



Extension of ADMM Algorithm in Solving Optimal Control Model Governed by Partial Differential Equation

K. A. Dawodu*

Department of Mathematical Sciences, Federal University of Technology Akure, Nigeria

Abstract

This paper presents an Algorithm for the numerical solution of the Optimal Control model constrained by Partial Differential Equation using the Alternating Direction Method of Multipliers (ADMM) accelerated with a parameter factor in the sense of Nesterov. The ADMM tool was applied to a partial differential equation-governed optimization problem of the one-dimensional heat equation type. The constraint and objective functions of the optimal control model were discretized using the Crank-Nicolson and Composite Simpson's Methods respectively into a derived discrete convex optimization form amenable to the ADMM. The primal-dual residuals were derived to ascertain the rate of convergence of the model for increasing iterates. An existing example was used to test the efficiency and degree of accuracy of the algorithm and the results were favorable when compared the existing method.

DOI:10.46481/jnsps.2021.159

Keywords: Optimal Control model, Alternating Direction Method of Multipliers, Partial Differential Equation constraint, Crank-Nicolson, Composite Simpson's, primal-dual.

Article History :

Received: 22 January 2021

Received in revised form: 05 May 2021

Accepted for publication: 10 May 2021

Published: 29 May 2021

©2021 Journal of the Nigerian Society of Physical Sciences. All rights reserved.
Communicated by: T. Latunde

1. Introduction

Optimal control problems (OCP) involving partial differential equations (PDE) are derivations arising from solid mechanics, fluid dynamics, engineering, Economics and sciences amongst others. Most of which are formulations into second order performance measures governed by PDE constraints of the elliptic type. The classical analytic techniques of the Hamilton -Jacobi -Bellman (HJB) or Euler Lagrange(E-L) may proof difficult because of the complexity of the formulated continuous nonlinear models. Betts and Campbell in [1] worked

on the "first discretized then optimize" technique, a novel approach in obtaining a discretized optimization problem with finite number of (large) vectors. Ghobadi *et al.* in [2] extended this approach to the one-dimensional heat equation and enumerated a lot of advantages when compared with the historical "first optimize then discretized" approach. These include (i) the circumvention of the need to establish certain properties of the problem such as the active constraint interval and the number of touch points a priori and (ii) its amenability to well-structured discretization schemes and optimization solvers for inequality constrained problems. The work of [2] was shrouded with conditional stability problem due to the explicit discretization method used. However, this draw-

* Corresponding author tel. no:

Email address: dawodukazeem@futa.edu.ng (K. A. Dawodu)

back of conditional stability will be circumvented in this research with the use of the implicit method of higher order in the discretization and the ADMM extended in obtaining the optimal solutions with better accuracy of results and faster rate of convergence.

Moreover, many authors have such as Olotu and Dawodu in ([3, 4, 5]) have worked extensively on proportional and multi-delay systems using the "first discretized then optimize" approach. However, He and Yuan in [6] and He *et. al* in [7] worked on block-convex programming problem synonymous to discretized continuous distributed systems, though the implementation on PDE-based optimal control problems were not explored.

2. Statement of problem

Considering a one-dimensional bar connected to a heating (cooling) device at its both ends with known initial and lower-bound conditions but with unknown boundary conditions, then the temperature functions on the boundaries are considered to be the control variables ($u_i(t), i = 1, \dots, q$) as in [1]. The heat equation is then given as;

$$\frac{\partial f(x, t)}{\partial t} = \frac{\partial^2 f(x, t)}{\partial x^2}, \quad 0 \leq t \leq T, \quad l_1 \leq x \leq l_2 \quad (1)$$

where $f(x, t)$ is the temperature at position x and at time t . It is then imperative to have the temperature profile $f(x, t)$ at each point of the bar that does not go below a given specific temperature function given by $g(x, t)$ (also known as a lower bound function), during the time interval. Setting the temperature of the bar very high makes sure the lower-bound condition (inequality constraint) below is satisfied.

$$f(x, t) \geq g(x, t) \quad \text{for all } x \text{ and } t \quad (2)$$

However, the objective of this modeling is to choose an appropriate temperature profile such that the energy consumed (objective function) is at minimum by reducing overheating, that is, to minimize the cost of energy consumed while keeping the temperature of the bar above the lower-bound profile. The appropriate choice of the objective function can be compared to the temperature profile of the bar with the lower-bound profile as closely as possible expressed mathematically as $\|f((x, t), y) - g(x, t)\|$. Assuming the temperature profile is the approximation of the temperature of the bar provided the heat equation and initial condition are satisfied. And suppose the squared values of all the temperature in time t and space x is expressed as $f^2(x, t)$, then the objective function to be minimized is the integral of $f^2(x, t)$ in time and space. Then the optimization model of the heat-transfer problem

can be mathematically expressed as

$$\begin{aligned} \min_{u, f} \quad & \int_0^T \int_{l_1}^{l_2} f^2(x, t) dx dt + \int_0^T [pu_1^2(t) + qu_2^2(t)] dt, \\ \text{s.t.} \quad & \frac{\partial^2 f(x, t)}{\partial x^2} = \frac{\partial f(x, t)}{\partial t}, \\ & f(x, t) > g(x, t), \\ & f(x, 0) = f_0(x), \\ & f(l_1, t) = u_1(t), \\ & f(l_2, t) = u_2(t), \\ & t \in [0, T], \quad x \in [l_1, l_2], \end{aligned} \quad (3)$$

where $x \in \mathbf{R}$, $u \in \mathbf{R}$ and $f_0(x)$ is a known initial function of temperature at the initial time ($t = 0$). While $u_1(t)$ and $u_2(t)$ are the temperatures at the two boundary points, l_1 and l_2 respectively, of the bar whose optimal values will be obtained.

3. Materials and methods

The continuous non-linear model (problem) formulated above can now be solved using the approach of "first-discretized-then-optimize" to obtain a quadratic optimization problem with finite number of (large) variables. The implicit discretization methods will be used for the discretization of the continuous-time OCP because its (i) stability is unconditional, (ii) time step-length can be chosen independent of the spatial step-length and (iii) coefficient matrix remains well-conditioned.

3.1. Discretization of the Constraints

In the transformation of the continuous time PDE constraints, the second-order Crank-Nicolson method derived from the averaging of the *forward Difference method* at the j th step in time t and the *Backward-Difference* at the $(j + 1)$ th step in t with truncation error $\frac{h}{2} \frac{\partial^2 f}{\partial t^2}(x_i, \mu_j) + O(h^2)$ will be used. And given that the grid points are equally-distributed for both the spatial interval (length of bar) $[l_1, l_2]$ and the time interval $[0, T]$ with step-lengths δ and h respectively, computed as $\delta = (l_2 - l_1)/n$ and $h = T/N$. Then the number of intervals n from the discretization in space and N from the discretization in time will generate $(n + 1)$ and $(N + 1)$ number of grid points respectively.

In generating the numerical sequence, let $x_i = l_1 + i\delta$ be the i th point in space (for $i = 0, 1, 2, \dots, n$) and $t_j = jh$ be the j th point in time (for $j = 0, 1, 2, \dots, N$), provided the usual differentiability conditions are satisfied, then the Crank-Nicolson (averaged difference) discretizations scheme in time and space with local truncation error of order $O(\delta^2 + h^2)$ expressed below as;

$$\frac{\partial f(x, t_j)}{\partial t} = f'_t(x, t_j) \approx \frac{f_{i,j+1} - f_{i,j}}{h}, \quad (4)$$

$$\frac{\partial^2 f(x, t_j)}{\partial x^2} = \frac{f''_x(x_i, t_j) + f''_x(x_i, t_{j+1})}{2},$$

$$\approx \frac{1}{2} \left[\frac{f_{i-1,j} - 2f_{i,j} + f_{i+1,j}}{\delta^2} + \frac{f_{i-1,j+1} - 2f_{i,j+1} + f_{i+1,j+1}}{\delta^2} \right] \quad (5)$$

where $f_{i,j} \approx f(x_i, t_j)$ and $u_{k,j} \approx u_k(t_j)$ for $i = 0, 1, 2, \dots, n$, $j = 0, 1, 2, \dots, N-1$ and $k = 1, 2$. The discretization of the PDE constraint function using eqns. (4) and (5), given that $f(0, t_j) = f_{0,j} = u_{1,j}$, $f(x_n, t_j) = f_{n,j} = u_{2,j}$ and $f(x_j, 0) = 0$ yields;

$$\frac{f_{i,j+1} - f_{i,j}}{h} = \frac{1}{2} \left[\frac{f_{i-1,j} - 2f_{i,j} + f_{i+1,j} + f_{i-1,j+1} - 2f_{i,j+1} + f_{i+1,j+1}}{\delta^2} \right], \quad (6)$$

for $i = 1, 2, \dots, (n-1)$ and $0 \leq j \leq N-1$. Substituting $\theta = h/\delta^2$ and collecting like-terms gives the recurrence formula below;

$$\begin{aligned} & -\lambda f_{i-1,j} - \lambda f_{i-1,j+1} + 2(\lambda - 1)f_{i,j} + 2(\lambda + 1)f_{i,j+1} \\ & - \lambda f_{i+1,j} - \lambda f_{i+1,j+1} = 0. \end{aligned} \quad (7)$$

Setting $i = 1$ and $f_{0,j} = u_{1,j}$ yields;

$$\begin{aligned} & -\theta u_{1,j} - \theta u_{1,j+1} + 2(\theta - 1)f_{1,j} + 2(\theta + 1)f_{1,j+1} \\ & - \theta f_{2,j} - \theta f_{2,j+1} = 0 \quad \text{for } 0 \leq j \leq N-1. \end{aligned} \quad (8)$$

For $j = 0$ given $f_{1,0} = f_{2,0} = u_{1,0} = 0$ yields;

$$-\lambda u_{1,1} + 2(\theta + 1)f_{1,1} - \theta f_{2,1} = 0 \quad (9)$$

For $j = 1$ yields;

$$\begin{aligned} & -\theta u_{1,1} - \theta u_{1,2} + 2(\theta - 1)f_{1,1} + 2(\theta + 1)f_{1,2} \\ & - \theta f_{2,1} - \theta f_{2,2} = 0 \end{aligned} \quad (10)$$

For $j = k$ ($2 \leq k \leq N-1$);

$$\begin{aligned} & -\theta u_{1,k} - \theta u_{1,k+1} + 2(\theta - 1)f_{1,k} + 2(\theta + 1)f_{1,k+1} \\ & - \theta f_{2,k} - \theta f_{2,k+1} = 0. \end{aligned} \quad (11)$$

When eqn. (11) is expanded for $i = 1$ and $1 \leq j \leq N-1$ gives the matrix equation below:

$$\begin{pmatrix} -\theta & 0 & \dots & \dots & \dots & 0 \\ -\theta & -\theta & 0 & \ddots & \ddots & \vdots \\ 0 & -\theta & -\theta & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 0 & -\theta & -\theta \end{pmatrix} \begin{pmatrix} u_{1,1} \\ u_{1,2} \\ u_{1,3} \\ \vdots \\ \vdots \\ u_{1,N} \end{pmatrix} +$$

$$\begin{pmatrix} 2(\theta + 1) & 0 & \dots & 0 \\ 2(\theta - 1) & 2(\theta + 1) & \dots & \vdots \\ 0 & \ddots & \dots & \vdots \\ \vdots & \dots & 2(\theta - 1) & 2(\theta + 1) \end{pmatrix} \begin{pmatrix} f_{1,1} \\ f_{1,2} \\ \vdots \\ f_{1,N} \end{pmatrix}$$

$$+ \begin{pmatrix} -\theta & 0 & 0 & \dots & 0 \\ -\theta & -\theta & 0 & \dots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\theta & -\theta \end{pmatrix} \begin{pmatrix} f_{2,1} \\ f_{2,2} \\ \vdots \\ \vdots \\ f_{2,N} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \quad (12)$$

compactly written as

$$LU_1 + \hat{G}F_1 + LF_2 = \mathbf{0}, \quad (13)$$

where $F_1 = [f_{1,1}, f_{1,2}, \dots, f_{1,N}]$, $F_2 = [f_{2,1}, f_{2,2}, \dots, f_{2,N}]$, $U_1 = [u_{1,1}, u_{1,2}, \dots, u_{1,N}]^T$ and $\mathbf{0}$ are column vectors of dimension $N \times 1$ while L and \hat{G} are bi-diagonal square-matrices of $N \times N$ dimensions with respective entries described below.

$$L = \begin{cases} -\theta & : 1 \leq i \leq N \text{ and } j = i \\ -\theta & : 2 \leq i \leq N \text{ and } j = i - 1 \\ 0 & : \text{otherwise} \end{cases}, \quad (14)$$

and

$$\hat{G} = \begin{cases} 2(\theta + 1) & : 1 \leq i \leq N \text{ and } j = i \\ 2(\theta - 1) & : 2 \leq i \leq N \text{ and } j = i - 1 \\ 0 & : \text{otherwise} \end{cases} \quad (15)$$

Set $i = 2$; this then gives the recurrence equation below;

$$\begin{aligned} & -\theta f_{1,j} - \theta f_{1,j+1} + 2(\theta - 1)f_{2,j} + 2(\theta + 1)f_{2,j+1} - \\ & \theta f_{3,j} - \theta f_{3,j+1} = 0 \quad \text{for } 0 \leq j \leq N-1. \end{aligned} \quad (16)$$

For $j = 0$; and equating $f_{1,0} = f_{2,0} = f_{3,0} = 0$ yields,

$$-\theta f_{1,1} + 2(\theta + 1)f_{2,1} - \theta f_{3,1} = 0 \quad (17)$$

For $j = 1$;

$$\begin{aligned} & -\theta f_{1,1} - \theta f_{1,2} + 2(\theta - 1)f_{2,1} + 2(\theta + 1)f_{2,2} \\ & - \theta f_{3,1} - \theta f_{3,2} = 0 \quad \text{for } 0 \leq j \leq N-1. \end{aligned} \quad (18)$$

For $j = 2$;

$$\begin{aligned} & -\theta f_{1,2} - \theta f_{1,3} + 2(\theta - 1)f_{2,2} + 2(\theta + 1)f_{2,3} \\ & - \theta f_{3,2} - \theta f_{3,3} = 0 \quad \text{for } 0 \leq j \leq N-1. \end{aligned} \quad (19)$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

For $j = N-1$;

$$\begin{aligned} & -\theta f_{1,N-1} - \theta f_{1,N} + 2(\theta - 1)f_{2,N-1} + 2(\theta + 1)f_{2,N} \\ & - \theta f_{3,N-1} - \theta f_{3,N} = 0 \quad \text{for } 0 \leq j \leq N-1. \end{aligned} \quad (20)$$

The matrix equation for eqns. (17) to (20) is then given as,

$$LF_1 + \hat{G}F_2 + LF_3 = \mathbf{0}. \quad (21)$$

For $i = 2, 3, \dots, n - 2$ and $1 \leq j \leq N - 1$ will then be expressed below as:

$$LF_{i-1} + \hat{G}F_i + LF_{i+1} = 0 \tag{22}$$

For $i = n - 1, 1 \leq j \leq N - 1$ and setting $U_2 = F_n$ gives;

$$LF_{n-2} + \hat{G}F_{n-1} + LU_2 = 0. \tag{23}$$

Where $F_i = [f_{i,1}, f_{i,2}, \dots, f_{i,N}]^T \in \mathbf{R}^{N \times 1}, U_2 = [u_{2,1}, u_{2,2}, \dots, u_{2,N}]^T \in \mathbf{R}^{N \times 1}, L$ and \hat{G} are described in eqns. (14) and (15) respectively. Combining eqns. (13), (22) and (23) gives the matrix formulation below:

$$\begin{pmatrix} L & 0 \\ 0 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & L \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} + \begin{pmatrix} \hat{G} & L & 0 & \dots & 0 \\ L & \hat{G} & L & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \dots & \ddots & L \\ 0 & \dots & 0 & L & \hat{G} \end{pmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ \vdots \\ F_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \tag{24}$$

compactly written as,

$$CU + DF = \mathbf{0}, \tag{25}$$

where C is a block-matrix of dimension $(n - 1)N \times 2N, D$ is a square block-matrix of dimension $(n - 1)N \times (n - 1)N, U$ is a column vector of dimension $2N \times 1$ while F and $\mathbf{0}$ are column vectors of dimension $(n - 1)N \times 1. u_1(t_j)$ and $u_2(t_j)$ are the control variables for which gives access to the heating (cooling) resources, while the temperature variables $f(x_i, t_j)$ of the other parts of the bar are computed based on the heat-transfer equations.

3.2. Discretization of the Boundaries

Given

$$f_{i,j} \geq g(x_i, t_j) \text{ for } i = 1, 2, \dots, n - 1 \text{ and } j = 1, \dots, N,$$

then

$$u_{1,j} \geq g(l_1, t_j) \text{ and } u_{2,j} \geq g(l_2, t_j) \text{ for } j = 1, 2, \dots, N.$$

Therefore,

$$u_{1,j} + u_{2,j} \geq g(l_1, t_j) - g(l_2, t_j) \text{ for } j = 1, 2, \dots, N. \tag{26}$$

For $j = 1;$

$$(u_{1,1} + u_{2,1}) - z_1 = g(l_1, t_1) + g(l_2, t_1) = g_1 \tag{27}$$

For $j = 2;$

$$(u_{1,2} + u_{2,2}) - z_2 = g(l_1, t_2) + g(l_2, t_2) = g_2 \tag{28}$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

For $j = N;$

$$(u_{1,N} + u_{2,N}) - z_N = g(l_1, t_N) + g(l_2, t_N) = g_N \tag{29}$$

Combining eqns. (27) to (28) generates the matrix-structure below:

$$\begin{pmatrix} \left[\begin{matrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{matrix} \right] \left[\begin{matrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{matrix} \right] \\ - \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ \vdots \\ z_N \end{pmatrix} \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ \vdots \\ g_N \end{pmatrix} \begin{pmatrix} u_{1,1} \\ \vdots \\ u_{1,N} \\ u_{2,1} \\ \vdots \\ u_{2,N} \end{pmatrix} \tag{30}$$

Equation (30) is compactly written as;

$$MU - Z = G, \tag{31}$$

where $I \in \mathbf{R}^{N \times N}$ is the identity matrix, $M = [I, I] \in \mathbf{R}^{N \times 2N}$ is the augmented double identity matrices, $Z = [z_1, z_2, \dots, z_N]^T \in \mathbf{R}^{N \times 1}$ is the positive slack-variable vector, $U = [U_1, U_2]^T = [u_{11}, u_{12}, \dots, u_{1N}, u_{21}, u_{22}, \dots, u_{2N}]^T \in \mathbf{R}^{2N \times 1}$ is the unknown control-variable vector and $G = [g_1, g_2, \dots, g_N]^T \in \mathbf{R}^{N \times 1}$ is the known coefficient vector.

3.3. Discretization of Objective Function

In the discretization of the continuous-time objective function(performance measure) both in space and time, the composite Simpson's rule and the right-Riemann rule will be used respectively as stated below:

$$\int_{l_1}^{l_2} f(x, t) dx \approx \frac{2\delta}{3} \left[\sum_{i=1}^{n/2-1} f(x_{2i}, t) + 2 \sum_{i=1}^{n/2} f(x_{2i-1}, t) \right] + \frac{\delta}{3} [f(x_0, t) + f(x_i, t)] \tag{32}$$

$$\int_0^T f(x, t) dt \approx h \sum_{j=1}^N f(x, t_j) \tag{33}$$

Then the discretized objective function of the optimal control problem in eqn. (3) using eqns. (32) and (33) is expressed below as;

$$\min_{u,f} \int_0^T \left[\frac{2\delta}{3} J_1 + \frac{\delta}{3} J_2 + J_3 \right] dt \tag{34}$$

where,

$$J_1 = \sum_{i=1}^{n/2-1} f(x_{2i}, t) + 2 \sum_{i=1}^{n/2} f(x_{2i-1}, t), \quad (35)$$

$$J_2 = f(x_0, t) + f(x_n, t), \quad (36)$$

$$J_3 = pu_1^2(t) + qu_2^2(t). \quad (37)$$

Discretizing eqn. (34) using (32) and collecting like-terms gives;

$$\begin{aligned} &\simeq \min_{u,f} \left[\frac{4h\delta}{3} \sum_{j=1}^N \sum_{i=1}^{n/2} f_{2i-1,j}^2 + \frac{2h\delta}{3} \sum_{j=1}^N \sum_{i=1}^{n/2-1} f_{2i,j}^2 \right. \\ &\quad \left. + \left(p + \frac{\delta h}{3} \right) \sum_{j=1}^N u_{1,j}^2 + \left(q + \frac{\delta h}{3} \right) \sum_{j=1}^N u_{2,j}^2 \right], \\ &\simeq \min_{u,f} \left[\sum_{j=1}^N \left[\frac{4h\delta}{3} (f_{1,j}^2 + f_{3,j}^2 + \dots + f_{n-1,j}^2) + \right. \right. \\ &\quad \left. \left. \frac{2h\delta}{3} (f_{2,j}^2 + f_{4,j}^2 + \dots + f_{n-2,j}^2) + m_1 u_{1,j}^2 + m_2 u_{2,j}^2 \right] \right] \quad (38) \end{aligned}$$

where $m_1 = \left(p + \frac{\delta h}{3} \right)$ and $m_2 = \left(q + \frac{\delta h}{3} \right)$. Recall that $f_{0,j}^2 = f^2(l_1, j) = u_1^2(t_j) = u_{1,j}^2$ and $f_{n,j}^2 = f^2(l_2, j) = u_2^2(t_j) = u_{2,j}^2$, setting the values of $j = 1, 2, \dots, N$ in eqn. (38) yields the matrix operator.

For $j = 1$;

$$\begin{aligned} &\frac{4h\delta}{3} (f_{1,1}^2 + f_{3,1}^2 + \dots + f_{n-1,1}^2) \\ &+ \frac{2h\delta}{3} (f_{2,1}^2 + f_{4,1}^2 + \dots + f_{n-2,1}^2) \\ &+ m_1 u_{1,1}^2 + m_2 u_{2,1}^2 \end{aligned}$$

for $j = 2$;

$$\begin{aligned} &\frac{4h\delta}{3} (f_{1,2}^2 + f_{3,2}^2 + \dots + f_{n-1,2}^2) \\ &+ \frac{2h\delta}{3} (f_{2,2}^2 + f_{4,2}^2 + \dots + f_{n-2,2}^2) \\ &+ m_1 u_{1,2}^2 + m_2 u_{2,2}^2 \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \end{aligned}$$

for $j = N$;

$$\begin{aligned} &\frac{4h\delta}{3} (f_{1,N}^2 + f_{3,N}^2 + \dots + f_{n-1,N}^2) \\ &+ \frac{2h\delta}{3} (f_{2,N}^2 + f_{4,N}^2 + \dots + f_{n-2,N}^2) \\ &+ m_1 u_{1,N}^2 + m_2 u_{2,N}^2. \end{aligned}$$

The matrix formation of the above equations is then written as follows:

$$J(F,U) = \frac{1}{2} F^T Q F + \frac{1}{2} U^T H U, \quad (39)$$

where,

$$Q = \begin{pmatrix} \frac{8}{3} h\delta I & 0 & 0 & \dots & 0 \\ 0 & \frac{4}{3} h\delta I & 0 & \dots & \vdots \\ \vdots & \ddots & \vdots & 0 & \vdots \\ \vdots & 0 & \vdots & \ddots & \vdots \\ 0 & \dots & \vdots & \ddots & \frac{8}{3} h\delta I \end{pmatrix}, \quad (40)$$

$$H = \begin{pmatrix} 2m_1 I & 0 \\ 0 & 2m_2 I \end{pmatrix} \quad (41)$$

$F_i^T = [f_{i,1}, f_{i,2}, \dots, f_{i,N}] \in \mathbf{R}^{1 \times N}$ for $i = 1, 2, \dots, (n-1)$ with $F^T = [F_1^T, F_2^T, \dots, F_{n-1}^T] \in \mathbf{R}^{1 \times (n-1)N}$. Similarly, $U_k^T = [u_{k,1}, u_{k,2}, \dots, u_{k,N}] \in \mathbf{R}^{1 \times N}$ for $k = 1, 2$ with $U^T = [U_1^T, U_2^T] \in \mathbf{R}^{1 \times 2N}$. However, the matrices $Q \in \mathbf{R}^{(n-1)N \times (n-1)N}$ and $H \in \mathbf{R}^{2N \times 2N}$ in the compactified convex quadratic eqn. (39) are real, symmetric and positive-definite with the diagonal coefficients described below as follows:

$$Q = \begin{cases} \frac{8}{3} h\delta I : (k-1)N + 1 \leq i \leq kN \text{ and } j = i, \\ \quad \text{for } k = 1, 3, \dots, (n-1) \\ \frac{4}{3} h\delta I : (k-1)N + 1 \leq i \leq kN \text{ and } j = i, \\ \quad \text{for } k = 2, 3, \dots, (n-2) \\ 0 : \quad \text{otherwise} \end{cases} \quad (42)$$

and

$$H = \begin{cases} 2m_1 & : 1 \leq i \leq N \text{ and } j = i \\ 2m_2 & : N + 1 \leq i \leq 2N \text{ and } j = i \\ 0 & : \text{otherwise} \end{cases} \quad (43)$$

Combining eqns. (25), (31) and (39) gives, in a matrix form, the compact convex quadratic optimization problem with linear constraints as expressed below:-

$$\begin{cases} \text{Min } J = \frac{1}{2} U^T H U + \frac{1}{2} F^T Q F \\ \text{s.t} \quad : CU + DF = 0, \\ \quad \quad \quad MU \geq G. \end{cases} \quad (44)$$

The above re-formulated discrete model is a well-structured, quadratic optimization problem whose objective function is convex and separable. While the constraint function is linear and coupled which made the model amenable to the ADMM. However, the mixed-inequality constraint model is re-formulated into an equality constraint form by introducing non-negative slack variables $Z(\geq 0)$ into the model.

3.4. Stability and Feasibility of the Model

With the assumption that the solution of the PDE satisfies the usual differentiability conditions, the second order Crank Nicholson implicit scheme, with local truncation error of order $O(h + \delta^2)$, was deployed for the discretization of

the PDE constraint to avoid the conditional stability and feasibility associated with explicit discretization methods. This stability of the implicit discretization method reflects in the consistency and stability of the coefficient matrix operators. However, the composite Simpson's scheme, used for the discretization of the objective function, has truncation error of order $O(h^4)$ with degree of accuracy (precision) of three. This helps to improve the accuracy and rate of convergence of the model, hence an improvement over the trapezoidal used in [1].

3.5. Sparsity of the Model

The very suitable approach for ascertaining the sparsity of the model is to compactify it as studied by Kameswaran and Biegler in [8]. The set of optimization eqns. (45) was then re-structured in a compact form as stated below.

$$\min J(U, F) = \begin{pmatrix} U^T & F^T \end{pmatrix} \begin{pmatrix} H & \hat{0} \\ \hat{0}^T & Q \end{pmatrix} \begin{pmatrix} U \\ F \end{pmatrix}$$

s.t

$$\begin{pmatrix} C & D \\ M & \bar{0} \end{pmatrix} \begin{pmatrix} U \\ F \end{pmatrix} \geq \begin{pmatrix} 0^* \\ G \end{pmatrix},$$

compactly written as,

$$\min J(W) = W^T P W \quad \text{s.t} \quad V W \geq \bar{G}, \quad (45)$$

$$\text{where, } W = (U F)^T, P = \begin{pmatrix} H & \hat{0} \\ \hat{0}^T & Q \end{pmatrix},$$

$$V = \begin{pmatrix} C & D \\ M & \bar{0} \end{pmatrix} \quad \text{and} \quad \bar{G} = \begin{pmatrix} 0^* \\ G \end{pmatrix}.$$

The matrices H and Q are the two coefficient sub-matrices of the objective functions while C and D are the two sub-matrices of the constraint; both correspondent to those of the variables U and F . The sizes of the augmented matrices, showing the number of zero and non-zero entries are expressed in the table 1 below. The sparsity ratio is defined here as the proportion of the zero entries in comparison with the size of the matrix.

	Obj. Coef.	Constr. Coef.
	J	W
size	$(n+3)N \times (n+3)N$	$nN \times (n+1)N$
zeros	$(n+3)N$	$3n(2N-1) + 3 - 4N$
Non-zeros	$(n+3)N + (nN + 3N - 1)$	$nN(nN + N - 6) + (3n - 3 + 4N)$

Table 1: Sparsity Table for the Compactified Model

3.6. Positive definiteness and Convexity of the Matrix operator

Definition 1. (Sylvester criterion): A square $(n \times n)$ matrix is positive definite if it is real, symmetric with all the eigenvalues $(\lambda_i > 0; i = 1, 2, \dots, n)$ or the leading principal minors $(M_i > 0; i = 1, 2, \dots, n)$ strictly positive.

Definition 2. (Corollary to Sylvester criterion): If the principal diagonal entries a_{ii} are the only nonzero entries of a square $(n \times n)$ matrix such that $a_{ij} = 0$ for $i \neq j$; then the leading principal minors $(M_i > 0; i = 1, 2, \dots, n)$ are given by $M_i = \prod_{j=1}^i a_{jj}$ for $i = 1, 2, \dots, n$.

The constructed matrix-operator P for the derived quadratic function (45) is an augmentation of the real symmetric matrices Q and H in (40) and (41) respectively whose principal diagonal entries are all strictly positive. Therefore, P is symmetric and positive definite by definitions 2 and 3 since all the leading principal minors are all strictly positive. This property is pertinent to guarantee convexity and minimal solution of the quadratic function.

4. ADMM Algorithm Formulation

In the formulation of the ADMM, the derived update formulas will be accelerated with an accelerator variant referred to as the relaxation factor. The results will then be extendable to the discretized PDE governed OCP (44). The associated augmented Lagrangian functional is expressed as

$$L_\rho(U, F, Z, \xi_1, \xi_2) = \frac{1}{2} U^T H U + \frac{1}{2} F^T Q F + l_+(Z) + \frac{\rho}{2} \|CU + DF + \xi_1\|_2^2 + \frac{\rho}{2} \|MU - G - Z + \xi_2\|_2^2 \quad (46)$$

with the M-ADMM iterations stated thus:

$$U^{k+1} \leftarrow \arg \min_U \left\{ \frac{1}{2} U^T H U + \frac{\rho}{2} \|CU + DF + \xi_1^k\|_2^2 + \frac{\rho}{2} \|MU - G - Z + \xi_2\|_2^2 \right\}, \quad (47)$$

$$F^{k+1} \leftarrow \arg \min_F \left\{ \frac{1}{2} F^T Q F + \frac{\rho}{2} \|CU + DF + \xi_1^k\|_2^2 \right\}, \quad (48)$$

$$Z^{k+1} \leftarrow \arg \min_Z \left\{ \frac{\rho}{2} \|MU - G - Z + \xi_2\|_2^2 \right\} \quad (49)$$

s.t $Z \geq 0$,

$$\xi_1^{k+1} = \partial_{\xi_1} L(U^{k+1}, F^{k+1}, Z^{k+1}, \xi_1^k) \quad (50)$$

$$\xi_2^{k+1} = \partial_{\xi_2} L(U^{k+1}, F^{k+1}, Z^{k+1}, \xi_2^k) \quad (51)$$

where $\mu \in \mathbf{R}^{(n-1)N}$ and $\lambda \in \mathbf{R}^N$ are Lagrange multipliers, $\rho > 0$ is the penalty parameter, $\|\cdot\|_2$ denotes the euclidean (spectral) norm of a vector (matrix) argument, $\xi_1 = \mu/\rho \in \mathbf{R}^{(n-1)N}$ and $\xi_2 = \lambda/\rho \in \mathbf{R}^N$ are the scaled dual variables, $Z \in \mathbf{R}^N$ is the

introduced slack vector and $l_+(Z)$ is the indicator function for the non-negative orthants defined as $l_+(Z) = 0$ for $z \geq 0$ and $l_+(Z) = +\infty$ otherwise. As the algorithm runs through the $(k+1)$ -th iteration, the primal variables U^k and F^k as well as the dual variables μ^k and λ^k are updated, hence the need to derive the update formulas. Applying the Karush-Kuhn-Tucker (KKT) optimality conditions on the Augmented Lagrangian Functional (46) for the sequential minimization of the variables requires updating all the critical variables as indicated in the ADMM iterations. Moreover the update formulas of the variables U^k and F^k will be accelerated using the Gauss-seidel relaxation factor $\alpha \in [0, 2]$ as clearly illustrated in the works of Nesterov [9] and Ghadimi *et. al* [10].

4.1. Derivations of the update formulas

Applying the Karush-Kuhn-Tucker (KKT) optimality conditions on the Augmented Lagrangian Function in eqn. (46) for the sequential minimization of the variables requires updating all the critical variables as indicated in the ADMM iterations.

U-Update: Setting $\partial_U L(U, \cdot, \cdot) = 0$ yields;

$$U^{k+1} = \rho B_1 M^T B_2 \quad (U - \text{Update}). \quad (52)$$

where,

$$B_1 = \left[H + \rho(C^T C + M^T M) \right]^{-1}$$

$$B_2 = (G + Z^k - \xi_2^k) - C^T (DF^k + \xi_1^k)$$

F-Update: Setting $\partial_F L(U^{k+1}, F, \cdot) = 0$ and replacing CU^{k+1} , to relax the algorithm, with

$$h(U^{k+1}, F^k, \alpha) = \alpha CU^{k+1} + (1 - \alpha) DF^k \quad (53)$$

to yield

$$F^{k+1} = -\rho B_3 B_4 \quad (F - \text{Update}). \quad (54)$$

where,

$$B_3 = \left[Q + \rho(D^T D) \right]^{-1}$$

$$B_4 = \left[\alpha CU^{k+1} + (1 - \alpha) DF^k + \xi_1^k \right]$$

Z-Update: - Setting $\partial_Z L(\cdot, \cdot, Z, \cdot) = 0$ and relaxing it with $h(U^{k+1}, F^{k+1}, \alpha)$ yields;

$$Z^{k+1} = \max \left\{ 0, -Z^* \right\} \quad (55)$$

where

$$Z^* = \left[\alpha CU^{k+1} + (1 - \alpha) DF^{k+1} - G \right] \quad (56)$$

Updating the Scaled-Dual variables - ξ_1, ξ_2 yields

$$\xi_1^{k+1} = \xi_1^k + CU^{k+1} + DF^{k+1}, \quad (57)$$

$$\xi_2^{k+1} = \xi_2^k + MU^{k+1} - G - Z^{k+1}, \quad (58)$$

4.2. Convergence Analysis

Considering an optimization problem analogous to the form in (45) above;

$$\begin{cases} \text{Min}_{u,y} & f_1(u) + f_2(y) + l_+(z) \\ \text{s.t} & h_1(u) + h_2(y) = 0 \\ & g_1(u) - z = 0 \end{cases} \quad (59)$$

where $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$ while $f_1 : \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$, $f_2 : \mathbf{R}^m \rightarrow \mathbf{R} \cup \{\infty\}$ and $f_3 : \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ are convex functions. The augmented lagrangian formulation is then stated thus:

$$L_\rho(u, y, \mu, \lambda) = f_1(u) + f_2(y) + l_+(z) + \mu^T (h_1(u) + h_2(y)) + \frac{\rho}{2} \|h_1(u) + h_2(y)\|_2^2 + \lambda^T (g_1(u) - z) + \frac{\rho}{2} \|g_1(u) - z\|_2^2 \quad (60)$$

where $\mu \in \mathbf{R}^n$ and $\lambda \in \mathbf{R}^m$ are Lagrange multipliers, $\rho > 0$ is the penalty parameter, $\|\cdot\|_2$ denotes the euclidean (spectral) norm of a vector (matrix) argument, $\xi_1 = \mu/\rho \in \mathbf{R}^n$ and $\xi_2 = \lambda/\rho \in \mathbf{R}^m$ are the scaled dual variables, z is the introduced slack vector and $l_+(z)$ is the indicator function for the non-negative orthants.

From duality theory, it is well-known that if u^*, y^*, μ^* and λ^* are the primal and dual variables that minimize the Lagrangian function L_ρ , then every pair (u^k, y^k) in the convex set,

$$L_\rho(u^*, y^*, \mu^*, \lambda^*) \leq L_\rho(x^k, y^k, \mu^*, \lambda^*) \quad (61)$$

. Let $f^k = f_1(x^k) + f_2(y^k)$ be the objective function value at the k -th iterate (x^k, y^k) and $f^* = f_1(x^*) + f_2(y^*)$ be the optimal function value. Let $r_1^{k+1} = h_1(u^k) + h_2(y^k)$ and $r_2^{k+1} = g_1(u^k) - z^k$ be the residuals of the convex equality constraints in the k -th iteration. The objective of this research is now to prove the following theorems.

Definition 3. (Dual Updates): At optimality, the update equations for the dual variables from the augmented Lagrangian in (60) are given as

$$\mu^{k+1} = \mu^k + \rho \left[h_1(u^{k+1}) + h_2(y^{k+1}) \right], \quad (62)$$

$$\lambda^{k+1} = \lambda^k + \rho \left[g_1(u^{k+1}) - z^{k+1} \right], \quad (63)$$

Definition 4. (Lyapunov function): Let V^k be a non-negative quantity for the algorithm such that $V^k - V^{k+1}$ decreases in each iteration by an amount that depends on the sum of the norms of the residuals and on the change in slack variable z over one iteration. Then the quantities are defined as;

$$(i) \quad V^k = \frac{1}{\rho} \|\mu^k - \mu^*\|_2^2 + \frac{1}{\rho} \|\lambda^k - \lambda^*\|_2^2 + \rho \|z^{k+1} - z^*\|_2^2 \quad (64)$$

$$(ii) \quad \rho V_q \leq V^k - V^{k+1} \leq V^0 \quad (65)$$

where

$$V_q = \sum_{k=0}^{\infty} (\|r_1^{k+1}\|_2^2 + \|r_2^{k+1}\|_2^2) + \|z^{k+1} - z^k\|_2^2$$

See proof in Boyd and Vandenberghe [11].

Theorem 1. Given the convex linearly constrained control problem (59) with the optimal dual and primal variables for the problem μ^* , λ^* , u^* , y^* and z^* . Then the following residual terms

$$(i) \quad s_1^k = \rho \left[\partial h_1^T(u^*) (h_2(y^*) - h_2(y^k)) - \partial g_1^T(y^*) (z^* - z^k) \right] \quad (66)$$

$$(ii) \quad s_2^k = \partial f_2^T(y^*) + \partial h_2(y^*) \mu^* \quad (67)$$

that converge to zero for a given penalty parameter $\rho > 0$.

Proof: From the augmented Lagrangian function (60), u^{k+1} minimizes the sub-differential $L_\rho(u, y^k, \mu^k, \lambda^k)$ with respect to u . Hence $L_\rho(u, y^k, \mu^k, \lambda^k)$ with respect to u at u^k is expressed as

$$\begin{aligned} 0 \in & \partial f_1(u^{k+1}) + \partial h_1^T(u^{k+1}) \mu^k \\ & + \partial h_1^T(u^{k+1}) (h_1(u^k) + h_2(y^k)) + \partial g_1^T(u^{k+1}) \lambda^k \\ & + \partial g_1^T(u^{k+1}) (g_1(u^{k+1}) - z^k) \end{aligned} \quad (68)$$

Substituting for μ^k and λ^k from eqns. (62) and (63) into (68) yields,

$$\begin{aligned} 0 \in & \partial f_1(u^{k+1}) \\ & + \partial h_1^T(u^{k+1}) \underbrace{\left[\mu^{k+1} - \rho [h_1(u^{k+1}) + h_2(y^{k+1})] \right]}_{\mu^k} \\ & + \partial h_1^T(u^{k+1}) (h_1(u^k) + h_2(y^k)) \\ & + \partial g_1^T(u^{k+1}) \underbrace{\left[\lambda^{k+1} - \rho [g_1(u^{k+1}) - z^{k+1}] \right]}_{\lambda^k} \\ & + \partial g_1^T(u^{k+1}) (g_1(u^{k+1}) - z^k) \end{aligned} \quad (69)$$

$$\begin{aligned} & \partial f_1(u^{k+1}) + \partial h_1^T(u^{k+1}) \mu^{k+1} + \partial g_1^T(u^{k+1}) \lambda^{k+1} \\ & = \rho \left[\partial h_1^T(u^{k+1}) (h_2(y^{k+1}) - h_2(y^k)) \right. \\ & \quad \left. - \partial g_1^T(y^{k+1}) (z^{k+1} - z^k) \right]. \end{aligned} \quad (70)$$

If $u^{k+1} = u^*$ minimizes the Lagrangian functional, then eqn. (70) becomes

$$\begin{aligned} L_\rho(x^*, y^*, \mu^*, \lambda^*) = & \rho \left[\partial h_1^T(u^*) (h_2(y^*) - h_2(y^k)) \right. \\ & \left. - \partial g_1^T(y^*) (z^* - z^k) \right], \end{aligned}$$

and converges to zero for any $u \in \Omega$ (convex) which completes the proof for (i). QED \square

Similarly, y^{k+1} minimizes the sub-differential $L_\rho(u^{k+1}, y, \mu^k, \lambda^k)$ with respect to y such that

$$\begin{aligned} 0 \in & \partial f_2(y^{k+1}) + \partial h_2^T(y^{k+1}) \mu^k \\ & + \partial h_2^T(y^{k+1}) (h_1(u^{k+1}) + h_2(y^{k+1})), \\ 0 \in & \partial f_2(y^{k+1}) + \\ & + \partial h_2^T(y^{k+1}) \underbrace{\left[\mu^{k+1} - \rho [h_1(u^{k+1}) + h_2(y^{k+1})] \right]}_{\mu^k} \\ & + \partial h_2^T(y^{k+1}) (h_1(u^{k+1}) + h_2(y^{k+1})). \end{aligned}$$

This yields the equation below and completes the proof for (ii)

$$\partial f_2(y^{k+1}) + \partial h_2^T(y^{k+1}) \mu^{k+1}. \quad (71)$$

Hence, y^{k+1} minimizes the convex function

$$y \mapsto f_2(y) + \partial h_2^T(y) \mu^{k+1}. \quad (72)$$

The alternating direction method of multipliers algorithm can now be applied to the derived optimization problem that was obtained from the discretized PDE. By extension, the following vanishing dual residual obtained as the iterations get to optimality is stated below.

$$s_1^k = \rho \left[C^T D(F^{k+1} - F^k) - M^T (Z^{k+1} - Z^k) \right]. \quad (73)$$

Algorithm: Extension of ADMM Algorithm in Solving PDE Constrained Optimal Control Problem

Input parameters and operators ρ , α , M , G , C , D

Initialize U^0 , F^0 , Z^0 , $\xi_1^0 = \mu^0 / \rho$, $\xi_2^0 = \lambda^0 / \rho$

Set $H \geq 0$, $Q \geq 0$ (symmetric and positive definite)

For $k = 0, 1, 2, \dots$, **do**

| Compute U^{k+1} using eqn. (52)

| Compute F^{k+1} using eqn. (53)

| Compute Z^{k+1} using eqn. (54)

| Update ξ_1^{k+1} and ξ_2^{k+1} using eqns. (57) and (58) respectively.

| **Until** Convergence according to eqn. (73)

End (for)

Return U^{k+1} , F^{k+1} , Z^{k+1} , J^* , ξ_1^{k+1} , ξ_2^{k+1}

5. Numerical Experiments

Considering a PDE constrained optimal control described heat transfer model written as follows:

$$\begin{aligned} \min_{u,f} \quad & \int_0^5 \int_0^\pi f^2(x,t) dx dt + \int_0^5 [u_1^2(t) + 2u_2^2(t)] dt, \\ \text{s.t} \quad & \frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial f(x,t)}{\partial t}, \\ & f(x,t) > \sin(x) \sin\left(\frac{\pi t}{5}\right) - 0.7, \\ & f(x,0) = \sin(x) - 0.7, \\ & f(0,t) = u_1(t), \\ & f(\pi,t) = u_2(t), \\ & t \in [0,5], \quad x \in [0,\pi], \end{aligned}$$

This section illustrates, via a number of computational experiments, the effects of different penalty parameter selection on the number of iterations required to reach the stopping criteria on the above problem as posed by Ghobadi *et. al* in [2] and Betts and Campbell in [1]. The experiment involves selecting the termination (stopping) criteria $\epsilon^{rel} = 10^{-k}$ and $\epsilon^{abs} = 10^{-k}$, $k = 3, 4, 5$, as relative and absolute tolerances respectively for the convergence of the algorithm. The algorithm was implemented with a MATLAB subroutine running on a Pentium V Computer with 4.0 GB RAM and 1.7 GHz CPU. The discretization step in time is largely smaller than that of the spatial step to keep it feasible; though the implicit Crank-Nicholson numerical scheme used for the discretization of the PDE keeps it unconditionally stable. Below are the results of the experiment; Table 2 above shows the numbers of iter-

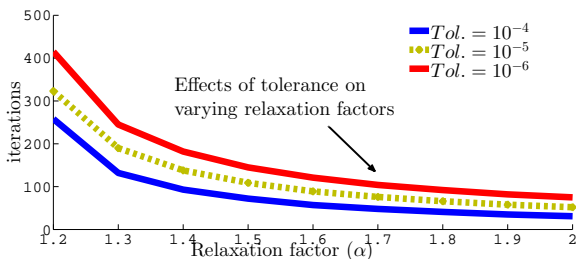


Figure 1: The effects of tolerance on varying relaxation factors

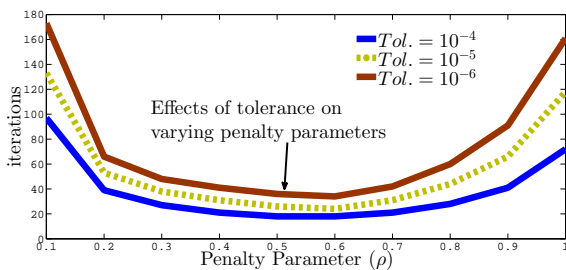


Figure 2: The effects of tolerance on varying penalty parameters

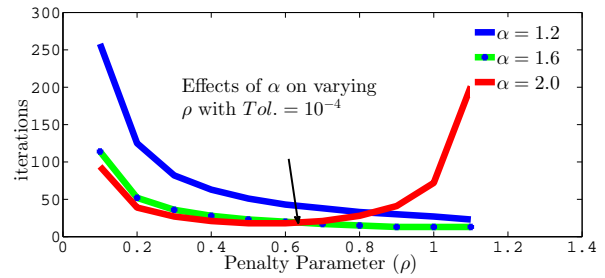


Figure 3: The effects of relaxation factor on varying penalty parameters

α	$h = 0.20$		$h = 0.10$		$h = 0.05$	
	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}
1.20	235	294	192	240	164	210
1.40	86	128	669	107	56	93
1.60	52	82	41	67	32	58
1.80	37	60	28	49	22	41
2.00	28	47	21	38	16	32

Table 2: Effects of tolerance & time-steps h on varying relaxation factors ($\rho = 0.1$, $\delta = 1/3$)

ations for the varying relaxation parameters α when the tolerance is fixed for a particular step size in time h and penalty parameter ρ . The step sizes were only kept for $h = 0.20$, $h = 0.10$ and $h = 0.05$. For a fixed choice of the relaxation factor $\alpha = 2.0$, Tolerance $Tol. = 10^{-3}$ and step-size $\rho = 0.1$, the dual sum $\|r_3^{k+1}\|_2 = \|r_1^{k+1} + r_2^{k+1}\|_2$, the distances $\|r_1^{k+1}\|_2 = \|\mu^k - \mu^*\|_2$ and $\|r_2^{k+1}\|_2 = \|\lambda^k - \lambda^*\|_2$ were plotted against the iteration as shown below in figures 4, 5 and 6 respectively. The convergence of the dual residual sum follows same as the monotonically decreasing quantity; that is $\|r_3^{k+1}\|_2 \leq V^k$.

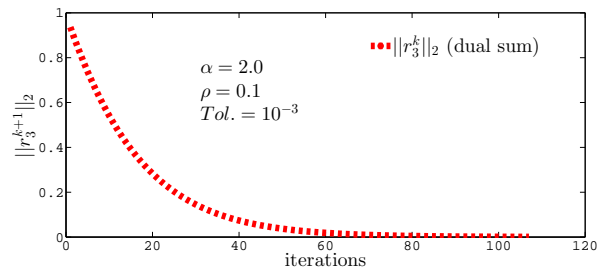


Figure 4: The response of dual residual sum to iteration

Table 3 above shows the varying penalty parameters ρ for same fixed α and various values of time step size h . The iteration was plotted against the dual residual $\|s^k\|_2$ and the primal variables ($\|U^{k+1}\|_2$, $\|F^{k+1}\|_2$, p^k) and the distance $\|Z^k - Z^*\|_2$ for a fixed choice of the relaxation factor $\alpha = 2.0$, Tolerance $Tol. = 10^{-3}$ and step-size $\rho = 0.1$. It was however observed, as expressed in figures 7, 8 and 9 respectively that the residuals of the algorithms decrease after subsequent it-

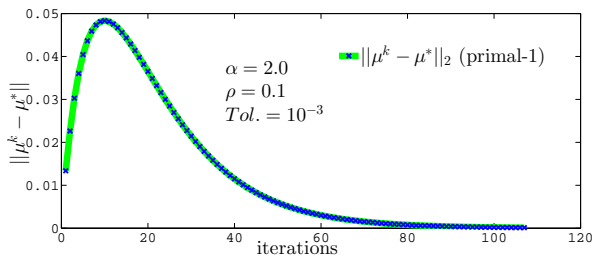


Figure 5: The response of first dual variable (μ) to iteration

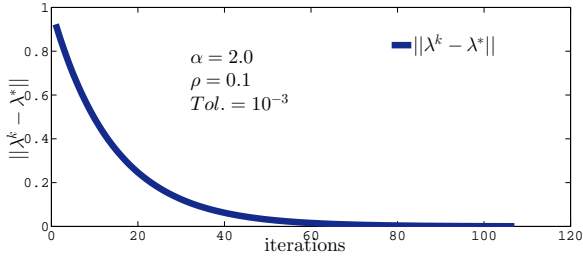


Figure 6: The response of second dual variable (λ) to iteration

erations. This then indicates convergence of the algorithm.

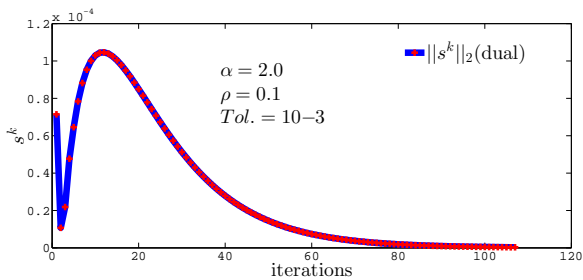


Figure 7: The response of dual residuals (s^k) to iteration

Table 4 below shows the sparsity of the compact form of the model for varying and fixed step-sizes in time and space respectively; with the optimal results obtained at few iterations of the ADMM. It was observed that the level of sparsity increases for reducing step sized which consequently reduces the computational burden in the iterative process. Table 5

ρ	$h = 0.20$		$h = 0.10$		$h = 0.05$	
	10^{-4}	10^{-5}	10^{-4}	10^{-5}	10^{-4}	10^{-5}
0.10	91	124	111	149	184	245
0.20	39	85	48	66	79	107
0.40	19	30	25	38	43	65
0.60	15	22	18	29	32	51
0.80	20	38	13	23	25	43
1.00	84	156	13	20	19	35

Table 3: Effects of tolerance & time-steps h on varying penalty parameters ($\alpha = 1.6, \delta = 1/3$)

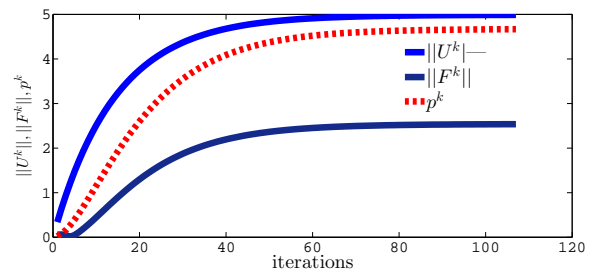


Figure 8: The response of primal variables to iteration

Figure 9: The effects of iterations on the distance norm $\|Z^k - Z^*\|_2$

shows the results of the problem for the varying values of the parameters and step size with the optimal results obtained at few iterations of the ADMM. The results obtained on MATLAB subroutine were compared with those of Ghobadi *et al.* in [2] for various time grid points (N) ranging from $N = 1000$ to 8000 ran on MOSEK software. Table 6 below shows the optimal objective functional values generated at varying numbers N of time grid points (TGP) for fixed number of grid-points in space kept at $n = 10$. The results in [2] when implemented on MOSEK is same as that obtained by the ADMM at a shorter Central Processing Unit (CPU) time though with higher iteration cycles due to high rate of convergence; using the tolerance of $tol. = 10^{-4}$. This clearly shows that the rate of convergence of the proposed algorithm is higher than that of MOSEK.

6. Conclusions

The extension of the ADMM to solving heat transfer problem as an example of PDE-based optimal control problems (OCP) with equality-Inequality constraints was well studied. The PDE-based OCP was transformed into a discretized convex quadratic optimization problem amenable to the ADMM.

(N,n)	$h = 0.20$	$h = 0.10$	$h = 0.05$
	(25,10)	(50,10)	(100,10)
Non-zero	325	650	1300
Zero	105300	421850	1688700
Size	105625	422500	1690000
Sparsity	99.69%	99.85%	99.92%
Non-zero	883	1783	3583
Zero	67867	273217	1096417
Size	68730	275000	1100000
Sparsity	98.74%	99.35%	99.67%

Table 4: Sparsity from the Compactification of the problem Model with $\delta = 1/3$

	$h = 0.50$		$h = 0.10$	
	10^{-4}	10^{-5}	10^{-4}	10^{-5}
$\ U\ $	0.476610	0.476655	0.476600	0.476627
$\ F\ $	0.245101	0.245661	0.239198	0.239730
$\ J^*\ $	0.476611	0.476713	0.476611	0.476659
Iter.(k)	70	96	33	49
Time	0.09282	0.11094	0.08385	0.11052
(N, n)	(10,4)	(10,4)	(25,4)	(25,4)

Table 5: Summary of Computational results at optimum parameters ($\rho^* = 1.60, \alpha^* = 2.00, \delta = 1/3$)

N	Obj. values		iterations k (time/sec)	
	[2]	proposed	[2]	proposed
1000	0.4741987	0.475758	16(0.26)	112(0.11)
2000	0.4741986	0.475781	18(0.43)	120(0.32)
4000	0.4744534	0.476010	18(0.84)	132(0.47)
8000	0.4745807	0.476620	24(2.04)	140(1.02)

Table 6: Comparison of Computational results at optimum

The implicit Crank Nicolson method and the 3rd order Simpsons rule were used for the discretization of the constraints and cost functional respectively with high feasibility, unconditionally stability and well-conditioned matrices. The ADMM algorithm converges faster with higher level of accuracy, low processing time and low memory requirements, when compared with that of Ghobadi *et. al* [2]. Moreover, this approach can as well be applied to other PDE-based optimization problems like that of wave equations or problems involving integral (entropy) functions.

Acknowledgments

I am grateful to Prof. Montaz Ali of *The University of The Witwatersrand, SA* for introducing the ADMM Algorithm to me and for his helpful comments in improving the presentation of this article.

References

- [1] J. T. Betts & S. L. Campbell, "Discretize then Optimize, in Mathematics for Industry: Challenges and Frontiers", SIAM (2005).
- [2] K. Ghobadi, N. Nedialkov & T. Terlaky, *On the Discretize then Optimize Approach*", Preprint for Industrial and Systems Engineering, 2009
- [3] O. Olotu & K. A. Dawodu, "On the Discretized Algorithm for Optimal Proportional Control Constrained by Delay Differential Equation", Journal of Math. Theory & Modeling **3** (2013), 157.
- [4] O. Olotu, K. A. Dawodu & T. T. Yusuf, "Algorithm for Solving Multi-Delay Optimal Control Problems Using Modified Alternating Direction Method of Multipliers", Journal of Applied & Comp. Mathematics **8** (2019) 445.
- [5] O. Olotu & K. A. Dawodu, "Numerical Solution to One –Dimensional Multi-Delay System using the Modified Alternating Direction Method of Multipliers", Journal of NAMP **50** (2019) 107.
- [6] B. He & X. Yuan, "Block-wise Alternating Direction Method of Multipliers for Multiple-block Convex Programming and Beyond", SIAM Journal of Comp. Mathematics, **691** (2015), 145.
- [7] B. He, X. Hong & X. Yuan, "On the proximal jacobian decomposition of alm for multiple-block separable convex minimization problems and its relationship to ADMM", Journal of Scientific Computing **66** (2016) 1204.
- [8] S. Kameswaran & L. T. Beigler, "Advantages of Nonlinear Programming-based Methodologies or Inequality Path Constrained Optimal Control Problem - a Numerical Study", SIAM Journal on Scientific Comp., **30** (2008) 957.
- [9] Y. E. Nesterov, "A method for Solving the Convex Programming Problem with Convergence Rate $O(1/k^2)$ ", Dokl. Akad. Nauk SSSR **269** (1983), 543.
- [10] E. Ghadimi, A. Teixeira, I. Shames and M. Johansson , "optimal Parameter Selection for the Alternating Direction Method of Multipliers (ADMM): Quadratic Problems", Linnaeus center, EE Royal Institute of Technology, **60** (2014) 644. <https://doi.org/10.1109/TAC.2014.2354892>
- [11] S. Boyd L. Vandenberghe, *Convex Optimization*, Cambridge University Press NY, USA 2004.