KYBERNETIKA — VOLUME *48* (2012), NUMBER 1, PAGES 123–129

COMPUTING MINIMUM NORM SOLUTION OF A SPECIFIC CONSTRAINED CONVEX NONLINEAR PROBLEM

Saeed Ketabchi and Hossein Moosaei

The characterization of the solution set of a convex constrained problem is a well-known attempt. In this paper, we focus on the minimum norm solution of a specific constrained convex nonlinear problem and reformulate this problem as an unconstrained minimization problem by using the alternative theorem.The objective function of this problem is piecewise quadratic, convex, and once differentiable. To minimize this function, we will provide a new Newton-type method with global convergence properties.

Keywords: solution set of convex problems, alternative theorems, minimum norm solution, residual vector

Classification: 90C05, 90C51

1. INTRODUCTION

In 1988, Mangasarian [6] has characterized the solution set of convex programs for twice continuously differentiable convex functions. This characterization described in the following theorem (see [6]), can be applied to obtain the minimum norm solution of convex optimization problem.

Theorem 1.1. Let $S \subseteq R^n$ be an open convex subset, $f : S \to R$ be a differentiable convex function and $X \subseteq S$ be any convex subset. Consider the problem: $\min_{x\in X} f(x)$. Take that the solution set of this problem is denoted by X^* and $x^* \in X^*$. Then,

$$X^* = \left\{x \in X : \nabla f(x^*)^T x^* = \nabla f(x^*)^T x, \nabla f(x^*) = \nabla f(x)\right\}. \tag{1}$$

At first, in section 2 of this paper we have a theorem characterizing the solution set of a specific convex nonlinear program. It continues by determining the solution of minimum norm for this problem, using alternative theorem and finally in section 3, we will give an example to test the results numerically.

In this article, all vectors will be column vectors and we denote the n-dimensional real space by R^n. By using $A^\dagger$, we denote the Moore–Penrose pseudoinverse of matrix A. The symbols A^T, $\|.\|$ and $\|.\|_\infty$ will denote transpose of matrix A, and Euclidean norm and ∞ norm respectively, and A_i will denote the ith row of matrix A. By using

$Null(A)$, we denote the null space of a matrix A, which is the set of all vectors x for which $Ax = 0$. For $x, y \in R^n$, $x > y$ means that $x_i > y_i$ for $i = 1, 2, \ldots, n$. In vector $a \in R^n$, the plus function a_+ is defined as $(a_+)_i = \max\{0, a_i\}$, $i = 1, 2 \ldots n$ and the scalar product of vectors c and x is denoted by $c^T x$. $\nabla f(x_0)^T d$ is called directional derivative of f in the direction d at x_0 where $\nabla f(x_0)$ is the gradient of f at x_0.

2. ALTERNATIVE THEOREMS AND MINIMUM NORM SOLUTION OF A CONVEX OPTIMIZATION PROBLEM

In this section, we use Theorem 1.1 for a convex programming to study the algorithm that looks for minimum norm solution to a specific constrained convex nonlinear programming. We start by considering the following programming.

$$\begin{aligned} \min_x f(x) = \frac{1}{2}\|(Qx-d)_+\|^2 & \\ \text{subject to} \qquad A_1 x \le b_1, & \\ A_2 x = b_2, & \\ x \ge 0, & \end{aligned} \tag{2}$$

where Q, A_1 and A_2 are $m \times n$, $m_1 \times n$ and $m_2 \times n$ full rank matrices respectively, and $d \in R^m$, $b_1 \in R^{m_1}$ and $b_2 \in R^{m_2}$ are fixed vectors.

Suppose that $X = \{x \in R^n : A_1 x \le b_1, A_2 x = b_2,\ x \ge 0\}$ and X^* is the solution set to problem (2).

Lemma 2.1. Let $X^* \neq \emptyset$ and $x^* \in X^*$, then there exist two submatrices Q_1, Q_2 of Q and two subvectors d_1, d_2 of d respectively, such that for each solution point $x \in X^*$, we have

$$\begin{aligned} (Q_1 x - d_1)_+ = 0, \qquad (Q_2 x - d_2) > 0, \\ {d_2}^T (Q_2 x - d_2) = d^T (Qx^* - d). \end{aligned} \tag{3}$$

P r o o f. By using Theorem 1.1, if $x \in X^*$ then $\nabla f(x^*) = \nabla f(x), \nabla f(x^*)^T x^* = \nabla f(x^*)^T x$ this implies:

$$Q^T (Qx - d)_+ = Q^T (Qx^* - d)_+, \tag{4}$$

$$x^{*T} Q^T (Qx^* - d)_+ = x^T Q^T (Qx^* - d)_+, \tag{5}$$

from (4), we obtain $(Qx - d)_+ = (Q^\dagger)^T Q^T (Qx^* - d)_+$. We use F to denote $(Q^\dagger)^T Q^T$. After some rearrangement, we can rewrite F and d in the following form

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \qquad d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \tag{6}$$

where $F_1(Qx^* - d)_+ = 0$ and $F_2(Qx^* - d)_+ > 0$. Thus, for each $x \in X^*$ we have $(Qx - d)_+ = \begin{bmatrix} F_1(Qx^* - d)_+ \\ F_2(Qx^* - d)_+ \end{bmatrix}$. This implies that there exist two submatrices Q_1, Q_2

of Q and two subvectors d_1, d_2 of d for which $(Q_1x - d_1)_+ = F_1(Qx^* - d)_+ = 0$ and $(Q_2x - d_2)_+ = F_2(Qx^* - d)_+$. Therefore, $(Q_1x - d_1) \leq 0$, and $(Q_2x - d_2) > 0$. Moreover, since $\|(Qx^* - d)_+\|^2 = \|(Qx - d)_+\|^2$, from formulas (4), (5) we obtain $d^T(Qx - d)_+ = d^T(Qx^* - d)_+$ and then we have ${d_2}^T(Q_2x - d_2) = d^T(Qx^* - d)_+$. Thus, $x \in X^*$ satisfies the system (3), and our proof is complete. □

The equation $(Q_1x - d_1)_+ = 0$ implies that $Q_1x \leq d_1$ and the equation $(Q_2x - d_2)_+ = F_2(Qx^* - d)_+$ implies that $Q_2x = F_2(Qx^* - d)_+ + d_2$. Also, from ${d_2}^T(Q_2x - d_2)_+ = d^T(Q_2x^* - d_2)_+$ we obtain ${d_2}^T(Q_2x - d_2) = d^T(Qx^* - d)_+$. On the other hand, if $x \in X$ and it also satisfies system (3), then by using $(Qx - d)_+ = (Q^\dagger)^T Q^T(Qx^* - d)_+$, we will have $\|(Qx - d)_+\| = \|(Qx^* - d)_+\|$. Hence, $x \in X^*$, and we have proven the following theorem.

Theorem 2.2. Suppose that the conditions of Lemma 2.1 hold, and assume that $p^* = (Qx^* - d)_+$. Then, X^* is characterized by the following system

$$\begin{array}{lll} A_1x \leq b_1, & Q_1x \leq d_1, & \\ A_2x = b_2, & Q_2x = F_2p^* + d_2 \quad , & {d_2}^T Q_2 x = d^T p^* + {d_2}^T d_2, \\ x \geq 0. & & \end{array} \tag{7}$$

We can, therefore, rewrite (7) as a linear system

$$\begin{aligned} Ax &\leq b, \\ A_{eq}x &= b_{eq}, \\ x &\geq 0, \end{aligned} \tag{8}$$

where $A = \begin{bmatrix} A_1 \\ Q_1 \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ d_1 \end{bmatrix}$, $A_{eq} = \begin{bmatrix} A_2 \\ Q_2 \\ {d_2}^T Q_2 \end{bmatrix}$ and $b_{eq} = \begin{bmatrix} b_2 \\ F_2p^* + d_2 \\ d^Tp^* + d_2^T d_2 \end{bmatrix}$.
The alternative system to (8) is

$${A_{eq}}^T u - A^T v \leq 0, \quad {b_{eq}}^T u - b^T v = \rho > 0, \quad v \geq 0, \tag{9}$$

where ρ is an arbitrary fixed positive number.

As we show below, the alternative system can be applied to obtain the minimum norm solution of (2). To do so, we introduce the following constrained minimization problem for the residual vector of system (9). Find

$$\min_{u,v \geq 0} g(u, v), \tag{10}$$

where

$$g(u, v) \equiv \frac{1}{2}\left(\|({A_{eq}}^T u - A^T v)_+\|^2 + |({b_{eq}}^T u - b^T v) - \rho|^2\right).$$

By combining the objective function of (10) and constraints $v \geq 0$ into a penalty function, we can define the penalty function as

$$\min_{u,v} \widetilde{g}(u,v), \tag{11}$$

where

$$\widetilde{g}(u,v) \equiv \frac{1}{2}\left(\|({A_{eq}}^T u - A^T v)_+\|^2 + |({b_{eq}}^T u - b^T v) - \rho|^2 + \mu\|(-v)_+\|^2\right).$$

We will minimize the objective function of (11) without constraints for a series of increasing values of μ.

Theorem 2.3. Suppose $X^* \neq \emptyset$ and $x^* \in X^*$, also suppose that the minimum norm solution of (2) is denoted by $\widehat{x}$, then

$$\widehat{x} = \frac{({A_{eq}}^T u^* - A^T v^*)_+}{(\rho - {b_{eq}}^T u^* + b^T v^*)}, \tag{12}$$

where $\begin{bmatrix} u^* \\ v^* \end{bmatrix}$ is the solution of (11).

Proof. The objective function to problem (10) is a convex piecewise quadratic and is bounded from below by 0. Thus, this problem has a global solution $\begin{bmatrix} u^* \\ v^* \end{bmatrix}$ and since g is convex, then $\frac{\partial g}{\partial u}(u^*,v^*) = 0$ and $\frac{\partial g}{\partial v}(u^*,v^*) \geq 0$. These relations imply that

$$A_{eq}({A_{eq}}^T u^* - A^T v^*)_+ - b_{eq}(\rho - {b_{eq}}^T u^* + b^T v^*) = 0, \tag{13}$$

$$-A({A_{eq}}^T u^* - A^T v^*)_+ + b(\rho - {b_{eq}}^T u^* + b^T v^*) \geq 0. \tag{14}$$

Since $X^* \neq \emptyset$ then the system (9) is infeasible. This implies that $g(u^*,v^*) > 0$ and therefore $(\rho - {b_{eq}}^T u^* + b^T v^*) > 0$ (see[2, 3]). Thus from (13) and (14) we obtain $\widehat{x} \geq 0$ and it satisfies the system (8). Moreover, since $\widehat{x} = \frac{({A_{eq}}^T u^* - A^T v^*)_+}{(\rho - {b_{eq}}^T u^* + b^T v^*)}$, then the following relation holds for $\widehat{x}$ (see[8]).

$$\widehat{x} - \frac{({A_{eq}}^T u^* - A^T v^*)_+}{(\rho - {b_{eq}}^T u^* + b^T v^*)} \geq 0, \tag{15}$$

$$\widehat{x}^T(\widehat{x} - \frac{({A_{eq}}^T u^* - A^T v^*)_+}{(\rho - {b_{eq}}^T u^* + b^T v^*)}) = 0, \tag{16}$$

but, from (15) and (16) we conclude that $\widehat{x}$ satisfies the KKT optimality conditions for minimum norm solution to the system (8) and this implies that $\widehat{x}$ is minimum-norm solution of (2) and so the proof is complete. □

3. AN EXAMPLE AND NUMERICAL RESULTS

In this section, we begin our discussion on method given in above algorithm, regarding convex programming, by considering the case where only equality constraints are present. Let us now consider the following equality-constrained nonlinear problem

$$\begin{aligned} & \min_x f(x) \\ \text{subject to} \quad & Sx = s, \\ & x \geq 0, \end{aligned} \tag{17}$$

where $f(x) \equiv \frac{1}{2}\|(Qx-d)_+\|^2$, Q and S are $m\times n$ and $k\times n$ full rank matrices respectively, and $d \in R^m$ and $s \in R^k$ are fixed vectors.

The objective function to problem (10) is $g(u,v)$ which is piecewise quadratic, convex and once differentiable. Suppose that u and $w \in R^m$. For gradient of $g(u,v)$ we have

$$\|\nabla g(u,v) - \nabla g(w,t)\| \leq \|B\|\|B^T\|(\|u-w\|^2 + \|v-t\|^2)^{\frac{1}{2}},$$

where $B=\begin{bmatrix} A_{eq} \\ -A \end{bmatrix}$. This means that ∇g is globally Lipschitz continues with a constant $K = \|B\|\|B^T\|$. Thus, for this function we can define the generalized Hessian matrix which is a symmetric positive semidefinite matrix of the form: [7, 8]

$$\nabla^2 g(u,v) = BD(z)B^T,$$

where $D(z)$ denotes a diagonal matrix with i-diagonal element, then z_i equals to 1 if $({A_{eq}}^T u - A^T v)_i > 0$, otherwise $z_i = 0$.

Similarly, $\nabla\widetilde{g}(u,v)$ is globally Lipschitz continuous and the generalized Hessian of $\widetilde{g}(u,v)$ is

$$\nabla^2\widetilde{g}(u,v) = B^T D(z)B + \mu\widetilde{D}(z),$$

where $\widetilde{D}(z)$ denotes a diagonal matrix whose ith diagonal entry z_i is equal to 1 if $(-v)_i > 0$; z_i is equal to 0 if $(-v)_i \leq 0$.
par Therefore, we can use the generalized Newton method to solve this problem [4, 5].

In the following algorithm we apply the generalized Newton method with a line-search based on the Armijo rule [1].

Algorithm: Generalized Newton Method with the Armijo Rule. Choose any $y_0 = [{u_0}^\top, {v_0}^\top]^\top$ and $\epsilon > 0$, $i = 0$;

while $\|\nabla\widetilde{g}(y_i)\|_\infty \geq \epsilon$

Choose $\alpha_i = \max\{c, c\delta, c\delta^2, \ldots\}$ *such that the next inequality holds.*

$\widetilde{g}(y_i) - \widetilde{g}(y_i + \alpha_i p_i) \geq -\delta\eta\nabla\widetilde{g}(y_i)^T p_i,$

where $p_i = -\nabla^2\widetilde{g}(y_i)^{-1}\nabla\widetilde{g}(y_i)$, $c > 0$ is a constant, $\delta \in (0,1)$ and $\eta \in (0,1)$.

$y_{i+1} = y_i + \alpha_i p_i$,

$i = i + 1$;

end———————————————————————

In this algorithm, the generalized Hessian may be singular, thus we use a modified Newton direction Choleski factorizations as follows:

$$M^T M = (\nabla^2 \widetilde{g}(y_k) + \gamma I),\ p_k = -(M^T M)^{-1} \nabla \widetilde{g}(y_k),$$

where M is an upper triangular matrix, γ is a small positive number and I is the identity matrix.

Now, we next give an example to illustrate this strategy:

Example 3.1. Consider problem (17) for which,

$$Q = \begin{bmatrix} 1 & 0 & -2 & 1 & 0 \\ 0 & -1 & 0 & 2 & -1 \end{bmatrix},\ d = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
$$S = \begin{bmatrix} 0 & 3 & 0 & -2 & -1 \\ 2 & 0 & -5 & 3 & 0 \end{bmatrix},\ s = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Define the vector $x^* = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$ and note that $Qx^* = 0$, $\|(Qx^* - d)_+\| = 0$ and $Sx^* = 0$. From these relations we may be able to tell that x^* is a global minimizer of (17). Also, it is obvious that $\widehat{x} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$ is minimum norm solution of (17).

Here, we will try to obtain the minimum norm solution of this problem by using our approach. Since $Null(Q) = \{0\}$ and $p^* = (Qx^* - d)_+ = 0$, we will have, $A = Q$, $b = d$ and $A_{eq} = S$, $b_{eq} = s$. Thus,

$$\begin{aligned} g(u,v) &= \frac{1}{2}(\|(S^T u - Q^T v)_+\|^2 + |(s^T u - d^T v) - \rho|^2) \\ &= \frac{1}{2}(\|(S^T u - Q^T v)_+\|^2 + |d^T v + \rho|^2). \end{aligned}$$

By solving the $\min_{u,v \geq 0} g(u,v)$ problem, we obtain
$u^* = \begin{bmatrix} 3.9096e-016 \\ 3.9501e-016 \end{bmatrix}$, $v^* = \begin{bmatrix} 1.1590e-015 \\ -6.9321e-016 \end{bmatrix}$. Therefore, we have

$$\widehat{x} = \frac{(S^T u^* - Q^T v^*)_+}{\rho + d^T v^*} = \frac{1}{\rho} \begin{bmatrix} 0 \\ 4.7969e-016 \\ 3.4289e-016 \\ 6.3055e-016 \\ 0 \end{bmatrix}.$$

It shows that $\|\widehat{x}\| = \frac{8.6328e-016}{\rho} \approx 0$. Thus, $\widehat{x}$ is the minimum norm solution of (17).

ACKNOWLEDGEMENT

The authors are very grateful to an anonymous referee for a very careful reading of the paper and helpful comments and suggestions which led to the improvement of the originally submitted version of this work.

(Received November 10, 2010)

REFERENCES

[1] L. Armijo: Minimazation of functions having Lipschitz-continuous first partial derivatives. Pacific J. Math. *16* (1966), 1–3.

[2] Yu. G. Evtushenko and A. I. Golikov: New perspective on the theorems of alternative. In: High Performance Algorithms and Software for Nonlinear Optimization, Kluwer Academic Publishers B.V., 2003, pp. 227–241.

[3] A. I. Golikov and Yu. G. Evtushenko: Theorems of the alternative and their applications in numerical methods. Comput. Math. and Math. Phys. *43* (2003), 338–358.

[4] C. Kanzow, H. Qi, and L. Qi: On the minimum norm solution of linear programs. J. Optim. Theory Appl. *116* (2003), 333–345.

[5] S. Ketabchi and E. Ansari-Piri: On the solution set of convex problems and its numerical application. J. Comput. Appl. Math. *206* (2007), 288–292.

[6] O. L. Magasarian: A simple characterization of solution sets of convex programs. Oper. Res. Lett. *7* (1988), 21–26.

[7] O. L. Magasarian: A Newton method for linear programming. J. Optim. Theory Appl. *121* (2004), 1–18.

[8] O. L. Magasarian: A finite Newton method for classification. Optim. Meth. Software *17* (2002), 913–930.

Saeed Ketabchi, Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, P.O. Box 1914, Rasht. Iran.
e-mail: sketabchi@guilan.ac.ir

Hossein Moosaei, Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, P.O. Box 1914, Rasht. Iran.
e-mail: hmoosaei@gmail.com