Journal of Computational Mathematics, Vol.18, No.1, 2000,1–12.

SOLVING TRUST REGION PROBLEM IN LARGE SCALE OPTIMIZATION[*1)]

Bing-sheng He
(*Department of Mathematics, Nanjing University, Nanjing 210093, China*)

Abstract

This paper presents a new method for solving the basic problem in the "model-trust region" approach to large scale minimization: Compute a vector x such that $\frac{1}{2}x^T Hx + c^T x = \min$, subject to the constraint $\|x\|_2 \le a$. The method is a combination of the CG method and a projection and contraction (PC) method. The first (CG) method with $x_0 = 0$ as the start point either directly offers a solution of the problem, or—as soon as the norm of the iterate greater than a, —it gives a suitable starting point and a favourable choice of a crucial scaling parameter in the second (PC) method. Some numerical examples are given, which indicate that the method is applicable.

Key words: Trust region problem, Conjugate gradient method, Projection and contraction method.

1. Introduction

Let H be a given $n \times n$ symmetric positive semidefinite matrix and $c \in R^n$. In this paper we consider the following quadratic programming with a simple quadratical constraint

$$\begin{aligned} &\frac{1}{2}x^T Hx + c^T x = \min \\ &\text{s.t} \quad \|x\|_2 \le a, \end{aligned} \tag{1}$$

where the parameter a is prescribed. This problem occurs frequently in *trust region* method for unconstrained optimization [1]. A number of approaches for solving (1) have been proposed in the literature [2,3,6,12–16]. One technique is to approximate a Lagrange multiplier λ by Newton's method. The approximation of this parameter may be quite delicate, however, and involves the computation of a sequence of singular value decompositions [5]. Since the SVD is too costly for large matrices, the method is applicable only for small problems. The advanced interior point methods could also be used to solve problem (1) and seem attractive, because one can show that these methods converge at a polynomial rate, see e.g. [11]; however, each iteration of an interior point method has to solve a system of linear equations and therefore is rather expensive.

For large and sparse problems, Golub and von Matt [6] presented a method, which uses a series incomplete decompositions and yields a sequence of upper and lower bounds on the Lagrange multiplier and enables them to compute an approximate solution x from these bounds via solving a system of linear equations.

---

* Received February 12, 1996.

1) This work was supported by the National natural Science Foundation of P.R. China and the Natural Science Foundation of Province Jiangsu of China.

Recently, we proposed some projection and contraction (PC) methods [7,8,9] for solving a class of linear variational inequalities, which can be applied to solve problem (1). One of the PC method is rather inexpensive and very simple to realize, because each iteration of the method consists essentially of a matrix vector product and the method does not need to solve any systems of linear equations. However, the performance of the PC method strongly depends on the scaling of H and c.

In this paper, we show—in theory as well as for some numerical examples—that the Lagrange multiplier must be small if the PC method converges slowly. We will pay particular attention to the problem of ill-conditioning and propose an alternative simple (CG-PC) method for solving problem (1). The method start with $x^0 = 0$, if $a \geq \|H^+c\|$, the CG method solves the problems, otherwise, as soon as a $\|x^k\| > a$, we get the information about a proper scaling parameter for H and c, and switch to use some simple projection and contraction (PC) methods [9]. For large trust region problems, the presented method is as simple as the Goldstein's fundamental projection method [4], in addition, as the numerical results will show, it is almost as powerful as the method proposed by Golub and von Matt [6].

1.1. Outline and notation

The equivalent linear projection equation is given in Section 2. In Section 3 we briefly quote some convergence facts of the CG method [10] and the PC method [7,8,9]. In Section 4 we study the convergence behaviour of the PC method and analyze how to treat ill-conditioned problems. Further details of our method are given in Section 5. In Section 6 we present some numerical results.

We use the following notations. A superscript such as in x^k refers to specific vectors and k usually denotes the iteration index. By $\|v\|$ we denote the Euclidean norm of some vector v, by $\|v\|_G$ and the norm $(v^TGv)^{1/2}$ induced by a positive definite matrix G, and by $\|H\|$ we denote the spectral norm $\lambda_{\max}(H)$ of some symmetric matrix H. H^+ denotes the pseudoinverse of a matrix H. Finally, by x^* we denote the solution of the problems.

1.2. Basic observations

An immediate observation regarding problem (1) tells us, if the dimensions of the matrix H are small and a singular value decomposition of H can be computed in moderate time, then for $a \geq \|H^+c\|$ the solution $x = H^+c$ can be computed from the singular value decomposition of H, and for $a < \|H^+c\|$ problem (1) is equivalent to finding a value $\lambda > 0$ such that

$$c^T(H + \lambda I)^{-2}c - a^2 = 0. \tag{2}$$

(Given such λ, the solution x is given by $x = -(H + \lambda I)^{-1}c$.) The derivative with respect to λ of the right hand side is $-2x^T(H+\lambda I)^{-3}x$, and is thus computable directly from the singular value decomposition of $H = V^T\Sigma V$ by observing that $(H + \lambda I)^{-1} = V^T(\Sigma + \mu I)^{-1}V$. Thus some modifications of Newton's method seem appropriate for solving (1), see e.g. [13]. In the following we will assume that H is large and sparse, and does not allow a singular value decomposition in moderate time.

2. The Equivalent Projection Equation

The Lagrange function of problem (1) is

$$L(x, \lambda) = x^THx + 2c^Tx + \lambda(x^Tx - a^2), \tag{3}$$

which is defined on $R^n \times R_+$. The Kuhn-Tucker Theorem of convex programming tells us that x^* is a solution of (1) if and only if there exists a $\lambda^* \geq 0$, such that (x^*, λ^*)