Saeed Ketabchi; Hossein Moosaei
Computing minimum norm solution of a specific constrained convex nonlinear problem

Kybernetika, Vol. 48 (2012), No. 1, 123--129

Persistent URL: http://dml.cz/dmlcz/142066

Terms of use:

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library http://project.dml.cz
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 $S \subseteq \mathbb{R}^n$ be an open convex subset, $f : S \rightarrow \mathbb{R}$ 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,

$$X^* = \{ x \in X : \nabla f(x^*)^T x = \nabla f(x^*)^T x, \nabla f(x^*) = \nabla f(x) \}.$$ (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 $\mathbb{R}^n$. By using $A^\dagger$, we denote the Moore–Penrose pseudoinverse of matrix $A$. The symbols $A^T$, $\| \cdot \|$ and $\| \cdot \|_{\infty}$ will denote transpose of matrix $A$, and Euclidean norm and $\infty$ norm respectively, and $A_i$ will denote the $i$th 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 ∈ R^n, x > y means that x_i > y_i for i = 1, 2, . . . , n. In vector a ∈ R^n, the plus function a_+ is defined as (a_+)_i = \max\{0, a_i\}, i = 1, 2 . . . n and the scalar product of vectors c and x is denoted by c^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.

\[
\begin{align*}
\min_x f(x) &= \frac{1}{2} \| (Qx - d)_+ \|^2 \\
\text{subject to} & \quad A_1 x \leq b_1, \\
& \quad A_2 x = b_2, \\
& \quad x \geq 0,
\end{align*}
\]

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

Suppose that \( X = \{ x \in R^n : A_1 x \leq b_1, A_2 x = b_2, x \geq 0 \} \) 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

\[
\begin{align*}
(Q_1 x - d_1)_+ &= 0, \quad (Q_2 x - d_2) > 0, \\
d_2^T (Q_2 x - d_2) &= d^T (Q x^* - d).
\end{align*}
\]

Proof. 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:

\[
\begin{align*}
Q^T (Q x - d)_+ &= Q^T (Q x^* - d)_+, \\
x^T Q^T (Q x^* - d)_+ &= x^T Q^T (Q x^* - d)_+.
\end{align*}
\]

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

\[
F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \quad d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix},
\]

where \( F_1 (Q x^* - d)_+ = 0 \) and \( F_2 (Q x^* - d)_+ > 0 \). Thus, for each \( x \in X^* \) we have

\[
(Q x - d)_+ = \begin{bmatrix} F_1 (Q x^* - d)_+ \\ F_2 (Q x^* - d)_+ \end{bmatrix}.
\]

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 $\|Qx^*-d\|^2_2 = \|Qx-d\|^2_2$, 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.

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

$$
\begin{align*}
A_1x & \leq b_1, \\
A_2x & = b_2, \\
Q_2x & = F_2p^* + d_2, \\
Q_1x & \leq d_1, \\
x & \geq 0.
\end{align*}
$$

(7)

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

$$
\begin{align*}
Ax & \leq b, \\
A_{eq}x & = b_{eq}, \\
x & \geq 0,
\end{align*}
$$

(8)

where $A = \begin{bmatrix} A_1 \\ Q_1 \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ d_1 \end{bmatrix}$, $A_{eq} = \begin{bmatrix} A_2 \\ Q_2 \\ d_2^TQ_2 \end{bmatrix}$ 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

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

(9)

where $\rho$ 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

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

(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

\[
\min_{u,v} \tilde{g}(u,v),
\]

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

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

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

\[
\hat{x} = \frac{(A_{eq}^T u^* - A^T v^*)_+}{(\rho - b_{eq}^T u^* + b^T v^*)},
\]

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

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

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

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 \( \hat{x} \geq 0 \) and it satisfies the system (8). Moreover, since \( \hat{x} = \frac{(A_{eq}^T u^* - A^T v^*)_+}{(\rho - b_{eq}^T u^* + b^T v^*)} \), then the following relation holds for \( \hat{x} \) (see [8]).

\[
\hat{x} - \frac{(A_{eq}^T u^* - A^T v^*)_+}{(\rho - b_{eq}^T u^* + b^T v^*)} \geq 0,
\]

\[
\hat{x}^T (\hat{x} - \frac{(A_{eq}^T u^* - A^T v^*)_+}{(\rho - b_{eq}^T u^* + b^T v^*)}) = 0,
\]

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

\[
\min_x f(x) \quad (17)
\]

subject to \( Sx = s, \)
\( x \geq 0, \)

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 \( d \in \mathbb{R}^m \) and \( s \in \mathbb{R}^k \) 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 \( w \in \mathbb{R}^m \). For gradient of \( g(u, v) \) we have

\[
\| \nabla_g(u, v) - \nabla_g(w, t) \| \leq \| B \| \| B^T \| (\| u - w \|^2 + \| v - t \|^2)^{\frac{3}{2}},
\]

where \( B = \left[ \begin{array}{c} A_{eq} \\ -A \end{array} \right] \). This means that \( \nabla g \) 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, \( \nabla\tilde{g}(u, v) \) is globally Lipschitz continuous and the generalized Hessian of \( \tilde{g}(u, v) \) is

\[
\nabla^2\tilde{g}(u, v) = B^T D(z)B + \mu \tilde{D}(z),
\]

where \( \tilde{D}(z) \) 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 \( y_0 = [u_0^T, v_0^T]^T \) and \( \epsilon > 0, i = 0; \)

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

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

\[
\tilde{g}(y_i) - \tilde{g}(y_i + \alpha_i p_i) \geq -\delta \eta \nabla\tilde{g}(y_i)^T p_i,
\]

where \( p_i = -\nabla^2\tilde{g}(y_i)^{-1} \nabla\tilde{g}(y_i), \) \( c > 0 \) is a constant, \( \delta \in (0, 1) \) and \( \eta \in (0, 1) \).
\[ y_{i+1} = y_i + \alpha ip_i, \]
\[ i = i + 1; \]
end

In this algorithm, the generalized Hessian may be singular, thus we use a modified Newton direction Choleski factorizations as follows:
\[
M^T M = (\nabla^2 g(y_k) + \gamma I), \quad p_k = -(M^T M)^{-1}\nabla g(y_k),
\]
where \( M \) is an upper triangular matrix, \( \gamma \) 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,
\[
Q = \begin{bmatrix} 1 & 0 & -2 & 1 & 0 \\ 0 & -1 & 0 & 2 & -1 \end{bmatrix}, \quad d = \begin{bmatrix} 1 \\ 1 \end{bmatrix},
\]
\[
S = \begin{bmatrix} 0 & 3 & 0 & -2 & -1 \\ 2 & 0 & -5 & 3 & 0 \end{bmatrix}, \quad s = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]

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 \( \hat{x} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T \) is minimum norm solution of (17).

Here, we will try to obtain the minimum norm solution of this problem by using our approach. Since \( \text{Null}(Q) = \{0\} \) and \( p^* = (Qx^* - d)_+ = 0 \), we will have, \( A = Q, \quad b = d \) and \( A_{eq} = S, \quad b_{eq} = s \). Thus,
\[
g(u, v) = \frac{1}{2}((S^T u - Q^T v)_+)^2 + |(s^T u - d^T v) - \rho|^2) \]
\[
= \frac{1}{2}((S^T u - Q^T v)_+)^2 + |d^T v + \rho|^2).
\]

By solving the \( \min_{u,v \geq 0} g(u, v) \) problem, we obtain
\[
u^* = \begin{bmatrix} 3.9096e - 016 \\ 3.9501e - 016 \end{bmatrix}, \quad v^* = \begin{bmatrix} 1.1590e - 015 \\ -6.9321e - 016 \end{bmatrix} \] .
Therefore, we have
\[
\hat{x} = \frac{(S^T u^* - Q^T v^*)_+}{\rho + d^T v^*} = \frac{1}{\rho} \begin{bmatrix} 0 \\ 4.7969e - 016 \\ 3.4289e - 016 \\ 6.3055e - 016 \\ 0 \end{bmatrix}.
\]

It shows that \( \|\hat{x}\| = \frac{8.6328e-016}{\rho} \approx 0 \). Thus, \( \hat{x} \) 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


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