



# Relatório Técnico

**Núcleo de  
Computação Eletrônica**

## **On Numerical Approximation of an Optimal Control Problem in Linear Elasticity**

**M. A. Rincon  
I-Shih Liu**

**NCE - 35/99**

**Universidade Federal do Rio de Janeiro**

# On Numerical Approximation of an Optimal Control Problem in Linear Elasticity

M. A. Rincon & I-Shih Liu\*

*Instituto de Matemática, Universidade Federal do Rio de Janeiro  
Caixa Postal 68530, Rio de Janeiro 21945-970, Brazil*

## Abstract

In this paper we apply the optimal control theory to a linear elasticity problem. An iterative method based on the optimality system characterizing the corresponding minimization of a cost functional is proposed. Convergence of the approximate solutions is proved provided that a parameter of penalization is not too small. Numerical solutions are presented to emphasize the role of this parameter. It is shown that the results are far from being good approximations of the expected ones, because the parameter can not be taken small enough in the iteration method. On the other hand, numerical results from a spectral analysis are shown without this limitation by the use of eigenfunction representations.

*Keywords:* Optimal control; Linear elasticity; finite element method; Iterative method; Eigenfunction expansion

## 1 Introduction

The theory of optimal control of systems governed by partial differential equations was essentially developed by Lions [6]. Many different type of control problems have been considered and solutions by numerical methods have been widely studied in the literature (see for examples, [4, 5, 6, 7]).

Theory of optimal control governed by a scalar equation of elliptic type, such as those involving heat conduction, are well-known in the literature. The problem is usually formulated as a minimization problem of a cost functional, involving a positive parameter for technical reasons. For the practical objective of the optimal control solution this parameter should be taken as small as possible.

---

\*E-mails: [rincon@dmm.im.ufrj.br](mailto:rincon@dmm.im.ufrj.br), [liu@dmm.im.ufrj.br](mailto:liu@dmm.im.ufrj.br)

In this paper we shall apply the theory of optimal control to equilibrium problems of linear elasticity, governed by an elliptic system of partial differential equations. An iterative method for the optimality system equivalent to the minimization problem is proposed and it is shown that the convergence of the iterative solutions is guaranteed if the technical parameter is not too small. The role of this parameter will be examined in the numerical solutions with finite elements based on the iterative method proposed. It is found that such solutions are quite unsatisfactory even with the smallest allowable parameter and hence the iterative method which does not allow the parameter to go to zero may hardly be able to deliver a satisfactory optimal solution as expected. However, this unpleasant situation seems to have been overlooked in many similar problems in the literature [5, 7, 9], where the convergence of the numerical solution is ensured by simply taking this parameter as 1 or a fixed convenient positive number. Therefore apparently those results may just as unsatisfactory as ours in this respect.

One may raise the question of whether the optimality system admits a solution in the limit when the parameter tends to zero. In order to see that a spectral analysis is considered and an explicit solution in Fourier series expansion is obtained, which clearly answers the question positively at least for the problem considered in this paper.

A bold-faced letter stands for a vector quantity and its components are represented by the corresponding normal letter with subindices ranging from 1 to  $n$ , the dimension of the physical space. The usual summation convention will be used to the component indices, *i.e.*, the repeated component indices indicate a sum over its range from 1 to  $n$ .

## 2 An Optimal Control Problem in Elasticity

We consider the following boundary value problem of Dirichlet type in linear elasticity:

$$\begin{cases} -\frac{\partial}{\partial x_j} \left( C_{ijkl} \frac{\partial u_k}{\partial x_l} \right) = f_i, & \text{in } \Omega \\ u_i = 0, & \text{in } \partial\Omega \end{cases} \quad (1)$$

where the elasticity tensor  $C_{ijkl}$  and the external force  $f_i$  are given functions of  $\mathbf{x} = (x_i)$  in a smooth region  $\Omega \subset R^n$  with smooth boundary  $\partial\Omega$ . The problem is to determine the displacement field  $\mathbf{u} = (u_i)$  satisfying the system (1). It is usually assumed that ([1, 3])

a) The elasticity tensor  $C_{ijkl}$  satisfy the condition:

$$C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}. \quad (2)$$

b) There are positive constants  $C$  and  $\bar{C}$ , such that for any symmetric matrices  $S_{ij}$ ,

$$CS_{ij}S_{ij} \leq C_{ijkl}S_{ij}S_{kl} \leq \bar{C}S_{ij}S_{ij}. \quad (3)$$

The assumption (a) follows from the existence of stored-energy function and the symmetry of stress and strain tensors, while the assumption (b) states that the elasticity tensor is bounded and strictly positive definite.

By virtue of these assumptions, the bilinear function  $\sigma$  defined by

$$\sigma(\mathbf{u}, \mathbf{v}) = \int_{\Omega} C_{ijkl} \frac{\partial u_i}{\partial x_j} \frac{\partial v_k}{\partial x_l} dx, \quad \forall \mathbf{u}, \mathbf{v} \in (H_0^1(\Omega))^n \quad (4)$$

is an inner product in  $(H_0^1(\Omega))^n$ , called the energy inner product. The energy norm  $\|\mathbf{u}\|_{\sigma} = \sigma(\mathbf{u}, \mathbf{u})^{1/2}$  is equivalent to the standard norm  $\|\mathbf{u}\|_1$  of the space  $(H^1(\Omega))^n$  for  $\mathbf{u} \in (H_0^1(\Omega))^n$ . Indeed, from (3) we have

$$C\|\mathbf{u}\|_1 \leq \|\mathbf{u}\|_{\sigma} \leq \bar{C}\|\mathbf{u}\|_1. \quad (5)$$

We denote the inner product in  $(L^2(\Omega))^n$  by

$$(\mathbf{u}, \mathbf{v})_0 = \int_{\Omega} \mathbf{u} \cdot \mathbf{v} dv, \quad \forall \mathbf{u}, \mathbf{v} \in (L^2(\Omega))^n,$$

where  $\mathbf{u} \cdot \mathbf{v} = u_i v_i$  is the usual inner product in  $R^n$ . The usual norm in  $(L^2(\Omega))^n$  will be denoted by  $\|\mathbf{u}\|_0 = (\mathbf{u}, \mathbf{u})_0^{1/2}$ .

The weak form of the problem (1) can now be stated as follows: For a given function  $\mathbf{f} \in (L^2(\Omega))^n$ , find the solution  $\mathbf{u} \in (H_0^1(\Omega))^n$  such that

$$\sigma(\mathbf{u}, \mathbf{v}) = (\mathbf{f}, \mathbf{v})_0, \quad \forall \mathbf{v} \in (H_0^1(\Omega))^n. \quad (6)$$

We can verify easily that the bilinear function  $\sigma(\cdot)$  is continuous and coercive in  $(H_0^1(\Omega))^n$ . Therefore, according to Lax-Milgram Theorem ([3, 8]) and by the use of elliptical regularity, for any  $\mathbf{f} \in (L^2(\Omega))^n$ , there is a unique solution  $\mathbf{u} \in (H_0^1(\Omega))^n \cap (H^2(\Omega))^n$ . The unique solution for the given  $\mathbf{f}$  will be denoted by  $\mathbf{u}(\mathbf{f})$ .

Now let us turn to the formulation of an optimal control problem to obtain a prescribed displacement by means of externally applied forces on the body. Suppose that the equilibrium position of the body has some prescribed form given by a function  $\mathbf{z}^o(\mathbf{x}) \in (L^2(\Omega))^n$ , called an *objective function*. We can ask the question of how to determine the function  $\mathbf{f}$  such that  $\mathbf{u}(\mathbf{f})$  is as close to the objective function  $\mathbf{z}^o$  as possible in  $(L^2(\Omega))^n$ .

We introduce the *cost functional* defined by

$$J(\mathbf{f}) = \|\mathbf{u}(\mathbf{f}) - \mathbf{z}^o\|_0^2 + N\|\mathbf{f}\|_0^2, \quad (7)$$

where the term with a positive constant  $N$  has been added for technical reasons. But for optimal result in approaching the objective function the constant  $N$  is to be taken as small as possible.

We can formulate the optimal control problem as a minimization problem of the cost functional:

*Determine a function  $\mathbf{g} \in U_{ad}$  such that*

$$J(\mathbf{g}) = \inf \{J(\mathbf{f}); \mathbf{f} \in U_{ad}\}, \quad (8)$$

where  $U_{ad}$  is a convex and closed set of  $(L^2(\Omega))^n$ .

The existence and uniqueness of the *optimal control*  $\mathbf{g} \in U_{ad}$  is a consequence of the following classical theorem ([6]):

**Theorem.** Let  $J$  be a functional defined on Hilbert space  $H$ , and suppose that it is lower semi-continuous and strictly convex, satisfying  $J(\mathbf{f}) \rightarrow \infty$  as  $\|\mathbf{f}\|_0 \rightarrow \infty$ ,  $\mathbf{f} \in U_{ad}$ , then there is only one element that minimize  $J$  in  $U_{ad}$ , where  $U_{ad}$  is a closed and convex set in  $H$ .

Since the norm of the Hilbert space is strictly convex and continuous, hence the above theorem applies, therefore, there exists a unique function  $\mathbf{g}$  of the problem.

### 3 Optimality System

For numerical calculation of the optimal control we shall establish a more convenient formulation in terms of a system of differential equations. By the use of the Gateaux-differential of the cost functional  $J$ , the the problem (8) can be characterized by

$$\int_{\Omega} (\mathbf{u}(\mathbf{g}) - \mathbf{z}^o) \cdot (\mathbf{u}(\mathbf{v}) - \mathbf{u}(\mathbf{g})) dx + N \int_{\Omega} \mathbf{g} \cdot (\mathbf{v} - \mathbf{g}) dx \geq 0, \quad \forall \mathbf{v} \in U_{ad}. \quad (9)$$

In order to obtain a more convenient form for the numerical calculation of the function  $\mathbf{g}$ , we shall reformulate the problem through the adjoint system. Let the operator defined on the left hand side of the system (1) be denoted by  $A : (H_0^1(\Omega))^n \rightarrow (L^2(\Omega))^n$ ,

$$Au_i = -\frac{\partial}{\partial x_j} \left( C_{ijkl} \frac{\partial u_k}{\partial x_l} \right). \quad (10)$$

Since by assumption, the elasticity  $C_{ijkl}$  is symmetric, the operator  $A$  is self-adjoint, i.e.,

$$(\mathbf{A}\mathbf{u}, \mathbf{v})_0 = (\mathbf{u}, \mathbf{A}\mathbf{v})_0, \quad \forall \mathbf{u}, \mathbf{v} \in (H_0^1(\Omega))^n.$$

If we define the adjoint system by

$$\begin{cases} \mathbf{A}\mathbf{p} = \mathbf{u}(\mathbf{g}) - \mathbf{z}^o & \text{in } \Omega, \\ \mathbf{p} = 0 & \text{in } \partial\Omega, \end{cases} \quad (11)$$

then by substituting  $(\mathbf{u}(\mathbf{g}) - \mathbf{z}^o)$  from (11) into (9), we have

$$\int_{\Omega} \mathbf{A}\mathbf{p} \cdot (\mathbf{u}(\mathbf{v}) - \mathbf{u}(\mathbf{g})) + N \int_{\Omega} \mathbf{g} \cdot (\mathbf{v} - \mathbf{g}) dx \geq 0.$$

Since  $A$  is self-adjoint and  $\mathbf{A}\mathbf{u}(\mathbf{f}) = \mathbf{f}$  from the system (1), we have

$$\int_{\Omega} (\mathbf{p} + N\mathbf{g}) \cdot (\mathbf{v} - \mathbf{g}) dx \geq 0, \quad \forall \mathbf{v} \in U_{ad}. \quad (12)$$

Therefore, we can define the *optimality system* by

$$\begin{cases} \mathbf{A}\mathbf{u} = \mathbf{g} & \text{in } \Omega \\ \mathbf{A}\mathbf{p} = \mathbf{u} - \mathbf{z}^o & \text{in } \Omega \\ \mathbf{u} = \mathbf{p} = 0 & \text{in } \partial\Omega, \end{cases} \quad (13)$$

where the function  $\mathbf{g}$  must satisfy the restriction (12). We shall consider two examples for which the restriction can be determined explicitly.

**Example 1:** Let  $U_{ad} = (L^2(\Omega))^n$ . In this case the relation (12) is equivalent to

$$\int_{\Omega} (\mathbf{p} + N\mathbf{g}) \cdot \mathbf{v} \, dx \geq 0, \quad \forall \mathbf{v} \in (L^2(\Omega))^n,$$

which implies that

$$\mathbf{g} = -\frac{\mathbf{p}}{N}.$$

**Example 2:** Let  $U_{ad} = (L^2(\Omega)_+)^n$ . In this case the relation (12) is equivalent to

$$\int_{\Omega} (\mathbf{p} + N\mathbf{g}) \cdot (\mathbf{v} - \mathbf{g}) \, dx \geq 0, \quad \forall \mathbf{v} \geq 0 \in (L^2(\Omega))^n,$$

which implies that

$$\mathbf{g} = \max \left\{ -\frac{\mathbf{p}}{N}, 0 \right\} = \frac{\mathbf{p}^-}{N}.$$

where  $\mathbf{p}^-$  stands for the negative part of the function  $\mathbf{p}$ .

An algorithm to solve the optimal control problem for both examples is given below. In the first example the problem is linear and in the second example the problem is non-linear. In general for an arbitrary convex set the problem is always non-linear.

## 4 Iterative Method

In order to get a numerical solution for the problem, the optimality system is uncoupled in the following way: Given  $\mathbf{p}^0 = 0$  then the values of  $\mathbf{u}^m$  and  $\mathbf{p}^m$  are iteratively calculated from the following algorithm:

$$\begin{cases} A\mathbf{u}^m = F(\mathbf{p}^{m-1}) & \text{in } \Omega, \\ A\mathbf{p}^m = \mathbf{u}^m - \mathbf{z}^o & \text{in } \Omega, \\ \mathbf{u}^m = \mathbf{p}^m = 0 & \text{in } \partial\Omega, \end{cases} \quad (14)$$

where  $F(\mathbf{p})$  stands for  $-\mathbf{p}/N$  or  $\mathbf{p}^-/N$ . In both cases  $F$  is a Lipschitz function, with constant  $1/N$ . We shall prove in the following theorem that the algorithm above is convergent if  $N$  is not too small.

**Theorem 1.** *There exists a positive constant  $\delta$ , such that for  $N > \delta$ ,*

$$\{\mathbf{p}^m, \mathbf{u}^m\} \rightarrow \{\mathbf{p}, \mathbf{u}\} \text{ strongly in } (H_0^1(\Omega))^n \cap (H^2(\Omega))^n,$$

where  $\mathbf{p}$  and  $\mathbf{u}$  are solutions of the optimality system (13).

**Proof.** With the notation,

$$\mathbf{P}^m = \mathbf{p}^m - \mathbf{p} \quad \text{and} \quad \mathbf{U}^m = \mathbf{u}^m - \mathbf{u},$$

from (13) and (14) we have

$$\begin{cases} AU^m = F(\mathbf{p}^{m-1}) - F(\mathbf{p}) & \text{in } \Omega, \\ AP^m = U^m & \text{in } \Omega, \\ U^m = P^m = 0 & \text{in } \partial\Omega. \end{cases} \quad (15)$$

Taking inner product of the first equation with  $U^m$  and the second with  $P^m$  and integrating over  $\Omega$ , we obtain

$$\int_{\Omega} AU^m \cdot U^m dx = \int_{\Omega} (F(\mathbf{p}^{m-1}) - F(\mathbf{p})) \cdot U^m dx,$$

$$\int_{\Omega} AP^m \cdot P^m dx = \int_{\Omega} U^m \cdot P^m dx.$$

After integration by parts, we get by the use of (4)

$$\sigma(U^m, U^m) = \int_{\Omega} (F(\mathbf{p}^{m-1}) - F(\mathbf{p})) \cdot U^m dx, \quad (16)$$

$$\sigma(P^m, P^m) = \int_{\Omega} U^m \cdot P^m dx. \quad (17)$$

Since  $F$  is a Lipschitz function, it follows from (16) that

$$\int_{\Omega} ((F(\mathbf{p}^{m-1}) - F(\mathbf{p})) \cdot U^m dx \leq \frac{1}{2N} \left\{ \int_{\Omega} (|\mathbf{P}^{m-1}|^2 + |U^m|^2) \right\} dx,$$

and by applying the elementary inequality,  $ab \leq \frac{1}{2}(\alpha a^2 + b^2/\alpha)$ , where  $\alpha$  is real positive number, we obtain from (17) that

$$\int_{\Omega} U^m \cdot P^m dx \leq \frac{1}{2} \left\{ \alpha \int_{\Omega} |U^m|^2 + \frac{1}{\alpha} \int_{\Omega} |P^m|^2 \right\} dx.$$

Substituting the above relations respectively into (16) and (17) and taking the energy norm, we have

$$\|U^m\|_{\sigma}^2 \leq \frac{1}{2N} \left\{ \|\mathbf{P}^{m-1}\|_0^2 + \|U^m\|_0^2 \right\},$$

$$\|P^m\|_{\sigma}^2 \leq \frac{1}{2} \left\{ \alpha \|U^m\|_0^2 + \frac{1}{\alpha} \|P^m\|_0^2 \right\}.$$

Since the norms  $\|\cdot\|_{\sigma}$  and  $\|\cdot\|_1$  are equivalent in  $(H_0^1(\Omega))^n$ , by (5) we have

$$C^2 \|U^m\|_1^2 \leq \frac{1}{2N} \left\{ \|\mathbf{P}^{m-1}\|_0^2 + \|U^m\|_0^2 \right\} \leq \frac{K}{2N} \left\{ \|\mathbf{P}^{m-1}\|_1^2 + \|U^m\|_1^2 \right\},$$

$$C^2 \|P^m\|_1^2 \leq \frac{1}{2} \left\{ \alpha \|U^m\|_0^2 + \frac{1}{\alpha} \|P^m\|_0^2 \right\} \leq \frac{K}{2} \left\{ \alpha \|U^m\|_1^2 + \frac{1}{\alpha} \|P^m\|_1^2 \right\}.$$

In the second part of the above relations, we have used the Poincaré inequality, where the constant  $K$  depends on  $\Omega$  only. Hence we have

$$(2NC^2 - K) \|U^m\|_1^2 \leq K \|\mathbf{P}^{m-1}\|_1^2,$$

$$(2\alpha C^2 - K)\|\mathbf{P}^m\|_1^2 \leq K\alpha^2\|\mathbf{U}^m\|_1^2.$$

Suppose that  $\{N, \alpha\} \geq r/2$ , where  $r = K/C^2$ , then by combining the above two inequalities we obtain

$$\|\mathbf{P}^m\|_1^2 \leq \gamma\|\mathbf{P}^{m-1}\|_1^2, \quad (18)$$

where

$$\gamma = \frac{\alpha^2 r^2}{(2\alpha - r)(2N - r)}.$$

The inequality (18) is valid for  $m = 1, 2, \dots$ , so we have

$$\|\mathbf{P}^m\|_1^2 \leq \gamma^m\|\mathbf{P}^0\|_1^2. \quad (19)$$

Since  $\mathbf{P}^0 \in (H_0^1(\Omega))^n$ , if  $\gamma < 1$  we conclude that as  $m \rightarrow \infty$ , the sequence  $\{\mathbf{P}^m\}$  converges strongly to zero in  $(H_0^1(\Omega))^n \cap (H^2(\Omega))^n$ , by the use of elliptic regularity. The convergence of the sequence  $\{\mathbf{U}^m\}$  also follows in exactly the same argument.

To conclude the proof of the theorem, we need to determine the constants  $\alpha$  and  $N$  so that the condition  $\gamma < 1$  is satisfied. Moreover, the value of  $\alpha$  will be chosen in such a way that the lower bound for  $N$  is as small as possible. Since  $\alpha > r/2$ , let  $\alpha = r/2 + \varepsilon$  for  $\varepsilon > 0$ , then the condition  $\gamma < 1$  is equivalent to

$$N > \frac{r}{2}\left(1 + \frac{r}{2\varepsilon}\left(\frac{r}{2} + \varepsilon\right)^2\right).$$

Let the right hand side be denoted by  $\delta$ , which takes its minimal value at  $\varepsilon = r/2$ . With this choice, we have  $\alpha = r$  and

$$N > \delta = \frac{r}{2}(1 + r^2), \quad r = \frac{K}{C^2}, \quad (20)$$

which ensures the condition  $\gamma < 1$  and the theorem is proved.  $\square$

**Remark.** *On the value of  $\delta$  :*

By the assumptions (2) and (3) on the elasticity tensor  $C_{ijkl}$ , the elliptic operator  $A$  defined in (10) is self-adjoint and positive definite. Hence by the spectral theorem, the eigenvectors of  $A$  form an orthonormal basis of the Hilbert space  $(H_0^1(\Omega))^n \cap (H^2(\Omega))^n$ , and the eigenvalues  $\{\lambda_m\}$  form an increasing sequence of positive real numbers, i.e.,  $0 < \lambda_m \leq \lambda_{m+1}$  for  $m = 1, 2, \dots$ . Moreover, one can show that

$$\lambda_1 = \inf_{\mathbf{u} \neq 0} \frac{\|\mathbf{u}\|_\sigma^2}{\|\mathbf{u}\|_0^2}.$$

Hence, it follows from (5) that

$$\|\mathbf{u}\|_0^2 \leq \frac{1}{\lambda_1}\|\mathbf{u}\|_\sigma^2 \leq \frac{\bar{C}}{\lambda_1}\|\mathbf{u}\|_1^2. \quad (21)$$

In order words, the constant  $K$  in the Poincaré inequality can be taken as  $\bar{C}/\lambda_1$ , and we arrive at an estimated value of  $\delta$ ,

$$N > \delta = \frac{\bar{C}}{2C^2\lambda_1} \left(1 + \frac{\bar{C}}{C^4\lambda_1^2}\right).$$

In particular, for  $A = -\Delta$ , the Laplace operator, and  $\Omega = (0, 1) \times (0, 1)$ , we have  $C = \bar{C} = 1$  and  $\lambda_1 = 2\pi^2$ , hence the constant  $N$  can be chosen as small as 0.0254 and since the above estimate is not optimal the lower limit of  $\delta$  can be even smaller than this value as we shall see in the numerical example later.

## 4.1 Finite Element Approximation

Let  $V^h$  be the finite dimensional subspace of  $(H_0^1(\Omega))^n$  in the finite element approximation with maximum mesh size  $h$ . The Galerkin formulation of the iterative problem (14) is given as follows:

Given  $\mathbf{z}^0 \in (L^2(\Omega))^n$  and  $\mathbf{p}_h^0 = 0$ , find  $\mathbf{u}_h^m \in V^h$  and  $\mathbf{p}_h^m \in V^h$  such that for all  $\mathbf{v}_h \in V^h$ ,

$$\begin{cases} \sigma(\mathbf{u}_h^m, \mathbf{v}_h) = (F(\mathbf{p}_h^{m-1}), \mathbf{v}_h)_0, \\ \sigma(\mathbf{p}_h^m, \mathbf{v}_h) = (\mathbf{u}_h^m - \mathbf{z}^0, \mathbf{v}_h)_0. \end{cases} \quad (22)$$

By employing finite element basis functions in  $V^h$ , (22) are systems of linear algebraic equations.

In Theorem 1, we have proved for  $N > \delta$  that the numerical solution  $\{\mathbf{u}^m, \mathbf{p}^m\}$  for uncoupled system (14) converges to the solution  $\{\mathbf{u}, \mathbf{p}\}$  of the optimality system. In the next theorem, we shall prove that the numerical solution of system (22),  $\{\mathbf{u}_h^m, \mathbf{p}_h^m\}$ , obtained by the finite element method also converges to  $\{\mathbf{u}, \mathbf{p}\}$ .

**Theorem 2.** *Let  $N > \delta$ , then for  $m \rightarrow \infty$  and  $h \rightarrow 0$ ,*

$$\{\mathbf{u}_h^m, \mathbf{p}_h^m\} \longrightarrow \{\mathbf{u}, \mathbf{p}\} \text{ in } (H_0^1(\Omega))^n \cap (H^2(\Omega))^n,$$

where  $\delta$  is the constant defined in Theorem 1.

**Proof.** Using a triangular inequality, we have

$$\|\mathbf{u}_h^m - \mathbf{u}\|_1 \leq \|\mathbf{u}^m - \mathbf{u}\|_1 + \|\mathbf{u}_h^m - \mathbf{u}^m\|_1, \quad (23)$$

$$\|\mathbf{p}_h^m - \mathbf{p}\|_1 \leq \|\mathbf{p}^m - \mathbf{p}\|_1 + \|\mathbf{p}_h^m - \mathbf{p}^m\|_1. \quad (24)$$

For the last two terms of the right hand side of (23) and (24), it is known in numerical analysis [2, 8, 10] that since  $\mathbf{u}^m$  and  $\mathbf{p}^m$  are solutions in  $(H_0^1(\Omega))^n \cap (H^2(\Omega))^n$  of the optimality system, we have as a consequence that

$$\|\mathbf{u}^m - \mathbf{u}_h^m\|_1 \leq C_1 h \|\mathbf{u}^m\|_2 \leq C_2 h \|F(\mathbf{p}^{m-1})\|_0 \leq C_3 h \|F(\mathbf{p})\|_0, \quad (25)$$

$$\|\mathbf{p}^m - \mathbf{p}_h^m\|_1 \leq \bar{C}_1 h \|\mathbf{p}^m\|_2 \leq \bar{C}_2 h \{\|\mathbf{u}^m - \mathbf{u}\|_0 + \|\mathbf{u} - \mathbf{z}^0\|_0\}, \quad (26)$$

where  $C_1, C_2, C_3, \bar{C}_1$  and  $\bar{C}_2$  are positive constants and independent from  $h$  and  $F$ . Substituting (25) and (26) into (23) and (24) respectively, and applying Theorem 1, we conclude that

$$\{\mathbf{u}_h^m, \mathbf{p}_h^m\} \longrightarrow \{\mathbf{u}, \mathbf{p}\} \text{ in } (H_0^1(\Omega))^n \cap (H^2(\Omega))^n.$$

for  $m \rightarrow \infty$  and  $h \rightarrow 0$ .  $\square$

## 5 Eigenfunction Expansion Method

In the previous section, we have proposed an iterative method for which the approximate solution  $\{\mathbf{u}_h^m, \mathbf{p}_h^m\}$  converges strongly to the solution  $\{\mathbf{u}, \mathbf{p}\}$  in  $(H_0^1(\Omega))^n \cap (H^2(\Omega))^n$ , provided that the constant  $N > \delta$ , meaning,  $N$  is not allowed to be arbitrary small. In our numerical calculations, we have found that such a restriction on  $N$  could be quite unsatisfactory in practical solutions. In order to find the solution of the optimality system free from such a restriction, in the following, we shall analyze the problem via the eigenfunction expansion method for the case  $U_{ad} = (L^2(\Omega))^n$ .

Since the operator  $A$  is positive-definite and self-adjoint defined in the Hilbert space  $H = (H_0^1(\Omega))^n \cap (H^2(\Omega))^n$ . By the spectral theorem,  $H$  admits a complete orthonormal basis of eigenfunctions  $\{\varphi_m\}$  of  $A$  and the corresponding eigenvalues  $\lambda_m$  can be arranged in ascending order,

$$0 < \lambda_m \leq \lambda_{m+1}, \quad m = 1, 2, 3, \dots$$

Consequently,  $\{\mathbf{u}, \mathbf{p}\} \in H$  can be expressed in an eigenfunction expansion of the form:

$$\mathbf{u}(x) = \sum_{m=1}^{\infty} u_m \varphi_m(x), \quad \mathbf{p}(x) = \sum_{m=1}^{\infty} p_m \varphi_m(x). \quad (27)$$

Therefore, we have

$$A\mathbf{u}(x) = \sum_{m=1}^{\infty} u_m \lambda_m \varphi_m(x), \quad A\mathbf{p}(x) = \sum_{m=1}^{\infty} p_m \lambda_m \varphi_m(x). \quad (28)$$

Substituting into the equation (13)<sub>1</sub>, we obtain

$$\sum_{m=1}^{\infty} u_m \lambda_m \varphi_m(x) = -\frac{1}{N} \sum_{m=1}^{\infty} p_m \varphi_m(x),$$

or

$$\sum_{m=1}^{\infty} \left( u_m \lambda_m + \frac{p_m}{N} \right) \varphi_m(x) = 0,$$

which implies that

$$p_m = -u_m \lambda_m N. \quad (29)$$

Similarly, from the equation (13)<sub>2</sub>, we obtain

$$\sum_{m=1}^{\infty} p_m \lambda_m \varphi_m(x) = \sum_{m=1}^{\infty} (u_m - z_m) \varphi_m(x),$$

or

$$\sum_{m=1}^{\infty} (p_m \lambda_m - (u_m - z_m)) \varphi_m(x) = 0,$$

where  $z_m$  are the Fourier coefficients in the eigenfunction expansion of the given objective function  $\mathbf{z}^0 \in (L^2(\Omega))^n$ :

$$\mathbf{z}^0(x) = \sum_{m=1}^{\infty} z_m \varphi_m(x). \quad (30)$$

Hence, we have

$$u_m = z_m + p_m \lambda_m. \quad (31)$$

From (29) and (31), we obtain the Fourier coefficients  $u_m$  and  $p_m$ ,

$$u_m = \frac{z_m}{N\lambda_m^2 + 1}, \quad p_m = -\frac{N\lambda_m z_m}{N\lambda_m^2 + 1},$$

and the solutions  $\mathbf{u}(x)$  and  $\mathbf{p}(x)$  of the optimality system (13) are given explicitly by

$$\mathbf{u}(x) = \sum_{m=1}^{\infty} \frac{z_m}{N\lambda_m^2 + 1} \varphi_m(x), \quad (32)$$

$$\mathbf{p}(x) = -\sum_{m=1}^{\infty} \frac{N\lambda_m z_m}{N\lambda_m^2 + 1} \varphi_m(x). \quad (33)$$

From these solutions we can easily see that when  $N \rightarrow 0$  the function  $\mathbf{p}(x)$  tends to zero, while the solution  $\mathbf{u}(x)$  converges to the objective function  $\mathbf{z}^0(x)$ . Moreover, by (32) it follows that  $\|\mathbf{u}\|_0 < \|\mathbf{z}^0\|_0$  for any  $N > 0$ .

On the other hand, the optimal control  $\mathbf{g} \in (L^2(\Omega))^n$  admits a representation in Fourier series, and since  $\mathbf{g} = -\mathbf{p}/N$  we obtain

$$\mathbf{g}(x) = \sum_{m=1}^{\infty} \frac{\lambda_m z_m}{N\lambda_m^2 + 1} \varphi_m(x), \quad (34)$$

which tends to the expected solution given by the eigenfunction expansion of  $A\mathbf{z}^0(x)$  in the limit as  $N \rightarrow 0$ .

## 6 Numerical Results

Due to the difficulties in the determination of eigenvalues  $\lambda_m$  and eigenfunctions  $\varphi_m$  for the operator  $A$ , for simplicity, we shall consider the Laplace operator

$$A = -\Delta$$

in the unit square  $\Omega = (0.1) \times (0.1)$  as an example. Mathematically, it is a very special case of the linear elasticity operator, in which the components of the equilibrium equation become independent of each other and the problem (1) can be regarded as two

independent problems of scalar equations for the individual components. But for the purpose of examining the qualitative behavior of the optimal control solutions, our numerical calculation, both in finite element iterative approximation and in Fourier series expansion, will be illustrated for the case of Laplace operator. For more general elliptic operators or more general domains, a numerical construction of eigenfunctions for spectral representations will be presented in the future.

Since  $u = p = 0$  on  $\partial\Omega$  the eigenvalues and the eigenfunctions of the Laplace operator are well-known and they are given by

$$\lambda_{mn} = (n^2 + m^2)\pi^2, \quad \varphi_{mn} = \sin m\pi x \sin n\pi y.$$

The objective function can then be represented by

$$z^0(x, y) = \sum_{m,n=1}^{\infty} z_{mn} \sin m\pi x \sin n\pi y.$$

Substituting into (32) and (34), we have the approximation of the objective function,

$$u(x, y) = \sum_{m,n=1}^{\infty} \frac{z_{mn}}{N(m^2 + n^2)^2\pi^4 + 1} \sin m\pi x \sin n\pi y,$$

and the function of optimal control,

$$g(x, y) = \sum_{m,n=1}^{\infty} \frac{(m^2 + n^2)\pi^2 z_{mn}}{N(m^2 + n^2)^2\pi^4 + 1} \sin m\pi x \sin n\pi y.$$

The exact solution of optimal control is then given by the function in the limit when  $N \rightarrow 0$ ,

$$g(x, y) = \sum_{m,n=1}^{\infty} (m^2 + n^2)\pi^2 z_{mn} \sin m\pi x \sin n\pi y.$$

For numerical calculations, we consider a prescribed objective function given by

$$z^0(x, y) = (16xy(1-x)(1-y)(x-y^2))^3.$$

In Fig. 1 and Fig. 2, the objective function  $z^0(x, y)$  and the exact optimal control  $g(x, y)$  given above are shown. In finite element approximation, we have taken a mesh of  $15 \times 15$  square elements. Remember that there is a lower limit of the parameter  $N$ . In the present case we have found that this limit is approximately equal to  $5 \times 10^{-3}$  (Note that it is much smaller than the estimated value given before) for which convergence is ensured in 20 iterations. Convergence is much faster for greater values of  $N$ . However, from Fig. 3 and Fig. 4, we can see that with this smallest allowable value of  $N$ , the numerical results are still quite far from a good approximation to the exact solutions by comparing the scales shown in the graphics. It is obvious from the graphics that  $\|u\|_0$  is much too small compared to the expected value  $\|z^0\|_0$ .

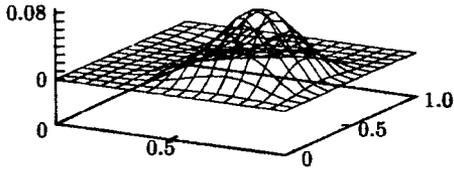


Fig. 1. Objective function  $z^0(x, y)$

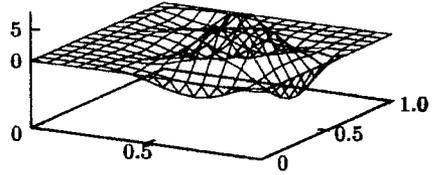


Fig. 2. Exact optimal control  $g(x, y)$

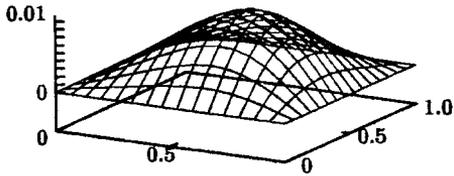


Fig. 3.  $u(x, y)$  for  $N = 0.005$

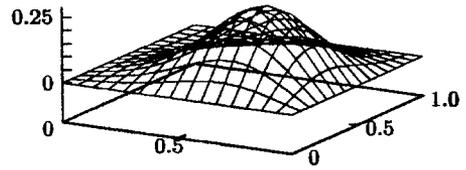


Fig. 4.  $g(x, y)$  for  $N = 0.005$

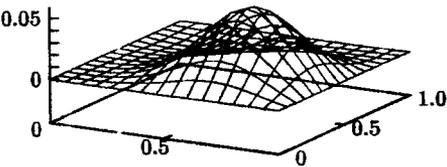


Fig. 5.  $u(x, y)$  for  $N = 0.00005$

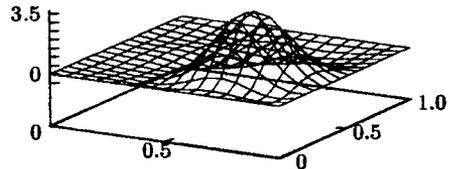


Fig. 6.  $g(x, y)$  for  $N = 0.00005$

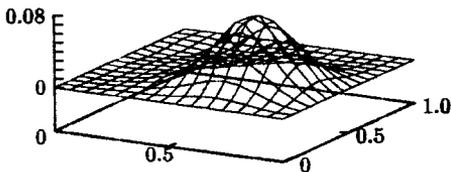


Fig. 7.  $u(x, y)$  for  $N = 0$

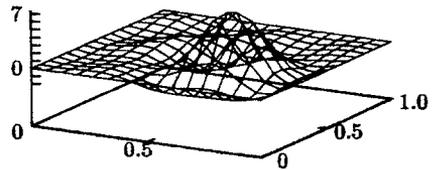


Fig. 8.  $g(x, y)$  for  $N = 0$

On the other hand, there is no restriction on the value of  $N$  for the method of Fourier series expansion. Approximation by Fourier series is calculated by a sum of 22 terms in eigenfunctions  $\varphi_{mn}$  for  $m^2 + n^2 < 6^2$ . Numerical solutions are obtained for values of  $N = 5 \times 10^{-3}$ ,  $5 \times 10^{-5}$  and also for  $N = 0$ . which represents the exact solution to the problem. The figures for  $N = 5 \times 10^{-3}$  are not shown here because they are practically identical to Figures 3 and 4 for the results of iterative approximation. Fig. 5, Fig. 6 and Fig. 7, Fig. 8 show the graphics for  $N = 5 \times 10^{-5}$  and  $N = 0$  respectively. We can see the gradual improvement of the approximation for decreasing values of  $N$  and an excellent agreement with the exact solutions shown in Fig. 1 and Fig. 2 for the case of  $N = 0$ , for which the series are calculated with a finite sum of the first 22 terms only.

**Acknowledgments:** *The author (ISL) acknowledges the partial support for research from CNPq of Brazil. The computer lab provided by FUJB is also appreciated.*

## References

- [1] Chou, S.-I., Wang, C.-C.: Error estimates of finite element approximations for problems in linear elasticity, Part I. Problems in elastostatics, Arch. Rational Mech. Anal., 72, 41-60 (1979).
- [2] Ciarlet, P. G.: The Finite Element Method for Elliptic Problems, North Holand, Amsterdam (1987).
- [3] Fichera, G.: Existence Theorems in Elasticity, in Handbuch der Physik, Band VIa/2, Edited by C. Truesdell, Springer-Verlag (1972).
- [4] Gill, S. I., Murray, W., Wright, M. H.: Practical Optimization, Academic Press (1988).
- [5] Haslinger, J.; Neittaanmäki, P.: Finite Element Approximation for Optimal Shape Design: Theory and Applications, John Wiley & sons (1988).
- [6] Lions, J. L.: Optimal control of systems governed by partial differential equations. Springer-Verlag, New York (1972).
- [7] Neittaanmäki, P., Tiba, D.: Optimal Control of Nonlinear Parabolic Systems, Marcel Dekker, Inc. New York (1994).
- [8] Oden, J. T., Reddy, J. N.: Variational Methods in Theoretical Mechanics, Springer-Verlag (1976).
- [9] Pironneau, O.: Optimal Shape Design for Elliptic Systems, Springer-Verlag, Berlin (1984).
- [10] Strang, G., Fix, G. J.: An Analysis of the Finite Element Method, Prentice-Hall (1973).