ARTICLE IN PRESS

Computers and Mathematics with Applications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A simple and feasible method for a class of large-scale l^1-problems

Jian-Xun Zhao

Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

Article history:
Received 12 October 2014
Received in revised form 15 April 2015
Accepted 6 June 2015
Available online xxxx

Keywords:
Elliptic optimal control problem
Alternating direction method
Global convergence
Large-scale problems

ABSTRACT

In this paper, we investigate a discretized version of an elliptic optimal control problem which is presented by Stadler (2009). An alternating direction method is proposed to solve this problem and demonstrated as globally convergent. This class of methods is attractive due to its simplicity and thus is adequate for solving large-scale problems. The preliminary numerical results present the efficiency of the proposed method.

1. Introduction

We consider a nonsmooth regularization optimization with constraints given by partial differential equation (PDE), which is described as follows.

The function to be minimized over a domain $\Omega \subset \Re^m$ is given by

$$J(y,u) := \frac{1}{2}\|y-\bar{y}\|_{L^2}^2 + \frac{\beta}{2}\|u\|_{L^2}^2 + \alpha\|u\|_{L^1},$$

where $\alpha, \beta > 0$ are two regularization parameters. The state variable y and the control variable u are linked via a PDF problem which will be taken throughout this paper:

$$-\Delta y = u + f \quad \text{in } \Omega,$$

here, $f \in L^2(\Omega)$. In addition, the control variable is bounded by so-called box constraints:

$$a(x) \le u(x) \le b(x), \quad \text{a.e in } \Omega,$$

where $a(x), b(x) \in L^2(\Omega)$ and $a(x) < 0 < b(x)$. The above presentation can be summarized in the following optimization system:

$$\begin{cases} \min \frac{1}{2}\|y-\bar{y}\|_{L^2}^2 + \frac{\beta}{2}\|u\|_{L^2}^2 + \alpha\|u\|_{L^1} & \text{s.t.} \\ -\Delta y = u + f & \text{in } \Omega, \\ y = \bar{y} & \text{on } \partial\Omega \\ a(x) \le u(x) \le b(x), & \text{a.e in } \Omega. \end{cases} \tag{1.1}$$

The above optimal control model is proposed by Georg Stadler in Ref. [1], which is a nonsmooth regularization problem due to the nonsmooth term $\|u\|_{L^1}$. Nonsmooth regularization for PDE-constrained optimization are mainly used for

E-mail address: tonnypanda@126.com.

http://dx.doi.org/10.1016/j.camwa.2015.06.004

processing inverse problems [2–5]. In particular, the use of the L^1-norm of the gradient as regularization has been verified that can achieve better results for the recovery of data from noisy measurements than smoothing regularization. The author in Ref. [1] introduces the original optimal control problem and discusses the advantages of nonsmooth regularization rather than smoothing regularization. For more details, please refer to Ref. [1].

In this paper, we mainly do research on the discretization of the above optimal problem, especially on the first-order conditions of it. One of the popular methods for discretizing the optimization problem is to use the five-point finite difference approximation. The final discretized version of (1.1) is presented by the following optimization problem:

$$\begin{cases} \min & \frac{1}{2}\|y-\bar{y}\|_2^2 + \frac{\beta}{2}\|u\|_2^2 + \alpha\|u\|_1 \\ \text{s.t.} & Ay = u + f, \\ & a \le u \le b. \end{cases} \tag{1.2}$$

where $A \in \Re^{n\times n}$ is nonsingular, $\bar{y}, f, a, b \in \Re^n$ with $a < 0 < b$ and $\alpha, \beta > 0$. Moreover, $\|\cdot\|_2$ denotes Euclidean norm in $\Re^n$ and $\|\cdot\|_1$ is the l^1-norm, i.e., $\|u\|_1 = \sum_{i=1}^n |u_i|$.

In the past decades, large amount of related works have been studied for solving optimization problems involved l^1-norm, especially, in compressive sensing, signal recovery, image process and so on [6–10]. Although these problems are convex programs, they do demand dedicated algorithm because standard methods, such as interior-point algorithm or Newton algorithm, are prohibitively inefficient for those large-scale problems. It should be noticed that the special structure of the matrix A which origins from the discretized Laplace operator would save the memory storage and enable fast matrix–vector multiplications. As a result, first-order algorithms which are able to take advantage of this feature can do better performance, and thus they are highly desirable. First-order methods which have been used to l^1-problems mainly include gradient projection methods, iterative shrinkage/thresholding methods, fixed-point continuation methods, block coordinate gradient descent methods [11–19] and so on.

It is well known that numerical algorithms have been so extensively researched on developing fast or large-scale algorithms for optimal control problems in the scientific literature. However, it seems still some way to go before efficient solvers can be achieved for the box-constrained optimal control problems governed by an elliptic equation.

In Ref. [1], the authors propose a semismooth Newton algorithm by reformulating the first-order optimality conditions and verify that this method has local convergence property with a superlinear rate. Their numerical experiments show the usefulness of this method for the location of control devices and the efficiency of this method for small-scale problems.

What we are interested in is whether an algorithm can be proposed to solve (1.2) which owns the following features:

(1) It is globally convergent from an arbitrary initial value.
(2) The performance speed of the method is faster than existing ones.
(3) It is computationally feasible for dealing with large-scale problems.

In this paper, we will propose an algorithm—Alternating Direction Method (ADM) and present a convergence theorem on this method. Moreover, we will conduct the numerical experiments to show that our algorithm owns the above features. Local convergent algorithms always need to choose a suitable initial point for its convergence and it is difficult in some cases. Comparing with it, the advantage of global convergent algorithms is that we can randomly set the initial point. It is well known that if we do each iteration of a system by using the semismooth Newton algorithm, it always needs to carry a heavy computational burden. Therefore, the semismooth algorithm cannot provide a desired performance on the computational speed for slightly large-scale problems. However, this difficulty can be overcome by exploiting the sparsity of the matrix A and the structure of ADM. The ADM only needs to compute a few matrix–vector multiplications at each iteration instead of a system of equations, which makes the small computational cost. In addition, using the semismooth method, we have to define some matrices and calculate inverse matrices. Even though the semismooth algorithm terminates in a small number of iterations, each iteration is very time-consuming due to the $O(n^3)$ computational complexity of solving an n-dimensional system of equations, or even fail due to the lack of enough memory to store a dense $n \times n$ matrix, especially when n is very large. It follows that this method cannot solve large-scale problems in practical computation. However, in practical computation, our method neither define matrices nor calculate inverse matrices, which implies that it turns to be possible to cope with large-scale problems by using our algorithm.

This paper is organized as follows. In Section 2, we simply introduce the framework of ADM approach and establish an important theorem which will be applied to solve a subproblem. In Section 3, we propose an ADM for solving (1.2) and verify that it is globally convergent. Some preliminary numerical results and some conclusions are given in Section 4.

In our notation, all vectors are column vectors. $\Re^n$ denotes the space of n-dimensional real column vectors, and $\Re^n_+$ (respectively, $\Re^n_{++}$) represents the nonnegative (respectively, positive) orthant in $\Re^n$. The symbol $\|\cdot\|_2$ behaves as the Euclidean norm in $\Re^n$ and as the induced matrix norm. The matrix I represents the identity matrix of appropriate dimension. Superscript T denotes the transpose operator. We denote the maximum eigenvalue of a matrix A as $\lambda_{\max}(A)$.

2. Preliminaries

Here, we will show a brief review on a general framework of ADM before starting our approach. The basic idea of ADM is from the work of Glowinski and Marrocco [20] and Gabay and Mercier [21]. Let $\theta_1 : \Re^s \to \Re$, $\theta_2 : \Re^t \to \Re$ be convex

Please cite this article in press as: J.-X. Zhao, A simple and feasible method for a class of large-scale l^1-problems, Computers and Mathematics with Applications (2015), http://dx.doi.org/10.1016/j.camwa.2015.06.004