Journal of mathematics and computer science 9(2014), 203-215

Contents list available at JMCS
Journal of Mathematics and Computer Science
Journal Homepage: www.tjmcs.com

Convergence analysis of gradient based iterative algorithm for solving PDE constrained optimization problems

Rasoul Naseri, A. Malek[1]
Department of Applied Mathematics, Faculty of Mathematical Sciences, Tarbiat Modares University, P. O. Box 14115-134, Tehran, Iran
(naserir@modares.ac.ir & mala@modares.ac.ir)

Article history:
Received August 2013
Accepted November 2013
Available online November 2013

Abstract

In this paper, by considering distributed optimal control over a PDE, a gradient based iterative Algorithm is proposed for solving is proposed and analyzed. Galerkin finite element method is used for solving underlying PDE, then the adjoint base technique for derivative computation to implementation of the optimal control issue in preconditioned Newton's conjugate gradient method isused.

The interface and connection between quadratic programming extracted from discretizing the problem and Newton's type method, as well as the convergence rate of the algorithm in each iteration is established.

Updating control values at discretization points in each iteration yields optimal control of the problem, where the corresponding state values at these points approximate the desired function. Numerical experiments are presented for illustrating the theoretical results.

Keywords: Diffusion equation, optimal control problem, finite element method, Newton's conjugate gradient method.

1 Introduction

Optimization problems, constraints with partial differential equation (PDE), arise in many areas such as mathematical finance [2, 3, 4], aerodynamics [13, 15], environmental engineering [11] and medicine [1, 10] and generally are infinite dimensional, large and complex. In order to solve a PDE-constrained optimization problem, the question about should I first discretize the optimization problem and then solve the discretized optimization problem (DO), or should first I optimize the continuous problem and obtain a set of equations to discretize (OD), is not avoidable. An important challenge in optimization problems is that, these two steps do not

[1]Corresponding author.

E-mail address: mala@modares.ac.ir.

commute. Thus, the two different approaches could lead to two different solutions. Our intention is to proposean efficient DO algorithm. The advantages and disadvantages of both approaches are summarized by Gunzburger in [6]. With OD approach, one can obtain inconsistent gradients of the objective functional. In other words, unless the grid is fine enough, the approximate gradient obtained with OD is not a true gradient [7]. In this paper the specific DO approach is used as follows: At first, discretization of the underlying PDE is done via Galerkin finite element.Since the discretized system of equation is sparse and so large, preconditioned Newton's conjugate gradient approach, is a good choice to solve the governed equation efficiently. In this way the control vector is computed, where the corresponding solution of PDE is nearest approximation of desired function and has low cost in the objective function [8]. The relation between quadratic programming extracted from discretizing the optimal control problem over PDE, and Newton's type method is established. It is shown that under certain conditions the algorithm is strictly convergent to its optimal vector. According to the above discussion, in Section 2, problem definition, derivatives computation and Newton's conjugate gradient method is discussed. In Section 3, distributed optimal control for elliptic equation using Galerkin finite element approach is proposed. Finally in Section 4, numerical results for some optimal control problems are presented.

2 Problem formulation

We consider optimal control problems of the form

$$min_{y\in Y,u\in U} \; J(y,u) \text{ subject to } e(y,u)=0, \quad (y,u)\in W_{ad} \tag{1}$$

where$J: Y\times U \to R$is the objective function, $e: Y\times U \to Z$is an operator between Banach spaces, and $W_{ad}\subset W := Y\times U$ is a nonempty closed set. Existence and uniqueness of the solution to these problems are ensured by implicit function theorem [16].It is considered that J and e are continuously F-differentiable and for each $u\,\epsilon\, U$ the state equation $e\,(y,u)=0$ possesses a unique corresponding solution $y(u)\,\epsilon\, Y$. So, in fact we have a solution operator $u\,\epsilon\, U\to y\,(u)\,\epsilon\, Y$. Furthermore, it is assumed that $e_y(y\,(u),u)\,\epsilon\,\mathcal{L}\,(Y,Z)$ is continuously invertible. Then, the continuously differentiability of $y(u)$ is ensured by implicit function theorem.
Differentiating the equation $e(y(u),u)=0$ with respect to u yields an equation for the derivative $y'(u)$:

$$e_y(y\,(u),u)y'(u)+e_u(y\,(u),u)=0 \tag{2}$$

Inserting $y(u)$ in (1), the reduced problem becomes

$$min_{u\epsilon U}\; \hat{J}(u) := J(y(u),u) \text{ subject to } u\,\epsilon\, \widehat{U}_{ad} := \{u\,\epsilon\, U:\ (y(u),u)\,\epsilon\, W_{ad}\} \tag{3}$$

It will be important to investigate the possibilities of computing the derivative of the reduced objective function$\hat{J}$.

2.1 Gradient and Hessian Computation

In order to investigate the gradient based optimization technique, computation of the derivatives of the reduced objective function $\hat{J}$ is not avoidable. In whole of this paperU^*, $Y^*$and Z^* denote the dual space of U, $Y$and Z respectively. For any $s \in U$, from

$$< \hat{J}'(u), s >_{U^*,U} = < J_y(y(u),u), y'(u)\, s >_{Y^*,Y} + < J_u(y(u),u), s >_{U^*,U}$$

$$= < y'(u)^T J_y(y(u),u), s >_{U^*,U} + < J_u(y(u),u), s >_{U^*,U}$$

we see that

$$\hat{J}'(u) = y'(u)^T J_y(y(u),u) + J_u(y(u),u).$$

Therefore, not the operator $y'(u) \in \mathcal{L}(U,Y)$, but only the vector $y'(u)^T J_y(y(u),u) \in U^*$ is ready required. Since by (2)

$$y'(u)^T J_y(y(u),u) = -e_u(y(u),u)^T e_y(y(u),u)^{-T} J_y(y(u),u),$$

it follows that

$$y'(u)^T J_y(y(u),u) = e_u(y(u),u)^T\, p(u),$$

where the adjoint state $p = p(u) \in Z^*$ solves the adjoint equation:

$$e_y(y(u),u)^T\, p = -J_y(y(u),u). \quad (4)$$

We thus have

$$\hat{J}'(u) = e_u(y(u),u)^T\, p(u) + J_u(y(u),u).$$

Now, consider (1) and define the Lagrange function $L: Y \times U \times Z^* \to R$,

$$L(y,u,p) = J(y,u) + < p, e(y,u) >_{Z^*,Z}$$

Inserting $y = y(u)$gives, for arbitrary$p \in Z^*$,

$$\hat{J}(u) = J(y(u),u)$$

$$= J(y(u),u) + < p, e(y(u),u) >_{Z^*,Z}$$

$$= L(y(u),u,p).$$

Differentiating this in the direction $S_1 \in U$ yields:

$$< \hat{J}'(u), S_1 >_{U^*,U} = < L_y(y(u),u,p), y'(u)S_1 >_{Y^*,Y} + < L_u(y(u),u,p), S_1 >_{U^*,U} \;,$$

Differentiating this once again in the direction $S_2 \in U$ gives:

$$< \hat{J}''(u)S_2, S_1 >_{U^*,U} = < L_y(y(u),u,p(u)) \,, y''(u)(S_1,S_2) >_{Y^*,Y}$$

$$+< L_{yy}(y(u),u,p(u))\, y'(u)S_2, y'(u)S_1 >_{Y^*,Y}$$

$$+< L_{yu}(y(u),u,p(u))S_2, y'(u)S_1 >_{Y^*,Y}$$

$$+< L_{uy}(y(u),u,p(u))\, y'(u)S_2, S_1 >_{U^*,U}$$

$$+< L_{uu}\big(y(u),u,p(u)\big)S_2, S_1 >_{U^*,U}$$

Now we choose $p = p(u)$, i.e., $L_y(y(u),u,p) = 0$. Then, the term containing $y''(u)$ drops out and we arrive at

$$< \hat{J}''(u)S_2, S_1 >_{U^*,U} = < L_{yy}(y(u),u,p(u))\, y'(u)S_2, y'(u)S_1 >_{Y^*,Y}$$

$$+< L_{yu}(y(u),u,p(u))S_2, y'(u)S_1 >_{Y^*,Y}$$

$$+< L_{uy}(y(u),u,p(u))\, y'(u)S_2, S_1 >_{U^*,U}$$

$$+< L_{uu}\big(y(u),u,p(u)\big)S_2, S_1 >_{U^*,U}$$

This shows that

$$\begin{aligned}\hat{J}''(u) &= y'(u)^T L_{yy}\big(y(u),u,p(u)\big)\, y'(u) + y'(u)^T L_{yu}(y(u),u,p(u)) \\ &\quad + L_{uy}\big(y(u),u,p(u)\big)\, y'(u) + L_{uu}\big(y(u),u,p(u)\big) \\ &= T(u)^T L_{\omega\omega}\big(y(u),u,p(u)\big)\, T(u)\end{aligned} \tag{5}$$

with

$$T(u) = \begin{pmatrix} y'(u) \\ I_U \end{pmatrix} \in \mathcal{L}(U, Y \times U)$$

$$L_{\omega\omega} = \begin{pmatrix} L_{yy} & L_{yu} \\ L_{uy} & L_{uu} \end{pmatrix}$$

Here, $I_U \in \mathcal{L}(U,U)$ is the identity. Note that by (2), $y^{'}(u) = -e_y(y(u),u)^{-1} e_u(y(u),u)$ and thus,

$$T(u) = \begin{pmatrix} y'(u) \\ I_U \end{pmatrix} = \begin{pmatrix} -e_y(y(u),u)^{-1} e_u(y(u),u) \\ I_U \end{pmatrix}$$

Usually, the Hessian representation (5) is not used to compute the whole operator $\hat{J}''(u)$. Rather, it is used to compute operator-vector-products $\hat{J}''(u)s$ that investigates the iterative solvers applied to the Newton's equation

$$\hat{J}^{''}(u^k)s^k = -\hat{J}^{'}(u^k). \tag{6}$$

The Newton's Eq. (6) is solved approximately using the conjugate gradient (CG) method. The CG method is truncated if the Newton system residual is sufficiently small. In practice some globalization technique for Newton's method should be employed. Line search techniques are popular choices for their case of implementation and relatively low computational cost. A line search algorithm attempts to find an optimal step size α_k and generates the iterate $u^{k+1} = u^k + \alpha^k S^k$. The step size is required to satisfy the Armijo condition (or sufficient decreas condition)

$$J(u^k + \alpha^k S^k) \leq J(u^k) + c\alpha^k J'(u^k)S^k$$

where $c \in (0,1)$ and is typically quite small, e.g. $c = 10^{-4}$ [9,12].

Algorithm.

1. Given$u^0$and$gtol > 0$, set$k = 0$
2. Compute the adjoint state by solving the adjoint equation

$$e_y(y(u),u)^* p = -J_y(y(u),u).$$

3. Compute$\hat{J}'(u)$ via

$$\hat{J}'(u) = e_u(y(u),u)^* p + J_u(y(u),u).$$

4. Compute the derivative:

$$y'(u)s = -e_y(y(u),u)^{-1} e_u(y(u),u)S$$

5. Compute

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} L_{yy}(y(u),u,p(u))\, y'(u)s + L_{yu}(y(u),u,p(u))s \\ L_{uy}(y(u),u,p(u))\, y'(u)s + L_{uu}\big(y(u),u,p(u)\big)s \end{pmatrix}.$$

6. Compute

$$h_3 = y'(u)^* h_1 = -e_u(y(u),u)^* e_y(y(u),u)^{-*} h_1$$

This requires an adjoint equation solve.

7. Set$\hat{J}''(u)s = h_2 + h_3.$
8. If$\|\nabla \hat{J}(u^k)\| < gtol$ stop.
9. Compute$\nabla^2 \hat{J}(u^k)$
10. Solve $\hat{J}''(u^k)s^k = -\hat{J}'(u^k)$using preconditioned CG method with preconditioner

$P = diag(\hat{J}''(u^k))$:

a. Set $x_0 = 0$ and $A = \hat{J}''(u^k)$ and $b = -\hat{J}'(u^k)$.
b. Set $r_0 \leftarrow Ax_0 - b$;
c. Solve $Py_0 = r_0$ for y_0;
d. Set $p_0 = -r_0$, $t \leftarrow 0$;
e. While $\|r_t\| > gtol$
   i. $\alpha_k \leftarrow \frac{r_t^T y_t}{p_t^T A p_t}$
   ii. $x_{t+1} \leftarrow x_t + \alpha_t p_t$
   iii. $r_{t+1} \leftarrow r_t + \alpha_t A p_t$;

iv. $Py_{t+1} \leftarrow r_{t+1}$;

v. $\beta_{t+1} \leftarrow \frac{r_{t+1}^T y_{t+1}}{r_t^T y_t}$;

vi. $p_{t+1} \leftarrow -y_{t+1} + \beta_{t+1} p_t$;

vii. $t \leftarrow t+1$;

End (while)

f. $s^k = x_{t+1}$

11. Perform Armijo line-search.

a. Set$\alpha^k = 1$andevaluate$J(u^k + \alpha^k S^k)$

b. While$J(u^k + \alpha^k S^k) > J(u^k) + 10^{-4} \alpha^k S^k \nabla \hat{J}(u^k)$ *do*

i. *Set* $\alpha^k = \alpha^k / 2$andevaluate$J(u^k + \alpha^k S^k)$.

12. Set$u^{k+1} = u^k + \alpha^k S^k$, $k \leftarrow k+1$. Goto *2.*

In the following, the connection between the Newton equation $\nabla^2 \hat{J}''(u)\, s = -\nabla \hat{J}'(u)$ and the solution of a quadratic program is established.

Theorem 2.Let $e_y(y(u), u)$ be invertible and let $\nabla^2 \hat{J}(u)$be symmetric positive semidefinite.

The Newton equation $\nabla^2 \hat{J}(u) s_u = -\nabla \hat{J}(u)$ is solved by the vectors_u if and only if solves the quadratic program:

$$min \begin{pmatrix} \nabla_y J(y,u)^T \\ \nabla_u J(y,u) \end{pmatrix}^T \begin{pmatrix} s_y \\ s_u \end{pmatrix} + \frac{1}{2} \begin{pmatrix} s_y \\ s_u \end{pmatrix}^T \begin{pmatrix} \nabla_{yy} \; L(y,u,p) & \nabla_{yu} \; L(y,u,p) \\ \nabla_{uy} \; L(y,u,p) & \nabla_{uu} \; L(y,u,p) \end{pmatrix} \begin{pmatrix} s_y \\ s_u \end{pmatrix}$$

$$s.t. \qquad e_y(y,u) s_y + e_u(y,u) s_u = 0 \qquad (7)$$

is solved by $(s_y, s_u)$with $s_y = e_y(y(u),u)^{-1} e_u(y(u),u) s_u$, where$y = y(u)$, $p = p(u)$.

Proof: For every feasible point in (7) we have

$$\begin{pmatrix} s_y \\ s_u \end{pmatrix} = \begin{pmatrix} e_y(y(u),u)^{-1} e_u(y(u),u) s_u \\ s_u \end{pmatrix}$$

$$= T(u)^T s_u$$

Since we have

$$\hat{J}'(u) = -e_u(y(u),u) e_y(y(u),u)^{-1} J_y(y(u),u) + J_u(y(u),u).$$

So, we can write $\hat{J}'(u)$ as

$$\hat{J}'(u) = \begin{pmatrix} e_y(y(u),u)^{-1} e_u(y(u),u) \\ I_U \end{pmatrix} \begin{pmatrix} J_y(y(u),u) \\ J_u(y(u),u) \end{pmatrix}$$

$$= T(u)^T \begin{pmatrix} J_y(y(u),u) \\ J_u(y(u),u) \end{pmatrix}$$

Now,

$$\begin{pmatrix} J_y(y(u),u) \\ J_u(y(u),u) \end{pmatrix}^T \begin{pmatrix} s_y \\ s_u \end{pmatrix} = \begin{pmatrix} e_y(y(u),u)^{-1} e_u(y(u),u) s_u \\ s_u \end{pmatrix} \begin{pmatrix} J_y(y(u),u) \\ J_u(y(u),u) \end{pmatrix}$$

$$= s_u^T T(u)^T \begin{pmatrix} J_y(y(u),u) \\ J_u(y(u),u) \end{pmatrix}$$

$$= s_u^T \nabla \hat{J}(u)$$

Also, using equation (5) we have,

$$\begin{pmatrix} s_y \\ s_u \end{pmatrix}^T \begin{pmatrix} \nabla_{yy} \ L(y,u,p) & \nabla_{yu} \ L(y,u,p) \\ \nabla_{uy} \ L(y,u,p) & \nabla_{uu} \ L(y,u,p) \end{pmatrix} \begin{pmatrix} s_y \\ s_u \end{pmatrix} = T(u) s_u^T L_{\omega\omega} s_u T(u)^T$$

$$= s_u^T T(u)^T L_{\omega\omega} T(u) s_u$$

$$= s_u^T \nabla^2 \hat{J}(u) s_u$$

Thus, the result follows from the equivalence between (7) and the following equation

$$\min_{s_u} s_u^T \nabla \hat{J}(u) + s_u^T \nabla^2 \hat{J}(u) s_u$$

Theorem 2.Suppose$\hat{J}$is twice continuously differentiable, $u^* \in U$is a point at which the second order sufficiency optimality conditions are satisfied, and consider the iterates $u_{k+1} = u_k + s_k$where s_k solves $\nabla^2 \hat{J}(u) s_u = -\nabla \hat{J}(u)$. In addition, assume that $\hat{J}''(z)$ is Lipschitz with constant L, then

$$\|u_{k+1} - u_*\| \leq \frac{L}{2} \left\| \hat{J}''(u_k)^{-1} \right\| \|u_k - u_*\|^2$$

Proof.Since the second order optimality conditions are satisfied at$u^* \in U$, thus$\hat{J}'(u_*) = 0$. Now, we can write

$$u_{k+1} - u_* = u_k + s_k - u_*$$

$$= u_k - u_* + \hat{J}''(u_k)^{-1} \hat{J}'(u_k)$$

$$= \hat{J}''(u_k)^{-1} \left[\hat{J}''(u_k)(u_k - u_*) - \left(\hat{J}'(u_*) - \hat{J}'(u_k) \right) \right]$$

$$= \hat{J}''(u_k)^{-1} \left[\hat{J}''(u_k)(u_k - u_*) + \int_0^1 \hat{J}''\left(u_k + t_k(u_k - u_*) \right)(u_k - u_*) dt \right]$$

$$= \hat{J}''(u_k)\left[\int_0^1 \left(\hat{J}''(u_k) + \hat{J}''\left(u_k + t_k(u_k - u_*)\right)\right)(u_k - u_*)dt\right]$$

Taking norms on both sides of the above equality and using the triangle inequality submultiplicativity and yields

$$\|u_{k+1} - u_*\| = \left\|\hat{J}''(u_k)\right\| + \int_0^1 \left\|\hat{J}''(u_k) + \hat{J}''\left(u_k + t_k(u_k - u_*)\right)\right\| \|u_k - u_*\| dt$$

$$\leq \left\|\hat{J}''(u_k)\right\| + \|u_k - u_*\| \int_0^1 L\|u_k + t_k(u_k - u_*)\| dt$$

$$= \left\|\hat{J}''(u_k)\right\| + \frac{L}{2}\|u_k - u_*\|^2$$

Thus, giving the desired inequality.

3 Distributed Control of Elliptic Equations

Now, we apply the result to distributed optimal control problem which consists of a cost functional to be minimized subject to a partial differential problem posed on a domain $\Omega \in R^2$ or R^3 [8,14]:

$$min \;\; J(y,u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \alpha\|u\|^2_{L^2(\Omega)} \tag{9}$$

$$s.t. \;\; -\Delta y = u \qquad on \;\; \Omega$$

$$Y = g \qquad on \;\; \partial\Omega$$

Since we want to use finite elements, the weak formulation of constraints in (9) is required. So, the problem is: find $y \epsilon H^1{}_g(\Omega) = \{u\colon \; u \epsilon H^1(\Omega), \; u = g \;\; on \;\; \partial\Omega \}$such that

$$\int_\Omega \nabla y.\, \nabla v = \int_\Omega vu \qquad \forall v \epsilon \; H^1_0(\Omega)(10)$$

Let$V^h_0 \epsilon H^1_0$be an n-dimensional vector space of test functions with $\{\phi_1, \dots, \phi_n\}$ as a basis. Then, in order to satisfy the boundary condition, the basis is extendedby defining functions $\phi_{n+1}, \dots, \phi_{n+\partial n}$ and coefficients Y_j so that$\sum_{\partial=n+1}^{n+\partial n} Y_j \phi_j$ interpolates the boundary data. Hence if$y_h \epsilon V^h_g \subset H^1_g(\Omega)$, then it is uniquely determined by $\boldsymbol{y} = (Y_1, \dots Y_n)^T$with

$$y_h = \sum_{J=1}^{n} Y_j \, \phi_j + \sum_{J=n+1}^{n+\partial n} Y_j \, \phi_j$$

Here the ϕ_i , $i = 1, \dots, n$, define a set of shape functions. Also it is assumed that this approximation is conforming, i.e. $V_g^h = span\{\phi_1, \dots, \phi_{n+\partial n}\} \subset H_g^1(\Omega)$. Thus the finite-dimensional analogue of (10) can be expressed as: Find $y_h \epsilon V_g^h$ such that

$$\int_\Omega \nabla y_h . \nabla v_h = \int_\Omega v_h u \quad \forall v_h \epsilon V_0^h .$$

Now the discretization of u, needed as it appears in (9), is done using the same basis used for y

$$u_h = \sum_{J=1}^{n} U_j \phi_j$$

Since it is well known that without loss of generality $u_h = 0 \;\; on \;\; \partial\Omega$. Thus the discrete analogue of minimization problem can be written as

$$min_{\,y_h,u_h} \frac{1}{2}\|y_h - \hat{y}\|_2^2 + \alpha\|u_h\|_2^2 \tag{11}$$

$$s.t. \quad \int_\Omega \nabla y_h . \nabla v_h = \int_\Omega v_h u_h \quad \forall v_h \epsilon V_0^h \tag{12}$$

The discrete cost functional can be written as

$$min_{y_h,u_h} \frac{1}{2}\|y_h - \hat{y}\|_2^2 + \alpha\|u_h\|_2^2 = min_{\boldsymbol{y},\boldsymbol{u}} \frac{1}{2}\boldsymbol{y}^T M\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{b} + \beta + \alpha\boldsymbol{u}^T M\boldsymbol{u} \tag{13}$$

where $\boldsymbol{y} = (Y_1, \dots, Y_n)^T$, $\boldsymbol{u} = (U_1, \dots, U_n)^T$, $\boldsymbol{b} = \{\int \hat{y}\phi i\}_{i=1,\dots,n}$, $\beta = \|\hat{y}\|_2^2$ and $M = \left\{\int \phi_i\phi_j\right\}_{i,j=1,\dots,n}$ is a mass matrix. We now turn our attention to the constraint: (12) is equivalent to finding $\boldsymbol{y}$ such that

$$\int_\Omega \nabla(\sum_{i=1}^{n} Y_i\phi_i) . \nabla\phi_j = \int_\Omega \nabla(\sum_{i=n+1}^{n+\partial n} Y_i\phi_i) . \nabla\phi_j = \int_\Omega (\sum_{i=1}^{n} U_i\phi_i)\,\phi_j \;, \quad j = 1, \dots, n$$

which is

$$\sum_{i=1}^{n} Y_i \int_\Omega \nabla\phi_i . \nabla\phi_j = \sum_{i=1}^{n} U_i \int_\Omega \phi_i\phi_j - \sum_{i=n+1}^{n+\partial n} Y_i \int \nabla\phi_i . \nabla\phi_j \;, \quad j = 1, \dots, n$$

or

$$A\boldsymbol{y} = M\boldsymbol{u} + \boldsymbol{d} \tag{14}$$

where the matrix $A = \left\{\int \nabla\phi_i . \nabla\phi_j\right\}_{i,j=1,\dots,n}$ is the discrete Laplacian (the stiffness matrix) and the vector **d** contains the terms coming from the boundary values of y_h. Thus (13) and (14) together are aquivalent to (11) and (12). In order to solve this minimization problem, one way is considering the Lagrangian

$$L := \frac{1}{2}\boldsymbol{y}^T M\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{b} + \beta + \alpha\boldsymbol{u}^T M\boldsymbol{u} + \lambda^T(A\boldsymbol{y} - M\boldsymbol{u} - \boldsymbol{d}). \tag{15}$$

By setting the partial derivatives of the Lagrangian with respect to $y_i$to be zero, the adjoint equations corresponding to (4) are obtained and are given by

$$M\boldsymbol{y} - \boldsymbol{b} + \lambda^T A = 0 \implies \lambda^T = (\boldsymbol{b} - M\boldsymbol{y})A^{-1} \tag{16}$$

Given the solution of (16), the gradient of the objective function $\hat{J}$ can be obtained by computing the partial derivatives with respect to $\boldsymbol{u}$ of the Lagrangian (15). The gradient and hessian are given by:

$$\nabla_u \hat{J}(u) = 2\alpha M\boldsymbol{u} - \lambda^T M = 2\alpha M\boldsymbol{u} - (\boldsymbol{b} - M\boldsymbol{y})A^{-1}M$$

$$\nabla_u^2 \hat{J}(u) = 2\alpha M$$

Considering$P = diag(A)$ as a preconditioner in the Algorithm, since A is a block diagonal matrix with the blocks consisting of mass matrices, P is guaranteed to be positive definite. So, the eigenvalues of $P^{-1}A$ satisfy

$$Mx = \lambda diag(M)x$$

and, since M is a mass matrix, the eigenvalues of $P^{-1}A$ will be bounded above and below by constant values, see [16].

4 Examples

Example 1. Let $\boldsymbol{\Omega} = [\boldsymbol{-2}, \boldsymbol{1}]^{\boldsymbol{2}}$, and consider the problem

$$\min_{y,u} \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \alpha\|u\|^2_{L^2(\Omega)}$$

$$s.t. \quad -\Delta y = u \text{in} \quad \Omega$$

$$y = y_d \text{on} \quad \partial\Omega$$

where $y_d = \sin(\pi x)\sin(\pi y)$.Taking $\alpha = 0.001$, and using 100 discretization point in each direction , the initial state in order to handling the Algorithm, is shown in Fig. 1. In Fig. 2, the desired function, the state of control problem as an approximation of desired function, the optimal control and its corresponding adjoint are illustrated and show the efficiency of the Algorithm.

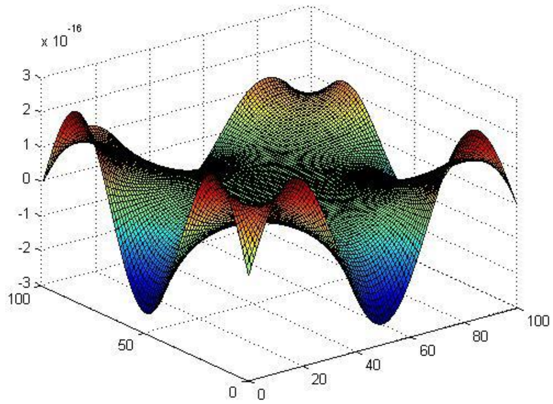

Fig . 1. Illustration of initial state considered to handling the Algorithm, for Example 1.

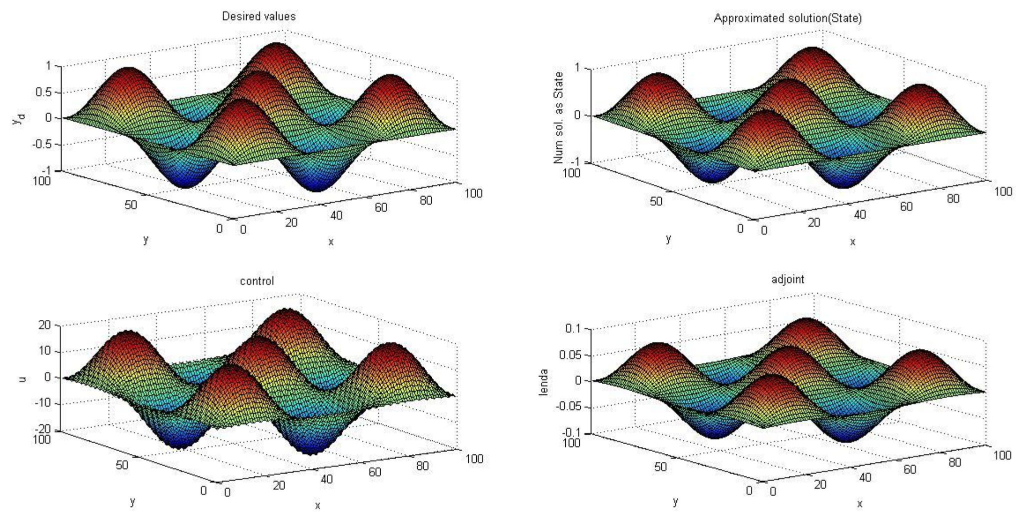

Fig. 2. Illustration of the solution with $y_d = \sin(\pi x) \sin(\pi y), \alpha = 0.001$ and 100 ×100 discretization point for Example 1.

Example 2. In Example 1, let $\boldsymbol{\Omega} = [-\mathbf{2}, \mathbf{2}]^{\mathbf{2}}$, and $y_d = \mathbf{1.1}\text{sign}(\sin(\pi x)\sin(\pi y))$**.** Taking $\boldsymbol{y} = \mathbf{0}$ on $\boldsymbol{\partial\Omega}$**,** the initial state for handling the Algorithm, is depicted in Fig. 3. Considering $\boldsymbol{\alpha} = \mathbf{0.0005}$, 50 ×50discretization points and solving the optimal control problem according to the proposed algorithm, yields the optimal control, state and adjoint solutions of the problem. Comparing the desired function and its approximation namely state of the control problem, as well as illustration of the optimal control and adjoint are depicted in Fig. 4.

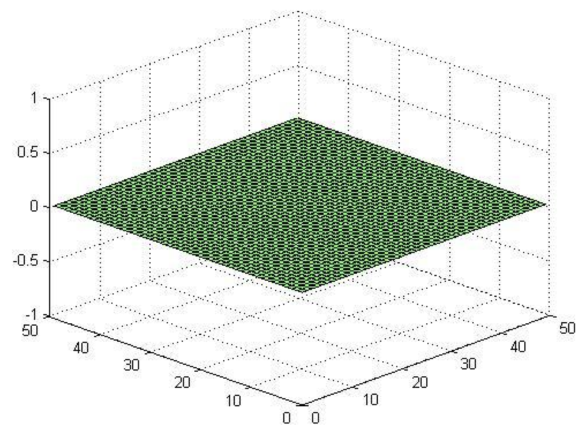

Fig . 3. Illustration of initial state considered to handling the Algorithm, for Example 2.

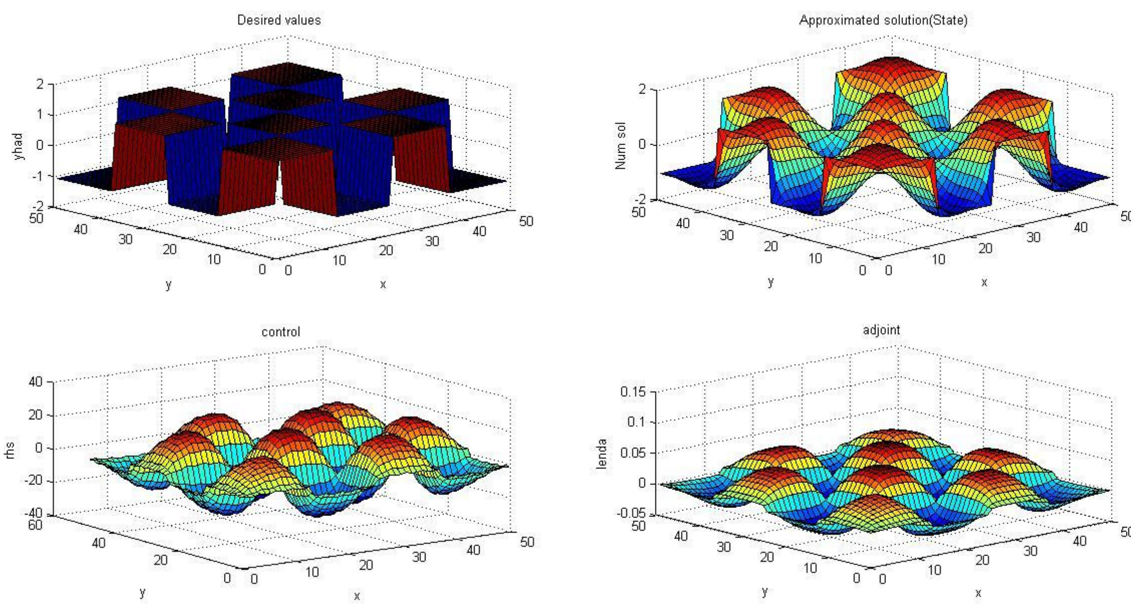

Fig. 4. Illustration of the solution with $y_d = 1.1\text{sign}(\sin(\pi x)\sin(\pi y))$,$\alpha = 0.0005$ and50 ×50discretization points for Example 2.

5 Conclusion

Here, a gradient based iterative algorithm for solving optimal control problem is proposed,where the analysis of the method in Theorem 1, shows relation between the Newton's type method used on the derivatives of the problem and the quadratic programming extracted from discretizing theproblem. As well as, Theorem 2 shows the convergence rate of control values in each iteration to the optimal control is proposed. In order to solve the large system of equations in Newton's equation, preconditioned conjugate gradient method is a good and efficient choice, where this fact is illustrated in some examples.

References

[1] S. R. Arridge Optical tomography in medical imaging, Inverse Problems, 15 (1999), pp. R41–R93.

[2] I. Bouchouev and V. Isakov, Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets, Inverse Problems, 15 (1999), pp. R95-R116.

[3] B. Dupire, Pricing with a smile, Risk, 7 (1994), pp. 32-39.

[4] H. Egger and H. W. Engl, Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates, Inverse Problems, 21 (2005), pp. 1027-1045.

[5]A. Fakharzadeh, S. Hashem, Solving a Class of Nonlinear Optimal Control Problems by Differential Transformation Method, Journal of mathematics and computer science, Vol .5, No.3 (2012) 146 -152.

[6] M. Gunzburger, Perspectives in Flow Control and Optimization, SIAM, 1987.

[7] Howard Elman, David Silvester, and Andy Wathen. Finite Elements and Fast Iterative Solvers withApplications in incompressible fluid dynamics. Numerical Mathematics and Scientific Computation.Oxford University Press, Oxford, 2005.

[8] Heinkenschloss and H. Nguyen. Neumann-Neumann domain decomposition preconditioners for linearquadratic elliptic optimal control problems. SIAM Journal on Scientific Computing, 28:1001–1028, 2006.

[9] C. T. Kelley. Iterative Methods for Optimization. SIAM, Philadelphia, 1999.

[10] M. V. Klibanov and T. R. Lucas, Numerical solution of a parabolic inverse problem in optical tomography using experimental data, SIAM J. Appl. Math., 59 (1999), pp. 1763-1789.

[11] C. D. Laird, L. T. Biegler, B. Waanders, and R. A. Bartlett, Time dependent Contaminant source determination for municipal water networks using large scale optimization, ASCE J. Water Res. Mgt. Plan., 131 (2005), pp. 125-134.

[12] J. Nocedal and S. J. Wright. Numerical Optimization. Springer Verlag, Berlin, Heidelberg, New York, second edition, 2006.

[13] C. Orozco and O. Ghattas, Massively parallel aerodynamic shape optimization, Comp. Syst. Eng., 1 (1992), pp. 311-320.

[14] T. Rees, H. S. Dollar, and A. J. Wathen: Optimal solvers for PDE-constrained optimization. SIAM J. Sci. Comput. 32 (2010), 271–298..

[15] A. Shenoy, M. Heinkenschloss, and E. M. Cliff, Airfoil design by an all-at-once method, Int. J. Comput. Fluid Dyn., 11 (1998), pp. 3-25.

[16] A. J. Wathen. Realistic eigenvalue bounds for the Galerkin mass matrix. IMA J Numer Anal, 7(4):449–457, 1987.

[17]A. Zakeri, M. Masjed-Jamei, A. H. Salehi Shayegan, A Numerical Method for Optimal Control-State Problem with Bivariate B-Spline Basis, Journal of mathematics and computer science, 8 (2014), 335-342.

[18] Zeidler, E., Nonlinear Functional Analysis and Its Applications. I, Fixed-Point Theorems. Springer, Berlin (1986).