Applied Mathematics and Computation 231 (2014) 9–15

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A decomposition method for large-scale box constrained optimization

Jing Yu [a,*], Mingqiang Li [b], Yongli Wang [b], Guoping He [b]

[a] School of Economics and Management, Beihang University, Beijing 100191, China
[b] College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China

ARTICLE INFO

Keywords:
Box constrained optimization
Decomposition method
Working set identification
Large-scale
KKT-violating index

ABSTRACT

A decomposition method for solving large-scale box constrained optimization is proposed. The algorithm is motivated by the successful use of the decomposition method presented by Joachims for training support vector machines. In particular, a new technique, based on the new definition "KKT-violating index", is introduced for working set identification. Finally, the numerical experiments and implementation details show that this method is practical for large-scale problems.

1. Introduction

In this paper, we mainly consider the following box constrained optimization problem

$$min \quad f(x), \tag{1}$$

$$s.t. \quad l \leqslant x \leqslant u, \tag{2}$$

where $f : R^n \rightarrow R$ is continuously differential, $x, l, u \in R^n$ with $l < u$. The gradient of f at x is denoted by $\nabla f(x) = [\nabla f(x)_1, \nabla f(x)_2, \ldots, \nabla f(x)_n]^T$. Let $\Omega = \{x \in R^n : l \leqslant x \leqslant u\}$.

The box constrained optimization problems are widely used in the elastic–plastic torsion problem [1], journal bearing lubrication [2], molecular conformation analysis [3], the obstacle problem [4], optimal design problems [5], inversion problems in elastic wave propagation [6] and so on. Thus, the development of numerical algorithms to efficiently solve (1) and (2), especially when the dimension is large, is important in both theory and application.

At present, many different approaches are developed for solving the box constrained optimization problems such as active set methods [7–11], trust region methods [12–14], interior-point methods [15,16] and so on. Among them, the active set methods always need to explore a working set (face) at each iteration, which is an estimation of active set at the solution. Then, with the variables in the working set being fixed, an unconstrained subproblem with reduced dimension is solved to obtain a new iterate. Hence, the efficiency and convergence property of active set methods mainly depend on the selection of working set and its updating strategy. Initially, only a single active constraint can be added to or deleted from the active set at each iteration which will slow down the convergence rate when dealing with large-scale problems [17]. To accelerate the convergence, various projection techniques and their improvements are proposed to identify the working set which can add or delete several constraints at one time [4,18,19]. Later, some active set identification functions are also proposed to identify the working set [8,20]. Nonetheless, although the dimension of subproblem in an active-set method is much lower than that

* Corresponding author.
E-mail address: yujingbuaa@163.com (J. Yu).

0096-3003/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amc.2013.12.169

of the original problem, it may be still very large if the original problem is of high dimension. This will probably lead to computational difficulties in some cases.

Recently, studies on decomposition methods for training support vector machines become very popular [21–23]. Numerical experiments also show that these methods are effective especially for large-scale SVM problems [24–26]. In a decomposition method, a working set is also determined at each iteration. However, different from the one defined in an active set method, the working set in the decomposition methods always include very little constraints, such as only two in the SMO methods [25]. This makes the subproblem easy to solve and significantly reduces the computation burden.

In this paper, motivated by the successful use of the decomposition method in training support vector machines, we provide a decomposition method for general nonlinear programs with box constraints. A new technique is proposed for the selection of working set based on a new definition of "KKT-violating index" in the following text. At each iteration, a small-scale subproblem with respect only to the variables in the working set needs to be solved. Moreover, the global convergence is established by changing the subproblem into a proximal item. And the numerical experiments show that the algorithm is practical for large-scale problems.

The paper is organized as follows. The decomposition method and its convergence property are presented in Section 2. In Section 3, we introduce a working set identification method which is based on the "KKT- violating index", and also give the theoretical basis. Section 4 is devoted to some preliminary numerical experiments for large scale problems and implementation details of working set identification. Conclusions are given in the last section.

2. The decomposition method

In this section, a new decomposition method is presented for solving box-constrained optimization problem based on the technique proposed by Joachims [24]. The main steps are as follows.

Algorithm 2.1.

Step 1. Initialization. Choose a feasible point $x^1 \in \Omega$. Let B and N denote the working set and non-working set, respectively.
Step 2. Identification. Find the indices corresponding to the nonzero components of the solution to the following problem

$$\min \quad \nabla f(x^k)^T d, \tag{3}$$
$$\text{s.t.} \quad d_i \geqslant 0 \quad x_i^k = l_i, \tag{4}$$
$$d_i \leqslant 0 \quad x_i^k = u_i, \tag{5}$$
$$-1 \leqslant d_i \leqslant 1, \tag{6}$$
$$|\{d_i | d_i \neq 0\}| \leqslant q. \tag{7}$$

Let B include these indices and N include the remaining indices.
Step 3. If $d = 0$, then x^k is a KKT point, stop.
Step 4. Solve the following subproblem

$$\min \quad f(x_B) + t\|x_B - x_B^k\|^2, \tag{8}$$
$$\text{s.t.} \quad l_B \leqslant x_B \leqslant u_B. \tag{9}$$

Let x_B^{k+1} be the optimal solution.

Step 5. Set $x^{k+1} = \begin{pmatrix} x_B^{k+1} \\ x_N \end{pmatrix}$, $k = k+1$, go to Step2.

Remark. The notation x_B indicate the subvector of x made up of the component x_i, with $i \in B$. Similarly, x_N, l_B, u_B are all defined in this way.

In the algorithm, it is desirable to select a set of variables such that the current iteration will make much progress toward the minimum of $f(x)$. Here, we use the strategy based on Zoutendijk's method [27], which uses a first-order approximation to the objective function. This idea is to find a feasible steepest descent direction d whose number of non-zero elements is less or equal to q. Then the current working set B is composed with the indices corresponding to these non-zero elements of d. In step 2, a linear programming problem need to be solved to identify working set, but actually, a much easier approach will be introduced in next section, and a rigorous theorem and reasoning process are presented to verify that the working set identification in our study is equivalent to solving the linear programming problem. Obviously, these are emphases of this paper.

Besides, in step 4, different from the decomposition method proposed by Joachims [24], the objective function of the subproblem is changed by $f(x_B) + t\|x_B - x_B^k\|^2$, where $t > 0$ is a constant [22]. Then we have for all k, $f(x^{k+1}) + t\|x^{k+1} - x^k\|^2 \leqslant f(x^k)$ which implies that

$$\|x^{k+1} - x^k\| \to 0. \tag{10}$$