A_{22}^2\right).$$

For a vector-valued function $u(x) = (u_1(x), u_2(x))^{\mathrm{T}}$, the L_2-norm and the H^1-norm are given by

$$\|u\|_{L_2}^2 = \|u\|_{L_2(\Omega)}^2 = \int_\Omega \left(u_1^2 + u_2^2\right)$$

[1] Feynman referred to the related equations of motion in linear elasticity as "horrible"[51].

and

$$\|u\|_{H^1}^2 = \|u\|_{H^1(\Omega)}^2 = \|u\|_{L_2}^2 + \|\nabla u\|_{L_2}^2,$$

respectively.

To maintain clarity in the subsequent formulations, we make the simplifying assumption that $g = 0$ in (I.1). We then take the space of test functions, $\widehat{V}$, given by:

$$\widehat{V} = \{\bar{v} \in H^1(\Omega) \times H^1(\Omega) : \ \bar{v} = 0 \text{ on } \Gamma_1\}$$

where $H^1(\Omega)$ is the familiar Sobolev space (see Section VIII.1 for more details).

Taking some $\bar{v} \in \widehat{V}$ and multiplying through (I.1a), we have the following relation from Green's identity:

$$\int_\Omega (\nabla \cdot \sigma) \cdot \bar{v} = -\int_\Omega \sigma : \varepsilon(\bar{v}) + \int_\Gamma \sigma n \cdot \bar{v}.$$

Applying boundary conditions (I.1c) and (I.1d), we then have

$$\begin{aligned} -\int_\Omega (\nabla \cdot \sigma) \cdot \bar{v} &= \int_\Omega \sigma : \varepsilon(\bar{v}) - \int_{\Gamma_1} gn \cdot \bar{v} - \int_{\Gamma_2} h\bar{v} \\ &= \int_\Omega \sigma : \varepsilon(\bar{v}) - \int_{\Gamma_2} h\bar{v}. \end{aligned}$$

Now, applying (I.1b) we have

$$\begin{aligned} \int_\Omega \sigma : \varepsilon(\bar{v}) &= \int_\Omega (2\mu\varepsilon(\bar{u}) + \lambda \operatorname{div} uI) : \varepsilon(\bar{v}) \\ &= \int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega \lambda \operatorname{div} \bar{u}\, I : \varepsilon(\bar{v}). \end{aligned}$$

Noting that from the definition of the tensor product we have

$$I : \varepsilon(\bar{v}) = \frac{\partial \bar{v}_1}{\partial x_1} + \frac{\partial \bar{v}_2}{x_2} = \operatorname{div}\ \bar{v},$$

we can then apply the above to obtain the following weak form of the elasticity system (I.1):
Find $\bar{u} \in \widehat{V}$ such that

$$\int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega \lambda(\operatorname{div} \bar{u})(\operatorname{div} \bar{v}) = \int_\Omega f\bar{v} + \int_{\Gamma_2} h\bar{v}, \quad \text{for every } \bar{v} \in \widehat{V}. \tag{I.2}$$

As noted in the previous section, our focus is on the case where (I.1) describes the response of an isotropic nearly incompressible elastic object, i.e. where $\lambda \gg \mu$. In cases such as these, large amounts of energy must be introduced to produce small changes in the material's density [30].

To get a sense of the complications that arise from such a seemingly innocuous restriction, consider the bilinear form $\widehat{K} : \widehat{V} \times \widehat{V} \to \mathbb{R}$ arising from (I.2):

$$\widehat{K}(\bar{u}, \bar{v}) = \int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega \lambda(\operatorname{div} \bar{u})(\operatorname{div} \bar{v}).$$

With

$$\widehat{K}(\bar{v},\bar{v}) = \int_\Omega (2\mu+\lambda)\left[\left(\frac{\partial v_1}{\partial x_1}\right)^2 + \left(\frac{\partial v_2}{\partial x_2}\right)^2\right] + \mu\left[\left(\frac{\partial v_1}{\partial x_2}\right)^2 + \left(\frac{\partial v_2}{\partial x_1}\right)^2\right] + 2\mu\left(\frac{\partial v_1}{\partial x_2}\frac{\partial v_2}{x_1}\right) + 2\lambda\left(\frac{\partial v_1}{\partial x_1}\frac{\partial v_2}{\partial x_2}\right)$$

and

$$\begin{aligned} C\|\bar{v}\|_{\widehat{V}}^2 &= C\|\bar{v}\|_{L^2}^2 + C\|\nabla\bar{v}\|_{L^2} \\ &= C\int_\Omega \left(v_1^2+v_2^2\right) + C\int_\Omega \left(\frac{\partial v_1}{\partial x_1}\right)^2 + \left(\frac{\partial v_1}{\partial x_2}\right)^2 + \left(\frac{\partial v_2}{\partial x_1}\right)^2 + \left(\frac{\partial v_2}{\partial x_2}\right)^2, \end{aligned}$$

if we take $C \geq (2\mu+\lambda)$ such that

$$C\int_\Omega \left(v_1^2+v_2^2\right) \geq 2\mu\left(\frac{\partial v_1}{\partial x_2}\frac{\partial v_2}{x_1}\right) + 2\lambda\left(\frac{\partial v_1}{\partial x_1}\frac{\partial v_2}{\partial x_2}\right),$$

then we obtain

$$\widehat{K}(\bar{v},\bar{v}) \leq C\|\bar{v}\|_{\widehat{V}}^2 \qquad \text{for all } \bar{v}\in\widehat{V}. \tag{I.3}$$

Thus $\widehat{K}(\cdot,\cdot)$ is bounded.

We now note that

$$2\min(\mu,\mu+\lambda)\left(\varepsilon(\bar{v}):\varepsilon(\bar{v})\right) \leq \left(2\mu\varepsilon(\bar{v}) + \lambda\operatorname{div}\bar{v}\right):\varepsilon(\bar{v}) \qquad \text{for all } \bar{v}\in\widehat{V}.$$

Furthermore, supposing that $\mu > 0$, it follows that $2\min(\mu,\mu+\lambda) = 2\mu$ since $\lambda \gg \mu$. From Korn's inequality, there exists a constant $\kappa > 0$ such that

$$\kappa\|\bar{v}\|_{\widehat{V}} \leq \|\varepsilon(\bar{v})\|_{L^2} \qquad \text{for all } \bar{v}\in\widehat{V}.$$

Combining these two inequalities then gives

$$\alpha\|\bar{v}\|_{\widehat{V}}^2 \leq \widehat{K}(\bar{v},\bar{v}) \qquad \text{for all } \bar{v}\in\widehat{V}$$

where $0 < \alpha \leq \min(2\kappa^2\mu,\mu)$. Therefore $\widehat{K}(\cdot,\cdot)$ is also coercive and subsequently $\widehat{V}$-elliptic.

However, if we consider a finite-dimensional subspace $\widehat{V}_h$ of $\widehat{V}$ and $u\in\widehat{V}$ a solution to (I.2) with $u_h\in\widehat{V}_h$ the corresponding finite-dimensional solution, then from Céa's lemma it follows that

$$\|u-u_h\|_{\widehat{V}} \leq \frac{C}{\alpha}\|u-v_h\|_{\widehat{V}} \qquad \text{for all } v_h\in\widehat{V}_h.$$

With $\lambda \gg \mu$, $C \geq 2\mu+\lambda$, and $\alpha \leq \mu$, the ratio $\frac{C}{\alpha}$ is large and thus the actual error is larger, perhaps significantly so, than the approximation error. This phenomenon, termed *Poisson locking*,

volume locking, or more generally simply *the locking effect* renders classical finite element methods ineffective for this problem (see [30]).

Throughout the literature, various strategies have been used to overcome the locking effect. Least-squares finite element methods [35], discontinuous Galerkin methods [70], and mixed finite element methods [31] have all been used to this end. One of the most popular of these approaches has been the mixed finite elements approach, which in the present context, introduces a pressure term $p \in Q = L^2(\Omega)$ such that

$$p = \lambda(\operatorname{div} \bar{u}). \tag{I.4}$$

This gives the weak form:

$$\int_\Omega (\operatorname{div} \bar{u}) q - \int_\Omega \frac{1}{\lambda} pq = 0, \quad \text{for every } q \in Q. \tag{I.5}$$

By using relation (I.4), the weak form of (I.2) now reads: Find $\bar{u} \in \widehat{V}$ such that

$$\int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega p(\operatorname{div} \bar{v}) = \int_\Omega f\bar{v} + \int_{\Gamma_2} \bar{v}h, \quad \text{for every } \bar{v} \in \widehat{V}, \tag{I.6}$$

where the pressure p is also an unknown satisfying (I.5).

Subsequently, the problem of finding only $\bar{u} \in \widehat{V}$ satisfying (I.2) has now been transformed into the problem of finding $(\bar{u}, p) \in \widehat{V} \times Q$ satisfying both (I.5) and (I.6). Equations (I.5) and (I.6) can either be studied in the framework of saddle point problems (see below and §II for a detailed development of this topic) or can be combined and studied as a single variational problem. We will now briefly examine this latter approach.

We define $V = \widehat{V} \times Q$ and consider the problem of finding $u = (\bar{u}, p) \in V$ such that for every $v = (\bar{v}, q) \in V$, we have

$$\int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega (\operatorname{div} \bar{v}) p - \int_\Omega (\operatorname{div} \bar{u}) q + \int_\Omega \frac{1}{\lambda} pq = \int_\Omega f\bar{v} + \int_{\Gamma_2} \bar{v}h.$$

This gives rise to the bilinear form $\widetilde{K} : V \times V \to \mathbb{R}$ by

$$\widetilde{K}((\bar{u}, p), (\bar{v}, q)) = \int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega (\operatorname{div} \bar{v}) p - \int_\Omega (\operatorname{div} \bar{u}) q + \int_\Omega \frac{1}{\lambda} pq.$$

Through the imposition of suitable conditions, it can be shown that $\widetilde{K}$ is both coercive and continuous. Thus the Lax-Milgram lemma ensures the existence of a unique solution $u \in V$. However, the coercivity of $\widehat{K}$, although sufficient to give uniqueness, is still affected by the value $\frac{1}{\lambda}$. Again, in the case of large λ, this indicates numerical instability.

We will conclude this section with a brief summary of the saddle point approach to the nearly incompressible linear elasticity problem (expanded and generalized in §II). A classic treatise on this approach using mixed finite element methods can be found in Brezzi and Fortin [31]. As they detail, the development of saddle point problems for nearly incompressible linear elasticity is rooted in convex analysis and duality theory.

For a given convex functional $F(\bar{v})$ on $\widehat{V}$, we define the *conjugate functional*, $F^*(\bar{v}^*)$ on the dual space $\widehat{V}'$ of $\widehat{V}$ such that

$$F^*(\bar{v}^*) = \sup_{\bar{v}\in\widehat{V}} \left\{ \langle \bar{v}, \bar{v}^* \rangle_{\widehat{V}\times\widehat{V}'} - F(\bar{v}) \right\}$$

or symmetrically

$$F(\bar{v}) = \sup_{\bar{v}^*\in\widehat{V}'} \left\{ \langle \bar{v}, \bar{v}^* \rangle_{\widehat{V}\times\widehat{V}'} - F^*(\bar{v}^*) \right\}. \tag{I.7}$$

We also note that the solution u to the underlying system of equations (I.1) can equivalently be viewed as a solution to the problem of minimizing the elastic system's energy:

$$\inf_{\bar{v}\in\widehat{V}} \left\{ \frac{1}{2} \left[\int_\Omega \lambda |\operatorname{div} \bar{v}|^2 + \int_\Omega 2\mu |\varepsilon(\bar{v})|^2 \right] - \int_\Omega f\bar{v} - \int_{\Gamma_2} h\bar{v} \right\}. \tag{I.8}$$

Now, taking $F(\bar{v})$ as

$$F(\bar{v}) = \frac{1}{2} \int_\Omega \lambda |\operatorname{div} \bar{v}|^2 \quad = \left\langle \frac{\lambda}{2} \operatorname{div} \bar{v}, \operatorname{div} \bar{v} \right\rangle$$

and applying (I.7) we then have

$$\begin{aligned} \frac{1}{2} \int_\Omega \lambda |\operatorname{div} \bar{v}|^2 &= \sup_{q\in Q} \left\{ \langle \operatorname{div} \bar{v}, q \rangle - \frac{1}{4} \left\langle \frac{2}{\lambda} q, q \right\rangle \right\} \\ &= \sup_{q\in Q} \left\{ \int_\Omega (\operatorname{div} \bar{v}) q - \frac{1}{2} \int_\Omega \frac{1}{\lambda} |q|^2 \right\}. \end{aligned}$$

Substituting the above into (I.8) then gives the problem

$$\inf_{\bar{v}\in\widehat{V}} \sup_{q\in Q} \left\{ \frac{1}{2} \int_\Omega 2\mu |\varepsilon(\bar{v})|^2 + \int_\Omega q \operatorname{div} \bar{v} - \frac{1}{2} \int_\Omega \frac{1}{\lambda} |q|^2 - \int_\Omega f\bar{v} - \int_{\Gamma_2} h\bar{v} \right\}. \tag{I.9}$$

Thus the solution $(\bar{u}, p) \in \widehat{V} \times Q$ to the saddle point problem (I.9) is the solution to the system

$$\begin{aligned} \int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) + \int_\Omega p \operatorname{div} \bar{v} &= \int_\Omega f\bar{v} + \int_{\Gamma_2} h\bar{v} \quad && \text{for every } \bar{v} \in \widehat{V}, \\ \int_\Omega q \operatorname{div} \bar{u} - \int_\Omega \frac{1}{\lambda} p\, q &= 0 && \text{for every } q \in Q. \end{aligned}$$

I.4 Literature Review

Before covering the approach to the the elasticity imaging problem developed in the remaining sections of this work, we will first consider the historical development of the problem through a review of literature.

As noted in the introduction, many authors have contributed significantly to the general approach of using varying elastic properties of identifying unhealthy tissue. In 1990, Parker et al. [118]applied low-frequency mechanical vibrations (generally < 1 kHz) to ultrasound phantoms[2], to measure the modulus of elasticity using a computational framework based on finite element methods. Bertrand et al. [23] developed a "speckle" tracking method, also based on ultrasound, for characterizing tissue dynamics with an application to the assessment of skeletal and cardiac muscle contraction. Bamber et al. [12] proposed a semi-quantitative technique for evaluating real-time breast characteristics using ultrasound imaging. Their approach used pattern recognition to classify and rank each diagnostic feature's utility in diagnosis. All of the above methods assumed known lesion locations and geometry, allowing for the selection of a particular tissue deformation that yielded the optimum elasticity images. Raghavan and Yagle [121] sought a more general approach, placing the problem in a inverse problem framework where elasticity is recovered using measured strains and the equilibrium equations such as those in(I.1). In contrast to some of the other approaches, theirs used finite difference methods to solve the equations of equilibrium. Using a model with finite element methods, Kallel and Bertrand [79] employed a Newton-Raphson algorithm to fit (in a least-squares sense) a set of axial displacement fields estimated using the correlation of ultrasound signals. The Hessian matrix's general ill-conditioning, something inherent in this approach, was mitigated using Tikhonov regularization (see IV for a more detailed discussion of regularization). In a related approach, Doyley et al. [46], applied an iterative scheme based on a modified Newton-Raphson method to compute the spatial distribution of the Young's modulus[3]. The overall feasibility of reconstructing the tissue's elastic parameters using knowledge of known displacement and stress boundary conditions was evaluated with computer simulations. We direct the reader to [65] and [93] for additional developments along these same lines.

Barbone and Bamber [14] showed that the elasticity imaging inverse problem does not necessarily have a unique solution without sufficient *a priori* information about the stiffness along the domain's boundary. This means that the validity of reconstructions where the stiffness was assumed to

[2] *Ultrasound phantoms* are materials, made of gelatin or other materials, which provide a simulation of body tissue.

[3] The *Young's* or *elastic modulus*, often denoted by E, is related to the Lamé parameters by $E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu}$.

be constant along the boundary actually depends on whether it truly was constant. Barbone and Bamber's work also examined other factors and situations that lead to the non-uniqueness of solutions to the inverse problem. In 2003, Oberai et. al. [114] considered the identification of the shear modulus in an incompressible elastic medium with linearity when both boundary data and a single component of the displacement field was provided over the entire domain. Using the adjoint elasticity operator, the authors also developed an efficient technique for calculating the gradient of the output least squares functional. In a related but more general direction, Barbone and Gokhale [15] investigated the uniqueness of solutions to the elasticity imaging problem when N linearly independent displacement fields in the domain are known, generalizing beyond the typical case where only a single field is known (i.e. $N = 1$). Taking the input data for the inverse problem as directional displacements in a planar domain under a small quasi-static force, Fehrenbach et al. [50] developed a method of recovering the spatial distribution of the Young's modulus up to some multiplicative factor as well demonstrated the compactness of the differential of their parameter-to-state (coefficient-to-solution) map.

We note that many of the analyses outlined above deal only with the static/quasi-static approaches to the elasticity imaging inverse problem. For its dynamic counterpart, we direct the reader to Ji and McLaughlin [78], McLaughlin and Yoon [102], Park and Maniatty [117], and the cited references therein. The additionally interesting topic of stochasticity in elasticity imaging is addressed in recent works by Aquilo et al. and by Bochud and Rus [4, 28].

We conclude the review with a consideration of more recent developments in elasticity imaging. In 2010, Arnold et al. [10] used computational clusters to develop efficient numerical methods for identification of the elasticity modulus. Also in 2010, Ammari et al. [7] gave a promising optimization approach based on a discrepancy functional. Additionally, they showed that to recover significantly complex inclusions within a homogeneous medium, the information in the wavefield can be decomposed into a *near field* portion in the region immediately around the anomaly and a *far field* portion further away. This differentiation allows for a more precise recovery of the parameter. For more in-depth examinations of modern approaches to the elasticity imaging problem, see [3, 44, 22, 43, 52, 53, 103, 137]. Additionally, a thorough and accessible account of recent developments in elasticity imaging inverse problems can be found in the excellent survey article by Doyley [45].

I.5 Objective and Approach

The main objective of this work is to present a detailed computational study of the elasticity imaging inverse problem outlined above. We show how to repose the inverse problem within an optimization framework and after developing several optimization approaches, we then implement, apply and compare a variety of iterative optimization schemes to each approach.

In the introductory section, §I, we have examined both the underlying forward and inverse problems in elasticity imaging along with a consideration of some of the inherent difficulties found in both problems. This section also provided a brief literature review on the elasticity imaging problem's background. In §II, we introduce and analyze the modified output least squares (MOLS), energy output least squares (EOLS), and equation error (EE) optimzation frameworks for solving the elasticity imaging inverse problem. This section also includes a thorough discussion of the problem's discretization using mixed finite element methods. §III covers the background of gradient-based and extragradient methods for solving convex optimization problems and detail their application to elasticity imaging. We give detailed implementations of the following gradient and extragradient methods:

- Gradient and Scaled Gradient Projection
- Fast Iterative Shrinkage-Thresholding Method
- Khobotov extragradient method with Marcotte variants
- Goldstein-Type Extragradient methods
- Two-step Extragradient Methods
- Hyperplane Projection Methods

In §IV, we discuss regularization in the context of the related optimization problem and develop iterative methods that use proximal regularization. These schemes include the Hager-Zhang proximal point method and variants using Bregman functions and φ-divergence. In §V, we take a representative example in elasticity imaging and perform numerical experiments comparing all of the previously discussed iterative optimization methods and for each of the developed optimization frameworks. §VI, §VII, §VIII give some concluding remarks, acknowledgments, and a brief description of some of the the basic notations and definitions, respectively.

II. Computational Framework

II.1 General Saddle Point Problem

Let $\widehat{V}$ and Q be real Hilbert spaces, let B be a real Banach space, and let A be a nonempty, closed, and convex subset of B. Here B is the coefficient/parameter space and A is the set of all admissible coefficients. Let $a : B \times \widehat{V} \times \widehat{V} \to \mathbb{R}$ be a trilinear map which we assume to be symmetric with respect to the second and the third arguments. That is, for every $\ell \in B$ and for all $\bar{u}, \bar{v} \in \widehat{V}$, we have

$$a(\ell, \bar{u}, \bar{v}) = a(\ell, \bar{v}, \bar{u}).$$

Let $b : \widehat{V} \times Q \to \mathbb{R}$ be a bilinear form, let $c : Q \times Q \to \mathbb{R}$ be a symmetric bilinear form, and let $m : \widehat{V} \to \mathbb{R}$ be a linear and continuous map. We assume that there are positive constants $\kappa_1, \kappa_2, \varsigma_1, \varsigma_2$, and κ_0 such that the following inequalities hold:

$$\begin{aligned}
a(\ell, \bar{v}, \bar{v}) &\geq \kappa_1 \|\bar{v}\|^2 && \text{for every } \bar{v} \in \widehat{V}, \text{ for every } \ell \in A, \\
a(\ell, \bar{u}, \bar{v}) &\leq \kappa_2 \|\ell\| \|\bar{u}\| \|\bar{v}\| && \text{for every } \bar{u}, \bar{v} \in \widehat{V}, \text{ for every } \ell \in A, \\
c(q, q) &\geq \varsigma_1 \|q\|^2, && \text{for every } q \in Q, \\
c(q, q) &\leq \varsigma_2 \|q\|^2, && \text{for every } q \in Q, \\
b(\bar{v}, q) &\leq \kappa_0 \|\bar{v}\| \|q\|, && \text{for every } \bar{v} \in \widehat{V}, \text{ for every } q \in Q.
\end{aligned} \tag{II.1}$$

Presently, we consider the following problem:

Given $\ell \in A$, find $u = (\bar{u}, p) \in V = \widehat{V} \times Q$ such that

$$\begin{aligned}
a(\ell, \bar{u}, \bar{v}) + b(\bar{v}, p) &= m(\bar{v}) && \text{for every } \bar{v} \in \widehat{V} \\
b(\bar{u}, q) - c(p, q) &= 0 && \text{for every } q \in Q
\end{aligned} \tag{II.2}$$

Thus (II.2) represents the optimality condition of the saddle point problem

$$\inf_{\bar{v} \in \widehat{V}} \sup_{q \in Q} \left\{ \frac{1}{2} a(\ell, \bar{v}, \bar{v}) + b(\bar{v}, q) - \frac{1}{2} c(q, q) - m(\bar{v}) \right\}.$$

Provided with complete information about ℓ and $m(\cdot)$, the direct problem is to find $(\bar{u}, p)$, i.e. the solution to (II.2). The inverse problem in the context of the general saddle point problem is then to find a parameter $\ell \in A$ such that (II.2) is satisfied for some given measurement $(\bar{z}, \hat{z})$ of $(\bar{u}, p)$.

The equations (I.5) and (I.6) arising from the elasticity imaging inverse problem of identifying a

variable parameter μ (presented in § I.3) can be placed into this framework by taking:

$$a(\mu, \bar{u}, \bar{v}) = \int_\Omega 2\mu\varepsilon(\bar{u}) : \varepsilon(\bar{v}) \tag{II.3a}$$

$$b(\bar{u}, q) = \int_\Omega (\operatorname{div} \bar{u})\, q \tag{II.3b}$$

$$c(p, q) = \int_\Omega \frac{1}{\lambda} pq \tag{II.3c}$$

$$m(\bar{v}) = \int_\Omega f\bar{v} + \int_{\Gamma_2} h\bar{v}. \tag{II.3d}$$

In view of the above forms and the discussion of locking effect in § I.3, for numerical stability, it is of importance to have error estimates independent of the coercivity of $c(\cdot, \cdot)$. We note that the Babuška-Brezzi condition that requires that there exists a constant $\beta > 0$ such that

$$\beta \leq \inf_{q \in Q} \sup_{\bar{v} \in \widehat{V}} \frac{b(\bar{v}, q)}{\|q\| \|\bar{v}\|} \tag{II.4}$$

plays a fundamental role in achieving this goal. This condition is an abstract condition of the angle between the spaces $\widehat{V}$ and Q.

For later use in calculations, we now examine the solution map $S : A \to \widehat{V} \times Q$ for II.2 and some of its derivatives where

$$S(\ell) = (\bar{u}(\ell), p(\ell)). \tag{II.5}$$

Taking $u = u(\ell) = (\bar{u}(\ell), p(\ell))$ in II.2, we take the first derivative in an arbitrary direction $\delta\ell$ giving

$$\begin{aligned} a(\ell, \delta\bar{u}, \bar{v}) + b(\bar{v}, \delta p) &= -a(\delta\ell, \bar{u}, \bar{v}) \quad \text{for all } \bar{v} \in \widehat{V} \\ b(\delta\bar{u}, q) - c(\delta p, q) &= 0 \qquad\qquad\quad \text{for all } q \in Q \end{aligned} \tag{II.6}$$

where

$$\begin{aligned} \delta u &= (\delta\bar{u}, \delta p) \\ &= (D\bar{u}(\ell)\delta\ell, Dp(\ell)\delta\ell) \\ &= DS(\ell)\delta\ell. \end{aligned}$$

Thus the first derivative of the solution map $S(\ell)$ is the unique solution to the saddle point problem (II.6).

Similarly, we consider the second derivative of the solution map,

$$\begin{aligned} D^2S(\ell)(\delta\ell_1, \delta\ell_2) &= \delta^2 u \\ &= (\delta^2\bar{u}, \delta^2 p) \\ &= (D^2\bar{u}(\ell)(\delta\ell_1, \delta\ell_2), D^2p(\ell)(\delta\ell_1, \delta\ell_2)), \end{aligned}$$

which is the unique solution to the following saddle point problem:

$$\begin{aligned} a(\ell, \delta^2\bar{u}, \bar{v}) + b(\bar{v}, \delta^2 p) &= -a(\delta\ell_2, D\bar{u}(\ell)\delta\ell_1, \bar{v}) - a(\delta\ell_1, D\bar{u}(\ell)\delta\ell_2, \bar{v}) \quad \text{for all } \bar{v} \in \widehat{V} \\ b(\delta^2\bar{u}, q) - c(\delta^2 p, q) &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for all } q \in Q. \end{aligned} \tag{II.7}$$

Even more generally, the k-th derivative of the solution map,

$$D^k S(\ell)(\delta\ell_1, \ldots, \delta\ell_k) = \delta^k u = (\delta^k\bar{u}, \delta^k p) = (D^k\bar{u}(\ell)(\delta\ell_1, \ldots, \delta\ell_k), D^k p(\ell)(\delta\ell_1, \ldots, \delta\ell_k)),$$

is the unique solution to the saddle point problem:

$$\begin{aligned} a(\ell, \delta^k\bar{u}, \bar{v}) + b(\bar{v}, \delta^k p) &= -\sum_{i=1}^{k} a(\delta\ell_i, D^{k-1}\bar{u}(\ell)(\delta\ell_1, \ldots, \widehat{\delta\ell_i}, \ldots, \delta\ell_k), \bar{v}) \quad \text{for all } \bar{v} \in \widehat{V} \\ b(\delta^k\bar{u}, q) - c(\delta^k p, q) &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{for all } q \in Q. \end{aligned} \tag{II.8}$$

II.2 The Modified Output Least Squares Functional

One of the most commonly used objective functionals to solve parameter identification inverse problems in partial differential equations is the output least-squares (OLS) functional. In the context of the general saddle point problem II.2, the OLS functional takes the form

$$\begin{aligned} J_{\text{OLS}}(\ell) &= \frac{1}{2}\|u(\ell) - z\|_V^2 \\ &= \frac{1}{2}\|\bar{u}(\ell) - \bar{z}\|_{\widehat{V}}^2 + \frac{1}{2}\|p(\ell) - \hat{z}\|_Q^2, \end{aligned} \tag{II.9}$$

where $z = (\bar{z}, \hat{z})$ represents the measured data and $u(\ell) = (\bar{u}(\ell), p(\ell))$ is the solution of the saddle point problem (II.2) that corresponds to the coefficient ℓ.

The output least squares solution to the inverse problem of identifying ℓ is the one that solves the following optimization problem:

$$\begin{aligned} &\text{Find } \bar{\ell} \in A \text{ such that} \\ J_{\text{OLS}}(\bar{\ell}) \leq J_{\text{OLS}}(\ell), &\quad \text{for every } \ell \in A \end{aligned} \tag{II.10}$$

We observe that above optimization problem is constrained where the implicit constraint is provided by the saddle point problem and the explicit constraint is the set of admissible coefficients A. However, we also note that for nonlinear inverse problems, J_{OLS} is nonconvex generally and therefore can be used to detect only local minimizers.

It is therefore preferable to pursue a functional that yields a convex optimization problem for nonlinear inverse problems. Knowles [91] introduced a convex objective functional to identify

variable coefficients in a scalar partial differential equation and this approach was further extended to general elliptic inverse problems by Gockenbach and Khan [58].

In [76], Jadamba et al. proposed the following modified output least squares objective functional (MOLS) for the inverse problem of identifying ℓ in the saddle point problem (II.2)

$$J_{\text{MOLS}}(\ell) = \frac{1}{2}a(\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) + b(\bar{u}(\ell) - \bar{z}, p(\ell) - \hat{z}) - \frac{1}{2}c(p(\ell) - \hat{z}, p(\ell) - \hat{z}), \tag{II.11}$$

where $z = (\bar{z}, \hat{z})$ represents measured data and $u(\ell) = (\bar{u}(\ell), p(\ell))$ is the unique solution of (II.2).

Using (II.11) and the formulas for the first and second derivatives of the solution map (II.6) and (II.8)), we can compute the first and second derivatives of J_{MOLS}. The calculation of the first derivative in an arbitrary direction $\delta\ell$ is thus given by

$$\begin{aligned} DJ_{\text{MOLS}}(\ell)(\delta\ell) &= \frac{1}{2}a(\delta\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) + a(\ell, \delta\bar{u}(\ell), \bar{u}(\ell) - \bar{z}) \\ &\quad - c(\delta p(\ell), p(\ell) - \hat{z}) + b(\delta\bar{u}(\ell), p(\ell) - \bar{z}) + b(\bar{u}(\ell) - \bar{z}, \delta p) \\ &= \frac{1}{2}a(\delta\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) - a(\delta\ell, \bar{u}(\ell), \bar{u}(\ell) - \bar{z}) \\ &= -\frac{1}{2}a(\delta\ell, \bar{u}(\ell) + \bar{z}, \bar{u}(\ell) - \bar{z}) \end{aligned} \tag{II.12}$$

where again $\delta\bar{u} = D\bar{u}(\ell)(\delta\ell)$ and $\delta p = Dp(\ell)(\delta\ell)$.

Continuing, the second derivative is given by

$$\begin{aligned} D^2J_{\text{MOLS}}(\ell)(\delta\ell_1, \delta\ell_2) &= -\frac{1}{2}a(\delta\ell_1, \delta\bar{u}_2, \bar{u}(\ell) - \bar{z}) - \frac{1}{2}a(\delta\ell_1, \bar{u}(\ell) + z, \delta\bar{u}_2) \\ &= -\frac{1}{2}a(\delta\ell_1, \bar{u}(\ell) - \bar{z}, \delta\bar{u}_2) - \frac{1}{2}a(\delta\ell_1, \bar{u}(\ell) + z, \delta\bar{u}_2) \\ &= -a(\delta\ell_1, \bar{u}(\ell), \delta\bar{u}_2) \end{aligned} \tag{II.13}$$

where $\delta\bar{u}_2 = D\bar{u}(\ell)(\delta\ell_2)$. Applying (II.6) and (II.1), we additionally note that

$$\begin{aligned} D^2J_{\text{MOLS}}(\ell)(\delta\ell_1, \delta\ell_2) &= a(\ell, \delta\bar{u}_2, \delta\bar{u}_2) + b(\delta\bar{u}_2, \delta p) \\ &= a(\ell, \delta\bar{u}_2, \delta\bar{u}_2) + c(\delta p, \delta p) \\ &\geq \kappa_1\|\delta\bar{u}_2\|^2 + \varsigma_1\|\delta p\|^2. \end{aligned} \tag{II.14}$$

Remark. *Considering the calculation of the first and second derivatives of J_{MOLS}, we make the following two observations:*

1. *The formulation of the first derivative of J_{MOLS} does not depend on the derivative of the solution map.*
2. *It follows directly from* (II.14) *that J_{MOLS} is convex.*

II.3 The Energy Output Least Squares Functional

We now note that the saddle point problem (II.2) admits a certain ambiguity in its formulation in that the following problem is equivalent:

Given $\ell \in A$, find $u = (\bar{u}, p) \in V = \widehat{V} \times Q$ such that

$$\begin{aligned} a(\ell, \bar{u}, \bar{v}) + b(\bar{v}, p) &= m(\bar{v}) \quad \text{for every } \bar{v} \in \widehat{V} \\ -b(\bar{u}, q) + c(p, q) &= 0 \qquad \text{for every } q \in Q. \end{aligned} \tag{II.15}$$

Drawing on this idea, Doyley et al. [44] proposed the following energy output least squares functional (EOLS) for the solution of (II.2):

$$J_{\text{EOLS}}(\ell) = \frac{1}{2} a(\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) + \frac{1}{2} c(p(\ell) - \hat{z}, p(\ell) - \hat{z}). \tag{II.16}$$

Again, here $u(\ell) = (\bar{u}(\ell), p(\ell))$ is the unique solution of the saddle point problem (II.2) (or (II.15)) for the given parameter ℓ and $\bar{z}$ and $\hat{z}$ are measurements for $\bar{u}$ and p, respectively.

In a manner similar to the MOLS functional, the first derivative can be calculated using (II.2), (II.6), and (II.8). From (II.16), the calculation of the first derivative of J_{EOLS} follows directly:

$$\begin{aligned} DJ_{\text{EOLS}}(\ell)(\delta\ell) &= \frac{1}{2} a(\delta\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) + a(\ell, \delta\bar{u}(\ell), \bar{u}(\ell) - \bar{z}) + c(\delta p(\ell), p(\ell) - \hat{z}) \\ &= \frac{1}{2} a(\delta\ell, \bar{u}(\ell) - \bar{z}, \bar{u}(\ell) - \bar{z}) - a(\delta\ell, \bar{u}(\ell), \bar{u}(\ell) - \bar{z}) + b(\delta u, p(\ell) - \hat{z}) - b(\bar{u}(\ell) - \bar{z}, \delta p) \\ &= -\frac{1}{2} a(\delta\ell, \bar{u}(\ell) + \bar{z}, \bar{u}(\ell) - \bar{z}) + b(\delta\bar{u}, p(\ell) - \hat{z}) - b(\bar{u}(\ell) - \bar{z}, \delta p). \end{aligned}$$

This formulation, however, retains a dependence on the derivative of the solution map, δu. In [34], Cahill et al. introduced an alternative adjoint-based method for computing the first derivative that avoids any direct computation of δu. The adjoint method begins with the recognition that II.2 can be formulated as a variational problem of finding $u = (\bar{u}, p) \in V = \widehat{V} \times Q$ such that

$$T(\ell, u, v) = m(\bar{v}) \quad \text{for all } v = (\bar{v}, q) \in V \tag{II.17}$$

where $T : B \times V \times V \to \mathbb{R}$ is the trilinear form defined by

$$T(\ell, u, v) = a(\ell, \bar{u}, \bar{v}) + b(\bar{v}, p) + b(\bar{u}, q) - c(p, q).$$

Continuing, we compute the derivative of T with respect to ℓ in an arbitrary direction $\delta\ell$:

$$\begin{aligned} DT(\ell, u, v)(\delta\ell) &= a(\delta\ell, \bar{u}, \bar{v}) + a(\ell, \delta\bar{u}, \bar{v}) \\ &\quad + b(\bar{v}, \delta p) + b(\delta\bar{u}, q) - c(\delta p, q) \\ &= a(\delta\ell, \bar{u}, \bar{v}) + T(\ell, \delta u, v). \end{aligned} \tag{II.18}$$

We now define

$$J(\ell, v) = J_{\text{EOLS}}(\ell) + T(\ell, u, v) - m(\bar{v}).$$

Through the application of (II.17), we obtain that $J(\ell, v)$ is in accord with J_{EOLS} for all $v \in V$. That is,

$$J(\ell, v) = J_{\text{EOLS}}(\ell) \qquad \text{for all } v \in V.$$

This, in turn, implies that

$$D_\ell J(\ell, v)(\delta\ell) = DJ_{\text{EOLS}}(\ell)(\delta\ell) \qquad \text{for all } \delta\ell \in A, v \in V \tag{II.19}$$

where D_ℓ is understood as the partial derivative with respect to ℓ

Now by carefully selecting v, we can avoid the computation of δu in finding the derivative of J_{EOLS} as follows. First, we calculate the partial derivative of J:

$$\begin{aligned} D_\ell J(\ell, v)(\delta\ell) = \frac{1}{2} a(\delta\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + a(\ell, \delta\bar{u}, \bar{u} - \bar{z}) + c(\delta p, p - \hat{z}) \\ + a(\delta\ell, \bar{u}, \bar{v}) + T(\ell, \delta u, v). \end{aligned}$$

Next, taking $w = w(\ell) = (\bar{w}(\ell), p_w)$ to be the unique solution to the saddle point problem

$$\begin{aligned} a(\ell, \bar{w}, \bar{v}) + b(\bar{v}, p_w) &= -a(\ell, \bar{u} - \bar{z}, \bar{v}) - b(\bar{v}, p - \hat{z}) \quad && \text{for every } \bar{v} \in V \\ b(\bar{w}, q) - c(p_w, q) &= 0 && \text{for every } q \in Q. \end{aligned} \tag{II.20}$$

Note that the unique solution to the "adjoint" equation (II.20) is guaranteed by the same arguments as for the original saddle point problem (II.2) where we have taken $m(\cdot) = -a(\ell, \bar{u}, \cdot) - b(\cdot, p)$.

We now select $v = w$ and substitute into the partial derivative calculation above to give

$$\begin{aligned} D_\ell J(\ell, w)(\delta\ell) &= \frac{1}{2} a(\delta\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + a(\ell, \delta\bar{u}, \bar{u} - \bar{z}) + c(\delta p, p - \hat{z}) \\ &\quad + a(\delta\ell, \bar{u}, \bar{w}) + T(\ell, \delta u, w) \\ &= \frac{1}{2} a(\delta\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + a(\ell, \delta\bar{u}, \bar{u} - \bar{z}) + c(\delta p, p - \hat{z}) \\ &\quad + a(\delta\ell, \bar{u}, \bar{w}) - a(\ell, \bar{u} - \bar{z}, \delta\bar{u}) - b(\delta\bar{u}, p - \hat{z}) \\ &= \frac{1}{2} a(\delta\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + a(\delta\ell, \bar{u}, \bar{w}) + a(\ell, \delta\bar{u}, \bar{u}) + c(\delta p, p) \\ &\quad - b(\delta\bar{u}, p - \hat{z}) + a(\delta\ell, \bar{u}, \bar{w}) - a(\ell, \delta\bar{u}, \bar{u} - \bar{z}) \\ &= \frac{1}{2} a(\delta\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + a(\delta\ell, \bar{u}, \bar{w}) \\ &\quad + c(\delta p, p - \hat{z}) - b(\delta\bar{u}, p - \hat{z}), \end{aligned}$$

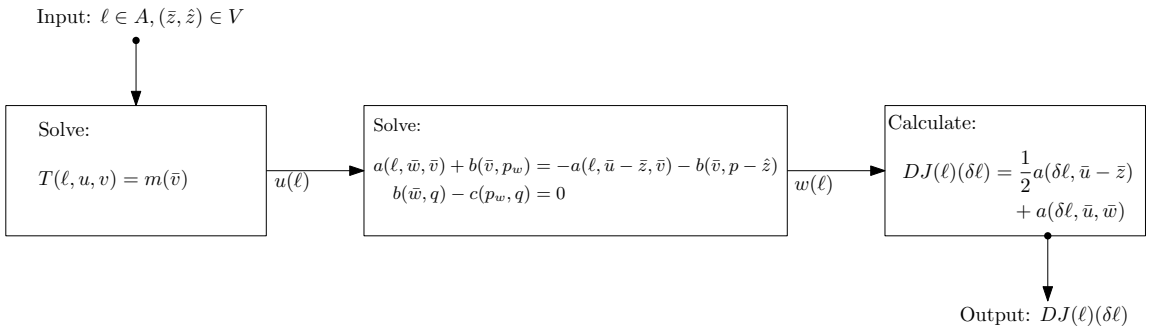


Figure 3: Block diagram of the adjoint derivative calculation for EOLS

where we have applied both the symmetry of T and $a(\ell, \cdot, \cdot)$.

Finally, using (II.6) and (II.19), we can simplify the above to obtain

$$D_\ell J(\ell, v)(\delta\ell) = DJ_{\text{EOLS}}(\ell) = \frac{1}{2}a(\delta\ell, \bar{u}, \bar{z}, \bar{u} - \bar{z}) + a(\delta\ell, \bar{u}, \bar{w}), \tag{II.21}$$

an expression for the derivative of J_{EOLS} that does not depend on δu.

We summarize this calculation in the block diagram in Figure 3.

As developed completely in [34], a similarly clever hybrid method yields a formulation for the second derivative of J_{EOLS}:

$$D^2 J_{\text{EOLS}}(\ell)(\delta\ell, \delta\ell) = 2a(\delta\ell, \delta\bar{u}, \bar{u} - \bar{z}) + a(\ell, \delta\bar{u}, \delta\bar{u}) + c(\delta p, \delta p) + 2a(\delta\ell, \delta\bar{u}, \bar{w}) \tag{II.22}$$

where again $w = w(\ell) = (\bar{w}(\ell), p_w(\ell))$ is the unique solution to the saddle point problem

$$\begin{aligned} a(\ell, \bar{w}, \bar{v}) + b(\bar{v}, p_w) &= a(\ell, \bar{v}, \bar{z} - \bar{u}) + b(\bar{v}, \hat{z} - p) \quad && \text{for all } \bar{v} \in \widehat{V} \\ b(\bar{w}, q) - c(p_w, q) &= 0 && \text{for all } q \in Q. \end{aligned} \tag{II.23}$$

The calculation of the second derivative of the EOLS functional is outlined in Figure 4.

Remark. *We make the following remarks regarding the EOLS functional:*

1. *From* (II.1)*, it follows that*

$$J_{EOLS}(\ell) \geq \frac{1}{2}\kappa_1 \|\bar{u}(\ell) - \bar{z}\|^2 + \frac{1}{2}\varsigma_1 \|p(\ell) - \hat{z}\|^2.$$

   Thus J_{EOLS} is positive for every $\ell \in A$.

2. *Unlike J_{MOLS}, J_{EOLS} is not necessarily convex.*

3. *The hybrid calculation for the second derivative alone has a significantly parallel structure. Additionally, when computed in tandem with the first derivative using an adjoint method, many of the calculations are shared, which lowers the total computational overhead.*

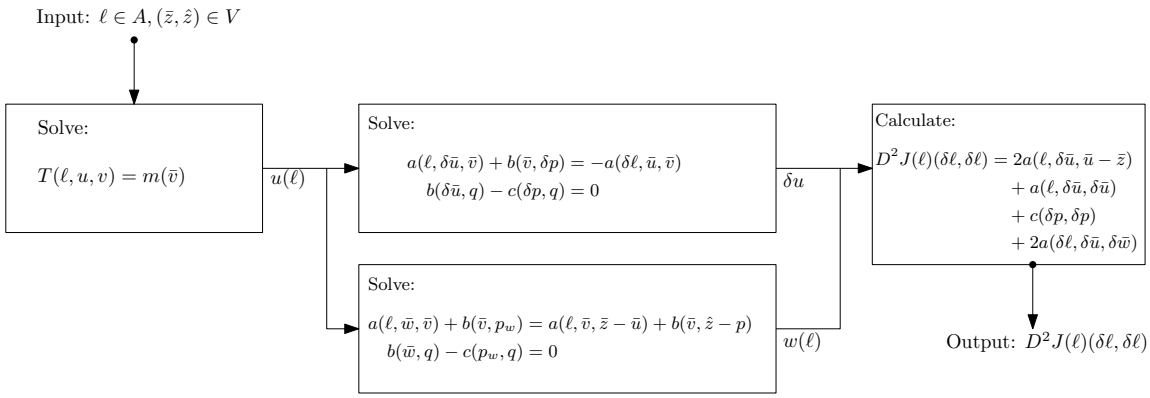


Figure 4: Block diagram of the hybrid second derivative calculation for EOLS

II.4 The Equation Error Functional

We now examine an approach to the tumor identification inverse problem that differs significantly from the OLS, EOLS, and MOLS approaches outlined thus far. This method is known as *equation error*.

To more succinctly explore the general equation error concept, we first consider an exemplar elliptic problem with suitable boundary conditions:

$$-\nabla \cdot (a\nabla u) = f \text{ in } \Omega. \tag{II.24}$$

The output least-squares (OLS) approach for determining the parameter a consists of minimizing the functional

$$J_{\text{OLS}}(a) = \|u(a) - z\|^2, \tag{II.25}$$

with z a measurement of u, $\|\cdot\|$ an appropriate norm, and where $u(a)$ solves the variational problem paired with (II.24). In contrast, the equation error (EE) method for identifying a would be to minimize the residual error between the left and right hand sides of (II.24) given a measurement z of u:

$$J_{\text{EE}}(a) = \frac{1}{2}\|\nabla \cdot (a\nabla\ z) + f\|^2_{H^{-1}(\Omega)}. \tag{II.26}$$

In Figure 5, we see a block diagram describing a generalized iterative optimization algorithm for solving the inverse problem using an OLS-based method. $\mathbb{P}$ here represents some general algorithmic operator for determining the next step in the iteration. Every iteration relies on the solution of a variational problem like the saddle point problem (II.2) to determine u. As can be seen in the diagram, this is often typically achieved through the solution of a large linear system, a costly

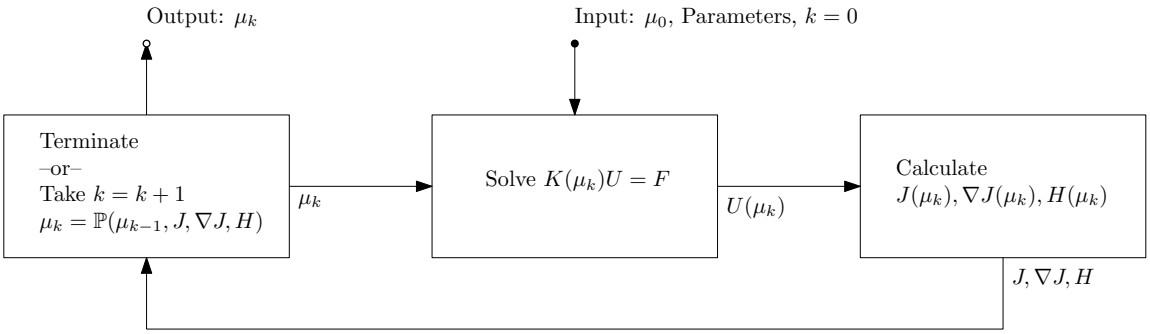


Figure 5: Generalized iterative approach to solving the inverse problem using OLS

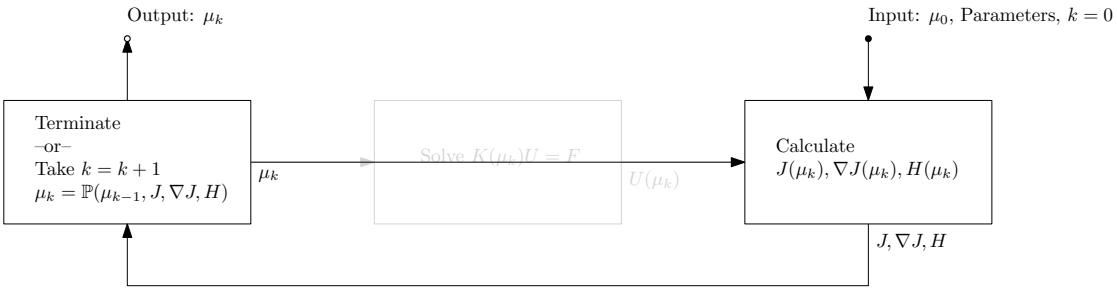


Figure 6: Generalized iterative approach to solving the inverse problem using EE

computation. We now contrast this with Figure 6 which depicts a similar generalized algorithm using equation error. Since the equation error functional does not depend on u the solution of the variational problem is not necessary, providing a significant computational savings.

For a general development of the equation error approach to inverse problems, we direct the reader to Acar [1], Gockenbach and Khan [55], Al-Jamal and Gockenbach [5] for more on (II.24), and Gockenbach, Jadamba, and Khan [56] for general elliptic inverse problems.

Given the potential advantages of the equation error approach, it makes sense to seek its extension to the tumor identification inverse problem.

For the solution of the inverse problem related to (II.2), we define the equation error functional as follows:

$$J_{\text{EE}}(\ell) = \|e(\ell, z)\|^2_{V^*}, \tag{II.27}$$

where again $z = (\bar{z}, \hat{z})$ is the data and where $e(\ell, \cdot) \in V := \widehat{V} \times Q$ is given by

$$\langle e(\ell, u), v\rangle = a(\ell, \bar{u}, \bar{v}) + b(\bar{v}, p) - b(\bar{u}, q) + c(p, q), \quad \text{for } u = (\bar{u}, p) \in V,\ v = (\bar{v}, q) \in V.$$

We then have $e(\ell, u) = (e_1(\ell, u), e_2(\ell, u)) \in \bar{V} \times Q$, where $e_1(\ell, u)$ and $e_2(\ell, u)$ are defined by

$$\langle e_1(\ell, u), \bar{v} \rangle = a(\ell, \bar{u}, \bar{v}) + b(\bar{v}, p) - m(\bar{v}) \qquad \text{for all } \bar{v} \in \overline{V} \tag{II.28}$$

$$\langle e_2(\ell, u), q \rangle = b(\bar{u}, q) - c(p, q) \qquad \text{for all } q \in Q \tag{II.29}$$

Clearly if $\ell = \ell^*$ and $u = u^* = (\bar{u}^*, p^*)$ satisfy (II.2), then $e(\ell^*, u^*) = 0$.

For a more comprehensive application of the equation error approach to the tumor identification inverse problem, we refer to Crossen et al. [43]. Derivative formulas for the discrete problem will be provided in the subsequent section.

Remark. *We make the following general remarks about the EE functional:*

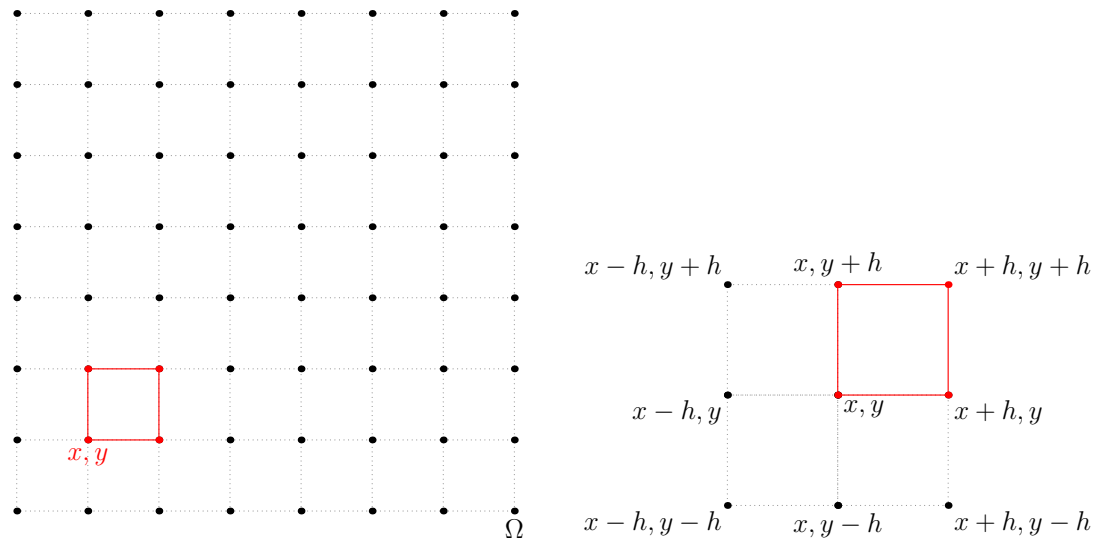
1. *The functional J_{EE} is quadratic in ℓ and thus the minimization of J_{EE} reduces to the solution of a positive (semi-)definite linear system after discretization, subsequently leading to a convex optimization problem.*

2. *The calculation of J_{EE} necessitates the differentiation of the measured data z, leaving the EE approach highly susceptible to noise.*

II.5 Discretization Using the Finite Element Method

In this section, we will derive discrete formulas for the saddle point problem (II.2) and for the MOLS, EOLS, and EE functionals for the solution of the associated inverse problem. We begin, therefore, with a triangulation $\mathcal{T}_h$ on $\Omega \subset \mathbb{R}^2$. L_h is taken as the space of all piecewise continuous polynomials of degree d_ℓ relative to $\mathcal{T}_h$, U_h is the space of all piecewise continuous polynomials of degree d_u relative to $\mathcal{T}_h$, and Q_h is the space of all piecewise continuous polynomials of degree d_q relative to $\mathcal{T}_h$.

To represent the discrete saddle point problem in a tractable form, we begin with a representation of bases for L_h, U_h and Q_h by $\{\varphi_1, \varphi_2, \ldots, \varphi_m\}$, $\{\psi_1, \psi_2, \ldots, \psi_n\}$, and $\{\chi_1, \chi_2, \ldots, \chi_k\}$, respectively. The space L_h is then isomorphic to $\mathbb{R}^m$ and for any $\ell \in L_h$, we define $L \in \mathbb{R}^m$ by $L_i = \ell(x_i), \quad i = 1, 2, \ldots, m$, where the nodal basis $\{\varphi_1, \varphi_2, \ldots, \varphi_m\}$ corresponds to the nodes $\{x_1, x_2, \ldots, x_m\}$. Conversely, each $L \in \mathbb{R}^m$ corresponds to $\ell \in L_h$ defined by $\ell = \sum_{i=1}^m L_i \varphi_i$. Analogously, $u \in U_h$ will correspond to $U \in \mathbb{R}^n$, where $\bar{U}_i = u(y_i),\ i = 1, 2, \ldots, n$, and $u = \sum_{i=1}^n \bar{U}_i \psi_i$, where $y_1, y_2, \ldots, y_n$ are the nodes of the mesh defining U_h. Finally, $q \in Q_h$ will correspond to $Q \in \mathbb{R}^k$, where $Q_i = q(z_i),\ i = 1, 2, \ldots, k$, and $q = \sum_{i=1}^k Q_i \chi_i$, where $z_1, z_2, \ldots, z_k$ are the nodes of the mesh defining Q_h. We note that L_h, U_h, and Q_h are all defined relative to the same elements, but their nodes will be different if $d_\ell \neq d_u \neq d_q$.



(a) Quadrangular mesh on Ω **(b)** Detail of reference convex with mesh size h

Figure 7: Two-dimensional quadrangular mesh

In Figure 7a, we consider an example discretization of a square domain Ω using a quadrangular mesh with a uniform "mesh size" h. Figure 7b gives the detail of a reference convex on the mesh. In Figure 8, we see example piecewise linear shape functions on the reference convex and the subsequent basis functions in Figure 9.

The discrete saddle point problem seeks the unique $(\bar{u}_h, p_h) \in U_h \times Q_h$ for each ℓ_h such that

$$a(\ell_h, \bar{u}_h, \bar{v}) + b(\bar{v}, p_h) = m(\bar{v}), \quad \text{for every } \bar{v} \in U_h, \tag{II.30a}$$

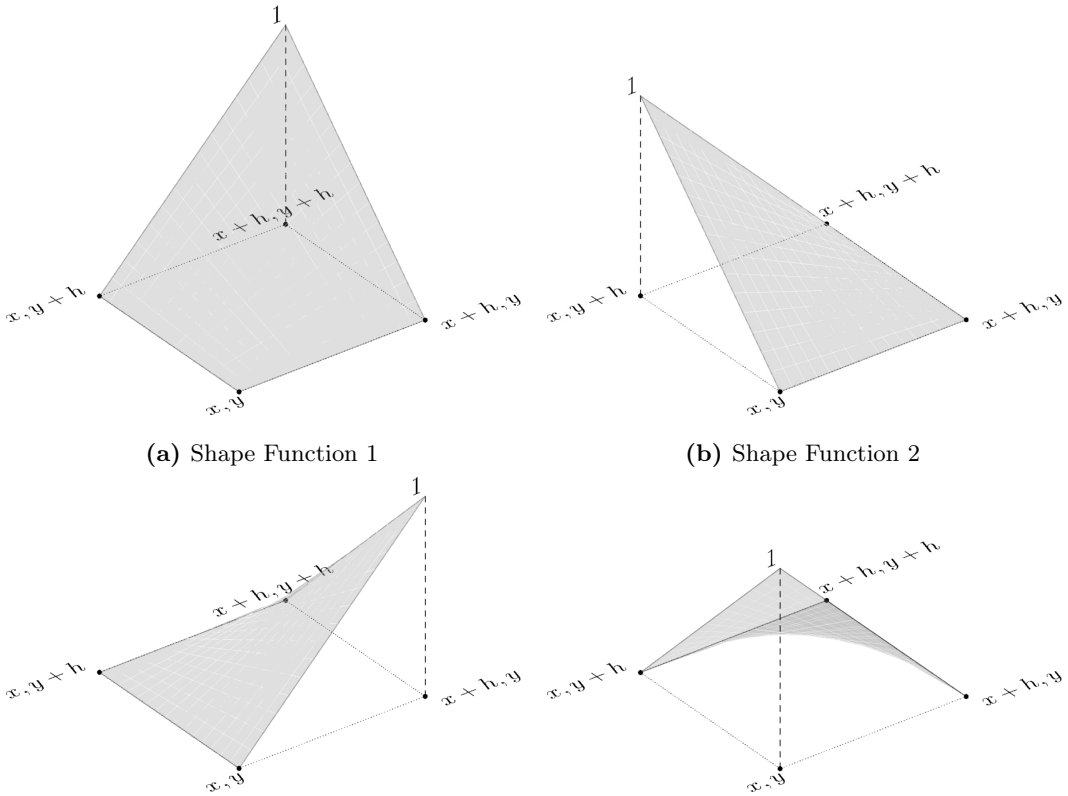
$$b(\bar{u}_h, q) - c(p_h, q) = 0, \quad \text{for every } q \in Q_h. \tag{II.30b}$$

We define $S : \mathbb{R}^m \to \mathbb{R}^{n+k}$ to be the finite element solution operator that assigns to each coefficient $\ell_h \in L_h$, the unique approximate solution $u_h = (\bar{u}_h, p_h) \in U_h \times Q_h$. Then $S(L) = U$, where U is defined by

$$K(L)U = F, \tag{II.31}$$

and where the stiffness matrix $K(L) \in \mathbb{R}^{(n+k)\times(n+k)}$ and the load vector $F \in \mathbb{R}^{n+k}$ are given by

$$K(L) = \begin{bmatrix} \widehat{K}_{n\times n}(L) & B^{\mathrm{T}}_{n\times k} \\ B_{k\times n} & -C_{k\times k} \end{bmatrix}$$



(a) Shape Function 1

(b) Shape Function 2

(c) Shape Function 3

(d) Shape Function 4

Figure 8: Two-dimensional shape functions on the reference convex

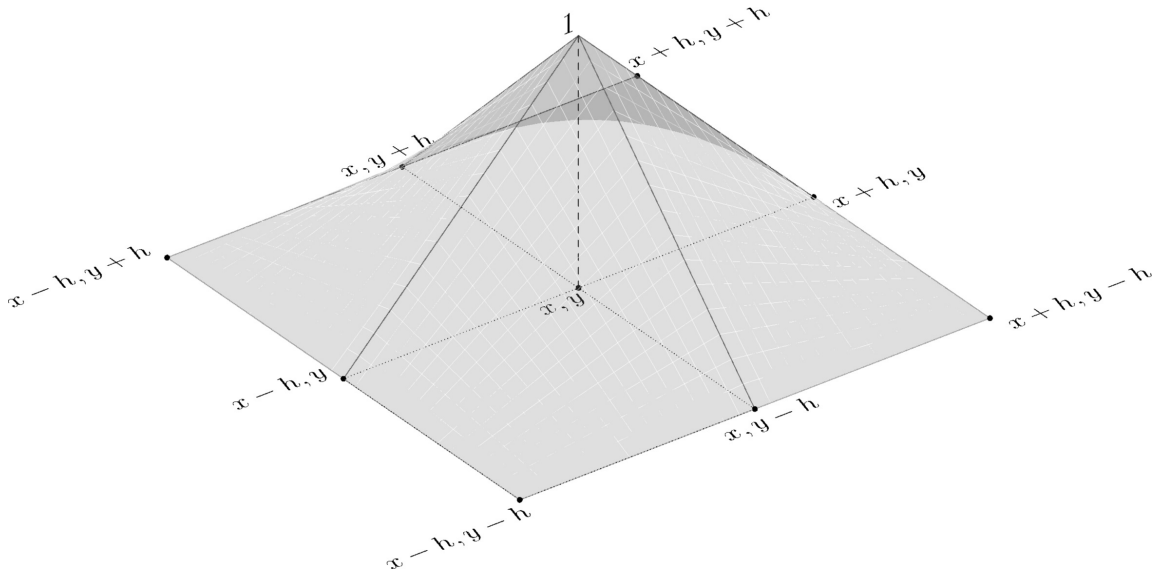


Figure 9: Basis function on a quadrangular mesh

with

$$\begin{aligned}
\widehat{K}(L)_{i,j} &= a(\ell, \psi_j, \psi_i), \quad i,j = 1,2,\ldots,n, \\
B_{i,j} &= b(\psi_j, \chi_i), \quad i = 1,2,\ldots,k, \quad n = 1,2,\ldots,n \\
C_{i,j} &= c(\chi_j, \chi_i), \quad i,j = 1,2,\ldots,k, \\
F_i &= m(\psi_i), \quad i = 1,2,\ldots,n, \\
F_j &= 0, \quad j = n+1, n+2, \ldots, n+k.
\end{aligned}$$

For future reference, we note that

$$\widehat{K}(L)_{ij} = T_{ijk} L_k,$$

where the summation convention is used and T is the tensor defined by

$$T_{ijk} = a(\varphi_k, \psi_i, \psi_j), \quad \text{for every } i,j = 1,\ldots,n,\ k = 1,\ldots,m.$$

It is convenient to approximate the components of U_h in a single finite element space $\widetilde{U}_h$ where $U_h = \widetilde{U}_h \times \widetilde{U}_h$. Therefore, if $\{\xi_1, \ldots, \xi_l\}$ are the basis of $\widetilde{u}_h$, then the vector-valued basis of U_h can be chosen as

$$\{\psi_i\}_{i=1}^n = \left\{ \begin{bmatrix} \xi_1 \\ 0 \end{bmatrix}, \begin{bmatrix} \xi_2 \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} \xi_l \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \xi_1 \end{bmatrix}, \begin{bmatrix} 0 \\ \xi_2 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \xi_l \end{bmatrix} \right\}$$

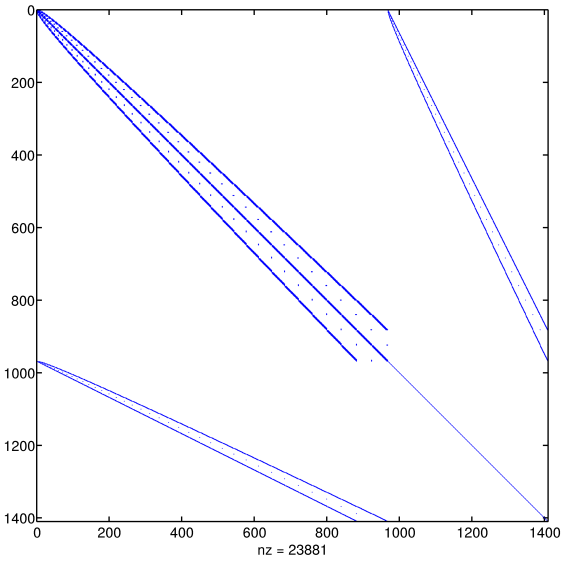


Figure 10: Sparsity structure of K

where $n = 2l$. In the specific case of the tumor identification problem, this suggests the following structure for $\widehat{K}$:

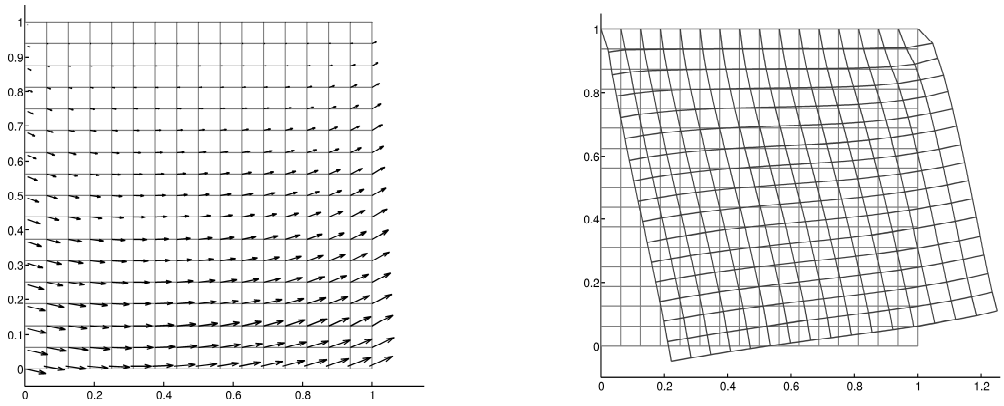
$$\widehat{K} = \begin{bmatrix} \bar{K}^{11} & \bar{K}^{12} \\ \bar{K}^{21} & \bar{K}^{22} \end{bmatrix}$$

where

$$\begin{aligned} \bar{K}^{11}_{i,j} &= a\left(\mu, [\xi_i, 0]^T, [\xi_j, 0]^T\right) = \int_\Omega 2\mu \frac{\partial \xi_i}{\partial x}\frac{\partial \xi_j}{\partial x} + \mu \frac{\partial \xi_i}{\partial y}\frac{\partial \xi_j}{\partial y} \\ \bar{K}^{12}_{i,j} = \bar{K}^{21}_{j,i} &= a\left(\mu, [\xi_i, 0]^T, [0, \xi_j]^T\right) = \int_\Omega \mu \frac{\partial \xi_i}{\partial y}\frac{\partial \xi_j}{\partial x} \\ \bar{K}^{22}_{i,j} &= a\left(\mu, [0, \xi_i]^T, [0, \xi_j]^T\right) = \int_\Omega \mu \frac{\partial \xi_i}{\partial x}\frac{\partial \xi_j}{\partial x} + 2\mu \frac{\partial \xi_i}{\partial y}\frac{\partial \xi_j}{\partial y} \end{aligned}$$

for $i, j = 1, \ldots, l$.

In Figure 10, we see the underlying sparsity structure of the matrix K for an example two-dimensional problem. Figure 11 shows the representation of an example solution $u(x, y)$ both as a discretized vector field and as a deformation of a discretized mesh.



(a) $u(x, y)$ as vector field on discretized Ω **(b)** Example discretized $u(x, y)$

Figure 11: Example discretization of $u(x, y)$

II.5.1 Discrete MOLS

Using details of the above discretization, we will now lay out the discrete form of the MOLS functional and its derivatives. Using the above notations, the discrete form of (II.11), i.e.

$$J_{\text{MOLS}}(\ell) = \frac{1}{2}a(\ell, \bar{u} - \bar{z}, \bar{u} - \bar{z}) + b(\bar{u} - \bar{z}, p - \hat{z}) - \frac{1}{2}c(p - \hat{z}, p - \hat{z}),$$

is given by

$$\begin{aligned} J_{\text{MOLS}}(L) &= \frac{1}{2}\begin{bmatrix} \overline{U}(L) - \overline{Z} \\ P(L) - \widehat{Z} \end{bmatrix}^T \begin{bmatrix} \widehat{K}(L) & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} \overline{U}(L) - \overline{Z} \\ P(L) - \widehat{Z} \end{bmatrix} \\ &= \frac{1}{2}(\overline{U}(L) - \overline{Z})^T \widehat{K}(L)(\overline{U}(L) - \overline{Z}) + (\overline{U}(L) - \overline{Z})^T B^T (P(L) - \widehat{Z}) \\ &\quad - \frac{1}{2}(P(L) - \widehat{Z})^T C(P(L) - \widehat{Z}) \end{aligned} \tag{II.32}$$

We can now compute the gradient of the discrete MOLS functional. For this, we first compute the derivative of the coefficient-to-solution map in discrete terms. The matrix-vector form of the discrete saddle point problem (II.30) reads (cf. (II.31))

$$\begin{bmatrix} \widehat{K}(L) & B^{\mathrm{T}} \\ B & -C \end{bmatrix} \begin{bmatrix} \overline{U}(L) \\ P(L) \end{bmatrix} = \begin{bmatrix} F \\ 0 \end{bmatrix}.$$

After differentiating the above with respect to L, we have the discrete analogue of (II.6):

$$\begin{aligned} \widehat{K}(L)\delta\overline{U} + B^{\mathrm{T}}(\delta P) &= -\widehat{K}(\delta L)\overline{U}(L) \\ B\delta\overline{U} - C\delta P &= 0, \end{aligned} \tag{II.33}$$

where $\delta U = (\delta\overline{U}, \delta P) = DS(L)(\delta L)$.

Then from (II.33), it follows that

$$\begin{aligned} (\overline{U}(L) - \overline{Z})^{\mathrm{T}}\widehat{K}(L)\delta\overline{U} + (\overline{U}(L) - \overline{Z})^{\mathrm{T}}B^{\mathrm{T}}(\delta P) &= -(\overline{U}(L) - \overline{Z})^{\mathrm{T}}\widehat{K}(\delta L)\overline{U}(L) \\ \delta U^{\mathrm{T}}B^{\mathrm{T}}(P(L) - \widehat{Z}) - \delta P^{\mathrm{T}}C(P(L) - \widehat{Z}) &= 0. \end{aligned} \tag{II.34}$$

We can then use (II.32) and (II.34) to compute the derivative of J_{MOLS}:

$$\begin{aligned} DJ_{\mathrm{MOLS}}(L)(\delta L) &= \delta\overline{U}(L)^{\mathrm{T}}\widehat{K}(L)(\bar{U}(L) - \overline{Z}) + \frac{1}{2}(\overline{U}(L) - \overline{Z})^{\mathrm{T}}\widehat{K}(\delta L)(\overline{U}(L) - \overline{Z}) - \delta P^{\mathrm{T}}C(P(L) - \widehat{Z}) \\ &+ \delta\overline{U}^{\mathrm{T}}B^{\mathrm{T}}(P(L) - \widehat{Z}) + (\overline{U}(L) - \overline{Z})^{\mathrm{T}}B^{\mathrm{T}}(\delta P) \\ &= \frac{1}{2}(\overline{U}(L) - \overline{Z})K(\delta L)(\overline{U}(L) - \overline{Z}) - (\overline{U}(L) - \overline{Z})K(\delta L)\overline{U}(L) \quad \text{(using (II.34))} \\ &= -\frac{1}{2}(\overline{U}(L) - \overline{Z})^{\mathrm{T}}K(\delta L)(\overline{U}(L) + \overline{Z}) \\ &= -\frac{1}{2}\delta L^{\mathrm{T}}\mathbb{A}(\overline{U}(L) + \overline{Z})^{\mathrm{T}}(\overline{U}(L) - \overline{Z}), \end{aligned}$$

where the matrix $\mathbb{A}$ is the so-called adjoint stiffness matrix defined by the following condition:

$$\widehat{K}(L)\overline{V} = \mathbb{A}(\overline{V})L, \quad \text{for every } L \in \mathbb{R}^m, \text{ for every } \overline{V} \in \mathbb{R}^n.$$

Therefore, we have shown that

$$DJ_{\mathrm{MOLS}}(L)(\delta L) = -\frac{1}{2}\delta L^{\mathrm{T}}\mathbb{A}(\overline{U}(L) + \overline{Z})^{\mathrm{T}}(\overline{U}(L) - \overline{Z}),$$

which implies that

$$\nabla J_{\mathrm{MOLS}}(L) = -\frac{1}{2}\mathbb{A}(\overline{U}(L) + \overline{Z})^{\mathrm{T}}(\overline{U}(L) - \overline{Z}) = -\frac{1}{2}\mathbb{A}(\overline{U}(L))^{\mathrm{T}}\overline{U}(L) + \frac{1}{2}\mathbb{A}(\overline{Z})^{\mathrm{T}}\overline{Z}.$$

In a similar manner, we can go on to derive the formula for the Hessian of the MOLS functional by first considering the discrete version of (II.13):

$$\begin{aligned} D^2J_{\mathrm{MOLS}}(L)(\delta L, \delta L) &= -\delta L^T\nabla\overline{U}(L)^T\widehat{K}(\delta L)\overline{U}(L) \\ &= -\delta L^T\nabla\overline{U}(L)^T\mathbb{A}(\overline{U}(L))\delta L. \end{aligned}$$

This implies that

$$\nabla^2J_{\mathrm{MOLS}}(L) = -\nabla\overline{U}(L)^T\mathbb{A}(\overline{U}(L)).$$

II.5.2 Discrete EOLS

Using the same notation and proceeding in a similar fashion to the development of the discrete MOLS functional, we can now present formulas for the discrete EOLS functional and its derivatives. Directly, we have the discrete matrix form of (II.16):

$$J_{\text{EOLS}}(L) = \frac{1}{2}(\overline{U}(L) - \overline{Z})^{\text{T}}\widehat{K}(L)(\bar{U}(L) - \overline{Z}) + \frac{1}{2}(P(L) - \widehat{Z})^{\text{T}}C(P(L) - \widehat{Z}).$$

Subsequently, the discrete form of the derivative of J_{EOLS} is given by

$$\begin{aligned} DJ_{\text{EOLS}}(L)(\delta L) = &-\frac{1}{2}(\overline{U}(L) - \overline{Z})^{\text{T}}\widehat{K}(\delta L)(\overline{U}(L) - \overline{Z}) \\ &+ \delta U^{\text{T}}B^{\text{T}}(P(L) - \widehat{Z}) - (\overline{U}(L) - \overline{Z})^{\text{T}}B^{\text{T}}\delta P \end{aligned} \tag{II.35}$$

Denoting the respective Jacobian matrices by $\nabla\overline{U}(L)$ and $\nabla P(L)$, (II.35) can be expressed in the following equivalent form:

$$\begin{aligned} DJ_{\text{EOLS}}(L)(\delta L) = &-\frac{1}{2}(\overline{U}(L) + \overline{Z})^{\text{T}}\widehat{K}(\delta L)(\overline{U}(L) - \overline{Z}) \\ &+ (P(L) - \widehat{Z})^{\text{T}}B\nabla\overline{U}(L)\delta L - (\overline{U}(L) - \overline{Z})^{\text{T}}B^{\text{T}}\nabla P(L)\delta L \end{aligned} \tag{II.36}$$

Employing the adjoint stiffness matrix to make substitutions within (II.36), we then obtain

$$\begin{aligned} DJ_{\text{EOLS}}(L)(\delta L) = &-\frac{1}{2}(\overline{U}(L) + \overline{Z})^{\text{T}}\mathbb{A}(\overline{U}(L) - \overline{Z})\delta L \\ &+ (P(L) - \widehat{Z})^{\text{T}}B\nabla U(L)\delta L - (\overline{U}(L) - \overline{Z})^{\text{T}}B^{\text{T}}\nabla P(L)\delta L \end{aligned}$$

which implies that

$$\begin{aligned} \nabla J_{\text{EOLS}}(L) = &-\frac{1}{2}(\overline{U}(L) + \overline{Z})^{\text{T}}\mathbb{A}(\overline{U}(L) - \overline{Z}) \\ &+ (P(L) - \widehat{Z})^{\text{T}}B\nabla\overline{U}(L) - (\overline{U}(L) - \overline{Z})^{\text{T}}B^{\text{T}}\nabla P(L). \end{aligned} \tag{II.37}$$

For the adjoint method outlined in Figure 3, we must first find the discrete solution $W(L) = \begin{bmatrix} \overline{W}(L) \\ P_W(L) \end{bmatrix}$ to the discrete analogue of the adjoint equations (II.20):

$$\begin{bmatrix} \widehat{K}(L) & B^{\text{T}} \\ B & -C \end{bmatrix} \begin{bmatrix} \overline{W}(L) \\ P_W(L) \end{bmatrix} = \begin{bmatrix} -\widehat{K}(L)(\overline{U} - \overline{Z}) - B^{\text{T}}(P - \widehat{Z}) \\ 0 \end{bmatrix}. \tag{II.38}$$

Taking (II.21) into account and applying the adjoint stiffness matrix, we have the following discretized expression for the derivative of J_{EOLS}:

$$\begin{aligned} DJ_{\text{EOLS}}(L)(\delta L) &= \frac{1}{2}\left(\overline{U}(L) - \overline{Z}\right)^{\text{T}}\widehat{K}(\delta L)\left(\overline{U}(L) - \overline{Z}\right) + \overline{U}(L)^{\text{T}}\widehat{K}(\delta L)\overline{W} \\ &= \frac{1}{2}\overline{U}(L)^{\text{T}}\mathbb{A}(\overline{U}(L) - \overline{Z})\delta L + \overline{U}(L)^{\text{T}}\mathbb{A}(\overline{W}(L))\delta L. \end{aligned}$$

Therefore the gradient $\nabla J_{\text{EOLS}}(L)$ can then be computed via

$$\nabla J_{\text{EOLS}}(L) = \frac{1}{2}\left(\overline{U}(L) - \overline{Z}\right)^{\text{T}} \mathbb{A}(\overline{U}(L) - \overline{Z}) + \overline{U}(L)^{\text{T}}\mathbb{A}(\overline{W}(L)). \tag{II.39}$$

Taking the solution $W(L)$ to the discrete adjoint equation II.38, we can develop the discrete hybrid calculation of the Hessian of J_{EOLS} using the following equations:

1. $a(\delta L, \delta\overline{U}, \overline{U}(L) - \overline{Z}) = \delta L^{\text{T}} \nabla\overline{U}(L)^{\text{T}} \widehat{K}(\delta L)\left(\overline{U}(L) - \overline{Z}\right) = \delta L^{\text{T}} \nabla\overline{U}(L)^{\text{T}} \mathbb{A}(\overline{U}(L) - \overline{Z})\delta L$
2. $a(L, \delta\overline{U}, \delta\overline{U}) = \delta L^T \nabla\overline{U}(L)^{\text{T}} \widehat{K}(L)\nabla\overline{U}(L)\delta L = \delta L^T \nabla\overline{U}(L)^{\text{T}} \widehat{K}(L)\nabla\overline{U}(L)\delta L$
3. $c(\delta P, \delta P) = \delta L^{\text{T}} \nabla P^{\text{T}} C \nabla P \delta L$
4. $a(\delta L, \delta\overline{U}(L), \overline{W}(L)) = \delta L^T \nabla\overline{U}(L)^{\text{T}} \widehat{K}(\delta L)\overline{W}(L) = \delta L^T \nabla\overline{U}(L)^{\text{T}} \mathbb{A}(\overline{W}(L))\delta L.$

Consequently, we have the following explicit formula for the Hessian:

$$\begin{aligned}\nabla^2 J_{\text{EOLS}}(L) = 2^{\text{T}} \nabla\overline{U}(L)^{\text{T}} \mathbb{A}(\overline{U}(L) - \overline{Z}) \ &+ \nabla\overline{U}(L)^{\text{T}} \hat{K}(L)\nabla\overline{U}(L) \\ &+ \nabla P(L)^{\text{T}} C \nabla P(L) + 2\nabla\overline{U}(L)^{\text{T}} \mathbb{A}(\overline{W}(L)).\end{aligned} \tag{II.40}$$

In the above equation, we find $\nabla U = (\nabla\overline{U}, \nabla P)$ through the direct solution of

$$\begin{aligned}\widehat{K}(L)\nabla\overline{U} + B^{\text{T}}\nabla P &= -\widehat{K}(\delta L)\overline{U}(L) \\ B\nabla\overline{U} - C\nabla P &= 0\end{aligned} \tag{II.41}$$

or equivalently

$$\begin{aligned}\widehat{K}(L)\nabla\overline{U} + B^{\text{T}}\nabla P &= -\mathbb{A}(U(L))\delta L \\ B\nabla\overline{U} - C\nabla P &= 0.\end{aligned} \tag{II.42}$$

II.5.3 Discrete EE

From (II.28) and (II.29), we have

$$\begin{aligned}(K + M)E_1 &= \widehat{K}(L)\overline{Z} + B^T P - F \\ M_Q E_2 &= B\overline{Z} - C\widehat{Z}.\end{aligned}$$

and consequently

$$\begin{aligned}E_1 &= (K + M)^{-1}\left(\widehat{K}(L)\overline{Z} + B^T\widehat{Z} - F\right) \\ E_2 &= M_Q^{-1}\left(B\overline{Z} - C\widehat{Z}\right)\end{aligned}$$

where K and M are the stiffness matrix and mass matrix corresponding to $\overline{V}_h$, respectively and M_Q is the mass matrix in Q_h. $Z = (\overline{Z}, \widehat{Z})$ is again the discrete measured data.

Using the above formulas, we then have

$$\begin{aligned} J_{\text{EE}}(L) &= \frac{1}{2}\langle E_1, (M+K)E_1\rangle + \frac{1}{2}\langle E_2, M_Q E_2\rangle \\ &= \frac{1}{2}\left\langle \widehat{K}(L)\overline{Z} + B^T\widehat{Z} - F, (K+M)^{-1}\left(\widehat{K}(L)\overline{Z} + B^T\widehat{Z} - F\right)\right\rangle + \frac{1}{2}\left\langle B\overline{Z} - C\widehat{Z}, M_Q^{-1}\left(B\overline{Z} - C\widehat{Z}\right)\right\rangle \\ &= \frac{1}{2}\left\langle \mathbb{A}(\overline{Z})\mu + B^T\widehat{Z} - F, (K+M)^{-1}\left(\mathbb{A}(\overline{Z})\mu + B^T\widehat{Z} - F\right)\right\rangle + \frac{1}{2}\left\langle B\overline{Z} - C\widehat{Z}, M_Q^{-1}\left(B\overline{Z} - C\widehat{Z}\right)\right\rangle. \end{aligned}$$

The discrete first derivative is given by

$$DJ_{\text{EE}}(L)(\delta L) = \left\langle \mathbb{A}(\overline{Z})\delta L, (K+M)^{-1}\left(\mathbb{A}(\overline{Z})L + B^T\widehat{Z} - F\right)\right\rangle$$

and consequently,

$$\nabla J_{\text{EE}} = \mathbb{A}(\overline{Z})^T(K+M)^{-1}\left(\mathbb{A}(\overline{Z})L + B^T\widehat{Z} - F\right).$$

For the second derivative, we then have

$$\begin{aligned} D^2J_{\text{EE}}(L)(\delta L, \delta L) &= \left\langle L(Z)\delta L, (K+M)^{-1}L(Z)\delta L\right\rangle \\ &= \left\langle L(Z)^T(K+M)^{-1}L(Z)\delta L, \delta L\right\rangle, \end{aligned}$$

which implies that the Hessian of $J_{\text{EE}}(L)$ is

$$\nabla^2 J_{\text{EE}}(L) = L(Z)^T(K+M)^{-1}L(Z). \tag{II.43}$$

III. Gradient and Extragradient Methods

III.1 Introduction

In this section, we introduce several variants of both projected gradient and extragradient methods to solve the tumor identification/elastography inverse problem by posing its solution as a variational inequality:

$$\begin{aligned} &\text{Find } \mu^* \in A \text{ such that:} \\ &\langle \nabla J(\mu^*), \mu - \mu^* \rangle \geq 0 \quad \text{for all } \mu \in A \end{aligned} \tag{III.1}$$

where $A \subset B$ is our set of admissible coefficients (nonempty, closed, and convex), B is our coefficient space (a Banach space), and J is the MOLS, EOLS, or EE functional. Although an abuse of notation, we take these spaces as stand-ins for their discretized counterparts for the remainder of the discussion.

The above variational inequality has a unique solution if ∇J is strongly monotone. That is, if

$$\langle \nabla J(\mu_1) - \nabla J(\mu_2), \mu_1 - \mu_2 \rangle \geq \ell \|\mu_1 - \mu_2\|^2, \quad \forall\ \mu_1, \mu_2 \in A, \quad \ell > 0,$$

and if ∇J is Lipschitz continuous:

$$\|\nabla J(\mu_1) - \nabla J(\mu_2)\| \leq L \|\mu_1 - \mu_2\|, \quad \forall \mu_1, \mu_2 \in A, \quad L > 0.$$

Projected gradient methods are constrained extensions of the classical unconstrained gradient-based iterative methods such as the well-known steepest descent method,

$$\mu_{k+1} = \mu_k - \alpha \nabla J(\mu_k), \tag{III.2}$$

where $\alpha > 0$ is a parameter determining the "step length" of the algorithm. The projected methods subsequently incorporate a projection mapping, $P_A : B \to A$, such that $y = P_A(x)$ if and only if

$$\|x - y\|_B^2 \leq \|x - z\|_B^2 \quad \text{for all } z \in A. \tag{III.3}$$

This gives rise to iterative methods such as:

$$\mu_{k+1} = P_A(\mu_k - \alpha \nabla J(\mu_k)) \tag{III.4}$$

where each iterate μ_k lies within A and where $\mu_k \to \mu^*$ under the strong monotonicity of ∇J and where the step length α satisfies

$$\alpha \in \left(0, \frac{2\ell}{L^2}\right).$$

Extragradient methods were initially introduced to solve minimization and saddle point problems and have received considerable interest recently, particularly in the context of variational inequalities (see [8, 9, 11, 20, 25, 26, 27, 29, 32, 33, 36, 37, **?**, 38, 42, 64, 69, 68, 71, 72, 87, 90, 88, 89, 94, 92, 95, 96, 99, 98, 101, 104, 105, 106, 108, 110, 120, 123, 124, 125, 130, 132, 136, 140, 138, 139, 141, 142, 143, 144, 148]).

Korpelevich [94], the method's progenitor, gave the name "extragradient" to the method due to the second or "extra" evaluation of the objective function f being minimized per algorithm iteration. When solving variational inequalities, this extra evaluation of f corresponds to an extra evaluation of the gradient of f. These methods are also known by more descriptive names such as double-projection [69] and prediction-correction [67] methods.

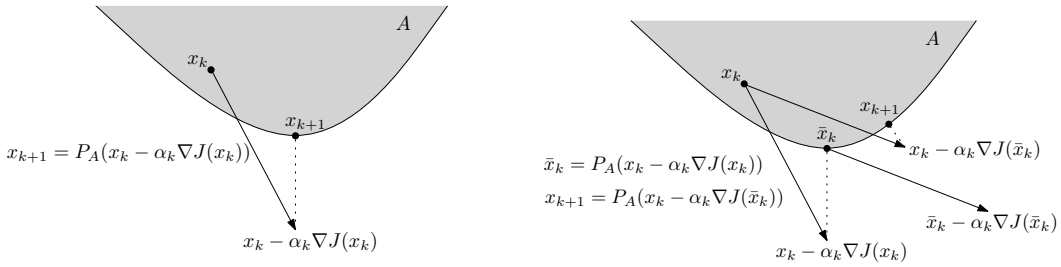
In the remainder of this section, we detail the implementation of numerous gradient and extragradient algorithms and in a subsequent section, we present a thorough numerical comparison of the application of all of these methods to the tumor identification inverse problem.

III.2 Extragradient and Gradient Methods

We analyze the following iterative schemes for solving the tumor identification/elastography inverse problem:

1. Gradient Projection Using Armijo Line Search
2. Fast Gradient Projection Using Armijo Line Search
3. Scaled Gradient Projection Using Barzilai-Borwein Rules
4. Khobotov Extragradient Method Using Marcotte Rules (3 Variants)
5. Solodov-Tseng Projection-Contraction Method
6. Improved He-Goldstein Type Extragradient Method
7. Two-step Extragradient Method
8. Hyperplane Extragradient Method

In Figure 12, we see a geometric comparison of the simplest examples of both a projected gradient and extragradient method. It is clear to see here the "extra" step in the extragradient method as well the sense in which the first step is a "prediction" followed by a secondary "correction".



(a) Projected Gradient Method **(b)** Khobotov Extragradient Method

Figure 12: Geometric comparison of gradient and extragradient iterative methods

III.2.1 Gradient Projection Method

The projected gradient algorithm (as outlined in the introduction) takes the form:

$$\mu^{k+1} = P_A(\mu^k - \alpha \nabla J(\mu^k)).$$

Again, the strong convergence of the iterative method can be established by assuming that

$$\alpha \in \left(0, \frac{2\ell}{L^2}\right),$$

where ℓ and L are the modulus of strong monotonicity and Lipschitz continuity.

However, information about ℓ and L is unavailable, practically, and hence some estimation method must be used to determine the step length α.

An Armijo line search can be used to backtrack from an arbitrary step size until the condition of *sufficient decrease* is satisfied [85]:

$$J(\mu^{k+1}) - J(\mu^k) \leq -\alpha\lambda \|\nabla J(\mu^k)\|^2, \quad \text{for } \lambda \in (0, 1).$$

III.2.2 Fast Iterative Shrinkage-Thresholding Method

Beck and Teboulle [19], following the work of Nesterov [113], presented a fast version of the projected gradient method which is an "optimal" first order method (see [111] for more details). Let L be the Lipschitz constant of ∇J. The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) is as follows.

Algorithm: FISTA

Choose $B^1 = \mu^0$, $t_1 = 1$, and N, the maximum number of iterations.

For $k = 0, 1, 2, \ldots, N$, perform the following:

Step 1: $\mu^k = P_A(B^k - \frac{1}{L}\nabla J(B^k))$

Step 2: $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$

Step 3: $B^{k+1} = \mu^k + \frac{t_k - 1}{t_{k+1}}(\mu^k - \mu^{k-1})$

End

Convergence is guaranteed under the assumption of Lipschitz continuity. A more practical version of the method (and the one actually implemented) uses Armijo line search in order to estimate a step length $0 < \alpha \lesssim \frac{1}{L}$.

III.2.3 Scaled Gradient Projection

The scaled gradient projection (SGP) method is a variant of the projected gradient algorithm that scales the step direction to accelerate convergence. The iterative step of the SGP algorithm takes the form:

$$\mu^{k+1} = P_A(\mu^k - \alpha_k D_k \nabla J(\mu^k)),$$

where D_k is a scaling matrix and α_k is again some step length. It is a common practice to take the scaling matrix D_k as the inverse of the main diagonal of the Hessian of $J(\mu^k)$ with all other entries equal to zero. This can accelerate convergence by mimicking a full Newton-type method such as

$$\mu_{k+1} = \mu_k - \alpha_k \left[\nabla^2 J(\mu_k)\right]^{-1} \nabla J(\mu_k)$$

while reducing the total computational cost associated with the inversion of the complete Hessian.

Algorithm: SGP

Choose $\mu^0 \in \mu_m$, β, $\theta \in (0,1)$, $0 < \alpha_{min} < \alpha_{max}$, $M > 0$

For $k = 0, 1, 2, \ldots,$ perform the following steps:

Step 1: Choose $\alpha_k \in [\alpha_{min}, \alpha_{max}]$ and D_k

Step 2: Projection $Y^k = P_A(\mu^k - \alpha D_k \nabla J(\mu^k))$

If $Y^k = \mu^k$ Stop

Step 3: Descent direction: $d^k = Y^k - \mu^k$

Step 4: Set $\lambda_k = 1$ and $f_{max} = \max_{0 \le j \le \min(k, M-1)} J(\mu^{k-j})$

Step 5: Backtracking loop:

If $J(\mu^k + \lambda_k d^k) \le f_{max} + \beta \lambda_k \nabla J(\mu^k)^T d^k$

Go to Step 6

Else

Set $\lambda_k = \theta \lambda_k$ and go to Step 5

EndIf

Step 6: $\mu^{k+1} = \mu^k + \lambda_k d^k$

End

Here, α_k is chosen using Barzilai-Borwein rules. That is, for

$$\begin{aligned} r^{k-1} &= \mu^k - \mu^{k-1} \\ z^{k-1} &= \nabla J(\mu^k) - \nabla J(\mu^{k-1}), \end{aligned}$$

we compute:

$$\begin{aligned} \alpha_k^{(1)} &= \frac{r^{(k-1)^T} D_k^{-1} D_k^{-1} r^{(k-1)}}{r^{(k-1)^T} D_k^{-1} z^(k-1)} \\ \alpha_k^{(2)} &= \frac{r^{(k-1)^T} D_k z^{(k-1)}}{z^{(k-1)^T} D_k^2 z^{(k-1)}}. \end{aligned}$$

Determining α_k

If $\frac{\alpha_k^{(2)}}{\alpha_k^{(1)}} \leq \tau_k$ then

$\alpha_k = \min\left(\alpha_j^{(2)}, j = \max\left(1, k - M_\alpha\right), ..., k\right)$

$\tau_{k+1} = 0.9\tau_k$

Else

$\alpha_k = \alpha_k^{(1)}$

$\tau_{k+1} = 1.1\tau_k$

EndIf.

III.2.4 Korpelevich Extragradient Method

We now explore the extragradient method proposed by Korpelevich [94] to relax the conditions on convergence for the projection method:

$$\begin{aligned} \bar{\mu}^k &= P_A(\mu^k - \alpha \nabla J(\mu^k)) \\ \mu^{k+1} &= P_A(\mu^k - \alpha \nabla J(\bar{\mu}^k)), \end{aligned}$$

where α is constant for all iterations.

Convergence can be proven under the following conditions:

1. The solution set is non-empty
2. ∇J is monotone
3. ∇J is Lipschitz continuous with constant L
4. $\alpha \in \left(0, \frac{1}{L}\right)$

Without direct knowledge of L, it follows from the above conditions that it can be impossible, in practice, to select an α that guarantees convergence. The temptation would be to choose an arbitrarily small value for α, however, although a small enough α guarantees convergence, convergence can be slow enough to be practically indistinguishable from non-convergence. This suggests a direct improvement of the basic extragradient method, where α is replaced by an adaptive step length based on some estimation of L.

III.2.5 Khobotov Extragradient Method

Khobotov [88] introduced the method below that removes the Lipschitz continuity constraint on ∇J. The adaptive algorithm now has the following two-step iteration:

$$\begin{aligned} \bar{\mu}^k &= P_A(\mu^k - \alpha_k \nabla J(\mu^k)) \\ \mu^{k+1} &= P_A(\mu^k - \alpha_k \nabla J(\bar{\mu}^k)). \end{aligned}$$

Again, a reduced α_k guarantees convergence, but to prevent slow convergence, it is obvious that how the sequence of $\{\alpha_k\}$ reduces must be controlled in some sense.

Khobotov provides the following reduction rule for α_k in [88]:

$$\alpha_k > \beta \frac{\|A^k - \bar{A}^k\|}{\|\nabla J(A^k) - \nabla J(\bar{A}^k)\|},$$

where $\beta \in (0,1)$. Numerical results from [132] and [88] suggest that a β-value between 0.8 and 0.9, is sufficient to effectively control the reduction of α_k.

Khobotov's extragradient method is then as follows:

Algorithm: Khobotov Extragradient

Choose α_0, μ^0, and $\beta \in (0,1)$
While $\|\mu^{k+1} - \mu^k\| >$ TOL
  Step 1: Compute $\nabla J(\mu^k)$
  Step 2: Compute $\bar{\mu}^k = P_A(\mu^k - \alpha_k \nabla J(\mu^k))$
  Step 3: Compute $\nabla J(\bar{\mu}^k)$
    If $\nabla J(\bar{\mu}^k) = 0$, Stop
  Step 4: If $\alpha_k > \beta \frac{\|\mu^k - \bar{\mu}^k\|}{\|\nabla J(\mu^k) - \nabla J(\bar{\mu}^k)\|}$
    then reduce α_k and go to Step 5
  Step 5: Compute $\mu^{k+1} = P_A(\mu^k - \alpha_k \nabla J(\bar{\mu}^k))$
End.

III.2.6 Marcotte Reduction Rules for Step Length

Khobotov's algorithm gives one workable method for reducing α_k but does not rule out other, perhaps more desirable, methods. Marcotte developed a new rule for reducing α_k along with closely related variants [100, 132]. The first Marcotte rule is based on the sequence $a_k = \frac{1}{2} a_{k-1}$ and forces

α_k to satisfy Step 5 of Khobotov's algorithm by additionally taking:

$$\alpha_k = \min\left\{\frac{\alpha_{k-1}}{2}, \frac{\|\mu^k - \bar{\mu}^k\|}{\sqrt{2}\|\nabla J(\mu^k) - \nabla J(\bar{\mu}^k)\|}\right\}.$$

Both the Khobotov and Marcotte reduction rules can still run the risk of choosing an initial α small enough that α_k is never reduced, resulting in slow convergence. Ideally, α_k should then have the ability to increase if α_{k-1} is smaller than some optimal value. This leads to a modified version of Marcotte's rule where an initial α is selected using the rule

$$\alpha = \alpha_{k-1} + \gamma\left(\beta\frac{\|\mu^{k-1} - \bar{\mu}^{k-1}\|}{\|\nabla J(\mu^{k-1}) - \nabla J(\bar{\mu}^{k-1})\|} - \alpha_{k-1}\right)$$

where $\gamma \in (0, 1)$.

The reduction rule in Step 5 of Khobotov's algorithm is then replaced with

$$\alpha_k = \max\left\{\widehat{\alpha}, \min\left\{\xi \cdot \alpha, \beta\frac{\|\mu^{k-1} - \bar{\mu}^{k-1}\|}{\|\nabla J(\mu^{k-1}) - \nabla J(\bar{\mu}^{k-1})\|}\right\}\right\}$$

where $\xi \in (0, 1)$, and $\widehat{\alpha}$ is some lower limit for α_k (generally taken as no less than 10^{-4}).

III.2.7 Algorithmic stopping criteria

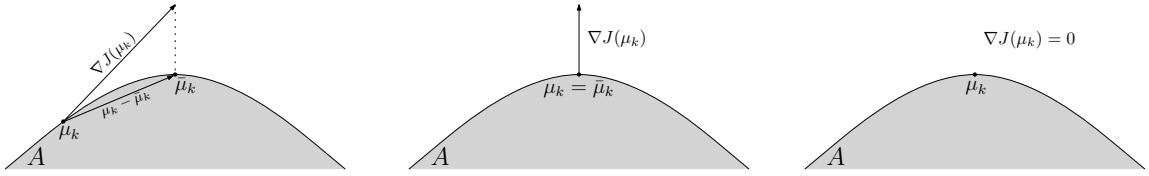
We also note that Marcotte provides the following useful stopping criteria for the Khobotov extragradient methods [100]. We consider the gap function

$$\gamma(\mu_k) := \langle \mu_k - \bar{\mu}_k, \nabla J(\mu_k)\rangle \tag{III.5}$$

and note that γ is both non-negative and that $\gamma(\mu_k) = 0$ implies that μ_k is a solution to the variational inequality (III.1). Thus the condition

$$\gamma(\mu_k) \leq \text{TOL},$$

where TOL is a suitably small constant, provides a reasonable stopping criteria for the extragradient method. When considered geometrically, as in Figure 13, the stopping criteria correlates with the intuitive notion of perpendicularity between ∇J and $\mu_k - \bar{\mu}_k$.



(a) $\gamma(\mu_k) \neq 0$ **(b)** $\gamma(\mu_k) = 0$ **(c)** $\gamma(\mu_k) = 0$

Figure 13: Geometric interpretation of the gap function stopping criteria

III.2.8 Scaled Extragradient Method

We now examine a representative algorithm from a family of a projection-contraction-type extragradient methods developed by Solodov and Tseng [127]. Similar to the SGP algorithm, this method employs a scaling matrix M to accelerate convergence. The two-step iteration has the form:

$$\begin{aligned} \bar{\mu}^k &= P_A(\mu^k - \alpha_k \nabla J(\mu^k)) \\ \mu^{k+1} &= \mu^k - \gamma M^{-1}(T_\alpha(\mu^k) - T_\alpha(P_A(\bar{\mu}^k)) \end{aligned}$$

where $\gamma \in \mathbb{R}^+$ and $T_\alpha = (I - \alpha \nabla J)$; here, I is the identity matrix, and α is chosen such that T_α is strongly monotone.

Additional discussion regarding the scaling matrix is provided by Tinti in [132]. In both [132] and [127], their numerical experiments consider only the case where $M = I$. We additionally consider a scaling matrix given by the diagonal of the Hessian with the remaining entries zero.

Algorithm: Solodov-Tseng

Choose $\mu^0, \alpha_{-1}, \theta \in (0,2), \rho \in (0,1), \beta \in (0,1), M \in \mathbb{R}^{m \times m}$

Initialize: $\bar{\mu}^0 = 0, k = 0, rx = \text{ones}(m,1)$

While $\|rx\| > \text{TOL}$

Step 1: if $\|rx\| < \text{TOL}$ then Stop

else $\alpha = \alpha_{k-1},\ flag = 0$

Step 2: if $\nabla J(\mu^k) = 0$ then Stop

Step 3: While $\alpha(\mu^k - \bar{\mu}^k)^T(\nabla J(\mu^k) - \nabla J(\bar{\mu}^k)) > (1-\rho)\|\mu^k - \bar{\mu}^k\|^2 \quad \text{or} \quad flag = 0$

If $flag \neq 0$ Then $\alpha = \alpha_{k-1}\beta$ endif

update $\bar{\mu}^k = P_A(\mu^k - \alpha \nabla J(\mu^k))$, compute $\nabla J(\bar{\mu}^k)$

$flag = flag + 1$

endwhile

Step 4: update $\alpha_k = \alpha$

Step 5: compute $\gamma = \theta\rho\|\mu^k - \bar{\mu}^k\|^2 / \|M^{1/2}(\mu^k - \bar{\mu}^k - \alpha_k \nabla J(\mu^k) + \alpha_k \nabla J(\bar{\mu}^k))\|^2$

Step 6: compute $\mu^{k+1} = \mu^k - \gamma M^{-1}(\mu^k - \bar{\mu}^k - \alpha_k \nabla J(\mu^k) + \alpha_k \nabla J(\bar{\mu}^k))$

Step 7: $rx = \mu^{k+1} - A^k,\ k = k+1$ go to Step 3

End

We note that the Solodov-Tseng method suggests a generalized extragradient methods:

$$\begin{aligned} \bar{\mu}^k &= P_\mu(\mu^k - \alpha_k \nabla J(\mu^k)) \\ \mu^{k+1} &= P_\mu(\mu^k - \eta_k \nabla J(\bar{\mu}^k)), \end{aligned}$$

where α_k and η_k are chosen according to different rules.

III.2.9 Goldstein-Type Extragradient Methods

The classical Goldstein projection method [24, 97] improves upon the simpler projected gradient method with an iteration step like the following:

$$\mu^{k+1} = P_A(\mu^k - \beta_k \nabla J(\mu^k))$$

where

$$0 < \varepsilon \leq \beta_k \leq \frac{2(1-\varepsilon)}{L}$$

with $0 < \varepsilon \leq \frac{2}{(2+L)}$ and where L is again the Lipschitz constant.

The He-Goldstein method, an extragradient method based on the classical Goldstein method has a two-step iteration of the form

$$\begin{aligned} \bar{\mu}^k &= P_A(\nabla J(\mu^k) - \beta_k \mu^k) \\ \mu^{k+1} &= \mu^k - \frac{1}{\beta_k}\{\nabla J(\mu^k) - \bar{\mu}^k\}. \end{aligned}$$

It can also be expressed:

$$\begin{aligned} r(\mu^k, \beta_k) &= \frac{1}{\beta_k}\{\nabla J(\mu^k) - P_A[\nabla J(\mu^k) - \beta_k \mu^k]\} \\ \mu^{k+1} &= \mu^k - r(\mu^k, \beta_k). \end{aligned}$$

The algorithm below is based on an improved version of the He-Goldstein algorithm in [97] that provides control over the second projection (choosing η_k).

Algorithm: Improved He-Goldstein

Initialize: choose $\beta_U > \beta_L > \frac{1}{(4\tau)}, \gamma \in (0,2), \varepsilon > 0, \mu^0, \beta_0 \in [\beta_L, \beta_U], k = 0$

Step 1: Compute:

$r(\mu^k, \beta_k) = \frac{1}{\beta_k}\{\nabla J(\mu^k) - P_A[\nabla J(\mu^k) - \beta_k \mu^k]\}$

If $\|r(\mu^k, \beta_k)\| \leq \varepsilon$ then Stop

Step 2: $\mu^{k+1} = \mu^k - \gamma \alpha_k r(\mu^k, \beta_k) \quad$ where $\alpha_k := 1 - \frac{1}{4\beta_k \tau}$

Step 3: Update β_k

$\omega_k = \frac{\|\nabla J(\mu^{k+1} - \nabla J(\mu^k)\|}{\beta_k \|\mu^{k+1} - \mu^k\|}$

If $\omega_k < \frac{1}{2}$ Then $\beta_{k+1} = \max\{\beta_L, \frac{1}{2}\beta_k\}$

Else if $\omega_k > \frac{3}{2}$ Then $\beta_{k+1} = \min\{\beta_U, \frac{6}{5}\beta_k\}$

Step 4: $k = k + 1$, go to Step 1

III.2.10 Two-step Extragradient Method

Zykina and Melenchuk consider a three step projection method which they called a two-step extragradient method[148]. Some numerical experiments with saddle point problems for a bilinear function (given in [148]) show that the convergence of this method is faster compared to the standard extragradient method. A simple adaptive version of the algorithm they present takes the form:

$$\begin{aligned} \bar{\mu}^k &= P_A(\mu^k - \alpha_k \nabla J(\mu^k)) \\ \widetilde{\mu}^k &= P_A(\bar{\mu}^k - \eta_k \nabla J(\bar{\mu}^k)) \\ \mu^{k+1} &= P_A(\mu^k - \xi_k \nabla J(\widetilde{\mu}^k)) \end{aligned}$$

where either $\xi_k = \eta_k = \alpha_k$ or all three step lengths are determined independently. Our current implementation of this method takes the simplest approach with all three step lengths the same.

III.2.11 Hyperplane Projection Method

Hyperplane-type method have the following general geometric outline (from [48]). Beginning with some $\mu_k \in A$, we consider (in the common extragradient fashion) the point

$$\bar{\mu}_k = P_A(\mu_k - \nabla J(x)).$$

Backtracking along the line segment connecting μ_k and $\bar{\mu}_k$ using a line search method, we can find some $\tilde{\mu}_k \in A$ such that the hyperplane H^k of μ's satisfying

$$\langle \nabla J(\tilde{\mu}_k), \tilde{\mu}_k - \mu \rangle = 0$$

separates our original iterate μ_k from the solution μ^* for the variational inequality (III.1). Now, we project μ_k onto H^k to get

$$\omega_k = P_{H^k}(\mu_k).$$

Finally, we project back on to A to get the next iterate

$$\mu_{k+1} = P_A(\omega_k).$$

In Figure 14, we see a graphic representation of one step of the hyperplane method outlined above.

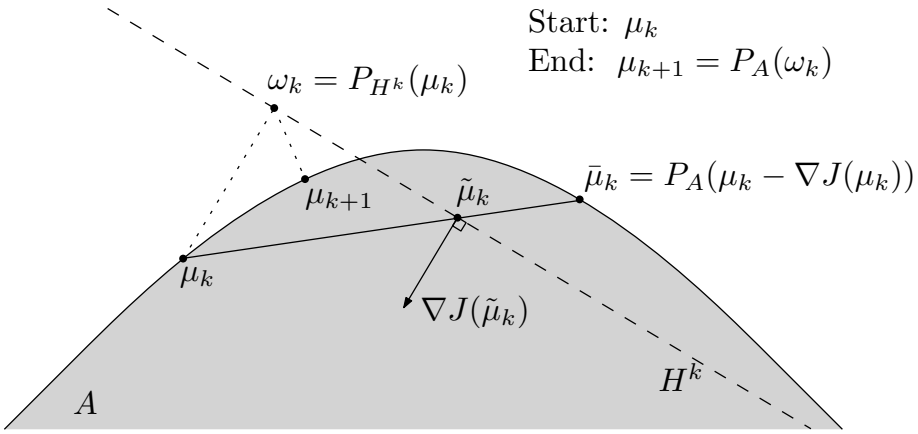


Figure 14: Geometric interpretation of a single iteration of the hyperplane method

The hyperplane method outlined below was presented by Iusem and requires three constants, $\varepsilon \in (0, 1)$ and $\tilde{\alpha} \geq \hat{\alpha} > 0$ such that the sequence α_k is computed where

$$\langle \nabla J(\bar{\mu}^k), \bar{\mu}^k - \mu^k \rangle \leq 0$$

when $\alpha_k \in [\hat{\alpha}, \tilde{\alpha}]$.

Algorithm: Hyperplane (Iusem)

Choose: $\mu^0, \varepsilon, \hat{\alpha}, \tilde{\alpha}$

Initialize: $k = 0, rx = \text{ones}(m, 1)$

While $\|rx\| > \text{TOL}$

Step 1: Choose $\tilde{\alpha}_k$ using a finite bracketing procedure

Step 2: Compute $K^k = P_A(\mu^k - \tilde{\alpha}_k \nabla J(\mu^k))$ and $\nabla J(K^k)$

Step 3: If $\nabla J(K^k) = 0$ then Stop

Step 4: If $\|\nabla J(\tilde{\mu}^k) - \nabla J(\mu^k)\| \leq \frac{\|K^k - \mu^k\|^2}{2\tilde{\alpha}_k^2 \|\nabla J(\mu^k)\|}$

Then $\bar{\mu}^k = K^k$

Else find $\alpha_k \in (0, \tilde{\alpha}_k)$ such that

$\varepsilon \frac{\|K^k - \mu^k\|^2}{2\tilde{\alpha}_k^2 \|\nabla J(\mu^k)\|} \leq \|\nabla J(P_A(\mu^k - \alpha_k \nabla J(\mu^k))) - \nabla J(\mu^k)\| \leq \frac{\|K^k - \mu^k\|^2}{2\tilde{\alpha}_k^2 \|\nabla J(\mu^k)\|}$

Step 5: Compute $\bar{\mu}^k = P_A(\mu^k - \alpha_k \nabla J(\mu^k))$

Step 6: If $\nabla J(\bar{\mu}^k) = 0$ then Stop

Step 7: Compute η_k

Step 8: Compute $\mu^{k+1} = P_A(\mu^k - \eta_k \nabla J(\bar{\mu}^k))$

Step 9: $rx = \mu^{k+1} - \mu^k, k = k + 1$; go to Step 3;

End

IV. Proximal Point Methods

IV.1 Introduction

Generally, parameter identification inverse problems in partial differential equations are highly ill-posed and this is particularly true of the inverse problem under consideration in this work. The general process of overcoming this ill-conditioning is known as *regularization* and perhaps the most commonly used method is Tikhonov regularization [131].

In the context laid out in § II, we recall that the solution to our inverse problem is given by the solution to an optimization problem:

$$\begin{aligned} &\text{Find } \mu^* \in A \text{ such that:} \\ &\qquad J(\mu^*) \leq J(\mu) \quad \text{for all } \mu \in A \end{aligned} \tag{IV.1}$$

where J is either the OLS, MOLS, EOLS, or EE functional.

The application of Tikhonov regularization to this problem begins with a perturbation of the original objective functional:

$$\mathcal{J}(\mu;\beta) = J(\mu) + \beta\|G\mu\|^2 \tag{IV.2}$$

where $\beta \geq 0$ is a constant known as the regularization parameter and where G is some weighting operator on μ (typically G is taken as the identity operator, which we do throughout). We then can then consider the perturbed analogue of the optimization problem[4]:

$$\begin{aligned} &\text{Find } \mu' \in A \text{ such that:} \\ &\qquad \mathcal{J}(\mu';\beta) \leq \mathcal{J}(\mu;\beta) \quad \text{for all } \mu \in A. \end{aligned} \tag{IV.3}$$

In (IV.3), the regularization term $\beta\|\mu\|^2$ acts to penalize those potential minimizers of $J(\mu)$ that are "large" with respect to the given norm. Thus it is not necessarily the case that $\mu' = \mu^*$, but instead, in a certain sense, μ' is a solution to IV.1 with minimal norm. We note that the typical norms used are the H^1, L^2 or H^1-seminorm.

For convex frameworks, like EE or MOLS, Tikhonov regularization gives a unique solution to the associated variational inequality:

$$\langle \nabla J(\mu^*), \mu^* - \mu \rangle \geq 0 \qquad \text{for all } \mu \in A, \tag{IV.4}$$

which in turn is a necessary and sufficient optimality condition for solving the inverse problem. For nonconvex frameworks, like EOLS and OLS, the variational inequality is only a necessary condition,

[4] Truly, a family of optimization problems since $\mathcal{J}(\cdot;\beta)$ is a family of functionals.

but a large enough regularization parameter can be chosen to ensure the uniqueness of solutions. This suggests the glaring downside of the Tikhonov approach.

It can be shown that as $\beta \to 0$, the Tikhonov-perturbed problem (IV.3) approaches the original problem (IV.1) [48]. However, assuming $\mu \neq 0$, it follows that unless $\beta = 0$, the Tikhonov problem (IV.3) and the original problem (IV.1) will never coincide. Therefore, to find a solution as close to the original problem as possible, we must choose β as small as possible but that still guarantees the well-conditioning of the original problem. Thus the proper selection of regularization parameter is of high importance for practical applications and given the typically heuristic nature of the choice, prone to the introduction of error through either over- or under-regularization.

Proximal methods (outlined in detail in the remaining subsections) are another approach to regularization that seem well-suited to the tumor identification inverse problem and which overcome this shortcoming of the Tikhonov regularization approach. Their general outline consists of the progressive replacement of a single convex optimization by a sequence of strictly or strongly convex optimization problems. To briefly illustrate the advantage of proximal methods, we consider the "proximal" perturbed functional

$$\mathcal{J}_P(\mu) = J(\mu) + c_k \|\mu - \mu^{k-1}\|^2 \tag{IV.5}$$

where $\{c_k\}$ is a bounded sequence in $\mathbb{R}$. Then, if the sequence $\{\mu_k\}$ converges, the proximal regularization term $c_k\|\mu - \mu^{k-1}\|^2$ goes to zero even if $\{c_k\}$ does not [48].

For convex problems using Tikhonov regularization, algorithms are known to converge to a minimal-norm solution. Comparatively, for proximal point methods, no such characterization concerning the solution is available. Therefore it is natural to ask whether proximal point methods can be applied directly to inverse problem optimization frameworks, thus obviating the need for the selection of an optimal regularization parameter.

IV.2 Proximal Methods

In this section, we apply several proximal-like optimization algorithms to the optimization frameworks developed in §II. In particular, we examine several variations on the self-adaptive, inexact Hager and Zhang proximal-point algorithms developed in [61].

Our analysis begins with a review of the classical proximal-point algorithm. Here we seek the

solution to our fundamental constrained minimization problem:

$$\min_{\mu \in A} J(\mu) \tag{IV.6}$$

where A is a closed and convex set representing our feasible parameters and J is the $J_{\text{MOLS}}, J_{\text{EOLS}}$, or J_{EE} objective functional.

We now consider the functional

$$\mathcal{J}_P(\mu) = J(\mu) + \frac{1}{2\lambda^k}\|\mu - \mu^k\|^2, \tag{IV.7}$$

where λ_k is a positive number and $\mu^k \in A$. We note that $\mathcal{J}_P(\mu)$ is strictly convex in the case of J_{MOLS} and J_{EE} since J and the introduced quadratic term $\frac{1}{2\lambda^k}\|\mu - \mu^k\|_2^2$, known as the *proximal regularization term*, are both also strictly convex. Thus, we have the related, uniquely-solvable subproblem

$$\min_{\mu \in A} \mathcal{J}_P(\mu) \tag{IV.8}$$

for which the necessary and sufficient optimality conditions in turn yield the following variational inequality problem:

$$\begin{aligned} &\text{Find } \mu^* \in A \text{ such that:} \\ &\langle \nabla \mathcal{J}_P(\mu^*), \mu - \mu^* \rangle \geq 0 \quad \text{for all } \mu \in A. \end{aligned} \tag{IV.9}$$

In this context, the classical proximal-point algorithm generates a sequence $\{\mu^k\}$ such that

$$\mu^{k+1} = \arg\min_{\mu \in A} \left\{ J(\mu) + \frac{1}{2\lambda^k}\|\mu - \mu^k\|^2 \right\} \tag{IV.10}$$

where $\{\lambda^k\}$ is a sequence of positive numbers. It can be shown that, under certain conditions, the iterates $\{x^k\}$ converge to a solution of (IV.6) [80].

Subsequently, we will consider several variations on the above proximal approach coupled with the method of accelerated convergence outlined by Hager and Zhang [61]. For further details on these methods and their history, we refer the interested reader to [63, 81, 116, 122, 135] and the cited references therein.

IV.2.1 Hager and Zhang's Proximal Point Method

Hager and Zhang [61] introduce two criteria between subsequent iterates of (IV.10) for the solution of the subproblem (IV.8):

$$\mathcal{J}_P(\mu^{k+1}) \leq J(\mu^k) \tag{IV.11}$$

$$\|\nabla \mathcal{J}_P(\mu^{k+1})\| \leq \theta^k \|\nabla J(\mu^k)\|. \tag{IV.12}$$

As they detail, taking the proximal regularization parameter as

$$\theta^k = \tau \|\nabla J(\mu^k)\|^\eta,$$

where $\eta \in [0, 2)$ and $\tau > 0$ are constants, gives quadratic convergence of the iterates to the solution set of (IV.8). This gives rise to the following algorithm:

Algorithm: Hager-Zhang Proximal Point

Initialization Step: Choose an initial guess μ^0, initialize τ and η, and take $k = 0$.

Let $\theta^k = \tau \|\nabla J(\mu^k)\|^\eta$ and let $\gamma = 1$.

Step 1: Find μ^{k+1} satisfying $\|\nabla \mathcal{J}_P(\mu^{k+1})\| \leq \theta^k \gamma \|\nabla J(\mu^k)\|$

Step 2:

If μ^{k+1} satisfies $\mathcal{J}_P(\mu^{k+1}) \leq J(\mu^k)$

Go to Step 3.

Else,

Set $\gamma = 0.1\gamma$ and go to Step 1.

End.

Step 3: Let $\mu^k = \mu^{k+1}$.

Step 4: Set $k = k + 1$ and go to Step 1.

In Step 1, the subproblem was solved using an unconstrained conjugate-gradient trust-region method to find the subsequent iterate μ^{k+1}.

IV.2.2 Hager and Zhang's Proximal Point Method Using φ-Divergence

The first variant of the classical proximal algorithm we examine replaces the proximal regularization term in (IV.7) by what are known as φ-divergences (see Kanzow [80] for further details). For their definition, let Φ denote the class of closed, proper and convex functions $\varphi : \mathbb{R} \to (-\infty, \infty]$ which have domain$(\varphi) \subset [0, \infty)$ and which possess the following properties:

1. φ is twice continuously differentiable on int(domain$(\varphi)) = (0, +\infty)$.
2. φ is strictly convex on its domain.
3. $\lim_{t \to 0+} \varphi'(t) = -\infty$.
4. $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$.

5. There exists $\nu \in \left(\frac{1}{2}\varphi''(1), \varphi''(1)\right)$ such that

$$\left(1 - \frac{1}{t}\right)(\varphi''(1) + \nu(t-1)) \leq \varphi'(t) \leq \varphi''(1)(t-1), \quad \forall t > 0.$$

Then for some $\varphi \in \Phi$, the φ-divergence between two $x, y \in \mathbb{R}^n_+$ is given by

$$d_\varphi(x, y) = \sum_{i=1}^{n} y_i \varphi\left(\frac{x_i}{y_i}\right). \tag{IV.13}$$

A few examples of suitable φ functions are

$$\begin{aligned} \varphi_1(t) &= t\log t - t + 1 \\ \varphi_2(t) &= -\log t + t - 1 \\ \varphi_3(t) &= \left(\sqrt{t} - 1\right)^2. \end{aligned}$$

Taking φ_1 above yields the φ-divergence

$$d_{\varphi_1}(x, y) = \sum_{i=1}^{n} x_i \log\frac{x_i}{y_i} + y_i - x_i \tag{IV.14}$$

We now replace the proximal regularization term in (IV.7) with (IV.14) giving

$$\mathcal{J}_{\varphi_1}(\mu) = J(\mu) + \theta^k d_{\varphi_1}(\mu, \mu^k). \tag{IV.15}$$

and subsequently the proximal-like iteration

$$\mu^{k+1} = \arg\min_{\mu \in A} \mathcal{J}_{\varphi_1}(\mu) \tag{IV.16}$$

Substituting $\mathcal{J}_{\varphi_1}(\mu)$ for $\mathcal{J}_P(\mu)$ into Algorithm 1 yields the φ-divergence proximal-like algorithm.

IV.2.3 Hager and Zhang's Proximal Point Method Using Bregman Functions

Continuing in the manner of φ-divergences above, the next modified algorithm replaces the proximal regularization term with another strictly convex distance function defined by

$$D_\psi(x, y) = \psi(x) - \psi(y) - \nabla\psi(y)^T(x - y), \tag{IV.17}$$

where ψ is a so-called Bregman function.

We now review the definition of a Bregman function. Let S be an open and convex set and a let $\psi : \bar{S} \to \mathbb{R}$ be a given mapping. If ψ is a Bregman function, it must satisfy the following criteria:

1. ψ is strictly convex and continuous on $\bar{S}$.
2. ψ is continuously differentiable in S.
3. The partial level set
$$L_\alpha = \{y \in \bar{S} | D_\psi(x, y) \leq \alpha\}$$
is bounded for every $x \in \bar{S}$.
4. If $\{y^k\} \subset \bar{S}$ converges to x, then $\lim_{k\to\infty} D_\psi(x, y^k) = 0$.

A few examples of Bregman functions are given below:

$$\begin{aligned}
\psi_1(x) &= \frac{1}{2}\|x\|^2 \quad \text{with } S = \mathbb{R}^n, \\
\psi_2(x) &= \sum_{i=1}^{n} x_i \log x_i - x_i \quad \text{with } S = \mathbb{R}^n_+, \\
\psi_3(x) &= -\sum_{i=1}^{n} \log x_i \quad \text{with } S = \mathbb{R}^n_+.
\end{aligned}$$

We note that the corresponding distance function for ψ_1 is

$$D_{\psi_1}(x, y) = \frac{1}{2}\|x - y\|^2,$$

which recovers the original proximal regularization term introduced in (IV.7). Similarly, for ψ_2 we have

$$D_{\psi_2}(x, y) = \sum_{i=1}^{n} x_i \log \frac{x_i}{y_i} + y_i - x_i$$

which is equivalent to the φ-divergence given by (IV.14).

Taking

$$D_{\psi_3}(x, y) = \sum_{i=1}^{n} \frac{x_i}{y_i} - \log \frac{x_i}{y_i} - 1$$

we can again replace the proximal regularization term to get the functional

$$\mathcal{J}_{\psi_3}(\mu) = J(\mu) + \theta^k D_{\psi_3}(\mu, \mu^k) \tag{IV.18}$$

and the corresponding subproblem

$$\mu^{k+1} = \arg\min_{\mu \in A} \mathcal{J}_{\psi_3}(\mu). \tag{IV.19}$$

IV.2.4 Proximal-like Methods Using Modified φ-Divergence

To solve the subproblem in Step 1 of the Hager-Zhang algorithm using the φ-divergence method, we made use of a fast conjugate-gradient-based trust-region method. Applying a second-order Newton-type method to solve the subproblem requires the calculation of the Hessian of both the objective functional and proximal regularization term.

To avoid possible ill-conditioning of the Hessian (see [80] for a full discussion), we replace the original φ-divergence function with

$$\tilde{d}_{\varphi}(x,y) = \sum_{i-1}^{n} y_i^2 \varphi\left(\frac{x_i}{y_i}\right) \tag{IV.20}$$

giving

$$\nabla_{xx}^2 \tilde{d}_{\varphi}(x,y) = \sum_{i=1}^{n} \varphi''\left(\frac{x_i}{y_i}\right) e_i e_i^T \tag{IV.21}$$

where e_i is the i-th unit basis vector of $\mathbb{R}^n$.

Again taking

$$\mathcal{J}_{\tilde{\varphi}_1}(\mu) = J(\mu) + \tilde{d}_{\varphi_1}(\mu, \mu^k) \tag{IV.22}$$

we have the iteration

$$\mu^{k+1} = \arg\min_{\mu \in A} \mathcal{J}_{\tilde{\varphi}_1}(\mu). \tag{IV.23}$$

However, now equipped with the Hessian of $\mathcal{J}_{\tilde{\varphi}_1}(\mu)$, we can apply a full Newton-type method to solve the associated subproblem.

V. Numerical Experiments

In this section we consider a representative example of an elastography inverse problem for the recovery of a variable μ on a two dimensional isotropic domain $\Omega = (0,1) \times (0,1)$ with boundary $\partial\Omega = \Gamma_1 \cup \Gamma_2$. Γ_1, where the Dirichlet boundary conditions hold, is taken as the top boundary of the domain while Γ_2, where the Neumann conditions hold, is taken as the union of the remaining boundary points.

The inverse problem is solved on a 30×30 quadrangular mesh with 441 quadrangles and 1409 total degrees of freedom (see §II.5 for a complete discussion of the discretization).

In keeping with the near incompressibility inherent in the problem, λ is taken as a large constant, particularly $\lambda = 10^6$.

The functions defining the coefficient, load, and boundary conditions are as follows:

$$\begin{aligned} &\mu(x,y) = 2.5 + \frac{1}{4}\sin(2\pi xy)\cos(\pi y), && f(x,y) = \begin{bmatrix} 2.3 + \frac{1}{10}x \\ 2.3 + \frac{1}{10}y \end{bmatrix}, \\ &g(x,y) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ on } \Gamma_1, && h(x,y) = \begin{bmatrix} \frac{1}{2} + x^2 \\ \frac{1}{2} + y^2 \end{bmatrix} \text{ on } \Gamma_2. \end{aligned} \tag{V.1}$$

The examples using extragradient and gradient methods all employ Tikhonov regularization with a fixed regularization parameter $\beta = 10^{-5}$ to ensure comparable results. The proximal point algorithms use proximal regularization only. See §IV for a detailed discussion of the difference between both regularization approaches.

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Korpelevich	76,130	76,130	–	38,062	$1.292 \cdot 10^{-3}$
Khobotov	13,172	13,172	–	6,583	$1.176 \cdot 10^{-3}$
Marcotte 1	29,082	29,082	–	14,536	$1.273 \cdot 10^{-3}$
Marcotte2	13,966	13,966	–	5,234	$1.184 \cdot 10^{-3}$
He-Goldstein	41,759	41,759	–	41,753	$1.180 \cdot 10^{-3}$
Solodov-Tseng (I)	$1.29 \cdot 10^5$	$1.29 \cdot 10^5$	–	42,937	$1.199 \cdot 10^{-3}$
Solodov-Tseng (H)	3,286	3,286	1	1,090	$1.366 \cdot 10^{-3}$
Hyperplane	35,730	35,730	–	5,530	$1.195 \cdot 10^{-3}$
Two-step	16,146	16,146	–	5,312	$1.191 \cdot 10^{-3}$
Projected Gradient	26,698	10,007	–	10,000	$1.172 \cdot 10^{-3}$
Scaled Projected Gradient	2,119	917	1	905	$1.172 \cdot 10^{-3}$
FISTA	$1.08 \cdot 10^5$	5,005	–	5,000	$1.173 \cdot 10^{-3}$

Table 1: Comparison of Gradient and Extragradient Methods (EE)

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Hager-Zhang	4,150	3,970	–	178	$9.230 \cdot 10^{-9}$
φ-divergence	3,922	3,695	–	176	$1.510 \cdot 10^{-8}$
Bregman	3,575	3,378	–	159	$3.210 \cdot 10^{-8}$
Quadratic φ	54	50	16	32	$8.760 \cdot 10^{-9}$

Table 2: Comparison of Proximal Point Methods (EE)

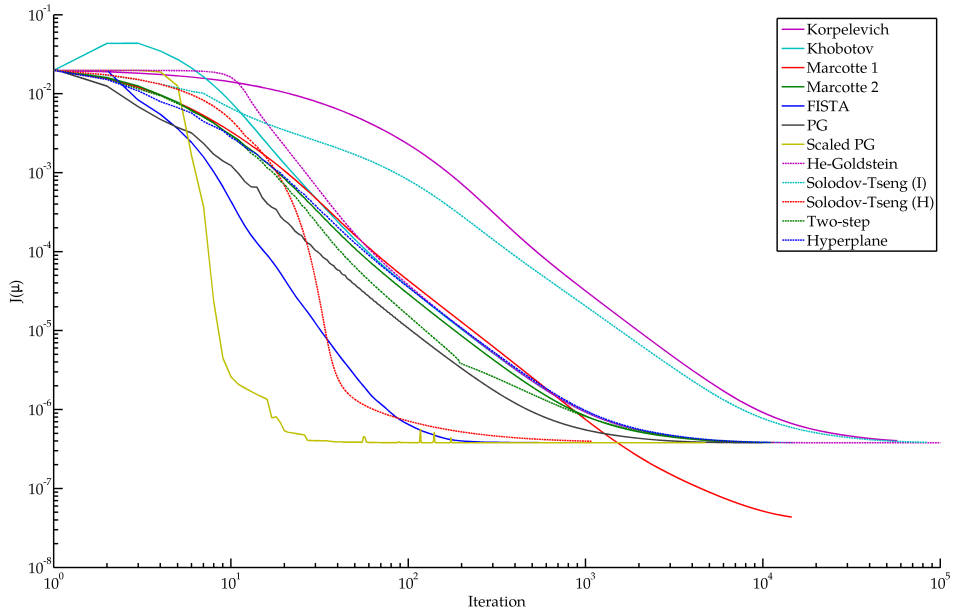


Figure 15: Convergence of $J(\mu)$ using EGM (EE)

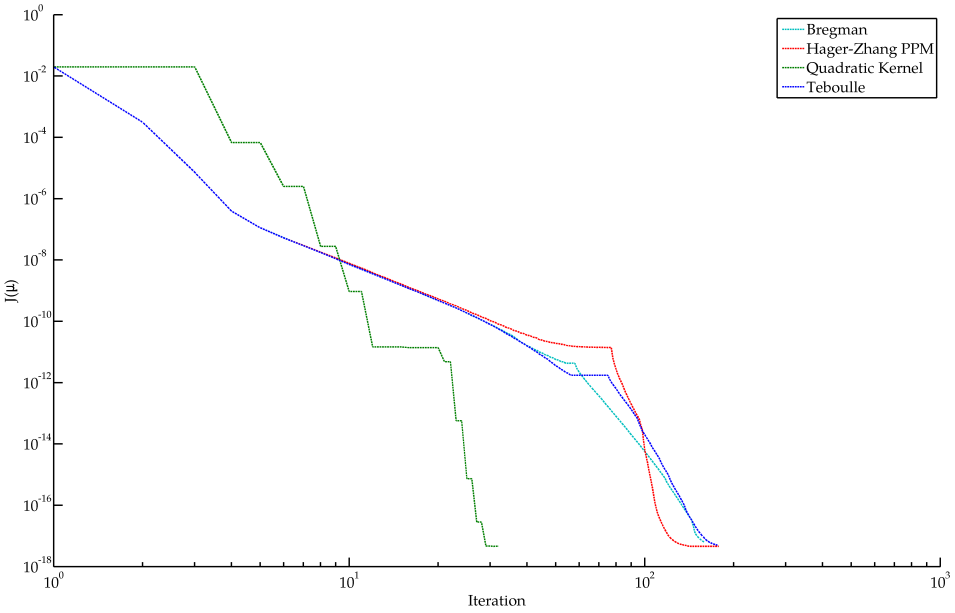


Figure 16: Convergence $J(\mu)$ using PPM (EE)

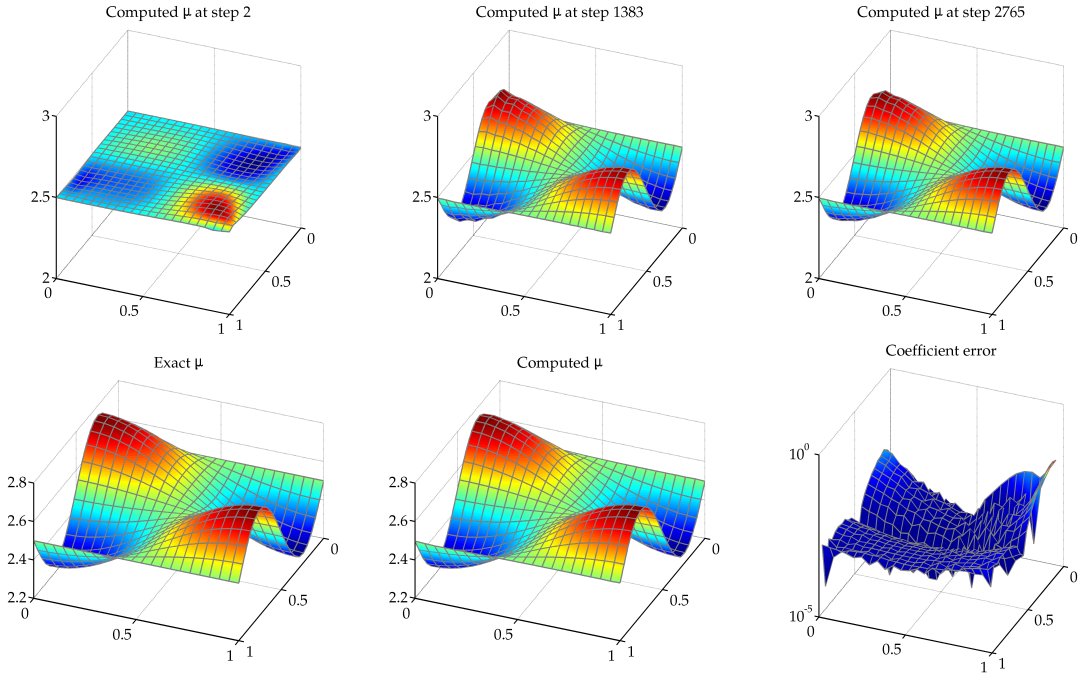


Figure 17: Hyperplane Method (EE) (EE)

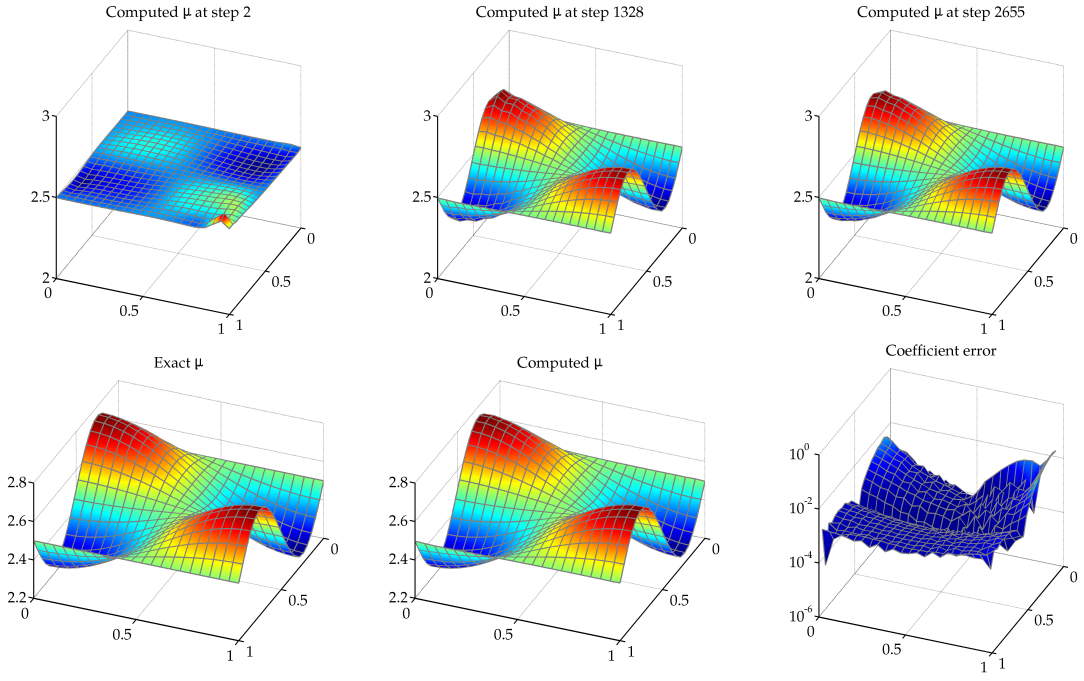


Figure 18: Two-step Method (EE) (EE)

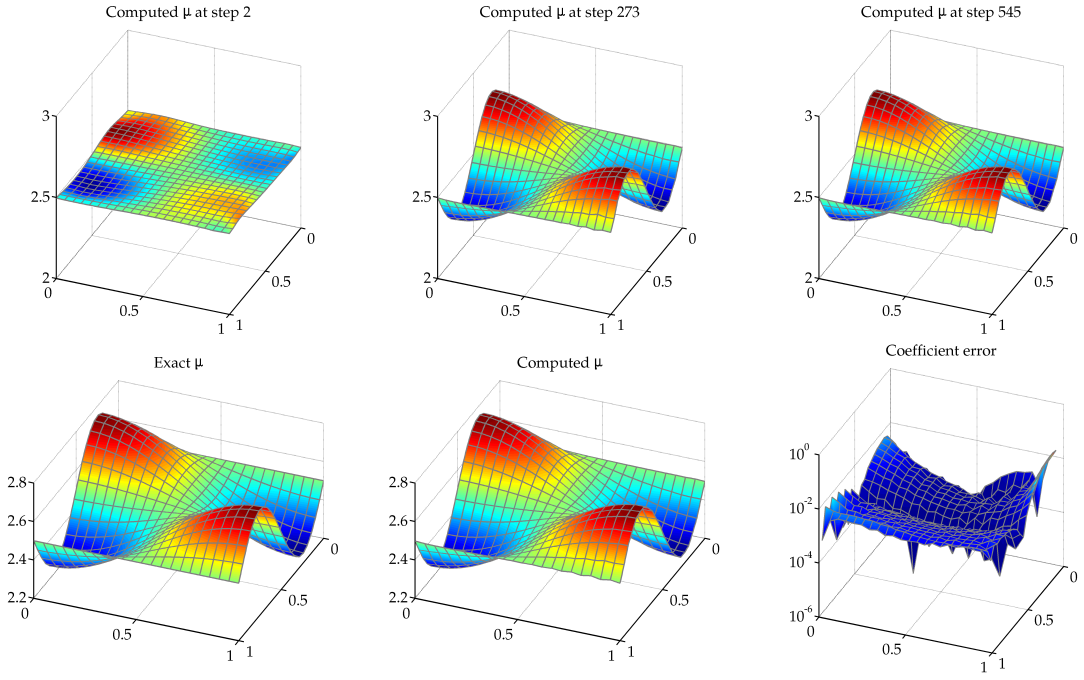


Figure 19: Solodov-Tseng Method using Hessian (EE)

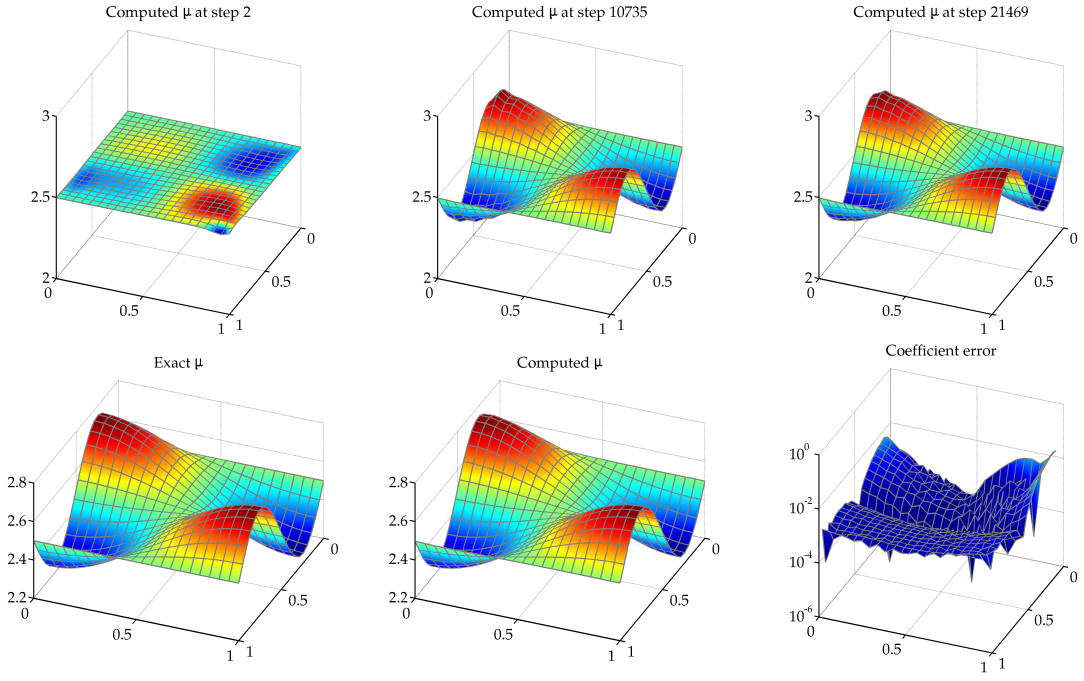


Figure 20: Solodov-Tseng Method using Identity (EE)

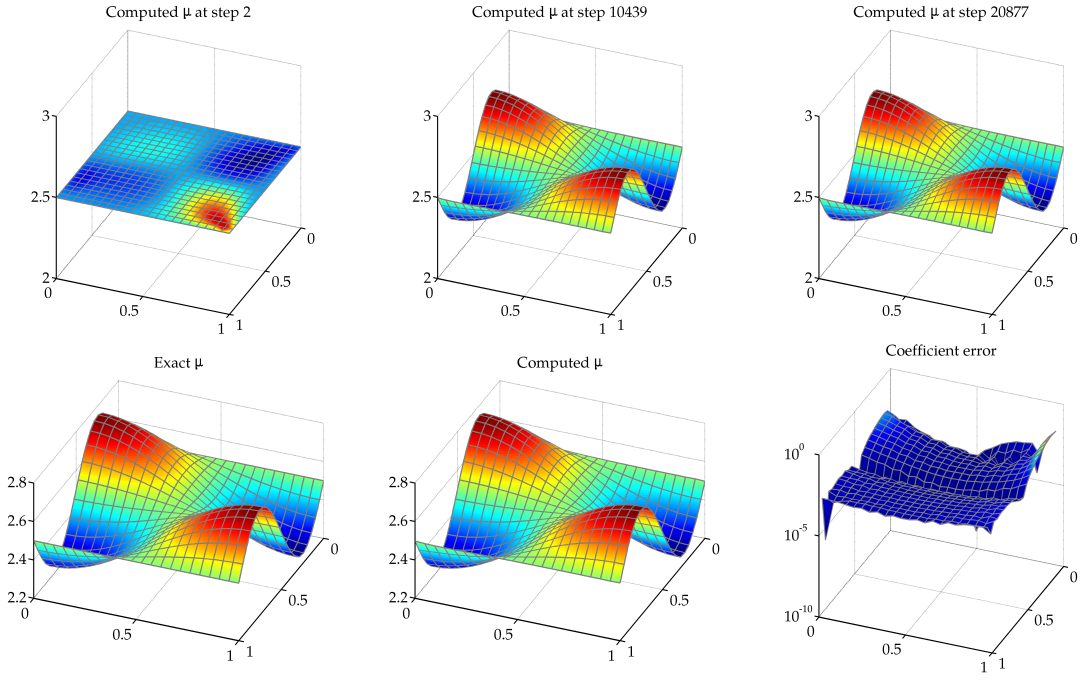


Figure 21: He-Goldstein Method (EE)

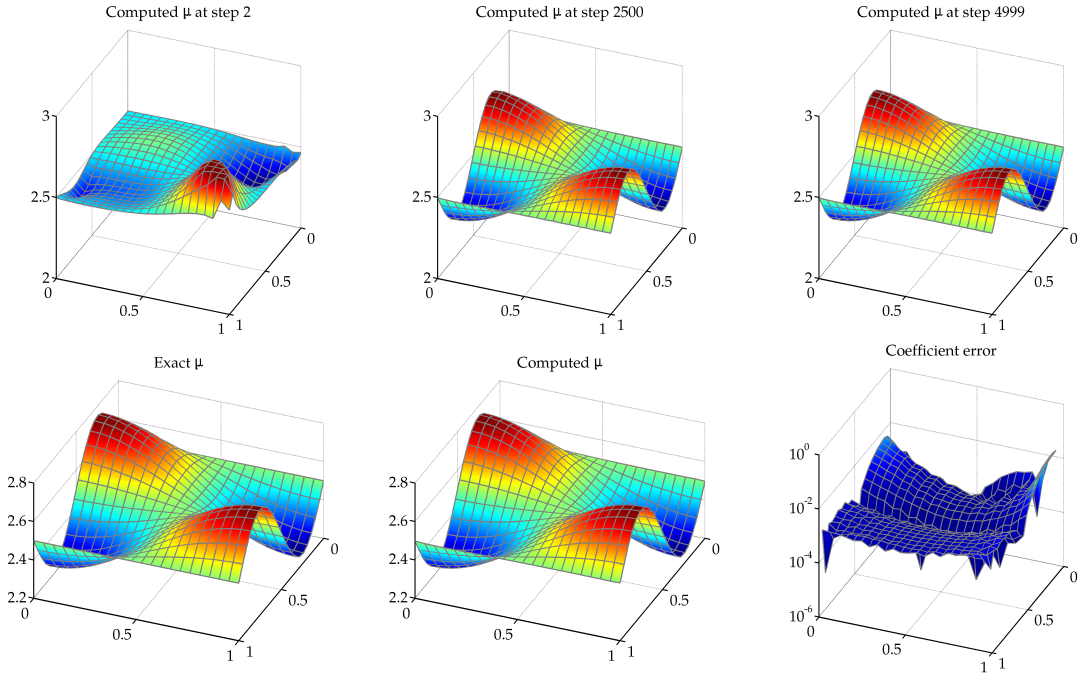


Figure 22: Projected Gradient with Hessian (EE)

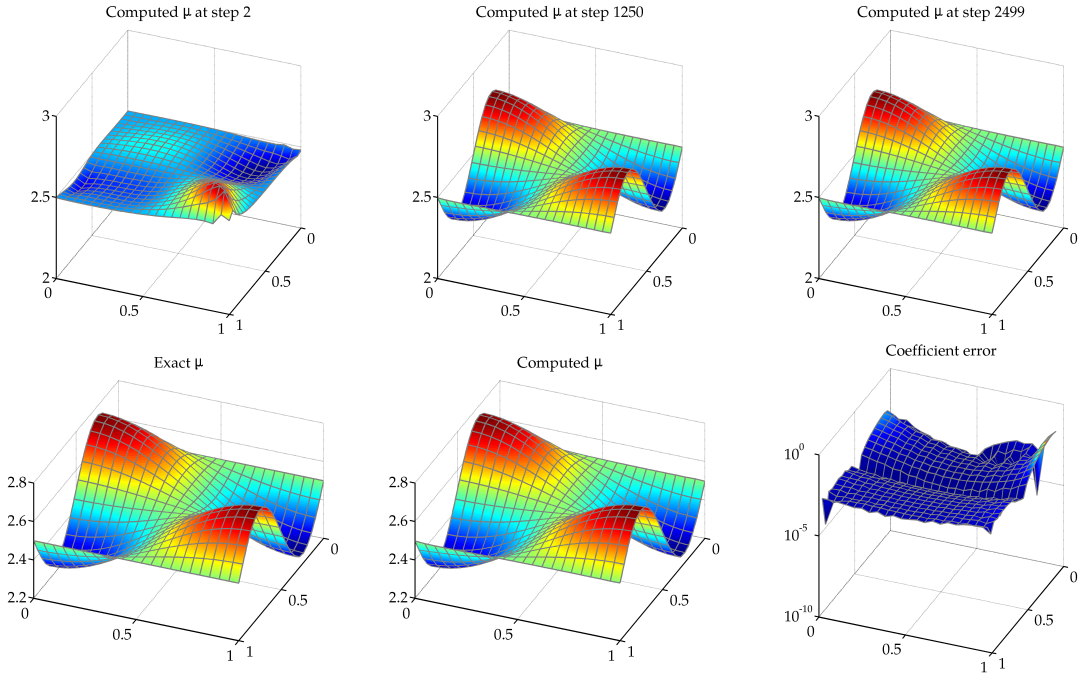


Figure 23: Fast Iterative Shrinkage-Thresholding (FISTA) (EE)

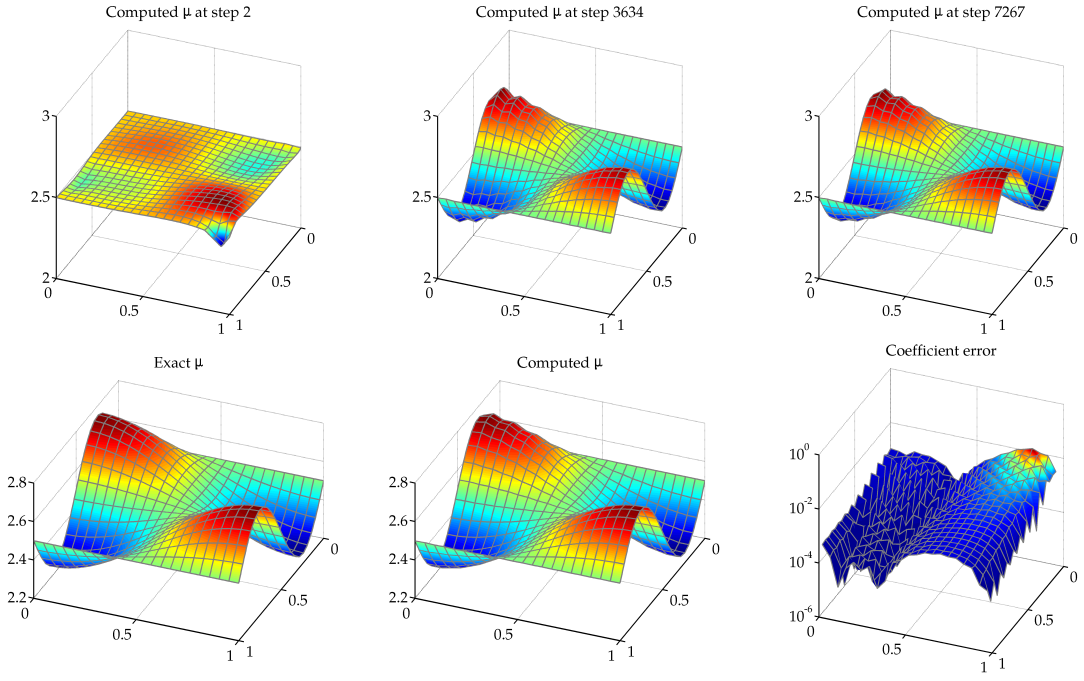


Figure 24: Marcotte – First Variant (EE)

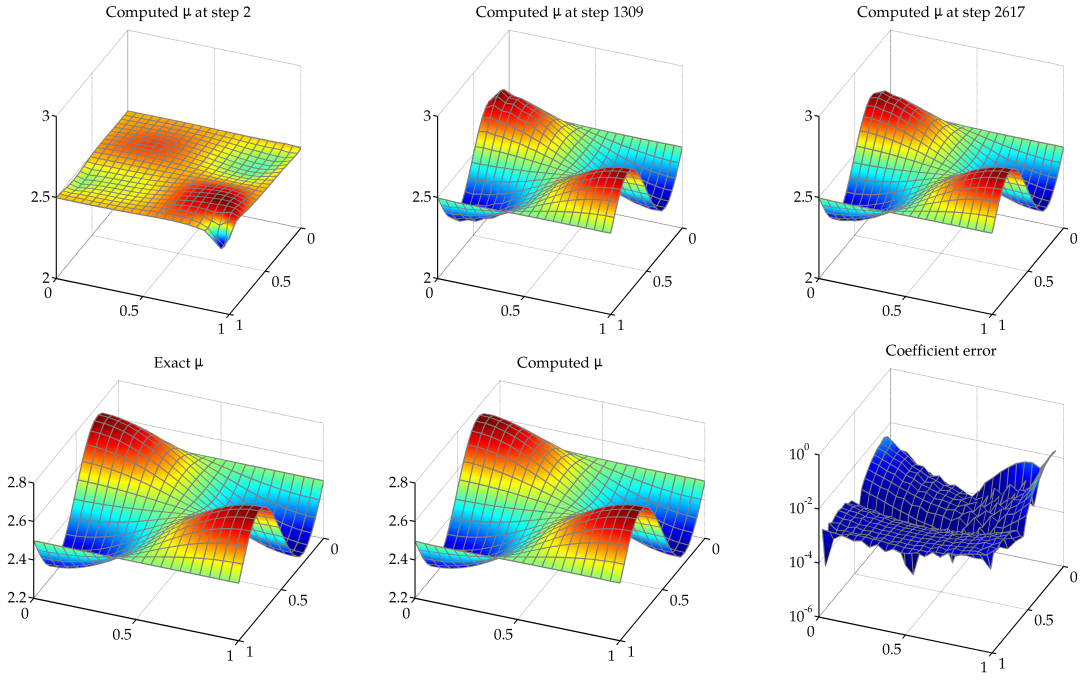


Figure 25: Marcotte – Second Variant (EE)

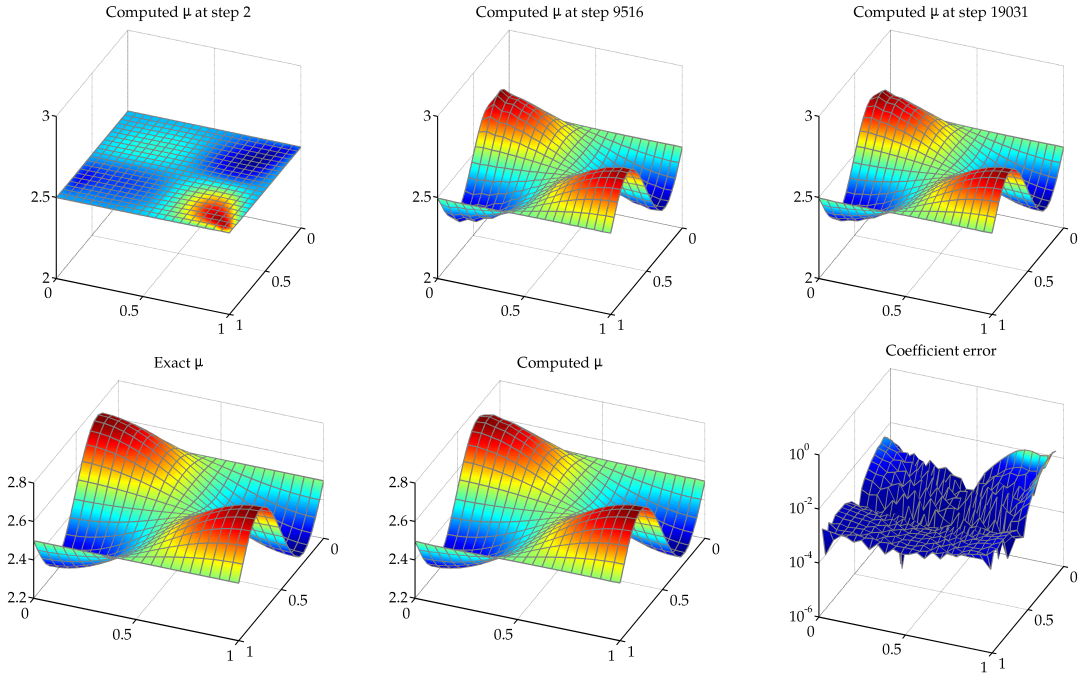


Figure 26: Korpelevich Method (EE)

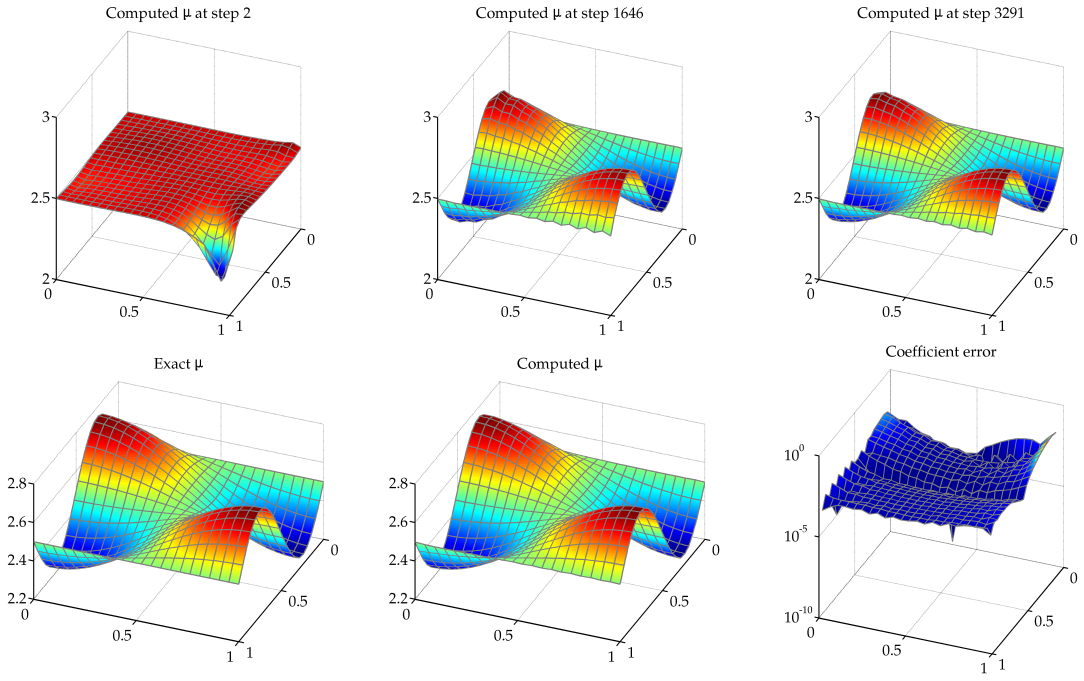


Figure 27: Khobotov Method (EE)

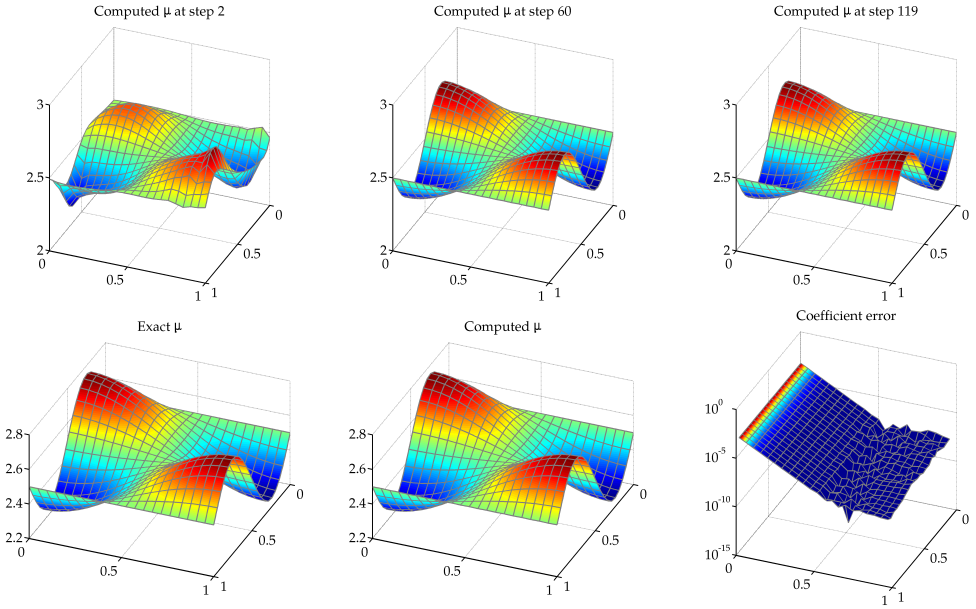


Figure 28: Hager-Zhang Method (EE)

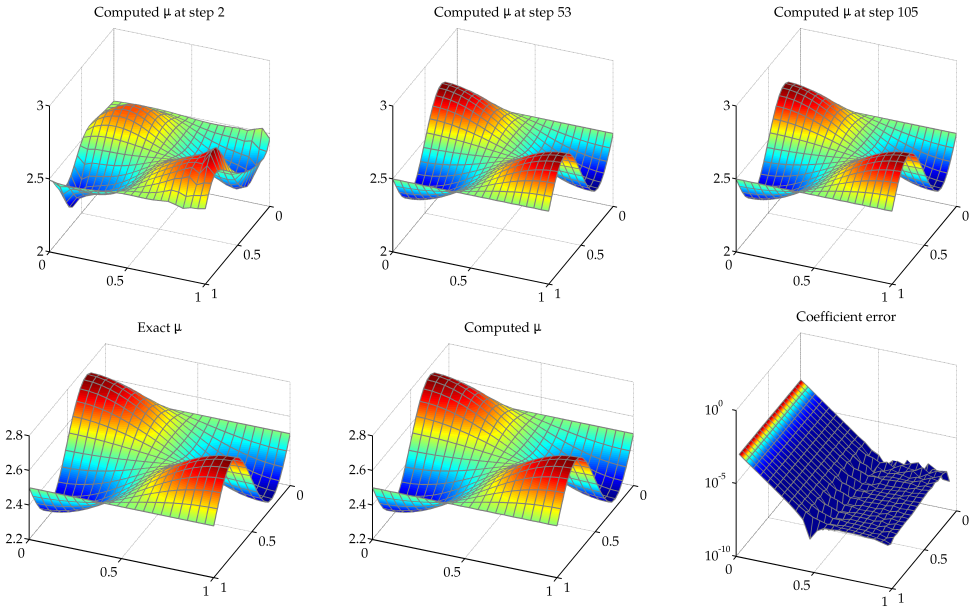


Figure 29: Hager-Zhang Method using Bregman Functions (EE)

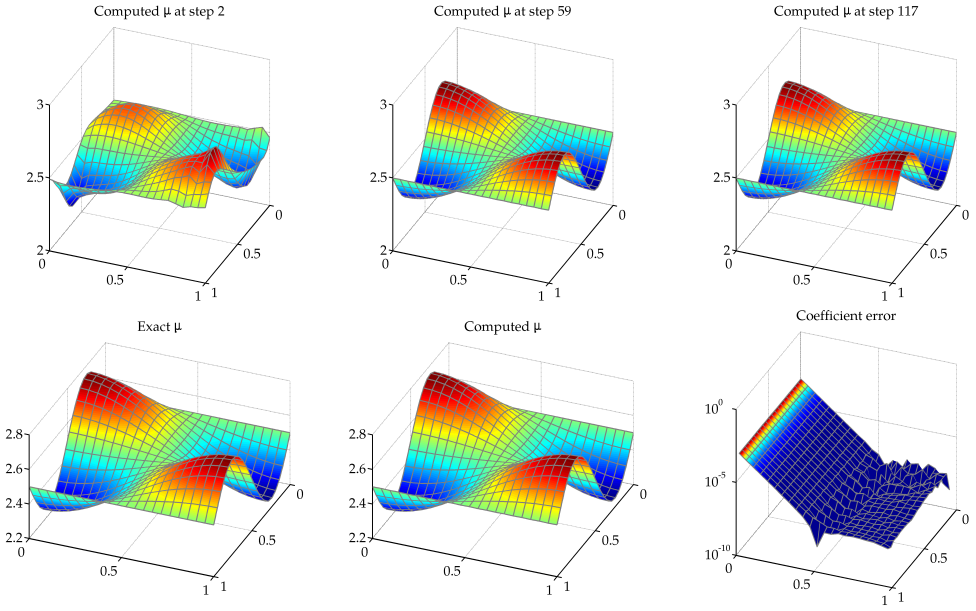


Figure 30: Hager-Zhang Method using φ-Divergence (EE)

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Korpelevich	$1.62 \cdot 10^5$	$1.62 \cdot 10^5$	–	80,751	$3.203 \cdot 10^{-3}$
Khobotov	12,836	12,836	–	6,415	$1.056 \cdot 10^{-3}$
Marcotte 1	7,972	7,972	–	3,982	$2.519 \cdot 10^{-3}$
Marcotte2	5,494	5,494	–	2,058	$2.523 \cdot 10^{-3}$
He-Goldstein	$1 \cdot 10^5$	$1 \cdot 10^5$	–	100,000	$9.332 \cdot 10^{-4}$
Solodov-Tseng (I)	$1.4 \cdot 10^5$	$1.4 \cdot 10^5$	–	46,767	$1.350 \cdot 10^{-3}$
Solodov-Tseng (H)	5,399	5,399	1	1,795	$2.057 \cdot 10^{-3}$
Hyperplane	15,813	15,813	–	5,266	$1.191 \cdot 10^{-3}$
Two-step	15,711	15,711	–	5,177	$1.190 \cdot 10^{-3}$
Projected Gradient	20,000	20,000	–	10,000	$9.532 \cdot 10^{-4}$
Scaled Projected Gradient	4,491	4,491	–	2,118	$9.090 \cdot 10^{-4}$
FISTA	30,864	30,864	–	1,000	$9.110 \cdot 10^{-4}$

Table 3: Comparison of Gradient and Extragradient Methods (MOLS)

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Hager-Zhang	$1.73 \cdot 10^5$	$1.66 \cdot 10^5$	–	7,650	$1.290 \cdot 10^{-3}$
φ-divergence	$1.62 \cdot 10^5$	$1.55 \cdot 10^5$	–	7,140	$1.300 \cdot 10^{-3}$
Bregman	$1.29 \cdot 10^5$	$1.24 \cdot 10^5$	–	5,712	$1.510 \cdot 10^{-3}$
Quadratic φ	276	274	274	550	$2.730 \cdot 10^{-7}$

Table 4: Comparison of Proximal Point Methods (MOLS)

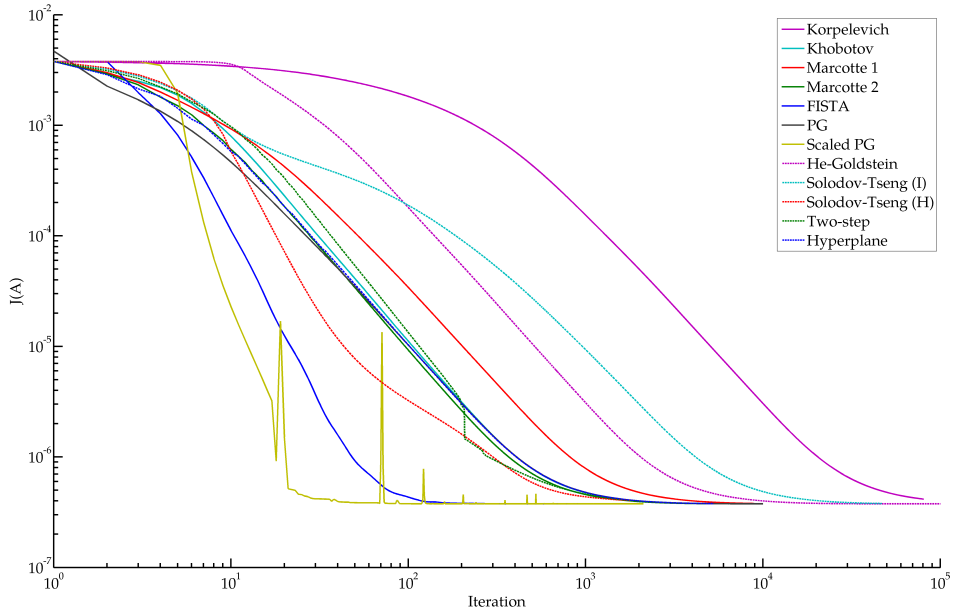


Figure 32: Convergence of $J(\mu)$ using EGM (MOLS)

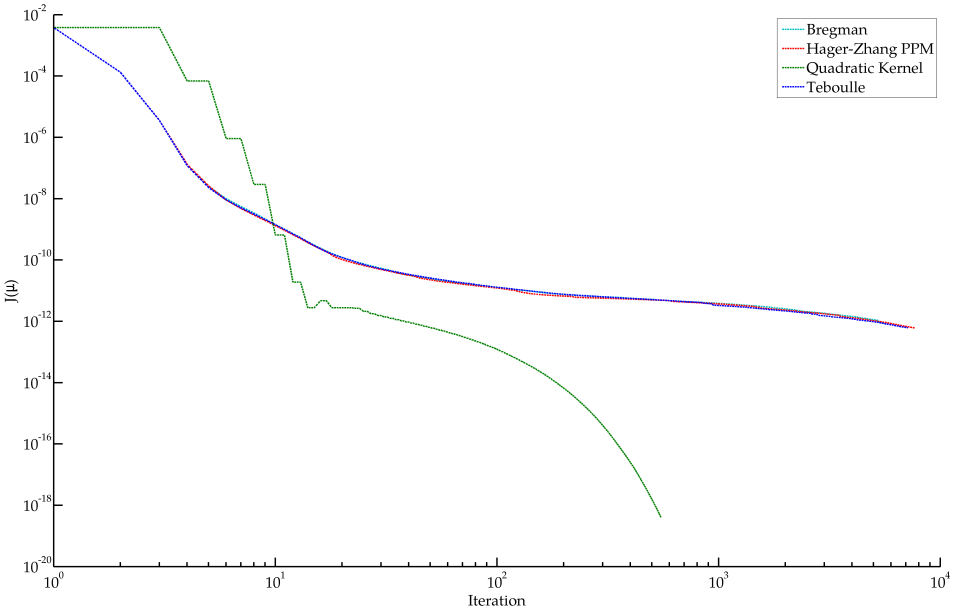


Figure 33: Convergence $J(\mu)$ using PPM (MOLS)

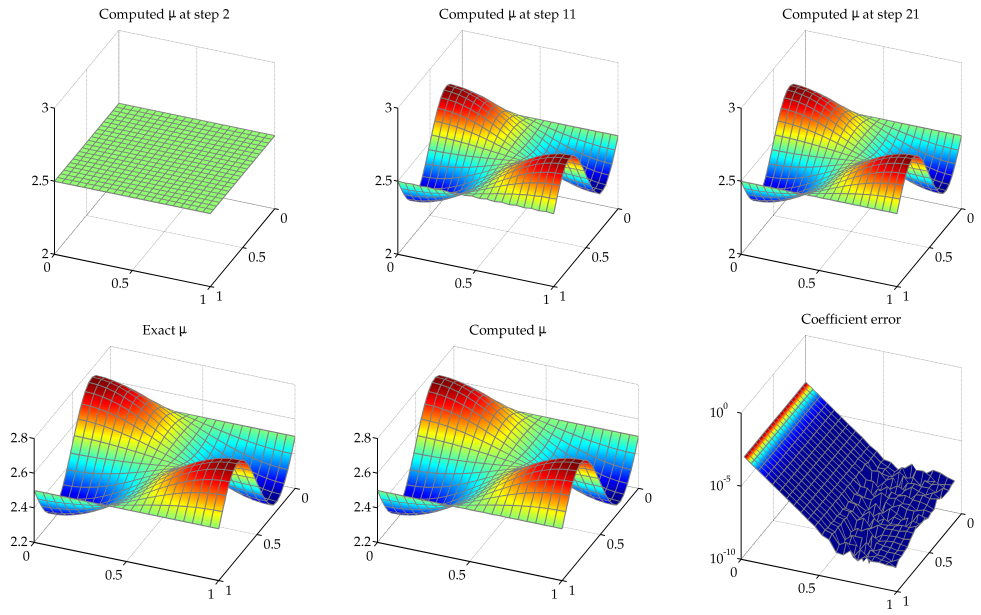


Figure 31: Hager-Zhang using Quadratic φ-Divergence (EE)

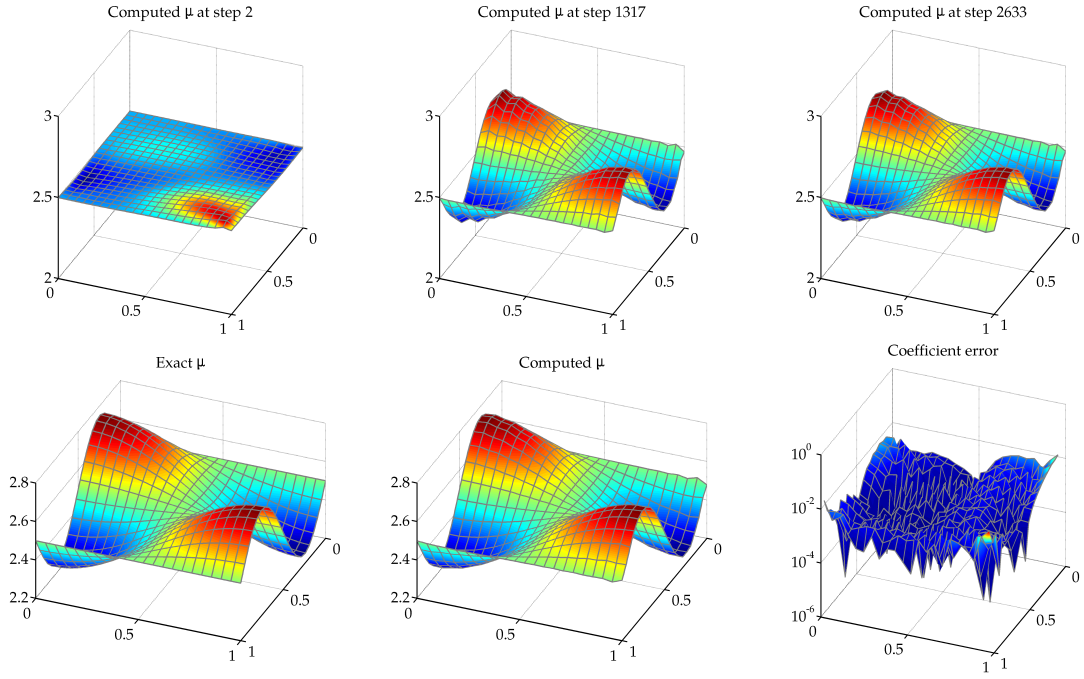


Figure 34: Hyperplane Method (MOLS)

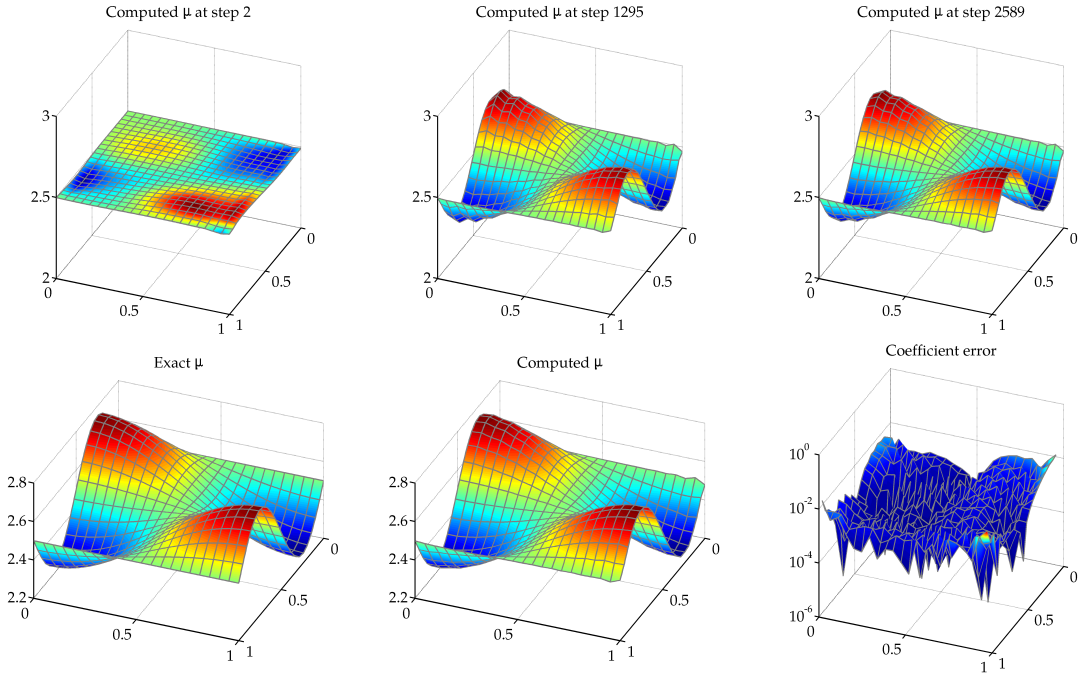


Figure 35: Two-step Method (MOLS)

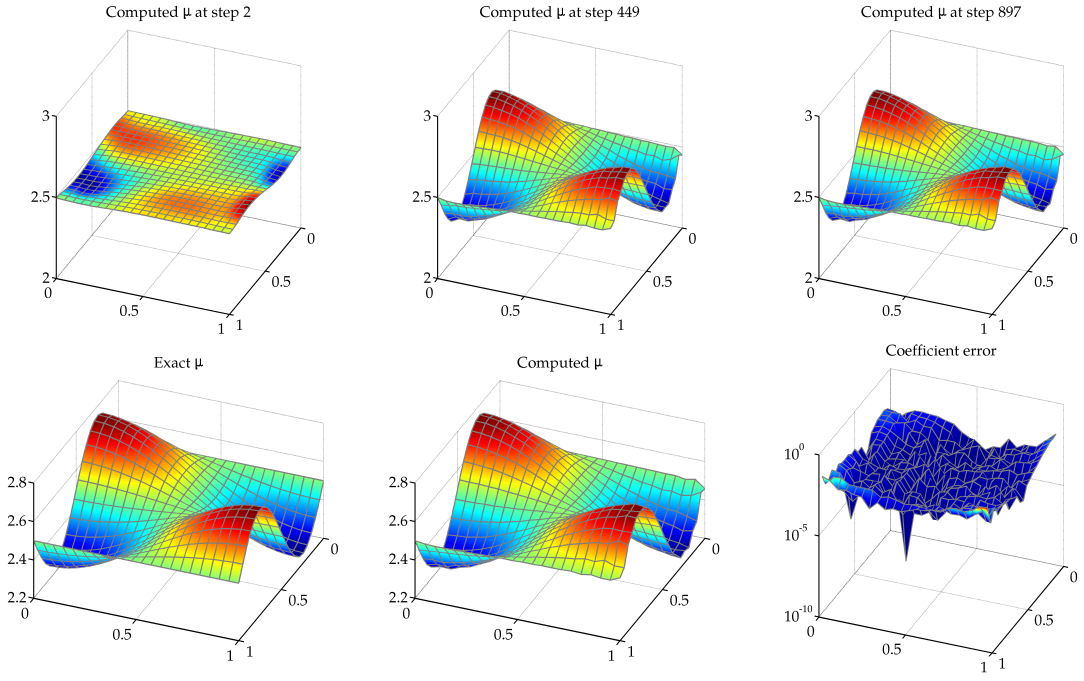


Figure 36: Solodov-Tseng Method using Hessian (MOLS)

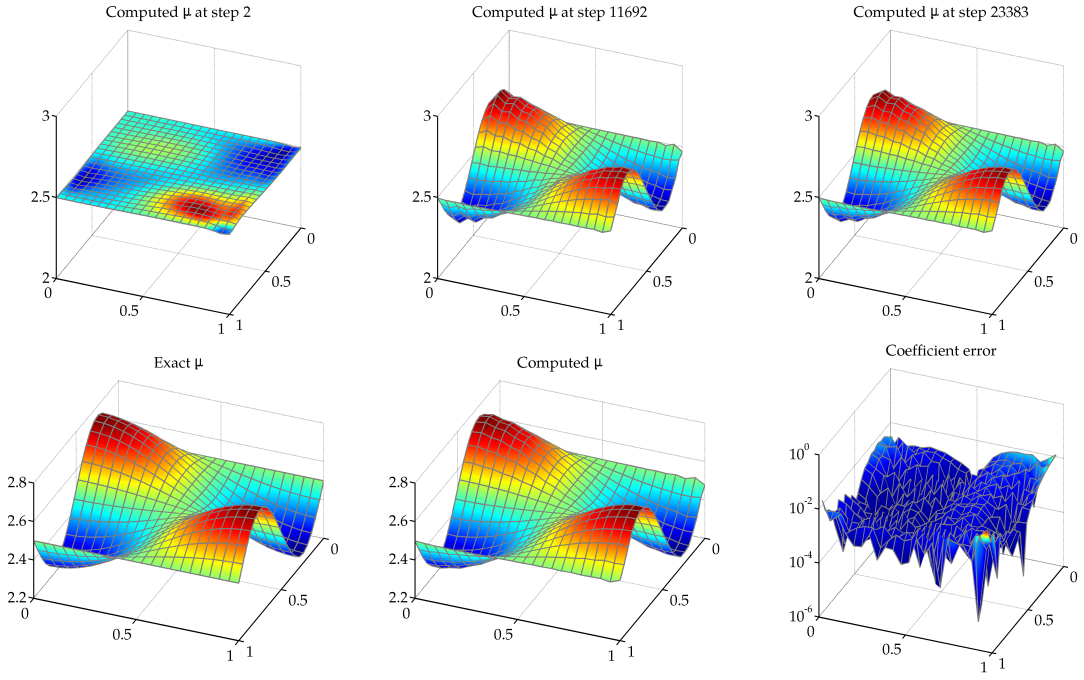


Figure 37: Solodov-Tseng Method using Identity (MOLS)

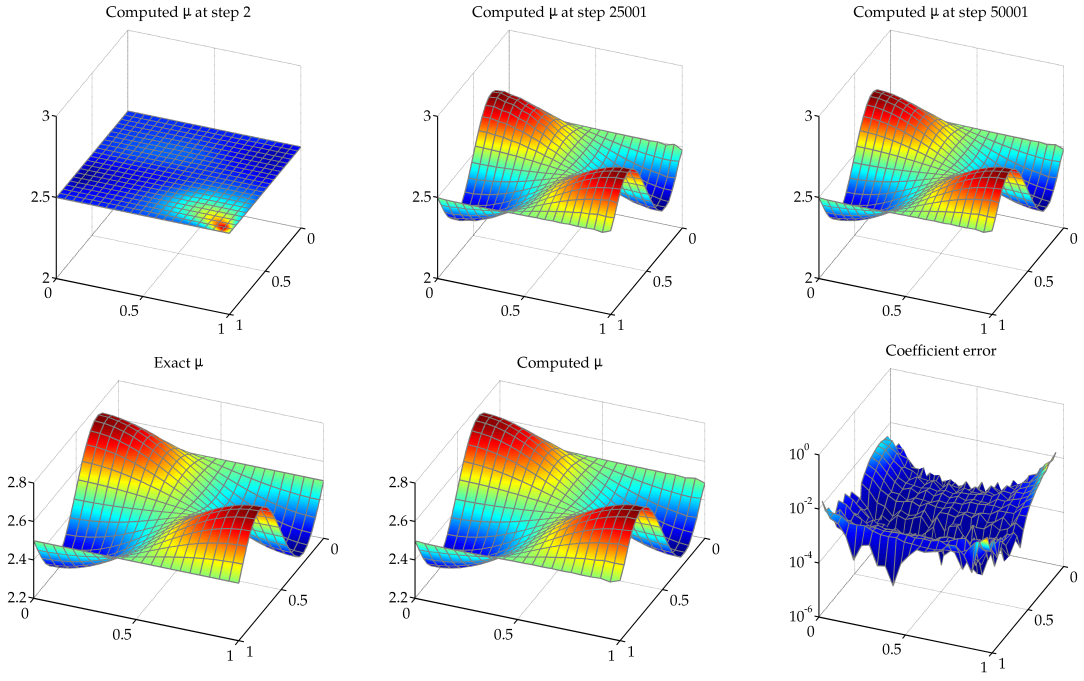


Figure 38: He-Goldstein Method (MOLS)

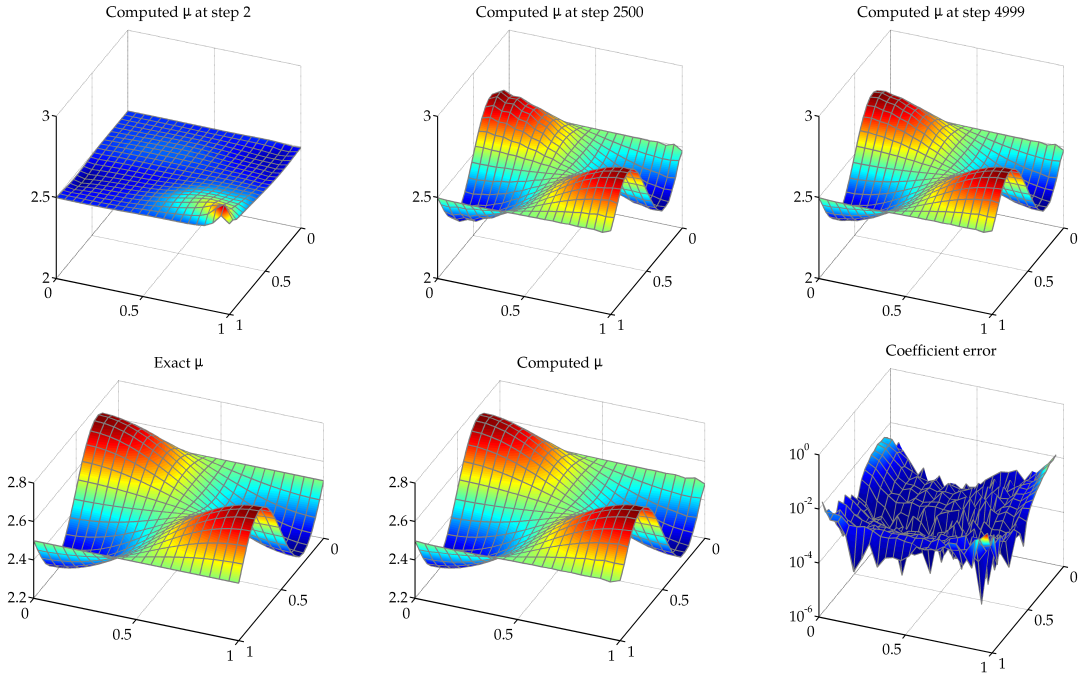


Figure 39: Projected Gradient with Hessian (MOLS)

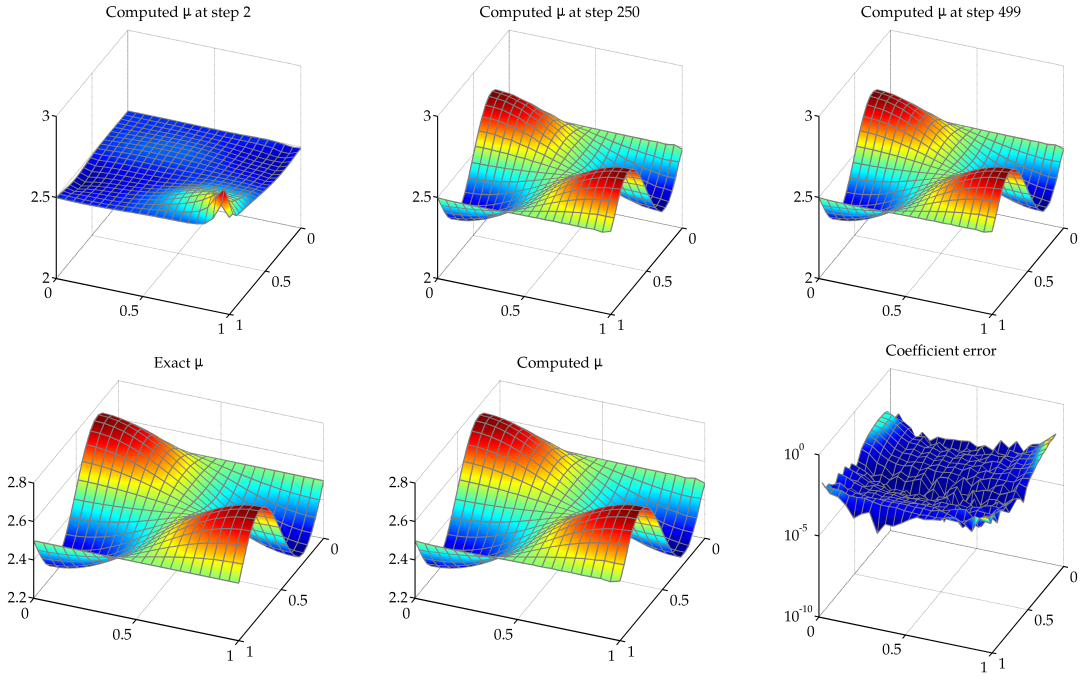


Figure 40: Fast Iterative Shrinkage-Thresholding (FISTA) (MOLS)

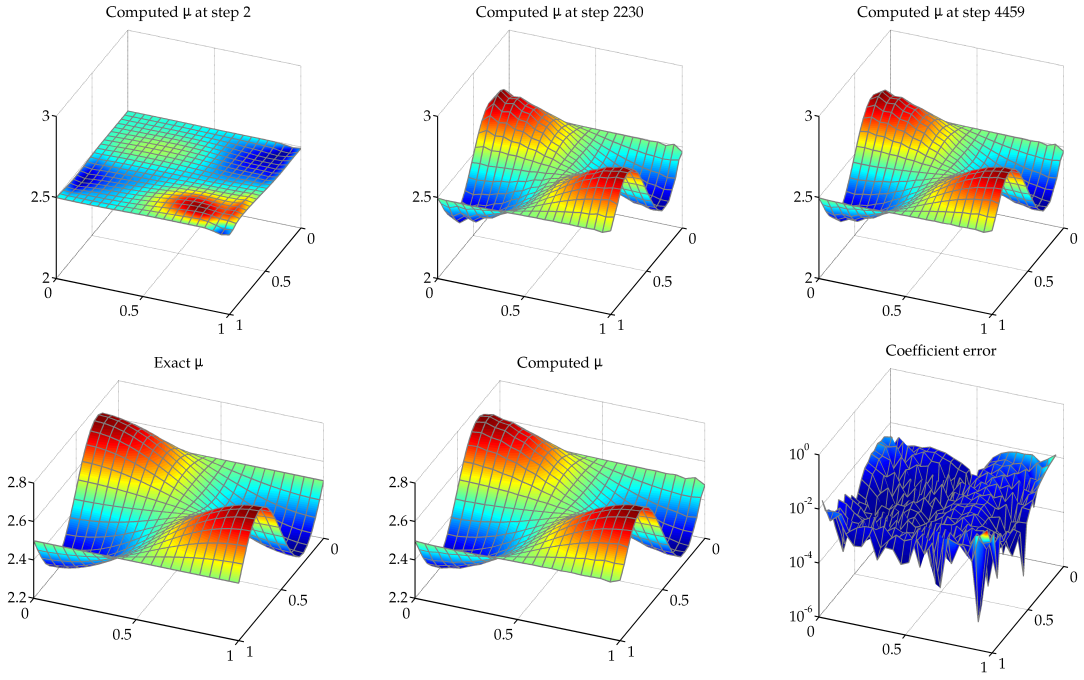


Figure 41: Marcotte – First Variant (MOLS)

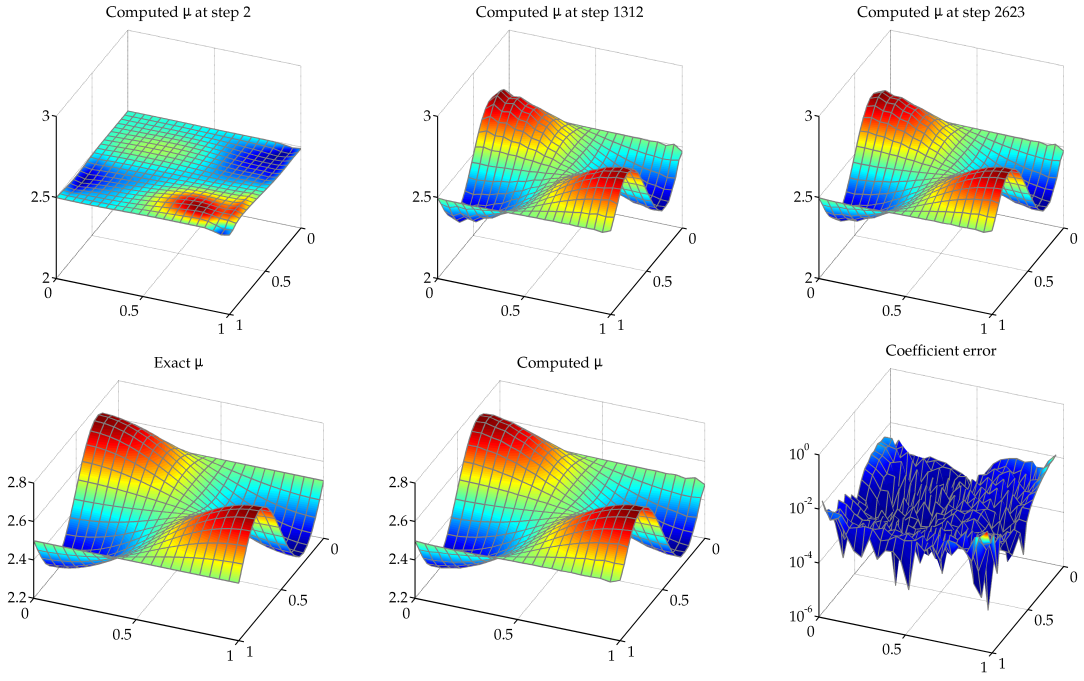


Figure 42: Marcotte – Second Variant (MOLS)

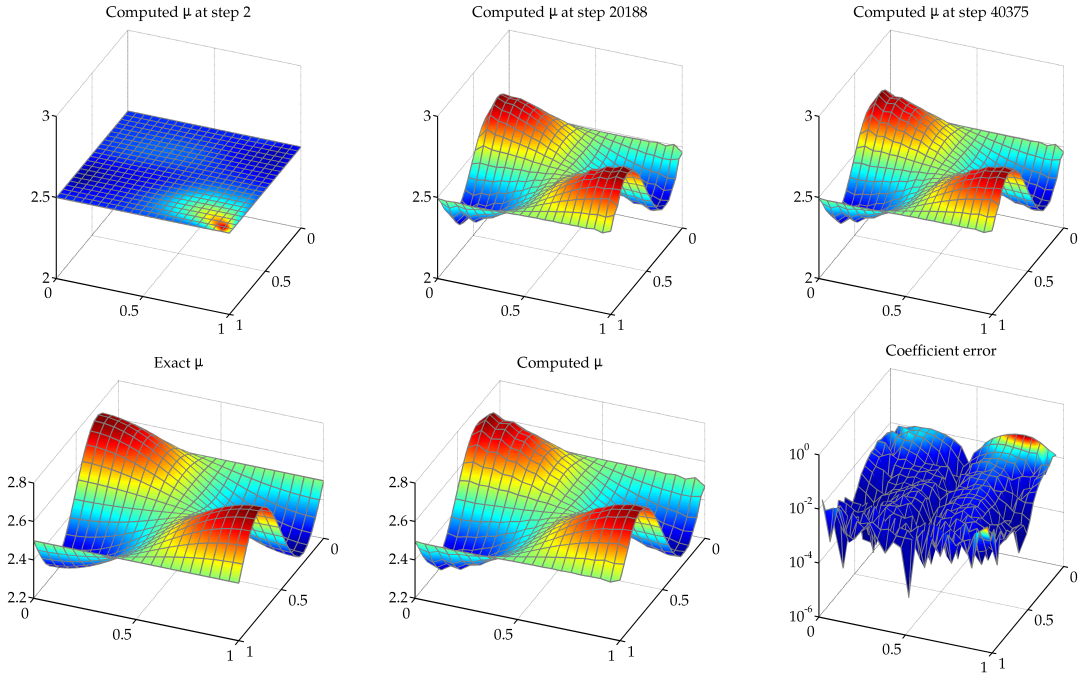


Figure 43: Korpelevich Method (MOLS)

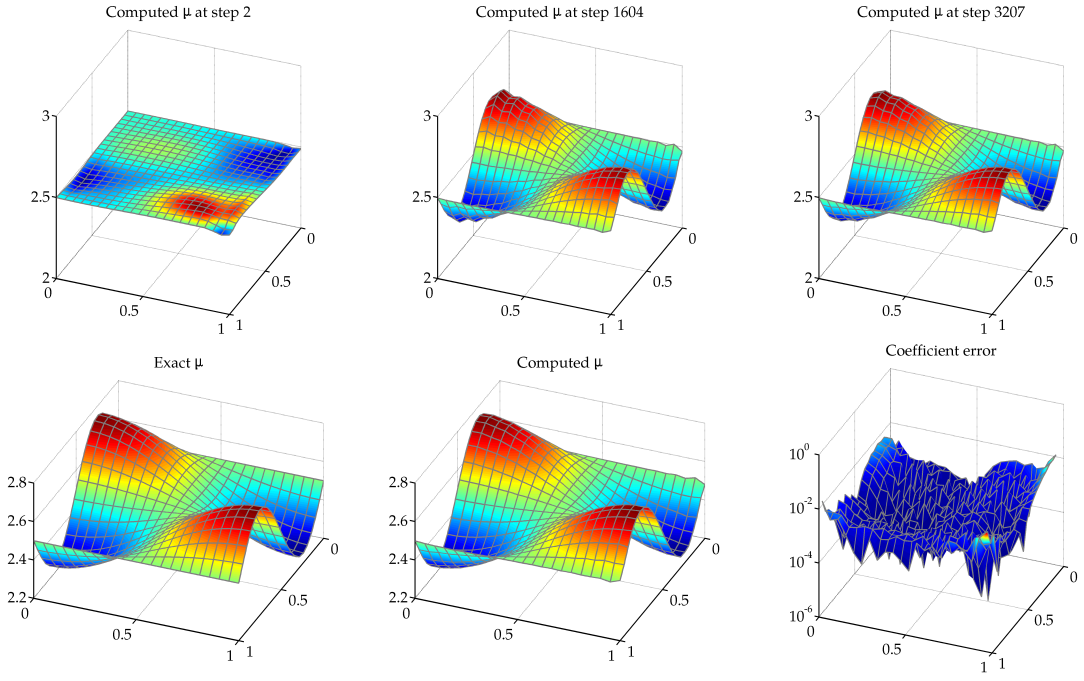


Figure 44: Khobotov Method (MOLS)

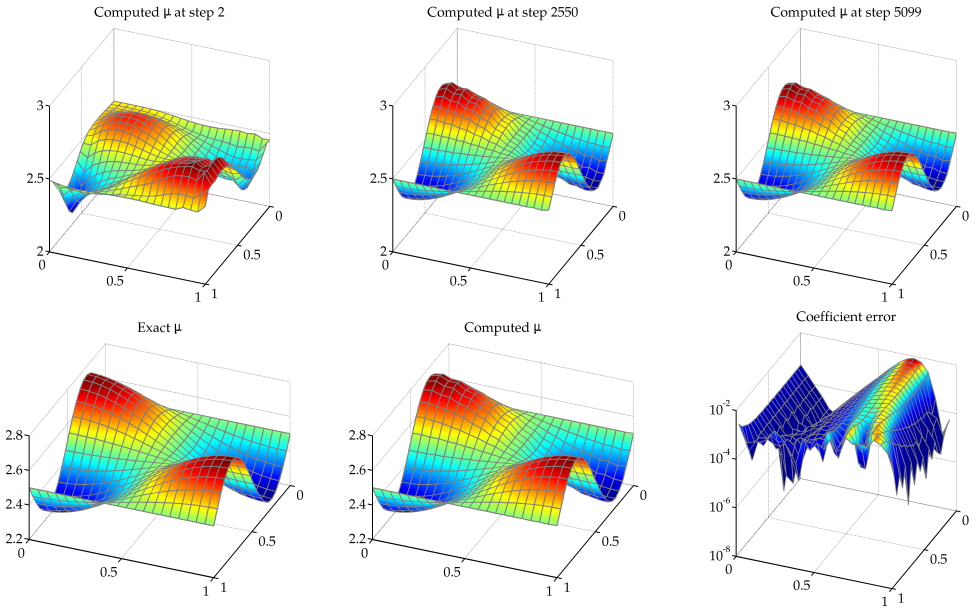


Figure 45: Hager-Zhang Method (MOLS)

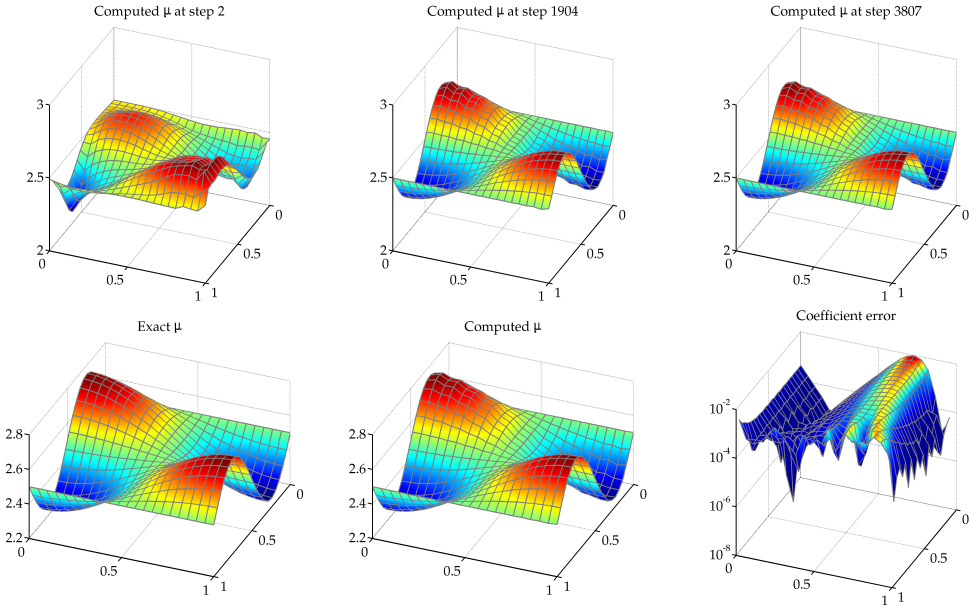


Figure 46: Hager-Zhang Method using Bregman Functions (MOLS)

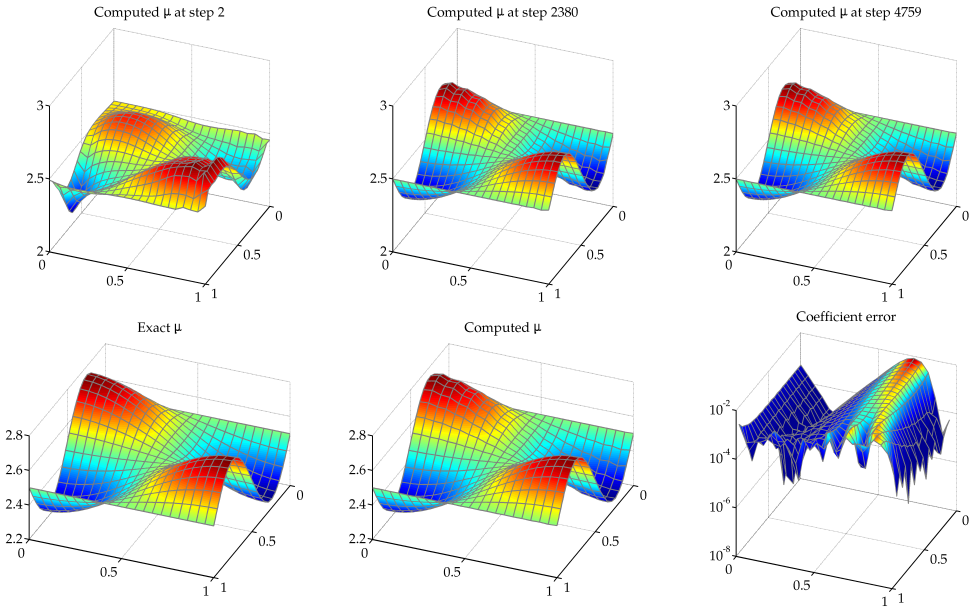


Figure 47: Hager-Zhang Method using φ-Divergence (MOLS)

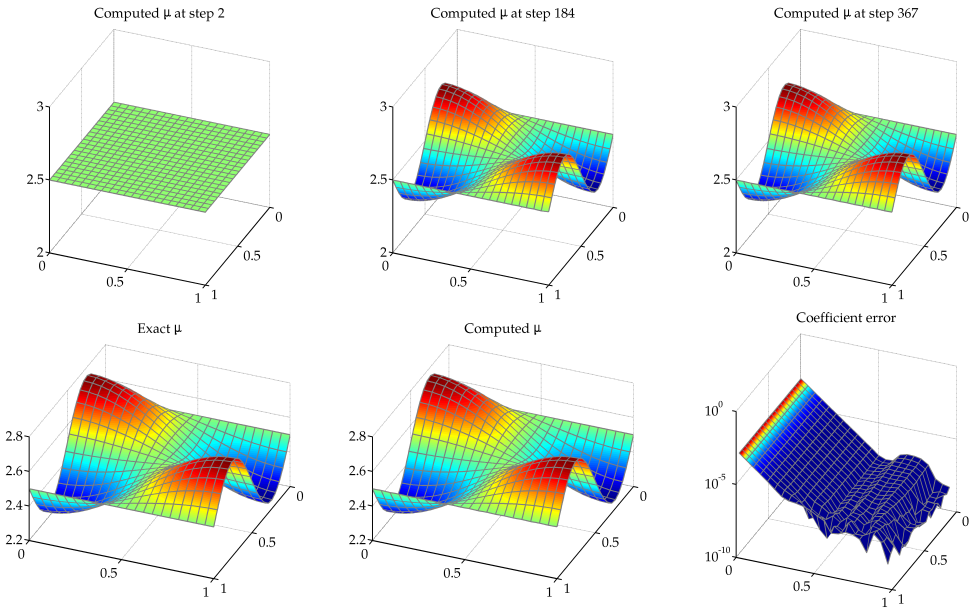


Figure 48: Hager-Zhang using Quadratic φ-Divergence (MOLS)

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Korpelevich	$1.62 \cdot 10^5$	$1.62 \cdot 10^5$	–	80,751	$3.203 \cdot 10^{-3}$
Khobotov	12,836	12,836	–	6,415	$1.056 \cdot 10^{-3}$
Marcotte 1	7,972	7,972	–	3,982	$2.519 \cdot 10^{-3}$
Marcotte2	5,494	5,494	–	2,058	$2.523 \cdot 10^{-3}$
He-Goldstein	$1 \cdot 10^5$	$1 \cdot 10^5$	–	100,000	$9.332 \cdot 10^{-4}$
Solodov-Tseng (I)	$1.4 \cdot 10^5$	$1.4 \cdot 10^5$	–	46,767	$1.350 \cdot 10^{-3}$
Solodov-Tseng (H)	5,399	5,399	1	1,795	$2.057 \cdot 10^{-3}$
Hyperplane	15,813	15,813	–	5,266	$1.191 \cdot 10^{-3}$
Two-step	15,576	15,576	–	5,118	$1.190 \cdot 10^{-3}$
Projected Gradient	20,000	10,000	–	10,000	$9.532 \cdot 10^{-4}$
Scaled Projected Gradient	7,037	3,403	1	3,391	$9.090 \cdot 10^{-4}$
FISTA	30,893	1,005	–	1,000	$9.110 \cdot 10^{-4}$

Table 5: Comparison of Gradient and Extragradient Methods (EOLS)

Method	J evals	∇J evals	$\nabla^2 J$ evals	Iterations	L^2 error
Hager-Zhang	$1.62 \cdot 10^5$	$1.55 \cdot 10^5$	–	7,140	$1.210 \cdot 10^{-3}$
φ-divergence	$1.37 \cdot 10^5$	$1.31 \cdot 10^5$	–	6,018	$1.562 \cdot 10^{-3}$
Bregman	$1.04 \cdot 10^5$	$1.09 \cdot 10^5$	–	4,794	$1.630 \cdot 10^{-3}$
Quadratic φ	278	276	276	554	$2.703 \cdot 10^{-7}$

Table 6: Comparison of Proximal Point Methods (EOLS)

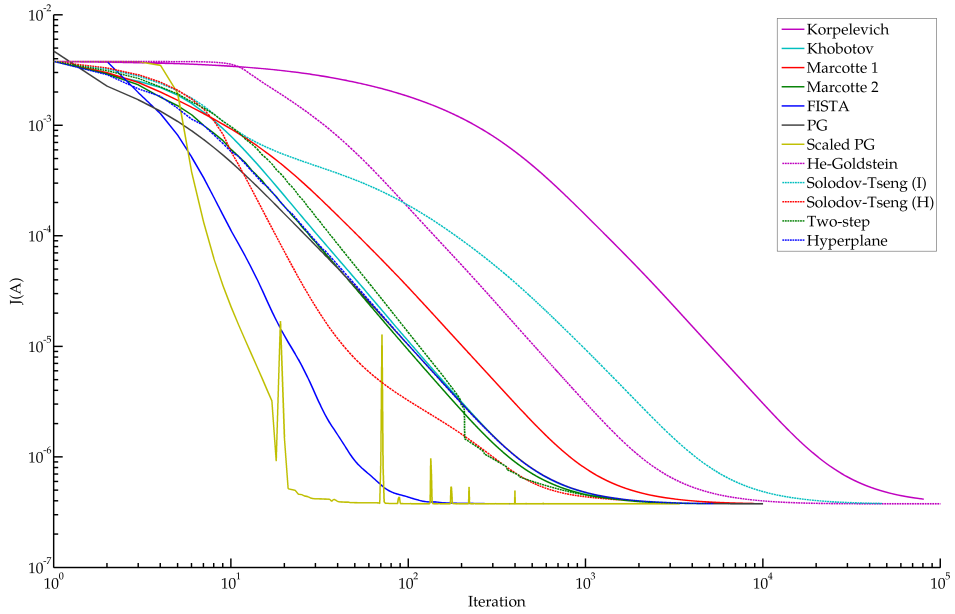


Figure 49: Convergence of $J(\mu)$ using EGM (EOLS)

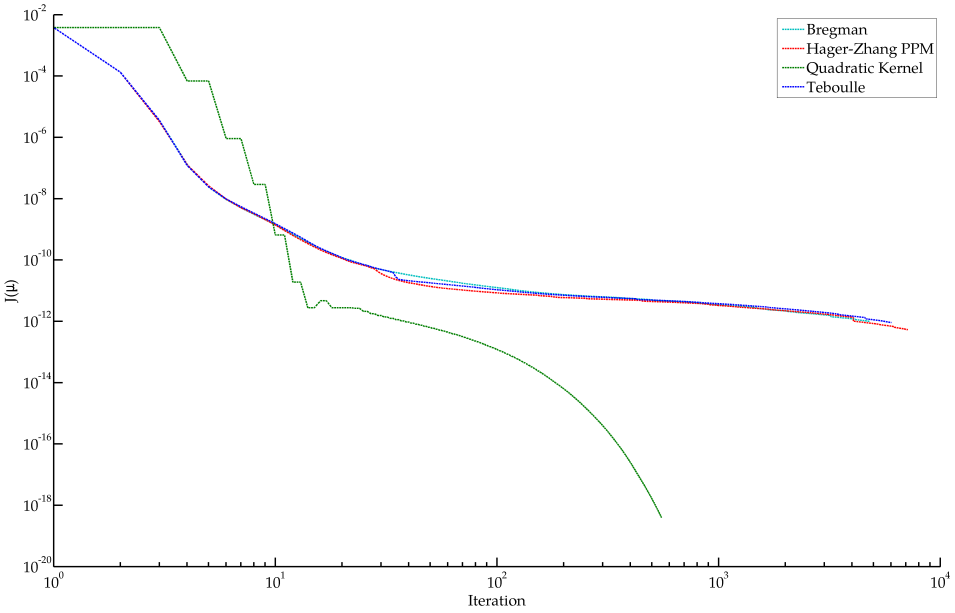


Figure 50: Convergence $J(\mu)$ using PPM (EOLS)

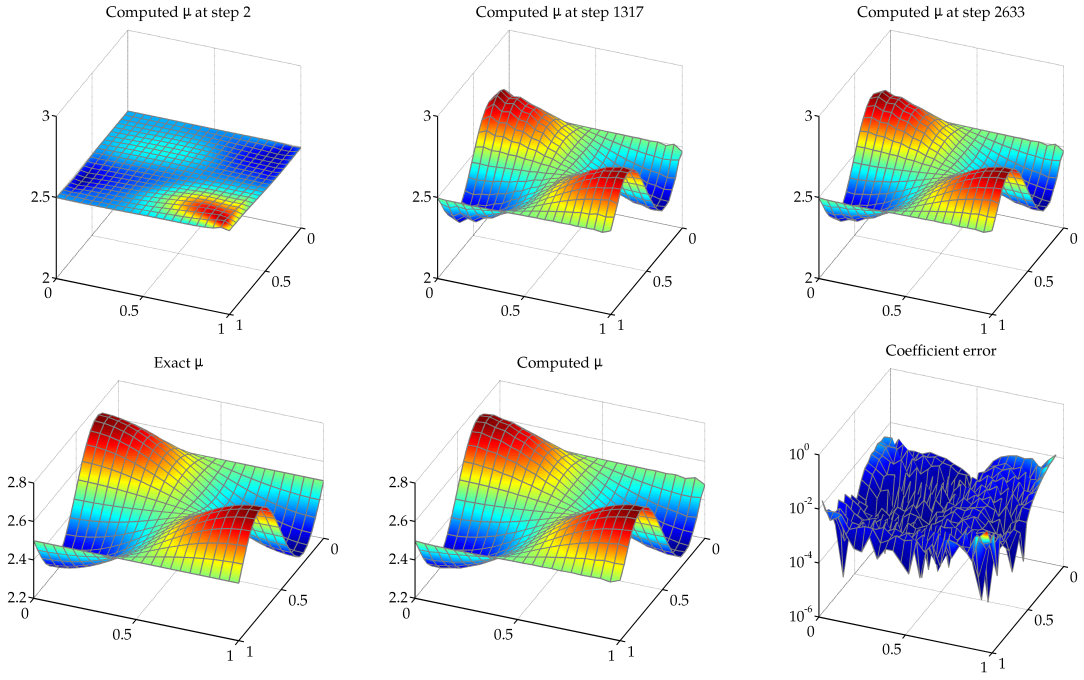


Figure 51: Hyperplane Method (EOLS)

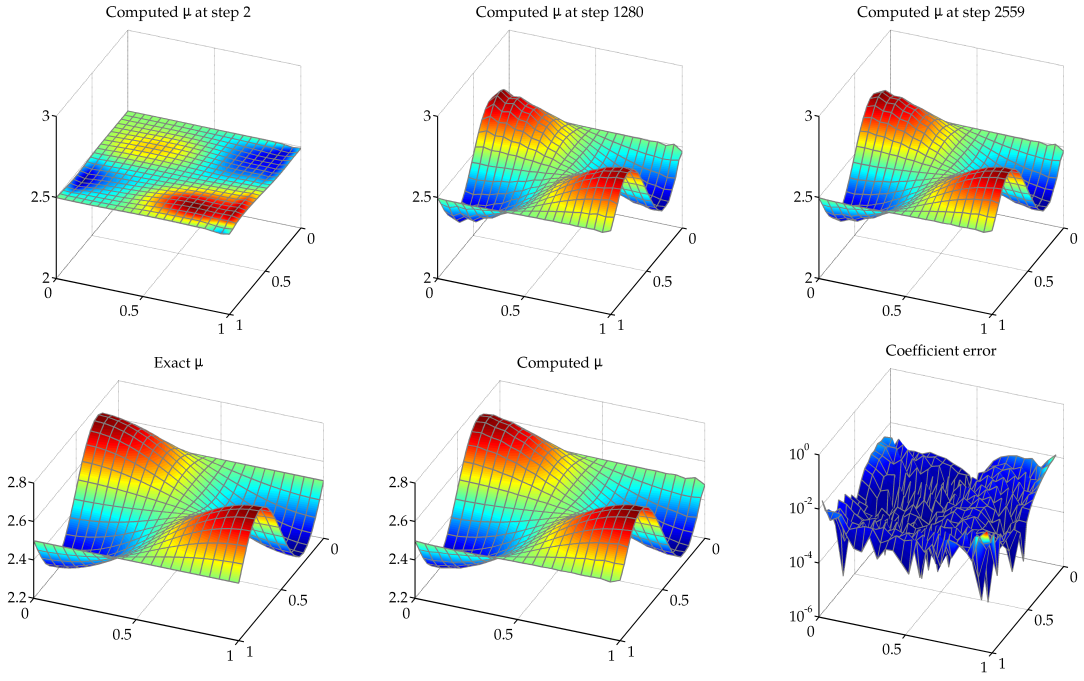


Figure 52: Two-step Method (EOLS)

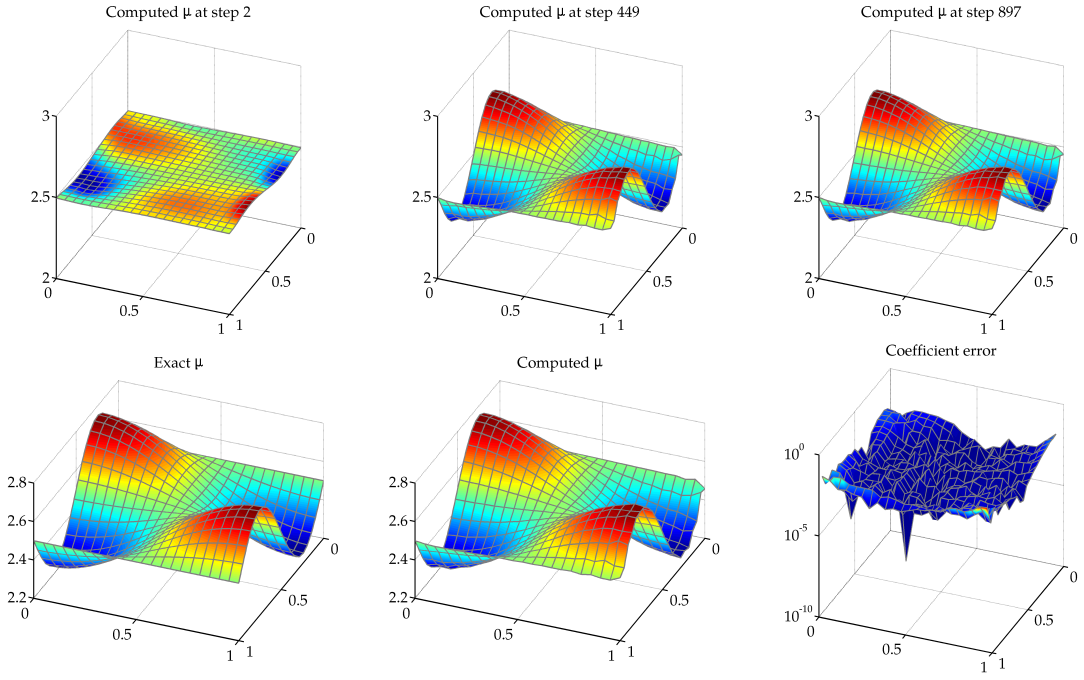


Figure 53: Solodov-Tseng Method using Hessian (EOLS)

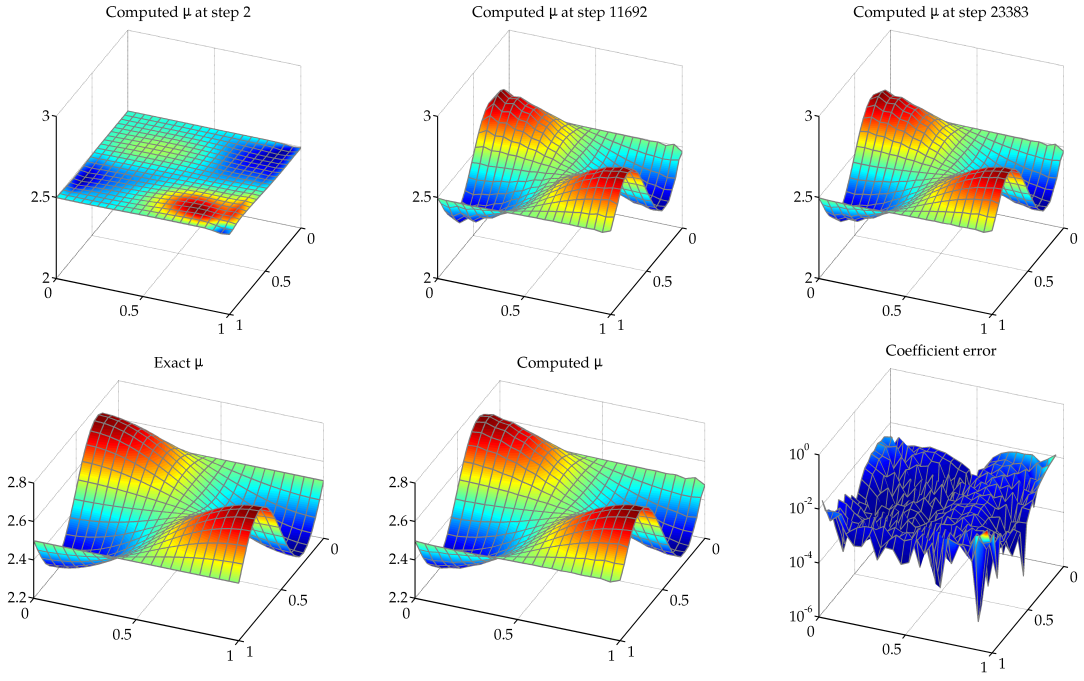


Figure 54: Solodov-Tseng Method using Identity (EOLS)

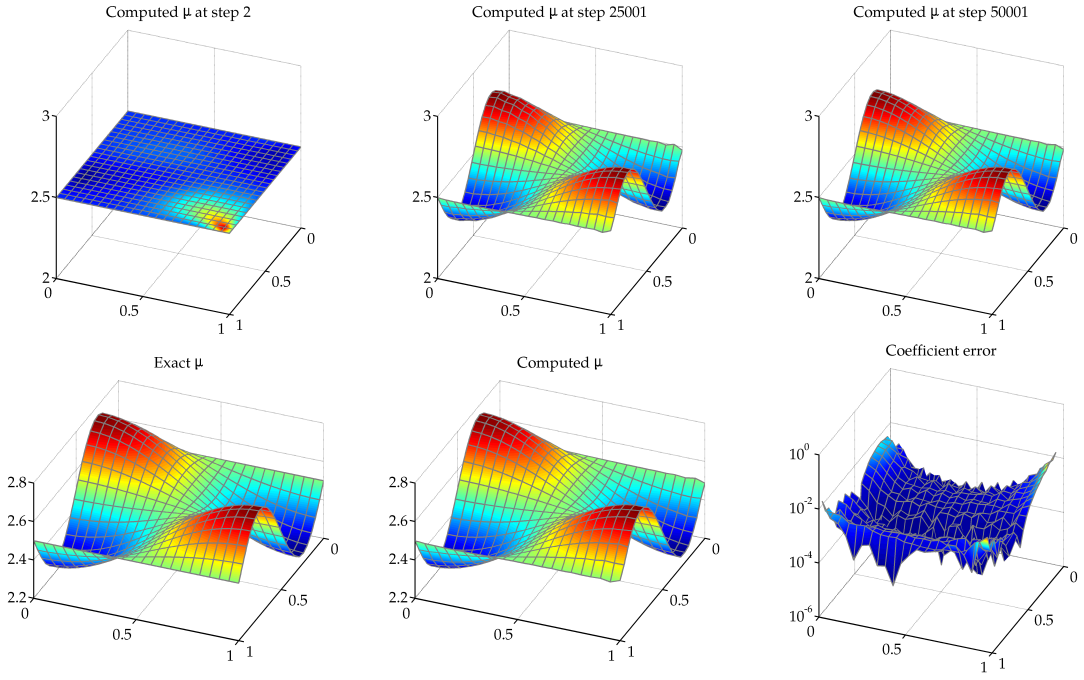


Figure 55: He-Goldstein Method (EOLS)

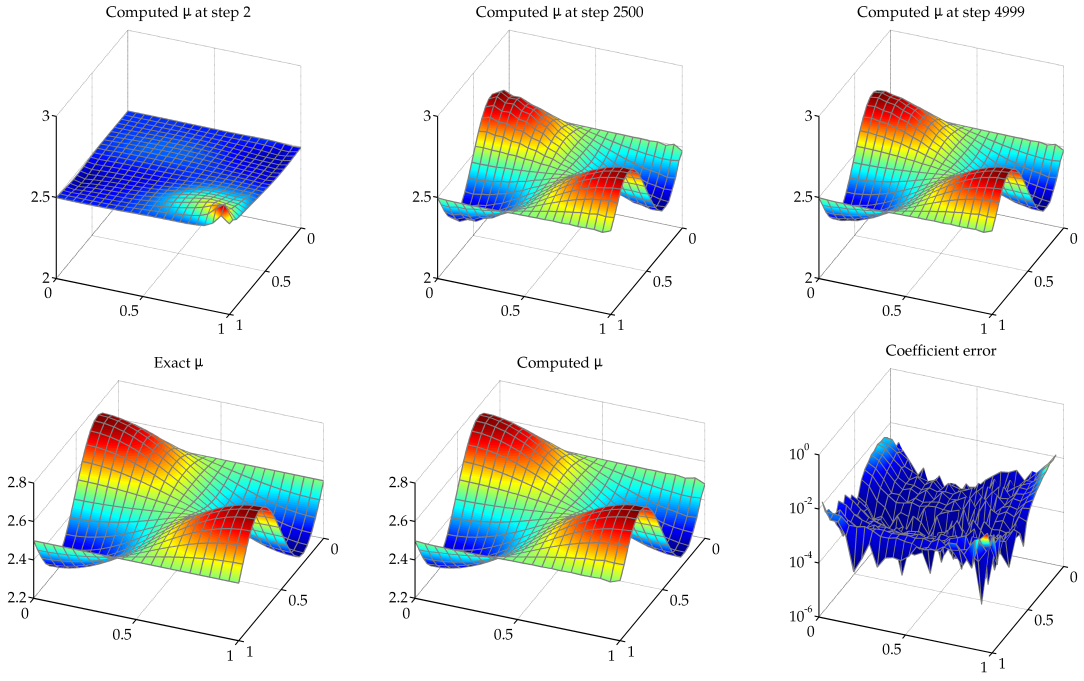


Figure 56: Projected Gradient with Hessian (EOLS)

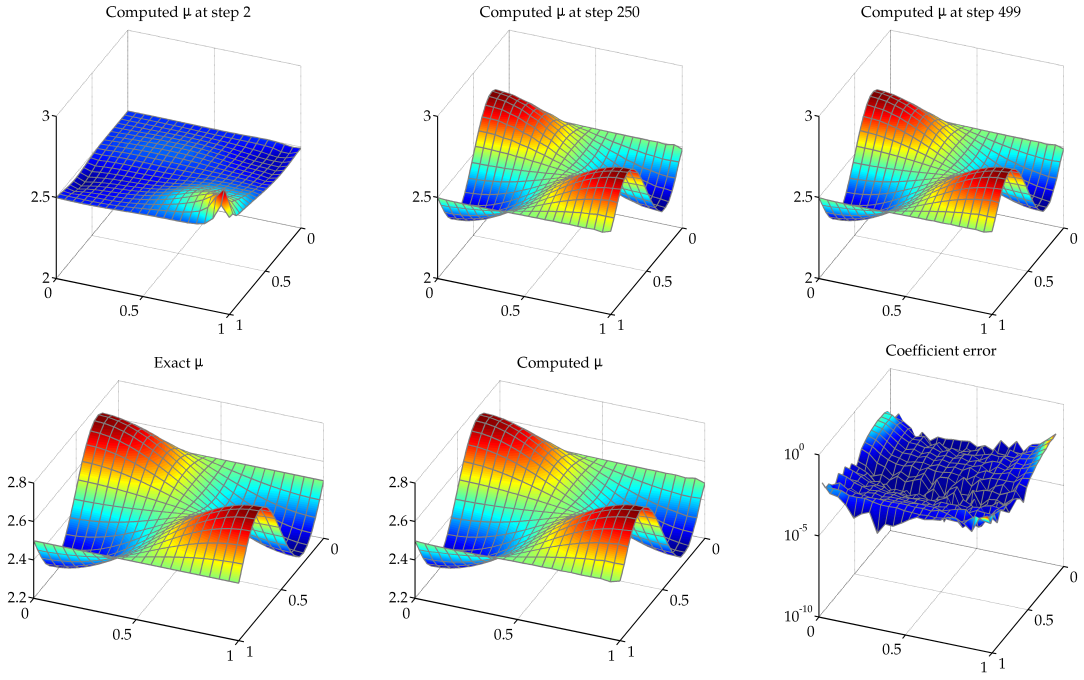


Figure 57: Fast Iterative Shrinkage-Thresholding (FISTA) (EOLS)

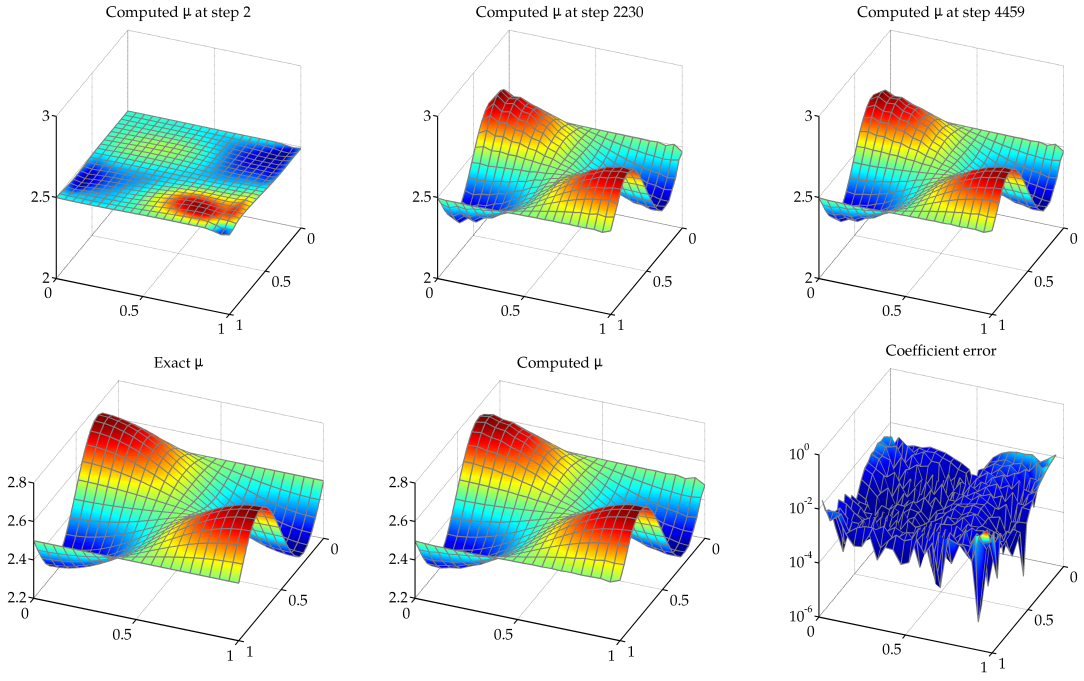


Figure 58: Marcotte – First Variant (EOLS)

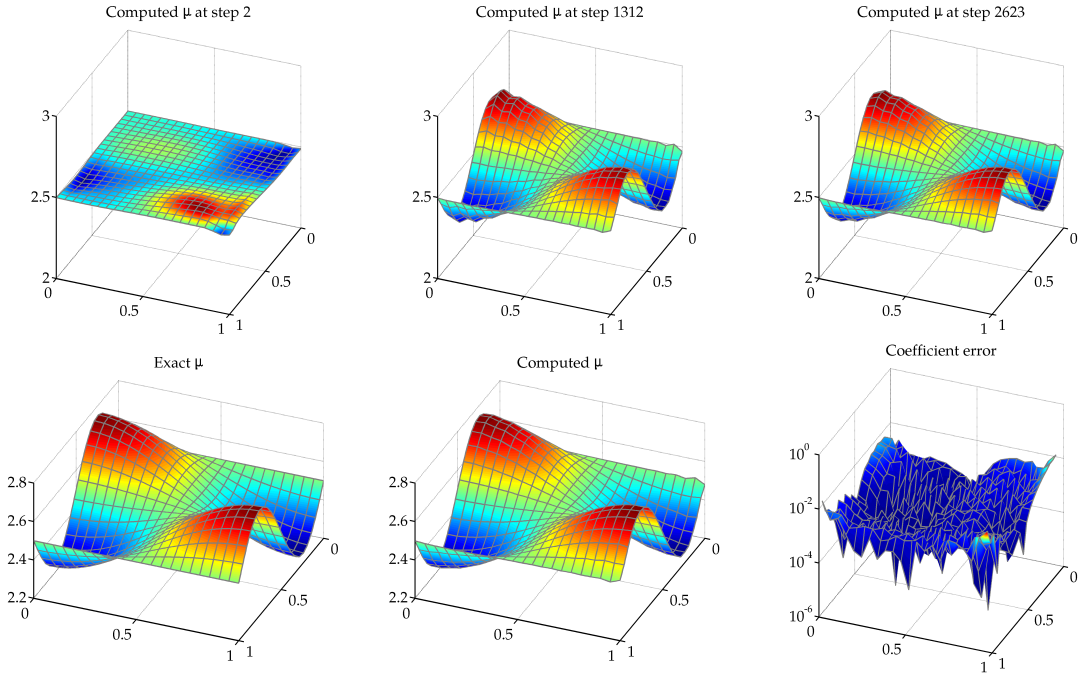


Figure 59: Marcotte – Second Variant (EOLS)

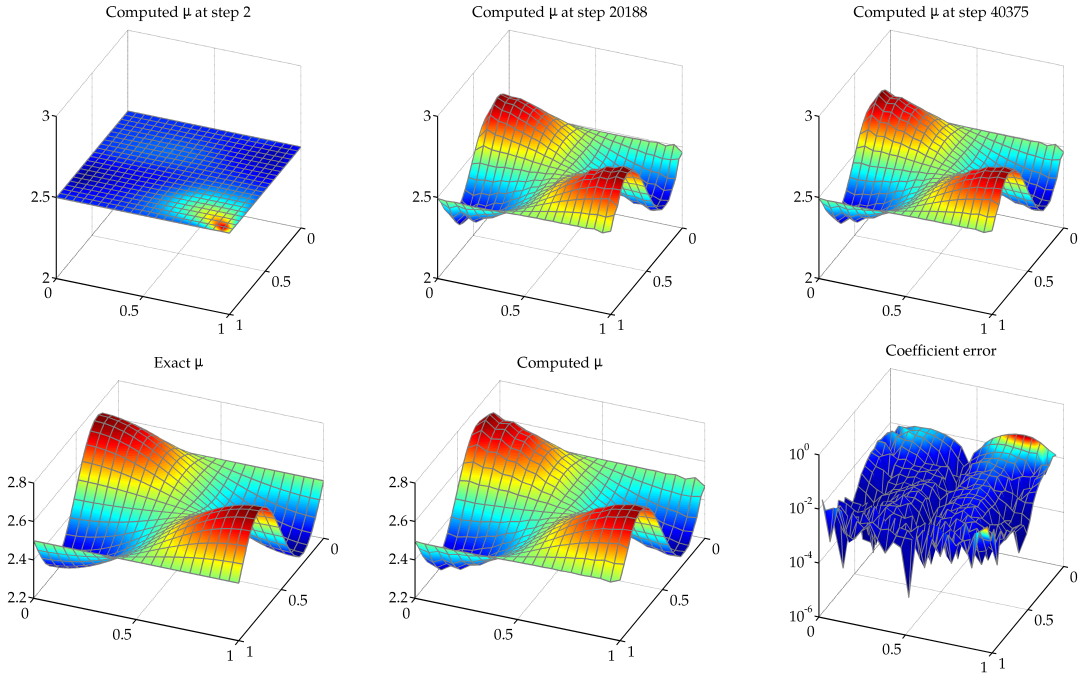


Figure 60: Korpelevich Method (EOLS)

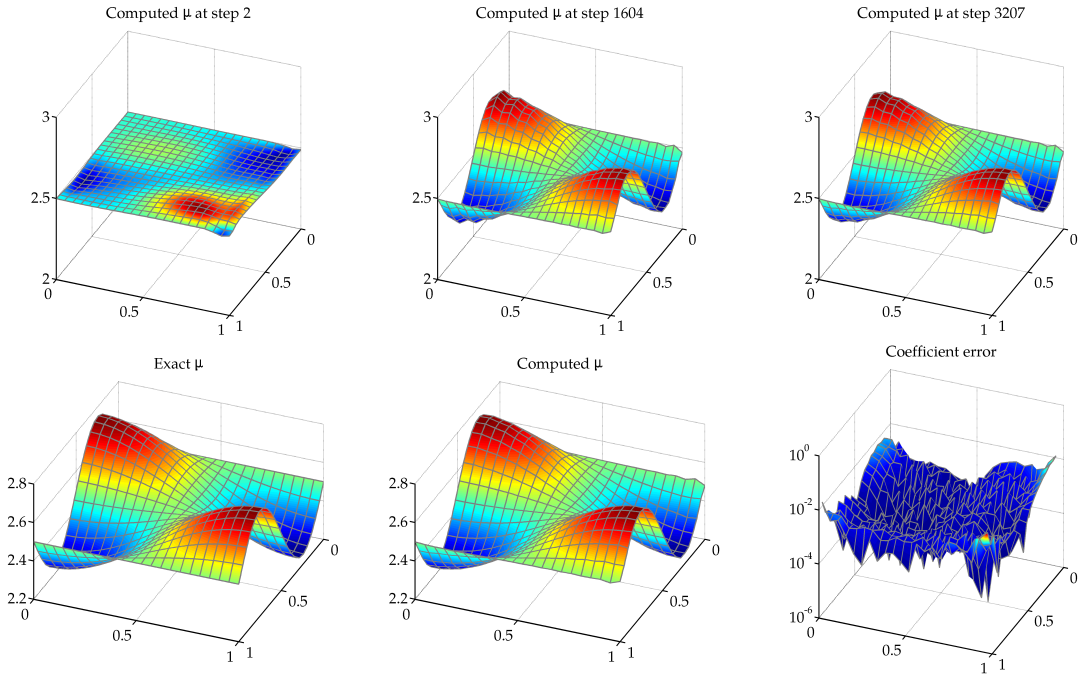


Figure 61: Khobotov Method (EOLS)

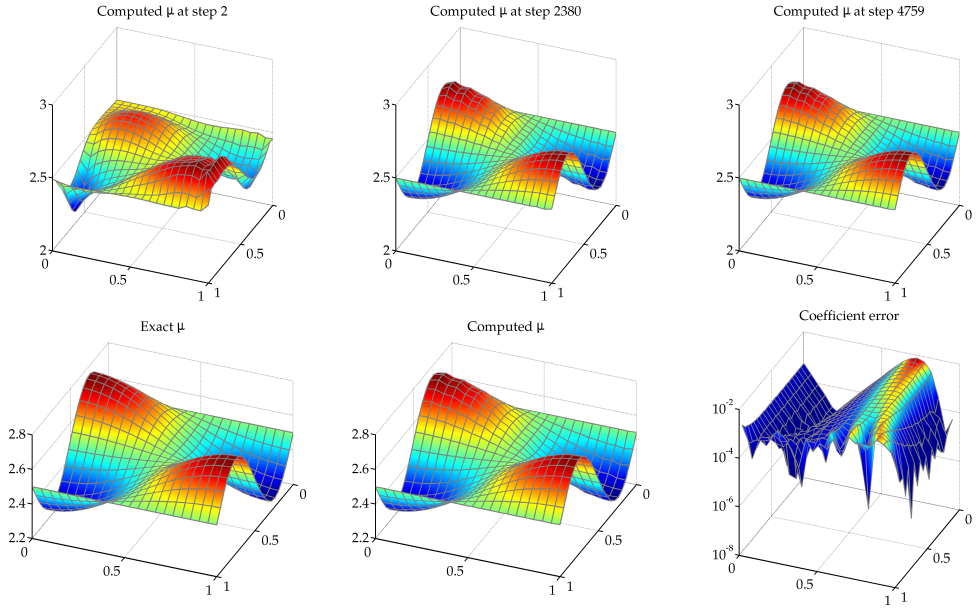


Figure 62: Hager-Zhang Method (EOLS)

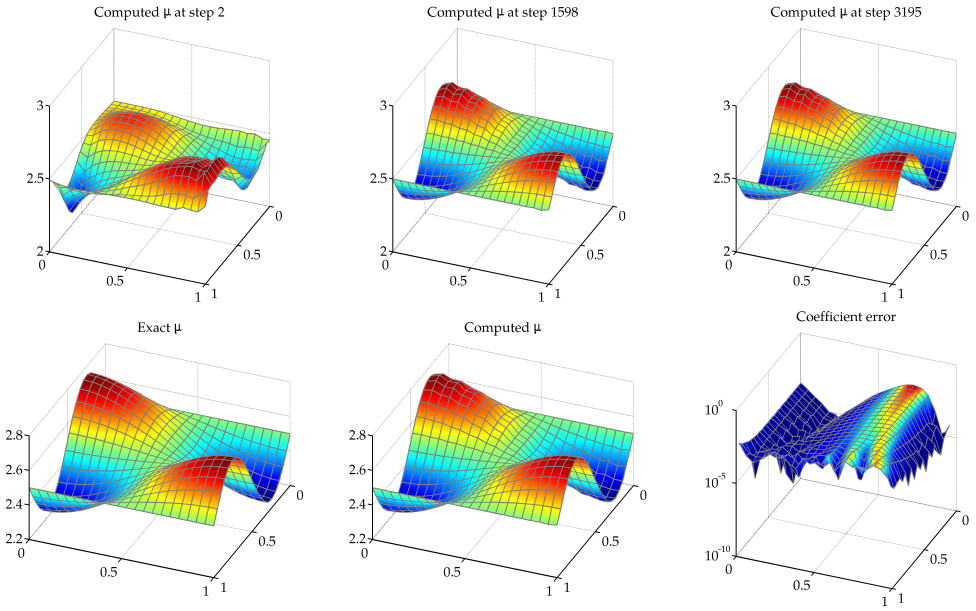


Figure 63: Hager-Zhang Method using Bregman Functions (EOLS)

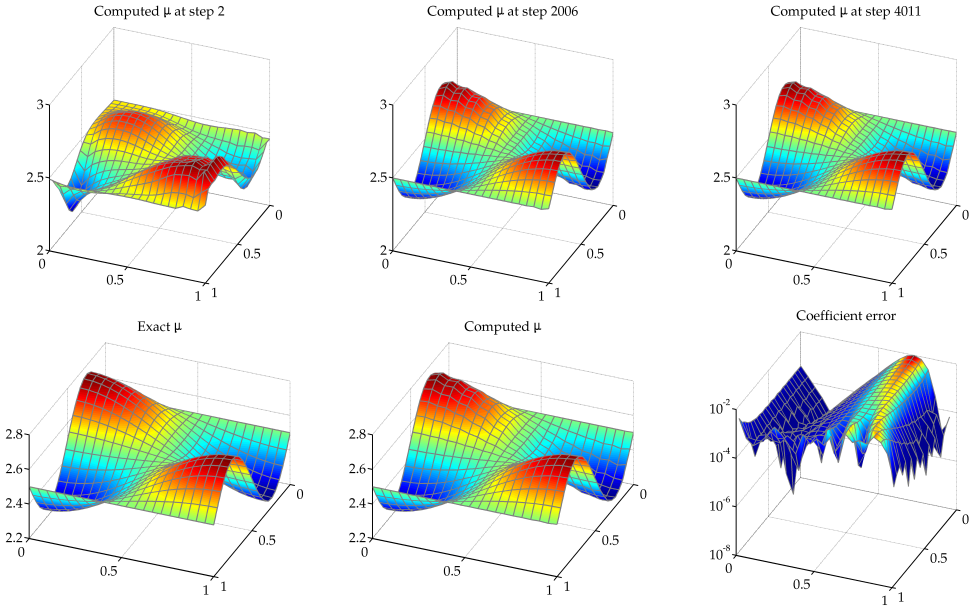


Figure 64: Hager-Zhang Method using φ-Divergence (EOLS)

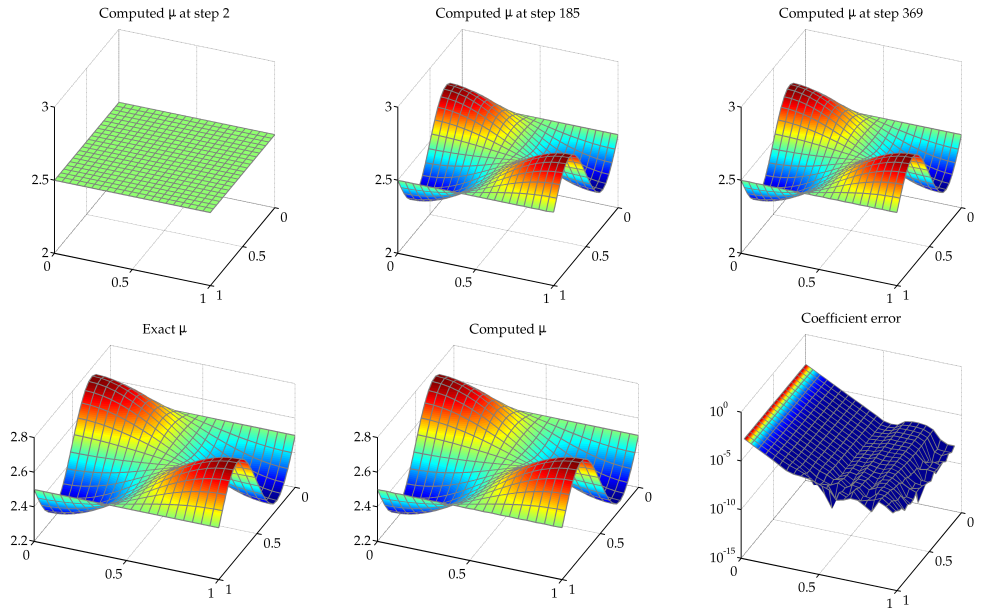


Figure 65: Hager-Zhang using Quadratic φ-Divergence (EOLS)

VI. Conclusion

In this work, we have examined a variety of computational methods for the solution of the tumor identification inverse problem in linear elasticity. The challenges inherent in applying numerical methods to the underlying boundary value problem, primarily the locking effect, have been analyzed and a method of overcoming this obstacle using a general saddle point problem approach has been presented. Based in this approach we have also presented three optimization frameworks, the modified output least squares (MOLS), energy output least squares (EOLS), and equation error (EE) for solving the general inverse parameter identification problem. The relative merits of each approach, convexity in the case of MOLS and EE, have been compared along with the full development of a mixed finite element discretization scheme for the elasticity imaging inverse problem.

For the solution of the related optimization problem, we have explored a variety of gradient-based, extragradient, and proximal point iterative optimization methods. Extragradient methods are appealing for the problem at hand due to the relaxed conditions for convergence on the objective functional. Alternatively, proximal point methods are appealing since they obviate the need for the selection of an optimal regularization parameter, a drawback of using Tikhonov regularization in conjunction with other methods.

To compare the differing approaches, an example problem in linear elasticity with near incompress-

ibility was developed and a mixed finite element discretization was applied to all of the optimization frameworks under consideration. Subsequently, all of the optimization methods were implemented and applied to each of the frameworks in turn.

The gradient-based algorithms all had certain performance advantages over all of the extragradient methods given that they only require a single gradient evaluation per iteration. Although, they exhibited some occasional numerical instability, all the gradient-based methods performed well when compared with the extragradient methods (see Figures 49,32,15). Of particular note is the performance of the FISTA algorithm, which, given its simplicity, performed admirably. We remark that this is likely the first time this promising algorithm has been applied to the solution of inverse problem like the one in elasticity imaging.

In general, all of the extragradient methods were capable of solving the inverse problem, but only one could be said to excel remarkably in terms of performance: the Solodov-Tseng scaled extragradient method using a Hessian-based scaling matrix. One potential cause of such poor performance is that many of the methods suffer from an overabundance of parameters (including the Tikhonov regularization parameter), all of which are selected heuristically and all of which may have a considerable impact on the overall convergence of the method. Comparatively, all of the proximal point methods outperformed the extragradient methods both in their rates of convergence as well as in their ease of use, given that little to no parameter selection was necessary. Compared to each other, the three first-order proximal methods performed similarly for all three functionals. As would be expected, any method that included second-order information about the objective functional (SPG, Soldov-Tseng extragradient using the Hessian, and the Quadratic φ-divergence proximal point method) outperformed all other first-order methods.

Another potential advantage of the proximal point methods regarding regularization can be seen in the comparison of Figures 49 and 50. Here we see that proximal point methods are able to find far "lower" minima of the functional J than either the gradient or extragradient approaches in (generally) far fewer iterations. This is possibly attributable to slower convergence but more likely to over-regularization through the selection of too large a Tikhonov regularization parameter on the part of the gradient and extragradient methods. Even if we were equipped with a method of determining an optimal Tikhonov regularization parameter, the proximal point methods would still have an advantage in this regard (for the theoretical reasons laid out in §IV) and this is in evidence in the results of our numerical experiments.

VII. Acknowledgments

I would like to thank the members of my thesis committee, Patricia Clark and Tamas Wiandt, for their helpful insight and suggestions throughout the creation of this work as well as thank Nathan Cahill for his support and guidance throughout the entire process. I would also like to express my heartfelt gratitude to Akhtar Khan and Baasansuren Jadamba for their support, collaboration, and kindness over the years. Truly, none of this would have been possible without you.

I would also like to thank my wife Jody for her love, support, understanding, and really, for everything, as well as thank my mother, Gale, and sister, Jodi, for their love, support and understanding.

VIII. Appendix

VIII.1 Notation and Definitions

The Lebesgue space $L^p(\Omega)$ with $\Omega \subset \mathbb{R}^n$ is the space of functions $u : \Omega \to \mathbb{R}$ for which

$$\left(\int_\Omega |u(x)|^p \, dx \right)^{\frac{1}{p}} < \infty.$$

In particular, we consider the space of square integrable functions $L^2(\Omega)$. $L^2(\Omega)$ is a Hilbert space with inner product

$$\langle u, v \rangle = \int_\Omega u(x)v(x) \, dx.$$

The Sobolev space $W^{m,p}(\Omega)$ is the space of functions $u : \Omega \to \mathbb{R}$ such that

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega) \text{ for } 0 \leq |\alpha| \leq m\}$$

where $D^\alpha u$ is understood as the α-th weak partial derivative of $u(x)$.

Taking $p = 2$, we define the space $H^m = W^{m,2}$ and again primarily consider

$$H^1 = \left\{ u \in L^2(\Omega) : Du \in L^2(\Omega) \right\}$$

along with

$$H_0^1 = \left\{ u \in H^1(\Omega) : u(x) = 0 \text{ on } \Gamma_1 \right\}$$

where Γ_1 is some boundary of Ω.

We represent a two-dimensional vector-valued function $u : \Omega \to \mathbb{R}^2$, that is $u(x) = \begin{bmatrix} u_1(x) \\ u_2(x) \end{bmatrix}$, in the product space

$$\widehat{V} = H^1(\Omega) \times H^1(\Omega) \qquad \text{or} \qquad \widehat{V} = H_0^1(\Omega) \times H_0^1(\Omega)$$

with the obvious extension for three dimensions.

For $u \in H^1(\Omega)$, we consider the H^1 norm $\| \cdot \|_{H^1} : H^1(\Omega) \to \mathbb{R}$ as follows:

$$\|u\|_{H^1}^2 = \langle u, u \rangle + \langle Du, Du \rangle = \int_\Omega |u|^2 \, dx + \int_\Omega |Du|^2 \, dx. \tag{VIII.1}$$

Additionally, we note that it also makes sense to consider the L^2 norm on $u \in H^1(\Omega)$, $\| \cdot \|_{L^2} : H^1(\Omega) \to \mathbb{R}$ where

$$\|u\|_{L^2}^2 = \langle u, u \rangle = \int_\Omega |u|^2 \, dx. \tag{VIII.2}$$

Likewise, we may also consider the H^1 semi-norm on $u \in H^1(\Omega)$, $|\cdot|_{H^1} : H^1(\Omega) \to \mathbb{R}$ where

$$|u|_{H^1}^2 = \langle Du, Du \rangle = \int_\Omega |u|^2 \, dx. \tag{VIII.3}$$

We remark that $|\cdot|_{H^1}$ is only a semi-norm since clearly $|u|_{H^1} = 0$ does not necessarily imply $u = 0$.

Taking V to be a Hilbert space, we say a form $a : V \times V \to \mathbb{R}$ is a *symmetric bilinear form* if it meets the following criteria:

1. $a(u, v) = a(v, u)$ for all $u, v \in V$. (Symmetry)
2. $a(\alpha u + \beta v, w) = \alpha a(u, w) + \beta a(v, w)$ for all $u, v, w \in V$ and for $\alpha, \beta \in \mathbb{R}$. (Bilinearity)

Additionally, $a(\cdot, \cdot)$ satisfies

3. $a(u, u) \geq 0$ for all $u \in V$ and where $u = 0$ implies $a(u, u) = 0$.

We call $a(\cdot, \cdot)$ *coercive* or V*-elliptic* if there exists some $\alpha \in \mathbb{R}, \alpha > 0$ such that

4. $a(u, u) \geq \alpha \|u\|_V^2$ for all $u \in V$.

$a(\cdot, \cdot)$ is *bounded* if there exists a $C \in \mathbb{R}, C > 0$ such that

5. $a(u, v) \leq C \|u\|_V \|v\|_V$ for all $u \in V$.

We denote the dual space of V, the space of all linear functionals on V, as V^*.

The Lax-Milgram Theorem states that for a symmetric bilinear form $a(\cdot, \cdot)$ that is both coercive and bounded, it follows that for any $m \in V^*$, there exists a unique solution $u \in V$ to

$$a(u, v) = m(v) \quad \text{for all } v \in V. \tag{VIII.4}$$

IX. Bibliography

[1] R. Acar. Identification of the coefficient in elliptic equations. *SIAM journal on control and optimization*, 31(5):1221–1244, 1993.

[2] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*. Academic Press, Amsterdam; Boston, 2003.

[3] M. A. Aguilo, W. Aquino, J. C. Brigham, and M. Fatemi. An inverse problem approach for elasticity imaging through vibroacoustics. *Medical Imaging, IEEE Transactions on*, 29(4):1012–1021, 2010.

[4] M. A. Aguiló, L. P. Swiler, and A. Urbina. First-order formulations for large-scale stochastic parameter estimation within the frameworks of steady state dynamics: the elastic and viscoelastic case. *Inverse Problems*, 28(7):075003, 2012.

[5] M. F. Al-Jamal and M. S. Gockenbach. Stability and error estimates for an equation error method for elliptic equations. *Inverse problems*, 28(9):095006, 2012.

[6] U. Albocher, A. A. Oberai, P. E. Barbone, and I. Harari. Adjoint-weighted equation for inverse problems of incompressible plane-stress elasticity. *Computer Methods in Applied Mechanics and Engineering*, 198(30):2412–2420, 2009.

[7] H. Ammari, P. Garapon, and F. Jouve. Separation of scales in elasticity imaging: a numerical study. *J. Comput. Math*, 28(3):354–370, 2010.

[8] P. N. Anh and N. D. Hien. The extragradient-Armijo method for pseudomonotone equilibrium problems and strict pseudocontractions. *Fixed Point Theory Appl.*, pages 2012:82, 16, 2012.

[9] A. S. Antipin, L. A. Artemeva, and F. P. Vasilev. A regularized extragradient method for solving a parametric multicriteria equilibrium programming problem. *Zh. Vychisl. Mat. Mat. Fiz.*, 50(12):2083–2098, 2010.

[10] A. Arnold, S. Reichling, O. T. Bruhns, and J. Mosler. Efficient computation of the elastography inverse problem by combining variational mesh adaption and a clustering technique. *Physics in medicine and biology*, 55(7):2035, 2010.

[11] A. Auslender and M. Teboulle. Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. *Math. Program.*, 120(1, Ser. B):27–48, 2009.

[12] J. C. Bamber, L. De Gonzalez, D. O. Cosgrove, P. Simmons, J. Davey, and J. A. McKinna. Quantitative evaluation of real-time ultrasound features of the breast. *Ultrasound in medicine & biology*, 14:81–87, 1988.

[13] B. Banerjee, T. F. Walsh, W. Aquino, and M. Bonnet. Large scale parameter estimation problems in frequency-domain elastodynamics using an error in constitutive equation functional. *Computer methods in applied mechanics and engineering*, 253:60–72, 2013.

[14] P. E. Barbone and J. C. Bamber. Quantitative elasticity imaging: what can and cannot be inferred from strain images. *Physics in medicine and biology*, 47(12):2147, 2002.

[15] P. E. Barbone and N. H. Gokhale. Elastic modulus imaging: on the uniqueness and nonuniqueness of the elastography inverse problem in two dimensions. *Inverse problems*, 20(1):283, 2004.

[16] P. E. Barbone, A. A. Oberai, and I. Harari. Adjoint-weighted variational formulation for a direct computational solution of an inverse heat conduction problem. *Inverse Problems*, 23(6):2325, 2007.

[17] P. E. Barbone, C. E. Rivas, I. Harari, U. Albocher, A. A. Oberai, and Y. Zhang. Adjoint-weighted variational formulation for the direct solution of inverse problems of general linear elasticity with full interior data. *International journal for numerical methods in engineering*, 81(13):1713–1736, 2010.

[18] H. H. Bauschke, P. L. Combettes, and D. R. Luke. A strongly convergent reflection method for finding the projection onto the intersection of two closed convex sets in a hilbert space. *Journal of Approximation Theory*, 141:63–69, 2006.

[19] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[20] J. Y. Bello Cruz and A. N. Iusem. Convergence of direct methods for paramonotone variational inequalities. *Comput. Optim. Appl.*, 46(2):247–263, 2010.

[21] F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero. Nonnegative least-squares image deblurring: improved gradient projection approaches. *Inverse Problems*, 26(2):025004, 18, 2010.

[22] E. Beretta, E. Bonnetier, E. Francini, and A. L. Mazzucato. Small volume asymptotics for anisotropic elastic inclusions. *Inverse Prob. Imaging*, 6:1–23, 2011.

[23] M. Bertrand, J. Meunier, M. Doucet, and G. Ferland. Ultrasonic biomechanical strain gauge based on speckle tracking. In *Ultrasonics Symposium, 1989. Proceedings., IEEE 1989*, page 859–863. IEEE, 1989.

[24] D. P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method. In *Proceedings of the 1974 IEEE Conference on Decision and Control including 13th Symposium on Adaptive Processes (Phoenix, Ariz., 1974)*, pages 47–52. Inst. Electrical Electron. Engrs., New York, 1974.

[25] A. Bnouhachem. An additional projection step to He and Liao's method for solving variational inequalities. *J. Comput. Appl. Math.*, 206(1):238–250, 2007.

[26] A. Bnouhachem, X.-L. Fu, M. H. Xu, and S. Zhaohan. New extragradient-type methods for solving variational inequalities. *Appl. Math. Comput.*, 216(8):2430–2440, 2010.

[27] A. Bnouhachem, M. A. Noor, E. Al-Said, M. Khalfaoui, and S. Zhaohan. Extragradient method for variational inequalities. *Hacet. J. Math. Stat.*, 40(6):839–854, 2011.

[28] N. Bochud and G. Rus. Probabilistic inverse problem to characterize tissue-equivalent material mechanical properties. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 59(7):1443–1456, 2012.

[29] S. Bonettini and V. Ruggiero. An alternating extragradient method for total variation-based image restoration from Poisson data. *Inverse Problems*, 27(9):095001, 26, 2011.

[30] D. Braess. *Finite elements: Theory, fast solvers, and applications in elasticity theory*. Cambridge University Press, Cambridge, third edition, 2007. Translated from the German by Larry L. Schumaker.

[31] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. Springer-Verlag New York, Inc., 1991.

[32] B. A. Budak. A second-order regularized continuous extragradient method with a variable metric for solving equilibrium programming problems with an inexactly specified set. *Differ. Uravn.*, 40(2):154–168, 285, 2004.

[33] R. S. Burachik, S. Scheimberg, and B. F. Svaiter. Robustness of the hybrid extragradient proximal-point algorithm. *J. Optim. Theory Appl.*, 111(1):117–136, 2001.

[34] N. D. Cahill, B. Jadamba, A. A. Khan, M. Sama, and B. C. Winkler. A first-order adjoint and a second-order hybrid method for an energy output least-squares elastography inverse problem of identifying tumor location. *Boundary Value Problems*, 2013(1):263, 2013.

[35] Z. Cai and G. Starke. Least-squares methods for linear elasticity. *SIAM Journal on Numerical Analysis*, 42(2):826–842, 2004.

[36] L.-C. Ceng and J.-L. Ho. Hybrid extragradient method with regularization for convex minimization, generalized mixed equilibrium, variational inequality and fixed point problems. *Abstr. Appl. Anal.*, pages Art. ID 436069, 27, 2014.

[37] Y. Censor, A. Gibali, and S. Reich. The subgradient extragradient method for solving variational inequalities in Hilbert space. *J. Optim. Theory Appl.*, 148(2):318–335, 2011.

[38] Y. Censor, A. Gibali, and S. Reich. Extensions of Korpelevich's extragradient method for the variational inequality problem in Euclidean space. *Optimization*, 61(9):1119–1132, 2012.

[39] T. F. Chan and X.-C. Tai. Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM Journal on Scientific Computing*, 25(3):881–904, 2003.

[40] P. G. Ciarlet. *Mathematical elasticity, Vol I: Three-Dimensional Elasticity*, volume 1 of *Studies in mathematics and its applications*. North-Holland ; Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co, Amsterdam ; New York : New York, N.Y., U.S.A, 1988.

[41] A. Cioaca, M. Alexe, and A. Sandu. Second-order adjoints for solving PDE-constrained optimization problems. *Optimization methods and software*, 27(4-5):625–653, 2012.

[42] P. L. Combettes. Systems of structured monotone inclusions: duality, algorithms, and applications. *SIAM J. Optim.*, 23(4):2420–2447, 2013.

[43] E. Crossen, M. S. Gockenbach, B. Jadamba, A. A. Khan, and B. Winkler. An equation error approach for the elasticity imaging inverse problem for predicting tumor location. *Computers & Mathematics with Applications*, 67(1):122–135, 2014.

[44] M. Doyley, B. Jadamba, A. A. Khan, M. SAMA, and B. Winkler. A new energy inversion for parameter identification in saddle point problems with an application to the elasticity imaging inverse problem of predicting tumor location. *Numerical Functional Analysis and Optimization*, 2014.

[45] M. M. Doyley. Model-based elastography: a survey of approaches to the inverse elasticity problem. *Physics in medicine and biology*, 57(3):R35, 2012.

[46] M. M. Doyley, P. M. Meaney, and J. C. Bamber. Evaluation of an iterative reconstruction method for quantitative elastography. *Physics in Medicine and Biology*, 45(6):1521, 2000.

[47] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*, volume 5. CRC press, 1991.

[48] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems II*, volume 2. Springer, New York, 2003.

[49] R. S. Falk. Error estimates for the numerical identification of a variable coefficient. *Mathematics of Computation*, 40(162):537–546, 1983.

[50] J. Fehrenbach, M. Masmoudi, R. Souchon, and P. Trompette. Detection of small inclusions by elastography. *Inverse problems*, 22(3):1055, 2006.

[51] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman lectures on physics. Mainly electromagnetism and matter*, volume 2. Addison-Wesley, Reading, Mass., 1977.

[52] I. N. Figueiredo and C. Leal. Physiologic parameter estimation using inverse problems. *SIAM Journal on Applied Mathematics*, 73(3):1164–1182, 2013.

[53] G. Geymonat and S. Pagano. Identification of mechanical properties by displacement field measurement: a variational approach. *Meccanica*, 38(5):535–545, 2003.

[54] M. S. Gockenbach. *Understanding and implementing the finite element method*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2006.

[55] M. S. Gockenbach, B. Jadamba, and A. A. Khan. Numerical estimation of discontinuous coefficients by the method of equation error. *International Journal of Mathematics and Computer Science*, 1(3):343–359, 2006.

[56] M. S. Gockenbach, B. Jadamba, and A. A. Khan. Equation error approach for elliptic inverse problems with an application to the identification of lamé parameters. *Inverse Problems in Science and Engineering*, 16(3):349–367, 2008.

[57] M. S. Gockenbach and A. A. Khan. Identification of lamé parameters in linear elasticity: a fixed point approach. *Journal of Industrial and Management Optimization*, 1(4):487, 2005.

[58] M. S. Gockenbach and A. A. Khan. An abstract framework for elliptic inverse problems: Part 1. an output least-squares approach. *Mathematics and mechanics of solids*, 12(3):259–276, 2007.

[59] M. S. Gockenbach and A. A. Khan. An abstract framework for elliptic inverse problems: Part 2. an augmented lagrangian approach. *Mathematics and Mechanics of Solids*, 14(6):517–539, 2009.

[60] D. Goeleven, D. Motreanu, Y. Dumont, and M. Rochdi. *Variational and hemivariational inequalities: Theory, methods and applications*, volume 1. Springer, 2003.

[61] W. W. Hager and H. Zhang. Self-adaptive inexact proximal point methods. *Computational Optimization and Applications*, 39(2):161–181, 2008.

[62] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73:957–961, 1967.

[63] D. Han. A generalized proximal-point-based prediction–correction method for variational inequality problems. *Journal of Computational and Applied Mathematics*, 221(1):183–193, 2008.

[64] D. Han and H. K. Lo. Two new self-adaptive projection methods for variational inequality problems. *Comput. Math. Appl.*, 43(12):1529–1537, 2002.

[65] T. P. Harrigan and E. E. Konofagou. Estimation of material elastic moduli in elastography: a local method, and an investigation of poisson's ratio sensitivity. *Journal of biomechanics*, 37(8):1215–1221, 2004.

[66] Y. Haugazeau. Sur les inéquations variationnelles et la minimisation de fonctionnelles convexes. In *Thèse*, pages 359–366. niversité de Paris, Paris, France, 1968.

[67] B. He, X. Yuanm, and J. J. Z. Zhang. Comparison of two kinds of prediction-correction methods for monotone variational inequalities. *Comput. Optim. Appl.*, 27(3):247–267, 2004.

[68] B. S. He, H. Yang, Q. Meng, and D. R. Han. Modified Goldstein-Levitin-Polyak projection method for asymmetric strongly monotone variational inequalities. *J. Optim. Theory Appl.*, 112(1):129–143, 2002.

[69] Y. He. A new double projection algorithm for variational inequalities. *J. Comput. Appl. Math.*, 185(1):166–173, 2006.

[70] P. Houston, D. Schötzau, and T. P. Wihler. An hp-adaptive mixed discontinuous galerkin FEM for nearly incompressible linear elasticity. *Computer methods in applied mechanics and engineering*, 195(25):3224–3246, 2006.

[71] A. N. Iusem and M. Nasri. Korpelevich's method for variational inequality problems in Banach spaces. *J. Global Optim.*, 50(1):59–76, 2011.

[72] A. N. Iusem and L. R. L. Pérez. An extragradient-type algorithm for non-smooth variational inequalities. *Optimization*, 48(3):309–332, 2000.

[73] B. Jadamba, A. A. Khan, M. Paulhamus, and M. Sama. Proximal point methods for the inverse problem of identifying parameters in beam models. In *EMERGING APPLICATIONS OF WAVELET METHODS: 7th International Congress on Industrial and Applied Mathematics-Thematic Minisymposia*, volume 1463, page 16–38. AIP Publishing, 2012.

[74] B. Jadamba, A. A. Khan, and F. Raciti. On the inverse problem of identifying lamé coefficients in linear elasticity. *Computers & Mathematics with Applications*, 56(2):431–443, 2008.

[75] B. Jadamba, A. A. Khan, and F. Raciti. On the inverse problem of identifying Lamé coefficients in linear elasticity. *Comput. Math. Appl.*, 56(2):431–443, 2008.

[76] B. Jadamba, A. A. Khan, G. Rus, M. Sama, and B. Winkler. A new convex inversion for parameter identification in saddle point problems with an application to the elasticity imaging inverse problem of prediction tumor location. 2013.

[77] B. Jadamba, A. A. Khan, and M. Sama. Inverse problems of parameter identification in partial differential equations. *Mathematics in Science and Technology, World Sci. Publ., Hackensack, NJ*, page 228–258, 2011.

[78] L. Ji and J. McLaughlin. Recovery of the lamé parameter μ in biological tissues. *Inverse Problems*, 20(1):1, 2004.

[79] F. Kallel and M. Bertrand. Tissue elasticity reconstruction using linear perturbation method. *Medical Imaging, IEEE Transactions on*, 15(3):299–313, 1996.

[80] C. Kanzow. *Proximal-Like Methods for Convex Minimization Problems*. Springer, 2005.

[81] A. Kaplan and R. Tichatschke. Proximal point method and elliptic regularization. *Nonlinear Analysis: Theory, Methods & Applications*, 71(10):4525–4543, 2009.

[82] T. Kärkkäinen. An equation error method to recover diffusion from the distributed observation. *Inverse Problems*, 13(4):1033, 1997.

[83] T. Kärkkäinen. Error estimates for distributed parameter identification in parabolic problems with output least squares and crank-nicolson method. *Applications of Mathematics*, 42(4):259–277, 1997.

[84] P. Kazemi and R. J. Renka. A Levenberg–Marquardt method based on sobolev gradients. *Nonlinear Analysis: Theory, Methods & Applications*, 75(16):6170–6179, 2012.

[85] C. T. Kelley. *Iterative methods for optimization*. Frontiers in applied mathematics. SIAM, Philadelphia, 1999.

[86] A. A. Khan and B. D. Rouhani. Iterative regularization for elliptic inverse problems. *Computers & Mathematics with Applications*, 54(6):850–860, 2007.

[87] P. D. Khanh and P. T. Vuong. Modified projection method for strongly pseudomonotone variational inequalities. *J. Global Optim.*, 58(2):341–350, 2014.

[88] E. N. Khobotov. A modification of the extragradient method for solving variational inequalities and some optimization problems. *Zh. Vychisl. Mat. i Mat. Fiz.*, 27(10):1462–1473, 1597, 1987.

[89] E. N. Khobotov. The use of the extragradient method for solving multistep games of various types. *Izv. Akad. Nauk Teor. Sist. Upr.*, (4):83–88, 1997.

[90] E. V. Khoroshilova. Extragradient-type method for optimal control problem with linear constraints and convex objective function. *Optim. Lett.*, 7(6):1193–1214, 2013.

[91] I. Knowles. Parameter identification for elliptic problems. *Journal of computational and applied mathematics*, 131(1):175–194, 2001.

[92] I. V. Konnov. A class of combined relaxation methods for decomposable variational inequalities. *Optimization*, 51(1):109–125, 2002.

[93] E. E. Konofagou, T. Harrigan, J. Ophir, T. Krouskop, and C. Fife. Poroelastography: Estimation and imaging of the poroelastic properties of tissues. In *IEEE Symp Proc, Lake Tahoe, NV*, page 1627–1630, 1999.

[94] G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, 12(4):747–756, 1976.

[95] N. Langenberg. Interior point methods for equilibrium problems. *Comput. Optim. Appl.*, 53(2):453–483, 2012.

[96] M. Li, L.-Z. Liao, and X. M. Yuan. Some Goldstein's type methods for co-coercive variant variational inequalities. *Appl. Numer. Math.*, 61(2):216–228, 2011.

[97] M. Li, L. L.Z., and X. . Yuan. Some goldstein-type methods for co-coervice variant variational inequalities. *Appl. Numer. Math.*, 61:216–228, 2011.

[98] P.-E. Maingé. A hybrid extragradient-viscosity method for monotone operators and fixed point problems. *SIAM J. Control Optim.*, 47(3):1499–1515, 2008.

[99] P.-E. Maingé and A. Moudafi. Coupling viscosity methods with the extragradient algorithm for solving equilibrium problems. *J. Nonlinear Convex Anal.*, 9(2):283–294, 2008.

[100] P. Marcotte. Application of khobotov's algorithm to variational inequalities and network equilibrium problems. *INFORM*, 29(4):258–270, 1991.

[101] J. Mashreghi and M. Nasri. Forcing strong convergence of Korpelevich's method in Banach spaces with its applications in game theory. *Nonlinear Anal.*, 72(3-4):2086–2099, 2010.

[102] J. R. McLaughlin and J.-R. Yoon. Unique identifiability of elastic parameters from time-dependent interior displacement measurement. *Inverse Problems*, 20(1):25, 2004.

[103] H. Mehrabian, G. Campbell, and A. Samani. A constrained reconstruction technique of hyperelasticity parameters for breast cancer assessment. *Physics in medicine and biology*, 55(24):7489, 2010.

[104] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM J. Optim.*, 20(6):2755–2787, 2010.

[105] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.*, 22(3):914–935, 2012.

[106] R. D. C. Monteiro and B. F. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *SIAM J. Optim.*, 23(2):1092–1125, 2013.

[107] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris*, 255:2897–2899, 1962.

[108] N. Nadezhkina and W. Takahashi. Modified extragradient method for solving variational inequalities in real Hilbert spaces. In *Nonlinear analysis and convex analysis*, pages 359–366. Yokohama Publ., Yokohama, 2004.

[109] M. Z. Nashed and O. Scherzer. Stable approximation of nondifferentiable optimization problems with variational inequalities. *Recent Developments in Optimization Theory and Nonlinear Analysis",(eds. Y. Censor and S. Reich), Contemp. Math*, 204:155–170, 1997.

[110] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.*, 15(1):229–251 (electronic), 2004.

[111] A. Nemirovski and D. B. Yudin. In *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics. John Wiley and Sons, New York, 1983.

[112] A. Nemirovski and D. B. Yudin. Problem complexity and method efficiency in optimization. *SIAM J. Optim.*, 15(1):229–251 (electronic), 2004.

[113] Y. E. Nesterov. A method for solving the convex programming problem with convergence rate o($1/k2$). *Dokl. Akad. Nauk SSSR*, 269:543–547 (in Russian), 1983.

[114] A. A. Oberai, N. H. Gokhale, and G. R. Feijoo. Solution of inverse problems in elasticity imaging using the adjoint method. *Inverse Problems*, 19(2):297, 2003.

[115] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the h 1. *Multiscale Modeling & Simulation*, 1(3):349–370, 2003.

[116] L. A. Parente, P. A. Lotito, and M. V. Solodov. A class of inexact variable metric proximal point algorithms. *SIAM Journal on Optimization*, 19(1):240–260, 2008.

[117] E. Park and A. M. Maniatty. Shear modulus reconstruction in dynamic elastography: time harmonic case. *Physics in medicine and biology*, 51(15):3697, 2006.

[118] K. J. Parker, S. R. Huang, R. A. Musulin, and R. M. Lerner. Tissue response to mechanical vibrations for "sonoelasticity imaging". *Ultrasound in medicine & biology*, 16(3):241–246, 1990.

[119] F. Pierron, S. Avril, and V. T. Tran. Extension of the virtual fields method to elasto-plastic material identification with cyclic loads and kinematic hardening. *International Journal of Solids and Structures*, 47(22):2993–3010, 2010.

[120] L. D. Popov. On schemes for the formation of a master sequence in a regularized extragradient method for solving variational inequalities. *Izv. Vyssh. Uchebn. Zaved. Mat.*, (1):70–79, 2004.

[121] K. R. Raghavan and A. E. Yagle. Forward and inverse problems in elasticity imaging of soft tissues. *Nuclear Science, IEEE Transactions on*, 41(4):1639–1648, 1994.

[122] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[123] C. Roland and R. Varadhan. New iterative schemes for nonlinear fixed point problems, with applications to problems with bifurcations and incomplete-data problems. *Appl. Numer. Math.*, 55(2):215–226, 2005.

[124] M. V. Solodov. Convergence rate analysis of iteractive algorithms for solving variational inquality problems. *Math. Program.*, 96(3, Ser. A):513–528, 2003.

[125] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM J. Control Optim.*, 37(3):765–776, 1999.

[126] M. V. Solodov and B. F. Svaiter. Forcing strong convergence of proximal point iterations in a hilbert space. *Math. Program. Ser.*, 87(1):189–202, 2000.

[127] M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM J. Control Optim.*, 34(5):1814–1830, 1996.

[128] T. Suzuki. A sufficient and necessary condition for halpern-type strong convergence to fixed points of nonexpansive mappings. *Proc. Amer. Math. Soc*, 135:99–106, 2007.

[129] W. Takahashi, Y. Takeuchi, and R. Kubota. Strong convergence theorems by hybrid methods for families of nonexpansive mappings in hilbert spaces. *J. Math. Anal. Appl.*, 341:276–286, 2008.

[130] B. Taskar, S. Lacoste-Julien, and M. I. Jordan. Structured prediction, dual extragradient and Bregman projections. *J. Mach. Learn. Res.*, 7:1627–1653, 2006.

[131] A. N. Tikhonov and V. Y. Arsenin. Solutions of ill-posed problems. 1977.

[132] F. Tinti. Numerical solution for pseudomonotone variational inequality problems by extragradient methods. In *Variational analysis and applications*, volume 79 of *Nonconvex Optim. Appl.*, pages 1101–1128. Springer, New York, 2005.

[133] D. A. Tortorelli and P. Michaleris. Design sensitivity analysis: overview and review. *Inverse problems in Engineering*, 1(1):71–105, 1994.

[134] P. Tossings. Mixing proximal regularization, penalization and parallel decomposition in convex programming. In *Advances in Optimization*, page 85–99. Springer, 1992.

[135] P. Tossings. The perturbed proximal point algorithm and some of its applications. *Applied Mathematics and Optimization*, 29(2):125–159, 1994.

[136] P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM J. Optim.*, 7(4):951–965, 1997.

[137] D. J. Wall, P. Olsson, and E. E. Van Houten. On an inverse problem from magnetic resonance elastic imaging. *SIAM Journal on Applied Mathematics*, 71(5):1578–1605, 2011.

[138] X. Wang. Improved steplength by more practical information in the extragradient method for monotone variational inequalities. *J. Optim. Theory Appl.*, 141(3):661–676, 2009.

[139] X. Wang, B. He, and L.-Z. Liao. Steplengths in the extragradient type methods. *J. Comput. Appl. Math.*, 233(11):2925–2939, 2010.

[140] Y. J. Wang, N. H. Xiu, and C. Y. Wang. Unified framework of extragradient-type methods for pseudomonotone variational inequalities. *J. Optim. Theory Appl.*, 111(3):641–656, 2001.

[141] Y. Yao, M. A. Noor, and Y.-C. Liou. Strong convergence of a modified extragradient method to the minimum-norm solution of variational inequalities. *Abstr. Appl. Anal.*, pages Art. ID 817436, 9, 2012.

[142] A. J. Zaslavski. The extragradient method for convex optimization in the presence of computational errors. *Numer. Funct. Anal. Optim.*, 33(12):1399–1412, 2012.

[143] A. J. Zaslavski. The extragradient method for solving variational inequalities in the presence of computational errors. *J. Optim. Theory Appl.*, 153(3):602–618, 2012.

[144] H. Zegeye and N. Shahzad. Extragradient method for solutions of variational inequality problems in Banach spaces. *Abstr. Appl. Anal.*, pages Art. ID 832548, 8, 2013.

[145] L.-C. Zeng and J.-C. Yao. strong convergence theorem by an extragradient method for fixed point problems and variational inequality problems. *Taiwan. J. Math.*, (5):1293–1303, 2006.

[146] Y. Zhu, T. J. Hall, and J. Jiang. A finite-element approach for young's modulus reconstruction. *Medical Imaging, IEEE Transactions on*, 22(7):890–901, 2003.

[147] J. Zou. Numerical methods for elliptic inverse problems. *International journal of computer mathematics*, 70(2):211–232, 1998.

[148] A. V. Zykina and N. V. Melenchuk. A two-step extragradient method for variational inequalities. *Izv. Vyssh. Uchebn. Zaved. Mat.*, (9):82–85, 2010.