Research Article
A New Augmented Lagrangian Method for Equality Constrained Optimization with Simple Unconstrained Subproblem

Hao Zhang\(^1,2\) and Qin Ni\(^1\)

\(^1\)College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
\(^2\)College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

Correspondence should be addressed to Hao Zhang; zhanghao@njau.edu.cn

Received 3 May 2017; Accepted 16 August 2017; Published 19 October 2017

Academic Editor: Seenith Sivasundaram

Copyright © 2017 Hao Zhang and Qin Ni. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a new method for equality constrained optimization based on augmented Lagrangian method. We construct an unconstrained subproblem by adding an adaptive quadratic term to the quadratic model of augmented Lagrangian function. In each iteration, we solve this unconstrained subproblem to obtain the trial step. The main feature of this work is that the subproblem can be more easily solved. Numerical results show that this method is effective.

1. Introduction

In this paper, we consider the following equality constrained optimization:

\[
\begin{align*}
\min_{x} & \quad f(x) \\
\text{s.t.} & \quad c_i(x) = 0, \quad i = 1, \ldots, m,
\end{align*}
\] (1)

where \(x \in \mathbb{R}^n, f(x) : \mathbb{R}^n \to \mathbb{R}\), and \(c_i(x) : \mathbb{R}^n \to \mathbb{R}\) are twice continuously differentiable.

The method presented in this paper is a variant of the augmented Lagrangian method (denoted by AL). In the late 1960s, AL method was proposed by Hestenes [1] and Powell [2]. Later Conn et al. [3, 4] presented a practical AL method and proved the global convergence under the LICQ condition. Since then, AL method attracted the attentions of many scholars and many variants were presented (see [5–11]). Up to now, there are many computer packages based on AL method, such as LANCELOT [4] and ALGENCAN [5, 6]. In the past decades, AL method was fully developed. Attracted by its well performance, there are still many scholars devoted to research AL method and its applications in recent years (see [7, 8, 11–15]).

For (1), we define the Lagrangian function

\[
l(x, \lambda) = f(x) - \lambda^T c(x)
\] (2)

and the augmented Lagrangian function

\[
L(x, \lambda, \sigma) = l(x, \lambda) + \frac{\sigma}{2} \|c(x)\|^2,
\] (3)

where \(\lambda\) is called the Lagrangian multiplier and \(\sigma\) is called the penalty parameter. In this paper, \(\|\cdot\|\) refers to the Euclidean norm.

In a typical AL method, at the \(k\)th step, for given \(\lambda_k\) and \(\sigma_k\), an unconstrained subproblem

\[
\min_{x \in \mathbb{R}^n} L(x, \lambda_k, \sigma_k)
\] (4)

is solved to find the next iteration point. Then, the multiplier and penalty parameter are updated by some rules. For convenience, for given \(\lambda_k\) and \(\sigma_k\), we define

\[
\Phi_k(x) = L(x, \lambda_k, \sigma_k) = l(x, \lambda_k) + \frac{\sigma_k}{2} \|c(x)\|^2.
\] (5)

Motivated by the regularized Newton method for unconstrained optimization (see [16–19]), we construct a new
subproblem of (1). At the kth iteration point \( x_k \), \( \Phi_k(x_k + s) \) is approximated by the following quadratic model:

\[
\bar{m}_k(s) = I(x_k, \lambda_k) + (\nabla \phi(x_k, \lambda_k))^T s + \frac{1}{2} s^T B_k s + \frac{\sigma_k}{2} \left\| c(x_k) + A(x_k) s \right\|^2
\]

\[
= I(x_k, \lambda_k) + (\nabla_x f(x_k) - A(x_k)^T \lambda_k)^T s + \frac{1}{2} s^T B_k s + \frac{\sigma_k}{2} \left\| c(x_k) + A(x_k) s \right\|^2
\]

\[
= \Phi_k(x_k) + \left[ \nabla_x f(x_k) - A(x_k)^T \lambda_k + \sigma_k A(x_k)^T c(x_k) \right]^T s + \frac{1}{2} s^T \left[ B_k + \sigma_k A(x_k)^T A(x_k) \right] s,
\]

where \( A(x_k) = [\nabla_v c_1(x_k), \ldots, \nabla_v c_m(x_k)]^T \) and \( B_k \) is a positive semidefinite approximation of \( \nabla_x^2 \phi(x_k, \lambda_k) \). Let

\[
\tilde{g}_k = \nabla \phi_k(x_k)
\]

\[
= \nabla_x f(x_k) - A(x_k)^T \lambda_k + \sigma_k A(x_k)^T c(x_k);
\]

\[
\tilde{B}_k = B_k + \sigma_k A(x_k)^T A(x_k),
\]

thus we have

\[
\bar{m}_k(s) = \Phi_k(x_k) + \tilde{g}_k^T s + \frac{1}{2} s^T \tilde{B}_k s.
\]

In [14, 15], \( \bar{m}_k(s) \) is minimized within a trust region to find the next iteration point. Motivated by the regularized Newton method, we add a regularization term to the quadratic model \( m_k(s) \) and define

\[
m_k(s) = \bar{m}_k(s) + \frac{1}{2} \mu_k \| s \|^2
\]

\[
= \Phi_k(x_k) + \tilde{g}_k^T s + \frac{1}{2} s^T \tilde{B}_k s + \frac{1}{2} \mu_k \| s \|^2,
\]

where \( \mu_k \) is called regularized parameter. At the kth step of our algorithm, we solve the following convex unconstrained quadratic subproblem:

\[
\min_{s \in \mathbb{R}^n} m_k(s) = \Phi_k(x_k) + \tilde{g}_k^T s + \frac{1}{2} s^T \tilde{B}_k s + \frac{1}{2} \mu_k \| s \|^2
\]

for finding the trial step \( s_k \). Then, we compute the ratio between the actual reduction and predicted reduction

\[
\rho_k = \frac{\text{Ared}_k}{\text{Pred}_k} = \frac{\Phi_k(x_k) - \Phi_k(x_k + s_k)}{\Phi_k(x_k) - m_k(x_k + s_k)}.
\]

When \( \rho_k \) is close to 1, we accept \( x_k + s_k \) as the next iteration point. At the same time, we think the quadratic model \( m_k(s) \) is a sufficiently "good" approximation of \( \Phi_k(x_k + s) \) and reduce the value of \( \mu_k \). Conversely, when \( \rho_k \) is close to zero, we set \( x_{k+1} = x_k \) and increase the value of \( \mu_k \), by which we wish to reduce the length of the next trial step. This technique is similar to the update rule of trust region radius. Actually, sufficiently large \( \mu_k \) indeed reduces the length of the trial step \( s_k \). However, the regularized parameter is different from trust region radius. In [14, 15], the authors construct a trust region subproblem

\[
\min_{s \in \Delta_k} \bar{m}_k(s) = \Phi(x_k) + g_k^T s + \frac{1}{2} s^T \tilde{B}_k s.
\]

The exact solution \( s_k \) of (12) satisfies the first-order critical conditions if there exists some \( \bar{m}_k \geq 0 \) such that \( \tilde{B}_k + \bar{m}_k I \) is positive semidefinite and

\[
\bar{g}_k + (\tilde{B}_k + \bar{m}_k I) s_k = 0
\]

\[
\| s_k \| - \Delta_k = 0,
\]

while the first-order critical condition of (10) is

\[
\bar{g}_k + (\tilde{B}_k + \mu_k) s_k = 0.
\]

Equations (13) and (15) can show the similarities and differences between regularized subproblem (10) and trust region subproblem (12). It seems that the parameter \( \nu_k \) plays a role similar to the multiplier \( \tilde{m}_k \) in the trust region subproblem. But, actually, the update rule of \( \nu_k \) (see (26)) shows that \( \nu_k \) is not the approximation of \( \tilde{m}_k \). The update of \( \nu_k \) depends on the quality of last trial step \( s_{k-1} \) and has no direct relation with system (13).

To establish the global convergence of an algorithm, some kind of constraint qualification is required. There are many well-known constraint qualifications, such as LICQ, MFCQ, CRCQ, RCR, CPLD, and RCPLD. In case there are only equality constraints, LICQ is equivalent to MFCQ in which \( \{ \nabla_c(x) \mid i = 1, \ldots, m \} \) has full rank; CRCQ is equivalent to CPLD in which any subset of \( \{ \nabla_c(x) \mid i = 1, \ldots, m \} \) maintains constant rank in a neighborhood of \( x \); RCR is equivalent to RCPLD in which \( \{ \nabla_c(x) \mid i = 1, \ldots, m \} \) maintains constant rank in a neighborhood of \( x \). RCPLD is weaker than CRCQ, and CRCQ is weaker than LICQ. In this paper, we use RCPLD which is defined in the following.

**Definition 1.** One says that RCPLD holds at a feasible point \( x^* \) of (1), if there exists a neighborhood \( U(x^*) \) of \( x^* \) such that \( \{ \nabla_c(x) \mid i = 1, \ldots, m \} \) maintains constant rank for all \( x \in U(x^*) \).

The rest of this paper is organized as follows. In Section 2, we give a detailed description of the presented algorithm. The global convergence is proved in Section 3. In Section 4, we present the numerical experiments. Some conclusions are given in Section 5.

**Notations.** For convenience, we abbreviate \( \nabla_x f(x_k) \) to \( g_k \), \( f(x_k) \) to \( f_k \), \( c(x_k) \) to \( c_k \), and \( A(x_k) \) to \( A_k \). In this paper, \( \lambda^{(t)} \) denotes the ith component of the vector \( \lambda \).
2. Algorithm

In this section, we give a detailed description of the proposed algorithm.

As mentioned in Section 1, we solve the unconstrained subproblem (10) to obtain the trial step \( s_k \). Since \( B_k \) is at least positive semidefinite and \( B_k = B_k + \sigma_k A_k^T A_k \) is positive definite as \( \mu_k > 0 \). Therefore, (10) is a strictly convex quadratic unconstrained optimization. \( s_k \) solves (10) if and only if

\[
\bar{g}_k + (\bar{B}_k + \mu_k I) s_k = 0
\]

holds. Global convergence does not depend on the exact solution of (15), although the linear system (15) is easy to be solved. For minimizer of (10) along the direction \(-\bar{g}_k\), specifically, we consider the following subproblem:

\[
\min_{\alpha \geq 0} m_k (-\alpha \bar{g}_k) = \Phi_k(x_k) - \alpha \|\bar{g}_k\|^2 + \frac{1}{2} \alpha^2 \bar{g}_k^T \bar{B}_k \bar{g}_k + \frac{1}{2} \bar{\sigma}_k \|\bar{g}_k\|^2.
\]

(16)

If \( \|\bar{g}_k\| > 0 \), then the minimizer of (16) is \( \alpha_k^* = \|\bar{g}_k\|^2 / (\bar{g}_k^T \bar{B}_k \bar{g}_k + \mu_k \|\bar{g}_k\|^2) \). Therefore, at the \( k \)th step, it follows that

\[
\text{Pred}_k = \Phi_k(x_k) - m_k (-\alpha_k^* \bar{g}_k) = \|\bar{g}_k\|^2 / \left( 2 \left( \bar{g}_k^T \bar{B}_k \bar{g}_k + \mu_k \|\bar{g}_k\|^2 \right) \right).
\]

(17)

By direct calculation, we have that

\[
\text{Pred}_k \geq \min_{s \in \mathbb{R}^n} \frac{1}{2} \|s\|^2 - \frac{1}{2} \left( \frac{1}{\|B_k\|} \right) \|\mu_k \| \|\bar{g}_k\|^2.
\]

(18)

In Section 3, we always suppose that (18) holds.

In a typical AL algorithm, the update rule of \( \sigma_k \) depends on the improvement of constraint violation. A commonly used update rule is that if \( \|s_k\| < \tau_k \|s_k\| \), where \( 0 < \tau_k < 1 \), one may think that the constraint violation is reduced sufficiently and thus \( \sigma_{k+1} = \sigma_k \) is a good choice. Otherwise, if \( \|s_k\| \geq \tau_k \|s_k\| \), one thinks that the current penalty parameter can not sufficiently reduce the constraint violation and increase it in the next iteration. In [20], Yuan proposed a different update rule of \( \sigma_k \) for trust region algorithm. Specifically, if

\[
\text{Pred}_k < \delta_k \sigma_k \min \left\{ \|s_k\| \min \left\{ \left. \frac{\|\bar{g}_k\|}{\mu_k}, \|\bar{g}_k\| \right\} \right\},
\]

(19)

\( \sigma_k \) is increased. In (19), \( \delta_k \) is an auxiliary parameter such that \( \delta_k \sigma_k \) tends to zero. We slightly modify (19) in our algorithm. Specifically, if

\[
\text{Pred}_k < \delta_k \sigma_k \min \left\{ \frac{\|\bar{g}_k\| \max \left\{ 1, \|\bar{g}_k\| \right\}}{\mu_k}, \|\bar{g}_k\| \right\},
\]

(20)

\( \sigma_k \) is increased.

In typical AL method, next iteration point \( x_{k+1} \) is obtained by minimizing \( L(x, \lambda_k, \sigma_k) \). In most AL methods, \( \lambda_{k+1} \) satisfies that \( \|\nabla L(x_{k+1}, \lambda_{k+1}, \sigma_k)\| < \epsilon_k \), where \( \epsilon_k \) is a controlling parameter which tends to zero. As

\[
\|\nabla L(x_{k+1}, \lambda_{k+1}, \sigma_k) - A_{k+1}^T (\lambda_{k+1} - \sigma_k c_k) \| = 0
\]

(21)

when \( \epsilon_k \) is sufficiently small, \( \lambda_{k+1} - \sigma_k c_k \) is a good estimate of the next multiplier \( \lambda_{k+1} \). As we obtain \( x_{k+1} \) by minimizing \( m_k(s) \), the critical point of \( m_k(s) \) has no direct relation to \( \|\nabla L(x_{k+1}, \lambda_{k+1}, \sigma_k)\| \). Therefore, the update rule \( \lambda_{k+1} = \lambda_k - \sigma_k c_k \) does not suit our algorithm. We obtain \( \lambda_{k+1} \) by approximately solving the following least squares problem:

\[
\min_{\lambda \in \mathbb{R}^n} \frac{1}{2} \|g_{k+1} - A_{k+1}^T \lambda\|^2.
\]

(22)

Most AL algorithms require that \( \{\lambda_k\} \) is bounded to ensure the global convergence. Hence, all components of \( \lambda_k \) are restricted to certain interval \([\bar{\lambda}, \bar{\lambda}]\). This technique is also used in our algorithm.

Now, we give the detailed algorithm in the following.

Algorithm 2.

Step 0 (initialization). Choose the parameters \( 0 < \gamma_1 < \gamma_2 < 1, 0 < \theta_1 < 1 \leq \theta_2 < \theta_3, \overline{\lambda} > \underline{\lambda} \). Determine \( x_0 \in \mathbb{R}^n, \lambda_0 \in \mathbb{R}^m, \sigma_0 > 0, \delta_0 > 0, B_0 \in \mathbb{R}^{m \times n} \). Let \( R_0 = \max \|c(x_0)\|, 1 \). Set \( k = 0 \).

Step 1 (termination test). If \( \|\bar{g}_k\| = 0 \) and \( \|\bar{c}_k\| = 0 \), return \( x_k \) as a KKT point. If \( \|\bar{g}_k\| = 0 \), \( \|\bar{c}_k\| > 0 \), and \( A_k^T \bar{c}_k = 0 \), return \( x_k \) as an infeasible KKT point.

Step 2 (determine the trial step). Evaluate the trial step \( s_k \) by solving

\[
\min_{s \in \mathbb{R}^n} m_k(s) = \Phi_k(x_k) + \frac{1}{2} s^T B_k s + \frac{1}{2} \mu_k \|s\|^2.
\]

(23)

such that (18) holds. Compute the ratio between the actual reduction to the predicted reduction

\[
\rho_k = \frac{\text{Pred}_k}{\text{Alred}_k}
\]

(24)

where \( \text{Pred}_k = \Phi_k(x_k) - \Phi_k(x_k + s_k), \text{Alred}_k = \Phi_k(x_k) - m_k(x_k + s_k) \). Set

\[
x_{k+1} = \begin{cases} x_k + s_k, & \rho_k \geq \gamma_1, \\ x_k, & \rho_k < \gamma_1. \end{cases}
\]

(25)

\[
\mu_{k+1} = \begin{cases} \theta \mu_k, & \rho_k \geq \gamma_2, \\ \theta \mu_k, & \gamma_1 \leq \rho_k < \gamma_2, \\ \theta \mu_k, & \rho_k < \gamma_1. \end{cases}
\]

(26)

Step 3 (update the penalty parameter). If

\[
\text{Pred}_k < \delta_k \sigma_k \min \left\{ \frac{\|\bar{g}_k\| \max \left\{ 1, \|\bar{g}_k\| \right\}}{\mu_k}, \|\bar{g}_k\| \right\},
\]

(27)
3. Global Convergence

In this section, we discuss the global convergence of Algorithm 2. We assume that Algorithm 2 can find an infinite set \( \{x_k\} \) and give some assumptions in the following.

**Assumptions 1.** (A1) \( f(x) \) and \( c(x) \) are twice continuously differentiable.

(A2) \( \{x_k\} \) and \( \{B_k\} \) are bounded, where \( B_k \) is positive semidefinite approximation of \( \nabla^2 f(x_k, \lambda_k) \).

Firstly, we give a result on the upper bound of the trial step.

**Lemma 4.** If \( s_k \) solves subproblem (23), then one has

\[
\|s_k\| \leq \frac{4\|g_k\|}{\mu_k} \tag{32}
\]

or \( \mu_k \leq 2\|B_k\| \) hold \( \forall k \geq 0 \).

**Proof.** Any approximate solution \( s_k \) of (23) satisfies \( \Phi_k(x_k) - m_k(s_k) \geq 0 \). Clearly,

\[
0 \leq \Phi_k(x_k) - m_k(s_k) = -g_k^T s_k - \frac{1}{2} s_k^T B_k s_k - \frac{1}{2} \mu_k \|s_k\|^2
\]

\[
\leq \|g_k\| \|s_k\| + \frac{1}{2} \|B_k\| \|s_k\|^2 - \frac{1}{2} \mu_k \|s_k\|^2
\]

\[
= \left( \|g_k\| - \frac{1}{2} \mu_k \|s_k\| \right) \|s_k\| + \frac{1}{2} \|B_k\| \|s_k\| - \frac{1}{2} \mu_k \|s_k\|^2 .
\]

If \( \|s_k\| = 0 \), then (32) holds. If \( \|s_k\| > 0 \), \((\|g_k\| - (1/4)\mu_k \|s_k\|) \|s_k\| + (1/2) \|B_k\| \|s_k\| - (1/2) \mu_k \|s_k\|^2 \geq 0 \) implies that \( \|g_k\| - (1/4)\mu_k \|s_k\| \geq 0 \) or \( \|B_k\| - (1/2) \mu_k \geq 0 \). Thus we can obtain (32).

Now, we discuss convergence properties in two cases. One is that the penalty parameter \( \sigma_k \) tends to \( \infty \) and the other is that \( \{s_k\} \) is bounded.

3.1. The Case of \( \sigma_k \to \infty \)

**Lemma 5.** Suppose that (A1)-(A2) hold and \( \sigma_k \to \infty \); then there exists a constant \( c^* \) such that \( \|c_k\| \to c^* \).

**Proof.** See Lemma 3.1 in Wang and Yuan [15].

In Lemma 5, if \( c^* > 0 \), then any accumulation point of \( \{x_k\} \) is infeasible. Sometimes (1) is naturally infeasible; in other words, the feasible set \( \{x \mid c(x) = 0\} \) is empty. In this case, we wish to find a minimizer of constraint violation. Specifically, we wish to solve

\[
\min_{x \in \mathbb{R}^n} \|c(x)\|^2 .
\]

The solution of this problem is characterized by

\[
A(x^*)^T c(x^*) = 0 .
\]

In the next theorem, we show that if \( \{c_k\} \) is not convergent to zero, at least one of the accumulation points of \( \{x_k\} \) satisfies (35).

**Theorem 6.** Suppose that (A1)-(A2) hold and \( \sigma_k \to \infty \). If \( \|c_k\| \to c > 0 \), then

\[
\liminf_{k \to \infty} \|A_k^T c_k\| = 0 .
\]

**Proof.** We prove this result by contradiction. Suppose that there exists some \( \tau > 0 \) such that

\[
\|A_k^T c_k\| > 2 \tau, \quad \forall k \geq 0 .
\]

By the definition of \( g_k \) in (7), we know that

\[
\|g_k\| \geq \sigma_k \|A_k^T c_k\| - \|g_k - A_k^T \lambda_k\| .
\]

As \( \{x_k\} \) and \( \{\lambda_k\} \) are bounded, we can deduce the boundedness of \( \|g_k - A_k^T \lambda_k\| \) by (A2); that is, there exists some \( \overline{\tau} > 0 \) such that

\[
\|g_k - A_k^T \lambda_k\| < \overline{\tau} .
\]

By (37), (38), and (39), we can conclude that

\[
\|g_k\| \geq 2 \tau \sigma_k - \overline{\tau} > \tau \sigma_k
\]
holds for all sufficiently large \( k \). By the boundedness of \( B_k \) and \( A_k \), we can conclude that there exists \( M > 0 \), such that

\[
\|B_k\| < M \sigma_k
\]
holds for all sufficiently large $k$, where $\tilde{B}_k$ is defined by (7). By (18), (40), and (41),

$$\text{Pred}_k \geq \frac{\|\tilde{g}_k\|^2}{4} \min \left\{ \frac{1}{\|B_k\|}, \frac{1}{\mu_k} \right\}$$

$$= \frac{\|\tilde{g}_k\|}{4} \min \left\{ \frac{\|\tilde{g}_k\|}{\|B_k\|}, \|\tilde{g}_k\| \right\}$$

$$\geq \frac{\tau \sigma_k}{4} \min \left\{ \frac{\tau}{M}, \frac{\|\tilde{g}_k\|}{\mu_k} \right\}$$

holds for all sufficiently large $k$. By the update rule of $\sigma_k$ and the fact that $\sigma_k \to \infty$, we have that

$$\text{Pred}_k < \delta_k \sigma_k \min \left\{ \|\tilde{g}_k\| \max \left\{ 1, \frac{\|\tilde{g}_k\|}{\mu_k} \right\}, \frac{\|\tilde{g}_k\|}{\mu_k} \right\} \tag{43}$$

holds for infinitely many $k$. As $\|\tilde{g}_k\| > 1$ holds for all sufficiently $k$ by (40), it is easy to see that (42) contradicts to (43) as $\delta_k \sigma_k \to 0$ and $\|\tilde{g}_k\| \to \infty$. Thus we can prove the desired result.

**Lemma 7.** Suppose that (A1)-(A2) hold, $\sigma_k \to \infty$, and $\|c_k\| \to 0$; then

$$\liminf_{k \to \infty} \|\tilde{g}_k\| = 0. \tag{44}$$

**Proof.** Assume that there exists $\tau > 0$ such that

$$\|\tilde{g}_k\| > \tau > 0, \quad \forall k \geq 0. \tag{45}$$

Then, by (18) and (41), we know that, for all sufficiently large $k$,

$$\text{Pred}_k \geq \frac{\tau}{4} \min \left\{ \frac{\tau}{M \sigma_k}, \|\tilde{g}_k\| \|c_k\| \right\}. \tag{46}$$

By the update rule of $\sigma_k$ and $\sigma_k \to \infty$,

$$\text{Pred}_k < \delta_k \sigma_k \min \left\{ \|c_k\| \max \left\{ 1, \|\tilde{g}_k\| \right\}, \frac{\|\tilde{g}_k\|}{\mu_k} \right\} \tag{47}$$

holds for infinitely many $k$. We will prove that (47) contradicts to (46). Let $K$ be the index set containing all $k$ such that (47) holds. Therefore, for all $k \in K$,

$$\text{Pred}_k < \delta_k \sigma_k \|\tilde{g}_k\| \max \left\{ 1, \|\tilde{g}_k\| \right\}, \frac{\|\tilde{g}_k\|}{\mu_k} \tag{48}$$

$$\text{Pred}_k < \delta_k \sigma_k \|\tilde{g}_k\|^2. \tag{49}$$

If there exists an infinite subset $K_1 \subseteq K$ such that $\|\tilde{g}_k\| \leq 1$ holds for all $k \in K_1$, then, by (48), it holds that

$$\text{Pred}_k < \delta_k \sigma_k \|\tilde{g}_k\|, \quad \forall k \in K_1. \tag{50}$$

As $\delta_k \sigma_k \to 0$, $\|c_k\| \to 0$, and $\|\tilde{g}_k\| > \tau$, (50) implies that

$$\text{Pred}_k < \frac{\tau}{4} \frac{\|\tilde{g}_k\|}{\mu_k} \tag{51}$$

holds for all sufficiently large $k \in K_1$. If there exists an infinite subset $K_2 \subseteq K$ such that $\|\tilde{g}_k\| > 1$ holds for all $k \in K_2$, then by (49) we have that

$$\text{Pred}_k < \delta_k \sigma_k \|\tilde{g}_k\| \tag{52}$$

for all $k \in K_2$. Equation (52) also implies (51) as $\delta_k \sigma_k \to 0$ and $\|\tilde{g}_k\| \to 0$. From (28) and (29), it follows that $\delta_k \sigma_k^2 = \delta_0 \sigma_0^2$ holds for all $k \geq 0$. Therefore, by (49) we know that, for all $k \in K$,

$$\text{Pred}_k < \delta_k \sigma_k \|c_k\|^2 = \delta_k \sigma_k^2 \|c_k\|^2 \frac{1}{\sigma_k} = \delta_0 \sigma_0^2 \|c_k\|^2 \frac{1}{\sigma_k} \tag{53}$$

As $\|c_k\| \to 0$, (53) implies that

$$\text{Pred}_k < \frac{\tau}{4} \frac{\tau}{M \sigma_k} \tag{54}$$

holds for all sufficiently large $k \in K$. Thus from (47), we obtain (51) and (54) which contradict to (46). Thus we can complete the proof.

**Theorem 8.** Suppose that (A1)-(A2) hold. If $c_k \to 0$ and $\sigma_k \to \infty$, then there exists one cluster point $x^*$ of $\{x_k\}$ such that $x^*$ is a KKT point of (1) or the RCPLD condition does not hold at $x^*$.

**Proof.** Under the assumptions of this theorem, Lemma 7 implies that there exists an index set $K$ such that $\{x_k \mid k \in K\}$ converge to some $x^*$,

$$\lim_{k \to \infty, k \in K} c_k = 0,$$

$$\lim_{k \to \infty, k \in K} \tilde{g}_k = \lim_{k \to \infty, k \in K} \left[ g_k - A_k^T (\lambda_k - \sigma_k c_k) \right] = 0, \tag{55}$$

where $\tilde{g}_k$ is defined in (7). With the help of Theorem 2 in Andreani et al. [21], (55) imply that $x^*$ is a KKT point or the RCPLD condition does not hold at $x^*$.

### 3.2. The Case of $\{\sigma_k\}$ Being Bounded

In this subsection, without loss of generality, we assume that $\sigma_k = \sigma_0$ for all $k \geq 0$. Thus by the update rule (29), we have that $\delta_k = \delta_0$ and

$$\text{Pred}_k \geq \delta_0 \sigma_0 \min \left\{ \|c_k\| \max \left\{ 1, \|\tilde{g}_k\| \right\}, \frac{\|\tilde{g}_k\|}{\mu_k} \right\} \tag{56}$$

holds for all $k \geq 0$. As $\sigma_k$ remains constant, it follows from (A1) and (A2) that $\{\tilde{g}_k\}$ and $\{B_k\}$ are all bounded. If we define the index set

$$S \triangleq \{ k \mid \rho_k \geq \gamma_1 \},$$

then $x_{k+1} = x_k$ for $k \notin S$.

**Lemma 9.** Suppose that (A1)-(A2) hold and $\sigma_k = \sigma_0$ for all $k \geq 0$. If $1/\mu_k \to 0$, as $k \to \infty$, and there exists some constant $\epsilon > 0$ such that

$$\text{Pred}_k > \epsilon \frac{\|\tilde{g}_k\|}{\mu_k} \tag{58}$$

holds for all sufficiently large $k$, then $\sum_{k \in S} (\|\tilde{g}_k\|/\mu_k)$ is divergent.
Proof. We prove this lemma by contradiction. We will show that if \( \sum_{k \in S} (\|\tilde{g}_k\|/\mu_k) \) is convergent, then \( \mu_{k+1} < \mu_k \) holds for all sufficiently large \( k \) which contradicts to the fact that \( 1/\mu_k \to 0 \), as \( k \to \infty \).

Suppose that
\[
\sum_{k \in S} \frac{\|g_k\|}{\mu_k} = s. \tag{59}\]

\( \lim_{k \to \infty} (1/\mu_k) = 0 \) and (32) imply that
\[
\|s_k\| \leq \frac{4 \|\tilde{g}_k\|}{\mu_k} \tag{60}\]
holds for all sufficiently large \( k \). By the definition of \( S \),
\[
\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| = \sum_{k \in S} \|s_k\|. \tag{61}\]
Equations (59)–(61) imply that \( \{x_k\} \) is convergent. Let
\[
r_k = (\rho_k - \gamma_k) \text{Pred}_k = (1 - \gamma_k) \text{Pred}_k + (\text{Ared}_k - \text{Pred}_k). \tag{62}\]
It is clear that \( r_k > 0 \Leftrightarrow \rho_k > \gamma_2 \). By Taylor’s theorem, it holds that
\[
\text{Ared}_k - \text{Pred}_k = m_k (s_k) - \Phi_k (x_k + s_k)
= -\left[\nabla \Phi_k (x_k + \xi_k) - \tilde{g}_k\right]^T s_k
+ \frac{1}{2} s_k^T \tilde{B}_k s_k + \frac{1}{2} \mu_k \|s_k\|^2
\geq -\|\nabla \Phi_k (x_k + \xi_k) - \tilde{g}_k\| \|s_k\|, \tag{63}\]
where \( \xi_k \) is a convex combination of \( x_k \) and \( x_k + s_k \). According to (60), we have that
\[
\text{Ared}_k - \text{Pred}_k \geq -4 \|\nabla \Phi_k (x_k + \xi_k) - \tilde{g}_k\| \frac{\|\tilde{g}_k\|}{\mu_k}. \tag{64}\]
holds for all sufficiently large \( k \) and thus
\[
r_k \geq \left[ c (1 - \gamma_2) - 4 \|\nabla \Phi_k (x_k + \xi_k) - \tilde{g}_k\| \frac{\|\tilde{g}_k\|}{\mu_k} \right]. \tag{65}\]
The convergence of \( \{x_k\} \) and the boundedness of \( \{\lambda_k\} \) imply that \( \|\nabla \Phi_k (x_k + \xi_k) - \tilde{g}_k\| \to 0 \). Therefore, for all sufficiently large \( k, r_k > 0 \). This implies that \( \rho_k > \gamma_2 \) and \( \mu_{k+1} < \mu_k \). \( \square \)

Lemma 10. Suppose that (A1)-(A2) hold and \( \sigma_k = \sigma_0 \) for all \( k \geq 0 \); then we have that
\[
\lim_{k \to \infty} \|c_k\| = 0. \tag{66}\]

Proof. Firstly, we prove that the sum of \( \text{Ared}_k \) is bounded. Define the indices set
\[
K = \{ k : \|c_{k+1}\| \leq R_k \}, \tag{67}\]
where \( R_k \) is defined by Steps 0 and 4 in Algorithm 2. From Step 4 of Algorithm 2, we know that if \( k \notin K \), then \( \|c_{k+1}\| > R_k \). Hence we have
\[
\sum_{k=0}^{\infty} (-\lambda_k^T c_k + \lambda_k^T c_{k+1}) = -\lambda_0^T c_0 + \sum_{k=0}^{\infty} (\lambda_k - \lambda_{k+1})^T c_{k+1}
= -\lambda_0^T c_0 + \sum_{k \in K} (\lambda_k - \lambda_{k+1})^T c_{k+1}
\leq -\lambda_0^T c_0 + 2 \|\lambda_{\max}\| \sum_{k \in K} c_{k+1}
\leq -\lambda_0^T c_0 + 2 \|\lambda_{\max}\| R_k, \tag{68}\]
where \( \|\lambda_{\max}\| \) is the upper bound of \( \{\lambda_k\} \). From Step 4 and (67), we have
\[
\sum_{k \in K} R_k \leq R_0 \left( 1 + \frac{1}{2} + \frac{1}{4} + \cdots \right) = 2R_0, \tag{69}\]
which implies
\[
\sum_{k=0}^{\infty} (-\lambda_k^T c_k + \lambda_k^T c_{k+1}) \leq -\lambda_0^T c_0 + 4R_0 \|\lambda_{\max}\|. \tag{70}\]
Then, we have
\[
\sum_{k=0}^{\infty} \text{Ared}_k = \sum_{k=0}^{\infty} [\Phi_k (x_k) - \Phi_k (x_{k+1})]
= \sum_{k=0}^{\infty} (f_k - f_{k+1}) + \sum_{k=0}^{\infty} (-\lambda_k^T c_k + \lambda_k^T c_{k+1})
+ \frac{\sigma_0}{2} \sum_{k=0}^{\infty} (\|c_k\|^2 - \|c_{k+1}\|^2)
\leq f_0 + (-\lambda_0^T c_0 + 4R_0 \|\lambda_{\max}\|) + \frac{\sigma_0}{2} \|c_0\|^2
\approx \overline{A}, \tag{71}\]
where \( \Phi_k (x) \) is defined by (5).

Secondly, we prove
\[
\lim_{k \to \infty} \inf \|c_k\| = 0 \tag{72}\]
by contradiction. Suppose that there exists some \( \tau > 0 \) such that
\[
\|c_k\| > \tau > 0, \quad \forall k \geq 0. \tag{73}\]
Equations (56) and (73) imply that
\[
\text{Pred}_k \geq \delta_\rho \sigma_0 \min \left\{ \frac{\max \left\{ \|\tilde{g}_k\| \right\}}{\mu_k} r, r^2 \right\}. \tag{74}\]
Considering the sum of \( \text{Pred}_k \) on the index set \( S \) (see (57)), we have by (71) that
\[
\sum_{k \in S} \text{Pred}_k \leq \frac{1}{\gamma_1} \sum_{k \in S} \text{Ared}_k = \frac{1}{\gamma_1} \sum_{k=0}^{\infty} \text{Ared}_k \leq \frac{1}{\gamma_1} \overline{A}. \tag{75}\]
It can be deduced by (74) and (75) that
\[ \sum_{k \in S} \max \{ 1, \frac{\| \tilde{g}_k \|}{\mu_k} \} < +\infty \] 
and thus
\[ \sum_{k \in S} \frac{\| \tilde{g}_k \|}{\mu_k} < +\infty, \] 
(76)
\[ \sum_{k \in S} \frac{1}{\mu_k} < +\infty. \] 
(77)

If \( S \) is a finite set, then it follows from (57) and Step 2 that \( \rho_k \geq \gamma_1 \) and \( \mu_{k+1} = \theta_3 \mu_k \) (\( \theta_3 > 1 \)) hold for all sufficiently large \( k \). Therefore, \( 1/\mu_k \to 0 \), as \( k \to \infty \). If \( S \) is an infinite set, the second inequality in (77) implies that \( 1/\mu_k \to 0 \), as \( k \to \infty \) and \( k \in S \). From Step 2, we know that if \( k \notin S \), then \( \mu_{k+1} \geq \mu_k \). Hence, we have \( 1/\mu_k \to 0 \), as \( k \to \infty \). The fact that \( 1/\mu_k \to 0 \) and (74) imply that
\[ \text{Pred}_k \geq \delta_0 \sigma_k \frac{\| \tilde{g}_k \|}{\mu_k} \] 
(78)
holds for all sufficiently large \( k \). Hence it can be deduced by Lemma 9 that \( \sum_{k \in S}(\| \tilde{g}_k \|/\mu_k) \) is divergent which contradicts the first part in (77).

Finally, we prove (66). If \( S \) is a finite set, then \( \{x_k\} \) is convergent. Thus, (72) implies (66). From now on we assume that \( S \) is an infinite set. Suppose that (66) does not hold; then there exist an infinite index set \( \mathcal{K} = \{k_i\} \) (\( \mathcal{K} \subset S \)) and a constant \( c > 0 \) such that
\[ \| c_k \| > 2c \quad \text{holds} \quad \forall k_i \in \mathcal{K}. \] 
(79)

By (72), there also exists an infinite index set \( \mathcal{K} = \{l_i\} \) (\( \mathcal{K} \subset S \)) such that \( k_i < l_i \),
\[ \| c_{l_i} \| > c \quad \text{holds} \quad \forall k_i \in \mathcal{K}. \] 
(80)
\[ \| c_{l_i} \| < c \quad \text{holds} \quad \forall l_i \in \mathcal{K}. \] 
(81)

Let \( \mathcal{K} = \{k \in S \| \| c_k \| \geq c\} \); then \( \mathcal{K} \subset \mathcal{K} \) and \( \mathcal{K} \) is an infinite index set,
\[ \sum_{k \in \mathcal{K}} \text{Pred}_k \leq \sum_{k \in S} \text{Pred}_k. \] 
(82)

Therefore, by (75), we have that
\[ \sum_{k \in \mathcal{K}} \text{Pred}_k \leq \frac{1}{\gamma_1} \] 
(83)

With the help of (56), (80), and (83), we obtain that
\[ \sum_{k \in \mathcal{K}} \max \{ 1, \frac{\| \tilde{g}_k \|}{\mu_k} \} < +\infty. \] 
(84)

A direct conclusion which can be drawn by (84) is
\[ \frac{1}{\mu_k} \to 0, \quad \text{as} \quad k \to \infty, \] 
(85)
Thus by Lemma 4, we have that, for all sufficiently large \( k \in \mathcal{K} \),
\[ \| s_k \| \leq 4 \frac{\| \tilde{g}_k \|}{\mu_k}. \] 
(86)

Therefore, for sufficiently large \( j \),
\[ \| x_j - x_k \| \leq \sum_{k \in \mathcal{K}} \| s_k \| \leq 4 \sum_{k \in \mathcal{K}} \frac{\| \tilde{g}_k \|}{\mu_k}. \] 
(87)

Equations (84) and (87) imply that \( \| x_j - x_k \| \to 0 \), as \( i \to \infty \). Therefore (79) contradicts (81). Thus we complete the proof.

Lemma 11. Suppose that (A1)-(A2) hold and \( \sigma_k = \sigma_0 \) for all \( k \geq 0 \); then we have
\[ \lim \inf_{k \to \infty} \frac{\| g_k \|}{\mu_k} = 0. \] 
(88)

Proof. Suppose that (88) does not hold. Then there exists \( \tau > 0 \), such that
\[ \frac{\| g_k \|}{\mu_k} > \tau \quad \text{holds} \quad \forall k. \] 
(89)

By (18), we have that
\[ \text{Pred}_k \geq \frac{\tau^2}{4} \min \left\{ \frac{1}{\| B_k \|}, \frac{1}{\mu_k} \right\} \quad \text{holds} \quad \forall k. \] 
(90)

As \( \{ B_k \} \) is bounded above, similar to the second part in the proof of Lemma 10, we can conclude that
\[ \frac{1}{\mu_k} \to 0, \quad \text{as} \quad k \to \infty, \] 
(91)

and thus
\[ \text{Pred}_k \geq \frac{\tau^2}{4} \frac{1}{\mu_k} \quad \text{holds for all sufficiently large} \quad k. \] 
(92)

By (75) and (92), we have that \( \sum_{k \in S}(1/\mu_k) \) is convergent and thus \( \sum_{k \in S}(\| \tilde{g}_k \|/\mu_k) \) is also convergent as \( \{ \tilde{g}_k \} \) is bounded. However, Lemma 9, (91), (92), and the boundedness of \( \{ \tilde{g}_k \} \) deduce the divergence of \( \sum_{k \in S}(\| \tilde{g}_k \|/\mu_k) \). This contradiction completes the proof.

With the help of Lemmas 10 and 11, we can easily obtain the following result.

Theorem 12. Suppose that (A1)-(A2) hold and \( \sigma_k = \sigma_0 \) for all \( k \geq 0 \); there exists an accumulation point of \( \{x_k\} \) at which the KKT condition holds.

Note that, in Theorem 12, we do not suppose that RCPLD holds.
Table 1: Results of Algorithm 2 and ALGENCAN.

<table>
<thead>
<tr>
<th>Name</th>
<th>n</th>
<th>m</th>
<th>$n_f$</th>
<th>$n_g$</th>
<th>$n_f$</th>
<th>$n_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRCRFTA</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>ARGTRIG</td>
<td>200</td>
<td>200</td>
<td>69</td>
<td>69</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>BOOTH</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>BROWNNALE</td>
<td>200</td>
<td>200</td>
<td>8</td>
<td>8</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>BROYDN3D</td>
<td>5000</td>
<td>5000</td>
<td>23</td>
<td>23</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>BYRDSPIR</td>
<td>3</td>
<td>2</td>
<td>19</td>
<td>19</td>
<td>76</td>
<td>78</td>
</tr>
<tr>
<td>BT1</td>
<td>4</td>
<td>2</td>
<td>18</td>
<td>18</td>
<td>106</td>
<td>100</td>
</tr>
<tr>
<td>BT5</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>15</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>BT7</td>
<td>5</td>
<td>3</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>CHNRSNBE</td>
<td>50</td>
<td>98</td>
<td>100</td>
<td>37</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>CLUSTER</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>11</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>CUBENE</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>DECONVNE</td>
<td>63</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>3982</td>
<td>1195</td>
</tr>
<tr>
<td>EIGENB</td>
<td>7</td>
<td>7</td>
<td>2550</td>
<td>2550</td>
<td>2451</td>
<td>608</td>
</tr>
<tr>
<td>GOTTFR</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>HATFLDF</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>HEART6</td>
<td>6</td>
<td>6</td>
<td>122</td>
<td>61</td>
<td>387</td>
<td>216</td>
</tr>
<tr>
<td>HEART8</td>
<td>8</td>
<td>8</td>
<td>29</td>
<td>10</td>
<td>87</td>
<td>43</td>
</tr>
<tr>
<td>HS39</td>
<td>4</td>
<td>2</td>
<td>65</td>
<td>61</td>
<td>106</td>
<td>100</td>
</tr>
<tr>
<td>HYDCAR6</td>
<td>29</td>
<td>29</td>
<td>267</td>
<td>85</td>
<td>212</td>
<td>154</td>
</tr>
<tr>
<td>INTEGREQ</td>
<td>502</td>
<td>500</td>
<td>6</td>
<td>6</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>MARATOS</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>11</td>
<td>42</td>
<td>43</td>
</tr>
<tr>
<td>Mwright</td>
<td>5</td>
<td>3</td>
<td>14</td>
<td>14</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>ORTHREGB</td>
<td>27</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>RECIPE</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>RSNBRNE</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>SINVALNE</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>TRIGGER</td>
<td>7</td>
<td>6</td>
<td>481</td>
<td>290</td>
<td>757</td>
<td>270</td>
</tr>
<tr>
<td>YFITNE</td>
<td>3</td>
<td>17</td>
<td>11</td>
<td>11</td>
<td>805</td>
<td>520</td>
</tr>
<tr>
<td>ZANGWIL3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

4. Numerical Experiment

In this Section, we investigate the performance of Algorithm 2. We compare Algorithm 2 with the famous Fortran package ALGENCAN. In our computer program, the parameters in Algorithm 2 are chosen as follows:

\[
\begin{align*}
\theta_1 &= 0.1, \\
\theta_2 &= 1, \\
\theta_3 &= 4, \\
\gamma_1 &= 0.1, \\
\gamma_2 &= 0.9, \\
\sigma_0 &= 1, \\
\delta_0 &= 1, \\
\lambda &= 10^{20}, \\
\lambda &= 10^{-20}. 
\end{align*}
\]

We set $B_k$ to be the exact Hessian of the Lagrangian $f(x) - \lambda^T c(x)$ at the point $x_k$. The Matlab subroutine minres is used to solve (15). All algorithms are terminated when one of the following conditions holds: (1) $\|g_k + A_k^T \lambda_k\| \leq 10^{-8}$ and $\|s_k\| \leq 10^{-8}$; (2) $\|g_k + A_k^T \lambda_k\| \leq 10^{-8}$ and $\|A_k^T s_k\| \leq 10^{-8}$; (3) $\|s_k\| \leq 10^{-8}$. All test problems are chosen from CUTEst collection [22].

The numerical results are listed in Table 1 where the name of problem is denoted by Name, the number of its variables...
is denoted by $n$, the number of constraints is denoted by $m$, the number of function evaluations is denoted by $n_f$, and the number of gradient evaluations is denoted by $n_g$. In Table 1, we list the results of 38 test problems. Considering the numbers of function evaluations ($n_f$), Algorithm 2 is better than ALGENCAN for 30 cases (78.9%). Considering the numbers of gradient evaluations ($n_g$), Algorithm 2 is better than ALGENCAN for 31 cases (81.6%).

5. Conclusions

In this paper, we present a new algorithm for equality constrained optimization. We add an adaptive quadratic term to the quadratic model of the augmented Lagrangian function. In each iteration, we solve a simple unconstrained subproblem to obtain the trial step. The global convergence is established under reasonable assumptions.

From the numerical results and the theoretical analysis, we believe that the new algorithm can efficiently solve equality constrained optimization problems.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by NSFC (11771210, 11471159, 11571169, and 61661136001) and the Natural Science Foundation of Jiangsu Province (BK20141409).

References

Submit your manuscripts at
https://www.hindawi.com