Journal of Computational Mathematics
Vol.36, No.3, 2018, 331–350.
http://www.global-sci.org/jcm
doi:10.4208/jcm.1705-m2016-0820

AN AUGMENTED LAGRANGIAN TRUST REGION METHOD WITH A BI-OBJECT STRATEGY*

Caixia Kou
School of Sciences, Beijing University of Posts and Telecommunications, Beijing 100876, China.
Email: koucx@lsec.cc.ac.cn.
Zhongwen Chen
School of Mathematical Sciences, Soochow University, Suzhou 215006, China.
Email: zwchen@suda.edu.cn
Yu-Hong Dai
LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
Email: dyh@lsec.cc.ac.cn
Haifei Han
School of Mathematical Sciences, Soochow University, Suzhou, 215006, China.
Email: 572852834@qq.com.

Abstract

An augmented Lagrangian trust region method with a bi-object strategy is proposed for solving nonlinear equality constrained optimization, which falls in between penalty-type methods and penalty-free ones. At each iteration, a trial step is computed by minimizing a quadratic approximation model to the augmented Lagrangian function within a trust region. The model is a standard trust region subproblem for unconstrained optimization and hence can efficiently be solved by many existing methods. To choose the penalty parameter, an auxiliary trust region subproblem is introduced related to the constraint violation. It turns out that the penalty parameter need not be monotonically increasing and will not tend to infinity. A bi-object strategy, which is related to the objective function and the measure of constraint violation, is utilized to decide whether the trial step will be accepted or not. Global convergence of the method is established under mild assumptions. Numerical experiments are made, which illustrate the efficiency of the algorithm on various difficult situations.

Mathematics subject classification: 90C55, 65K05, 90C30
Key words: Nonlinear constrained optimization; Augmented Lagrangian function; Bi-object strategy; Global convergence.

1. Introduction

The augmented Lagrangian (AL) method has become a class of important methods for constrained optimization [3] since its proposition by Hestenes [17] and Powell [22]. Specifically, Conn et al. [11] presented a practical augmented Lagrangian method, based on which a well-known package Lancelot [12] was released. The attractive features of AL methods are that, they can be implemented matrix-free [4,11] and possess fast local convergence guarantees under relatively weak assumptions [1]. Moreover, AL methods have successfully been applied in many

* Received December 28, 2016 / Revised version received April 26, 2017 / Accepted May 16, 2017 / Published online March 28, 2018 /

fields such as compressed sensing [2], distributed optimization and statistical learning [6] and signal reconstruction [25].

One efficient way to implement the AL method is to minimize some approximation model to the augmented Lagrangian function with some trust region at each iteration. On the other hand, it is known that if directly adding the trust region to the linearized constraints, this may lead to inconsistency (for example). To tackle the inconsistency, one may incorporates the linearized constraints into the objective function by virtue of some penalty factor, such as the l_1 exact penalty [13], the l_∞ exact penalty [26] and the l_2 exact penalty [20, 24]. The algorithms in these works can be regarded variants of the AL method. Meanwhile, we can also see that the AL method with the trust region technique circumvents the difficulty of dealing the inconsistency between the linearized constraints and the trust region. It should be noticed here that the use of the l_2 exact penalty will bring standard single-ball trust region subproblems for unconstrained optimization and hence can be solved efficiently by many mature methods.

Further, exact penalty methods, including the AL method, have proved to be effective techniques for solving difficult nonlinear programs. They are successful in solving certain classes of problems, in which standard constraint qualifications are not satisfied [7, 8]. However, there is still some trouble in choosing appropriate values of the penalty parameter. The performance of the AL method suffers when the choice of the penalty parameter and/or Lagrange multipliers is very poor. If the penalty parameter is too large and/or the estimate to the true Lagrangian multipliers is bad, there might be little or no progress in the primal space due to the iterates veering too far away from the feasible region. It is also not suitable to calculate a solution with higher precision if the choice of the penalty parameter and/or Lagrange multipliers is poor. Moreover, the penalty parameter in various approaches proposed in the literature may tend to infinity [7, 24]. Take the following infeasible problem as an example.

$$\begin{array}{ll} \min & (x_1-1)^2 \\ \text{s.t.} & x_2^2+1=0. \end{array} \tag{1.1}$$

It is easy to see that $x^* = (1, 0)^T$ is an infeasible stationary point. The augmented Lagrangian function of problem (1.1) is

$$P(x, \lambda, \sigma) = (x_1 - 1)^2 - \lambda(x_2^2 + 1) + \frac{1}{2}\sigma(x_2^2 + 1)^2.$$

The AL method aims to solve a sequence of unconstrained problems, $\min P(x, \lambda_k, \sigma_k)$, with $\{\sigma_k\}$, $\{\lambda_k\}$ determined in some way. If $\sigma_k > |\lambda_k|$, the augmented Lagrangian function $P(x, \lambda_k, \sigma_k)$ has a unique stationary point $x_k(\lambda_k, \sigma_k) = x^*$. If this happens, we see that the measure of the constraint violation remains unchanged; *i.e.*, $|c(x_k(\lambda_k, \sigma_k))| \equiv 1$. Thus by the mechanism of the AL method, the penalty parameter will be augmented ceaselessly until numerical overflows occur.

The above example shows that the penalty parameter in the AL method may monotonically tend to infinity and result in numerical overflows. It is known that the purpose of the penalization is to enforce the iterates approaching the feasible region. When the iterate is far away from the feasible region, it sounds reasonable to increase the penalty parameter to improve the feasibility. However, when the iterate becomes more and more feasible, we may think of reducing the penalty parameter. In other words, it is more reasonable to ask the penalty parameter to depend on the information about the current iterate. Based on this observation, we shall propose a new augmented Lagrangian trust region algorithm in this paper. The trial step in the new algorithm is determined by minimizing a quadratic approximation model of the augmented Lagrangian function where the trial step depends on the penalty parameter. To choose the penalty parameter, an auxiliary trust region subproblem is introduced related to the constraint violation. It turns out that the penalty parameter need not be monotonically increasing and will not tend to infinity. Moreover, the new algorithm can automatically detect the local infeasibility of the original problem. A bi-object strategy related to the objective function and the measure of constraint violation, which was first studied in [9], is utilized to decide whether the trial step will be accepted or not. The algorithm is proved to be globally convergent under weak assumptions. Further, we will see that the sequence of the penalty parameter does not tend to infinity, avoiding the occurrence of numerical overflows.

In general, numerical methods for constrained optimization can be divided into three classes. The first class uses some merit function, which is normally a combination of the objective function and the constraint violation by virtue of a penalty parameter. It includes the classical sequential quadratic programming methods [5] and the trust region SQP methods [27] and can be called as the penalty-type or merit-function-type method. The second class does not use any merit function nor a penalty parameter and hence can be called the penalty-free method or merit-function-free method. Examples are the filter method [14] and other methods without a filter nor a penalty [15,16,19,23]. The third class falls in between the penalty-type method and the penalty-free one. At each iteration of the method, the trial step is computed by minimizing a quadratic approximation of some merit function while the acceptance criterion does not involve any merit function. For instance, the penalty method with bi-object strategy [9]. It is obvious that the method proposed is of the third class.

The rest of the paper is organized as follows. In Section 2, we describe the details of the new algorithm. In Section 3, the well definedness of the algorithm is analyzed. In Section 4, we analyze the global convergence of the algorithm under weak assumptions. Finally, numerical experiments are reported in Section 5.

2. The Algorithm

Consider the nonlinear equality constrained optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c_i(x)=0,\ i\in\mathcal{E}=\{1,\cdots,m\}, \end{array} \tag{2.1}$$

where $f: R^n \to R$, $c_i : R^n \to R(i \in \mathcal{E})$ are twice continuously differentiable functions. The Lagrangian function of (2.1) is denoted as

$$L(x,\lambda)=f(x)-\lambda^T c(x),$$

where $c=(c_1(x),\cdots,c_m(x))^T$, $\lambda=(\lambda_1,\cdots,\lambda_m)^T$ is a Lagrange multiplier vector. The augmented Lagrangian function is denoted as

$$\Phi(x,\lambda,\sigma)=L(x,\lambda)+\sigma\, v(x)=f(x)-\lambda^T c(x)+\sigma\, v(x),$$

where $\sigma > 0$ is a penalty parameter and

$$v(x)=\frac{1}{2}\|c(x)\|^2.$$

Notice that $v(x)$ is some measure of the constraint violation at a point x.

By introducing the explicit Lagrange multiplier, the AL method avoids the possibility of ill-conditioning and nonsmoothness arising from simple penalty methods. Specifically, at the $k-$th iteration, the AL method minimizes the augmented Lagrangian function with pre-fixed Lagrange multiplier λ_k and penalty parameter σ_k,

$$\min \quad \Phi(x, \lambda_k, \sigma_k) = f(x) - \lambda_k^T c(x) + \sigma_k v(x). \tag{2.2}$$

Now we introduce the quadratic approximation to both the constraint violation $v(x)$ and the augmented Lagrangian function $\Phi(x, \lambda_k, \sigma_k)$. Define

$$m_k(d) = \frac{1}{2}\|c_k + A_k^T d\|^2,$$

where $c_k = c(x_k)$, $A_k = A(x_k)$ and $A(x) = \nabla c(x) = (\nabla c_1(x), \cdots, \nabla c_m(x))$. Further, define a quadratic approximation to $\Phi(x, \lambda_k, \sigma_k)$ (the constant term is omitted)

$$q_k(d, \lambda_k, \sigma_k) = (g_k - A_k\lambda_k)^T d + \frac{1}{2} d^T B_k d + \sigma_k m_k(d), \tag{2.3}$$

where $g_k = \nabla f(x_k)$ and B_k is the Hessian of the Lagrangian function $L(x, \lambda)$ at (x_k, λ_k) or its approximation. The algorithm in this paper involves the solution to two standard single-ball trust region subproblems. The first one is

$$\begin{aligned} \min \quad & q_k(d, \lambda_k, \sigma_k) \\ \text{s.t.} \quad & \|d\| \leq \Delta_k, \end{aligned} \tag{2.4}$$

where $\Delta_k > 0$ is the trust region radius at the current iteration. Since the solution to (2.4) is related to σ_k and Δ_k, we denote it as $d_k(\sigma_k, \Delta_k)$. The subproblem (2.4), which was first introduced by Niu and Yuan [20], is a standard single-ball trust region subproblem. One advantage of this subproblem is that, we do need not worry about the inconsistency between the linearized constraints and the trust region constraint, nor about the linear dependency of the gradients of the constraints. The second trust region subproblem is

$$\begin{aligned} \min \quad & m_k(d) \\ \text{s.t.} \quad & \|d\| \leq \Delta_k^c := \kappa_n \min\{1, (\Delta_k)^\mu\}\Delta_k, \end{aligned} \tag{2.5}$$

where κ_n, $\mu \in (0, 1)$ are some constants. Denote the solution to the subproblem (2.5) as $d_k^\infty = d_k(\sigma_\infty, \Delta_k)$.

Now we shall describe the basic idea of our new algorithm. As is known, Niu and Yuan [20] and Wang and Yuan [24] used the augmented Lagrangian function $\Phi(x, \lambda, \sigma)$ as a merit function to justify whether the trial step $d_k(\sigma_k, \Delta_k)$ to the subproblem (2.4) can be accepted or not. In the two algorithms, it is likely that the penalty parameter goes to infinity monotonically. Nevertheless, even when the penalty parameter goes to infinity, the feasibility cannot be improved if the original problem is not feasible (see the illustrative example (1.1)). Thus, at each iteration, we shall solve the subproblem (2.5) at first. If the following condition holds,

$$m_k(0) = m_k(d_k^\infty) > 0, \tag{2.6}$$

we see that the problem (2.1) is locally infeasible at x_k. We call x_k as an infeasible stationary point of (2.1) and the algorithm terminates. Otherwise, if the condition (2.6) does not hold, we shall look for a suitable Δ_k such that the solution $d_k(\sigma_k, \Delta_k)$ to the subproblem (2.4) is not only a feasible direction to improve the constraint violation but also a descent direction for the objective function $f(x)$. More exactly, we shall ask the penalty factor σ_k to satisfy both

$$m_k(0) - m_k(d_k(\sigma_k, \Delta_k)) \geq \epsilon_1(m_k(0) - m_k(d_k^\infty)), \quad (\epsilon_1 \in (0,1)), \tag{2.7}$$

$$q_k(0, \lambda_k, \sigma_k) - q_k(d_k(\sigma_k, \Delta_k), \lambda_k, \sigma_k) \geq \epsilon_2\sigma_k(m_k(0) - m_k(d_k^\infty)), \quad (\epsilon_2 \in (0, \epsilon_1)). \tag{2.8}$$

We will prove that there exists a suitable σ_k satisfying (2.7) and (2.8) (see Lemma 3.4).

Once a suitable penalty parameter σ_k is found, we shall adopt the bi-objective strategy to decide whether the solution $d_k(\sigma_k, \Delta_k)$ to the subproblem (2.4) is accepted or not. To this aim, we denote $ared_k^f(d_k)$ and $pred_k^f(d_k)$ to be the actual reduction and the predicted reduction of the objective function $f(x)$, respectively,

$$ared_k^f(d_k) = f(x_k) - f(x_k + d_k), \quad pred_k^f(d_k) = -g_k^T d_k - \frac{1}{2} d_k^T B_k d_k.$$

If $d_k = d_k(\sigma_k, \Delta_k)$ is such the following three conditions hold,

$$pred_k^f(d_k) \geq \delta(v_k)^\gamma \qquad (\delta > 0, \gamma > 1), \tag{2.9}$$

$$ared_k^f(d_k) \geq \eta_1^f pred_k^f(d_k), \quad (\eta_1^f \in (0,1)), \tag{2.10}$$

$$v(x_k + d_k) \leq v_k^{max}, \tag{2.11}$$

where v_k^{max} is the current upper bound on the constraint violation, we shall call d_k as $f-$type successful iterate step and set $x_{k+1} = x_k + d_k$. This indicates that if the predicted reduction of $f(x)$ is relatively greater than the constraint violation at the current iteration, we may mainly improve the optimality while keeping the feasibility not to be deteriorated too much. However, if the relation (2.9) does not hold, we need to mainly improve the feasibility. In this case, we turn to consider the solution $d_k = d_k^\infty$ to the subproblem (2.5). Denote $ared_k^c(d_k)$ and $pred_k^c(d_k)$ to be the actual reduction and the predicted reduction of the constraint violation, respectively,

$$ared_k^c(d_k) = v(x_k) - v(x_k + d_k), \quad pred_k^c(d_k) = m_k(0) - m_k(d_k).$$

If $d_k = d_k^\infty$ is such that

$$ared_k^c(d_k) \geq \eta_1^c pred_k^c(d_k) \quad (\eta_1^c \in (0,1)), \tag{2.12}$$

we shall call d_k as $c-$type successful iterate step and set $x_{k+1} = x_k + d_k$ as well. At the same time, we update v_k^{max} by

$$v_{k+1}^{max} := max\{\beta_1 v_k^{max}, v_{k+1} + \beta_2(v_k - v_{k+1})\} \quad (\beta_1, \beta_2 \in (0,1)). \tag{2.13}$$

In the case that (2.12) is not true, we shall decrease the trust region radius Δ_k and repeat the above procedure.

Now we shall give a detailed description of our algorithm.

Algorithm 2.1.

Initialization. Given $x_0 \in R^n$, $\lambda_0 \in R^m$, $B_0 \in R^{n\times n}$, $0 < \epsilon$, κ_n, μ, η_1^f, η_1^c, γ_1, β_1, $\beta_2 < 1$, $0 < \epsilon_2 < \epsilon_1 < 1$, $\delta > 0$, $\gamma > 1$, $\sigma_0 > 0$, $\Delta_{\max} > \Delta_{\min} > 0$, $\Delta_0 \in [\Delta_{\min}, \Delta_{\max}]$, $v_0^{max} = \max\{1, v(x_0)\}$, the maximal number of iteration $nmax$, $k := 0$, $j := 0$.

while $(k \leq nmax)$ **do**
 $\sigma_k^{(0)} = \sigma_k,\ \Delta_k^{(0)} = \Delta_k,\ j := 0.$
 while $(j \geq 0)$**do**
  Solve (2.5) and get a solution $d_k(\sigma_\infty, \Delta_k^{(j)})$.
  if $m_k(0) = m_k(d_k(\sigma_\infty, \Delta_k^{(j)})) > 0$ **then**
   x_k is an infeasible stationary point of (2.1). Stop.
  else
   Solve (2.4) and get a solution $d_k = d_k(\sigma_k^{(j)}, \Delta_k^{(j)})$.
  end if
  if $\max\{\|d_k\|, \|c_k\|\} \leq \epsilon$ **then**
   x_k is an approximate solution to (2.1). Stop.
  end if
  if $m_k(0) = 0$ **then**
   if $pred_k^f(d_k) \leq 0$ or (2.8) does not hold **then**
    $\sigma_k^{(j+1)} = 10\sigma_k^{(j)}$, $\Delta_k^{(j+1)} = \Delta_k^{(j)}$, $j := j+1$. Continue.
   end if
  else
   if (2.7) or (2.8) does not hold **then**
    $\sigma_k^{(j+1)} = 10\sigma_k^{(j)}$, $\Delta_k^{(j+1)} = \Delta_k^{(j)}$, $j := j+1$. Continue.
   end if
  end if
  if (2.9) holds **then**
   if (2.10) and (2.11) hold **then**
    $d_k = d_k(\sigma_k^{(j)}, \Delta_k^{(j)}), x_{k+1} = x_k + d_k, v_{k+1}^{max} = v_k^{max}$, $\Delta_k = \Delta_k^{(j)}$, $\sigma_k = \sigma_k^{(j)}$.
    Break.
  else
   $\Delta_k^{(j+1)} = \gamma_1 \Delta_k^{(j)}, \sigma_k^{(j+1)} = \sigma_k^{(j)}$, $j := j+1$. Continue.
  end if
 else
  if (2.12) holds **then**
   $d_k = d_k(\sigma_\infty, \Delta_k^{(j)})$, $x_{k+1} = x_k + d_k$, $\Delta_{k+1} = \Delta_k^{(j)}$, $\sigma_k = \sigma_k^{(j)}$. Update v_k^{max} by (2.13). Break.
  else
   $\Delta_k^{(j+1)} = \gamma_1 \Delta_k^{(j)}$, $\sigma_k^{(j+1)} = \sigma_k^{(j)}$, $j := j+1$. Continue.
  end if
 end if
end while
Update B_k and λ_k to B_{k+1} and λ_{k+1}, respectively. $\sigma_{k+1} = \|\lambda_{k+1}\|_\infty + 1$, $\Delta_{k+1} \in [\Delta_{min}, \Delta_{max}]$, $k := k+1$.
end while

Remark 2.1. If (2.9) holds, we call the current iteration as $f-$type iteration and the correspondent procedure of solving (2.4) as $f-$type inner cycle, whose main purpose is to improve the objective function with keeping the reasonable feasibility. If (2.9) does not hold, we call the current iteration as $c-$type iteration and the correspondent procedure of solving (2.5) as $c-$type inner cycle, whose main purpose is to improve the measure of the constraint violation. In this case, it is possible that the value of the objective function increases.

Remark 2.2. If $m_k(0) = 0$, then x_k is a feasible point, which implies that the current iteration is an $f-$type iteration. The current trial step $d_k(\sigma_k, \Delta_k)$ is required to reduce the value of the objective function. Therefore Algorithm 2.1 does not need the condition (2.7) in this case.

Remark 2.3. The Lagrangian multiplier estimation λ_k is defined as a least square solution of the following problem

$$\min \ \|g_k - A_k\lambda\|^2. \tag{2.14}$$

3. Well Definedness

We make the following assumptions about problem (2.1) and the iterates $\{x_k\}$ and the matrices $\{B_k\}$ generated by the algorithm.

Assumption A

A1. $f(x)$ and $c_i(x)(i \in \mathcal{E})$ are twice continuously differentiable in R^n.

A2. $\{x_k\}$ generated by Algorithm 2.1 is contained in a bounded convex closed set Ω.

A3. The matrix sequence $\{B_k\}$ is uniformly bounded; *i.e.*, there exists $M_B > 1$ such that $\|B_k\| \leq M_B$ for all k.

By Assumption A, there exist the constants $M_g > 0, M_{ff} > 0$, $M_{AA} > 0$ such that

$$\|g(x)\| \leq M_g \ , \quad \|\nabla^2 f(x)\| \leq M_{ff} \ , \ \|A_k A_k^T\| \leq M_{AA}, \qquad \forall x \in \Omega.$$

The following two lemmas present the fundamental relations in the trust region methods. We only give the results.

Lemma 3.1. *Under Assumption A, we have for all k,*

$$|ared_k^f(d_k) - pred_k^f(d_k)| \leq \frac{1}{2}(M_{ff} + M_B)\Delta_k^2,$$
$$|ared_k^c(d_k) - pred_k^c(d_k)| = O(\|d_k\|^2).$$

Lemma 3.2. *Let d_k^∞ and $d_k(\sigma_k, \Delta_k)$ be the solutions to subproblems* (2.5) *and* (2.4)*, respectively. Then*

$$pred_k^c(d_k^\infty) \geq \frac{1}{2}\|A_k c_k\| \min\left\{\Delta_k, \frac{\|A_k c_k\|}{1 + \|A_k A_k^T\|}\right\}, \tag{3.1}$$

$$-\tilde{g}_k^T d_k(\sigma_k, \Delta_k) - \frac{1}{2} d_k(\sigma_k, \Delta_k)^T (B_k + \sigma_k A_k A_k^T) d_k(\sigma_k, \Delta_k)$$
$$\geq \frac{1}{2}\|\tilde{g}_k\| \min\left\{\Delta_k, \frac{\|\tilde{g}_k\|}{\|B_k + \sigma_k A_k A_k^T\|}\right\}, \tag{3.2}$$

where $\tilde{g}_k = g_k - A_k\lambda_k$.

Lemma 3.3. *If Algorithm* 2.1 *does not terminate at* x_k*, then* $m_k(d_k(\sigma,\Delta_k))$ *is monotonically decreasing with* σ *and*

$$\lim_{\sigma\to+\infty} m_k(d_k(\sigma,\Delta_k)) = m_k(\tilde{d}_k^{\infty}), \tag{3.3}$$

where $\tilde{d}_k^{\infty}$ *is a solution to the following subproblem*

$$\begin{array}{ll} \min & m_k(d) \\ \text{s.t.} & \|d\| \le \Delta_k. \end{array}$$

Proof. Assume that $\sigma_2 > \sigma_1 > 0$, $d_k(\sigma_1,\Delta_k)$ and $d_k(\sigma_2,\Delta_k)$ are the solutions to (2.4) corresponding to σ_1 and σ_2, respectively. Denote

$$q_k^l(d,\lambda_k) = (g_k - A_k\lambda_k)^T d + \frac{1}{2}d^T B_k d.$$

Then $q_k(d,\lambda_k,\sigma) = q_k^l(d,\lambda_k) + \sigma m_k(d)$. Thus,

$$q_k^l(d_k(\sigma_2,\Delta_k),\lambda_k) + \sigma_2 m_k(d_k(\sigma_2,\Delta_k)) \le q_k^l(d_k(\sigma_1,\Delta_k),\lambda_k) + \sigma_2 m_k(d_k(\sigma_1,\Delta_k)),$$

$$q_k^l(d_k(\sigma_1,\Delta_k),\lambda_k) + \sigma_1 m_k(d_k(\sigma_1,\Delta_k)) \le q_k^l(d_k(\sigma_2,\Delta_k),\lambda_k) + \sigma_1 m_k(d_k(\sigma_2,\Delta_k)).$$

Adding these two relations above yields

$$(\sigma_2-\sigma_1)m_k(d_k(\sigma_1,\Delta_k)) \ge (\sigma_2-\sigma_1)m_k(d_k(\sigma_2,\Delta_k)).$$

Thus,

$$m_k(d_k(\sigma_1,\Delta_k)) \ge m_k(d_k(\sigma_2,\Delta_k)).$$

Now we prove the truth of (3.3). In fact, we assume that there exist $\varepsilon_0 > 0$ and $\sigma^{(j)} \to +\infty$ $(j\to\infty)$ such that

$$m_k(d_k(\sigma^{(j)},\Delta_k)) - m_k(\tilde{d}_k^{\infty}) \ge \varepsilon_0, \quad \text{for all } j = 1,2,\cdots,$$

where $d_k(\sigma^{(j)},\Delta_k)$ is a solution to (2.4) corresponding to $\sigma = \sigma^{(j)}$. It follows from

$$q_k(d_k(\sigma^{(j)},\Delta_k),\lambda_k,\sigma^{(j)}) \le q_k(\tilde{d}_k^{\infty},\lambda_k,\sigma^{(j)})$$

that

$$q_k^l(\tilde{d}_k^{\infty},\lambda_k) - q_k^l(d_k(\sigma^{(j)},\Delta_k),\lambda_k) \ge \sigma^{(j)}(m_k(d_k(\sigma^{(j)},\Delta_k)) - m_k(\tilde{d}_k^{\infty})) \ge \sigma^{(j)}\varepsilon_0.$$

Noticing that the left hand term in the above relation is bounded, we obtain a contradiction. Thus the lemma is proved. □

Lemma 3.4. *Assume that Algorithm* 2.1 *does not terminate at* x_k*. If* $m_k(0) = 0$*, there must exist sufficiently large* $\sigma > 0$ *such that* $pred_k^f(d_k(\sigma,\Delta_k)) > 0$ *and* (2.8) *holds. If* $m_k(0) > 0$*, there must exist sufficiently large* $\sigma > 0$ *satisfying* (2.7) *and* (2.8).

Proof. Consider the case that $m_k(0) = 0$. Since Algorithm 2.1 does not terminate at x_k, $\tilde{g}_k \ne 0$. It follows from Lemma 3.3 and (3.2) that $pred_k^f(d_k(\sigma,\Delta_k)) > 0$ and (2.8) holds if $\sigma > 0$ is sufficiently large.

Now we assume that $m_k(0) > 0$, which implies that $m_k(0) > m_k(d_k^\infty)$. Pick any $\sigma_+ > 0$ and let $d_+ = d_k(\sigma_+, \Delta_k)$ be the solution to the subproblem (2.4) corresponding to σ_+. Define

$$\hat{\sigma} = \max\left\{\sigma_+,\ \frac{(g_k - A_k\lambda_k)^T d_+ + \frac{1}{2}d_+^T B_k d_+}{(1-\epsilon_2)m_k(0) - m_k(d_+) + \epsilon_2 m_k(d_k^\infty)}\right\}, \tag{3.4}$$

where $0 < \epsilon_2 < \epsilon_1 < 1$. Let $\hat{d} = d_k(\hat{\sigma}, \Delta_k)$. By the choice of $\hat{\sigma}$ and the optimality of d_+, we can get that

$$\begin{aligned}\epsilon_2\hat{\sigma}[m_k(0) - m_k(d_k^\infty)] &\le -\tilde{g}_k^T d_+ - \frac{1}{2}d_+^T B_k d_+ + \hat{\sigma}[m_k(0) - m_k(d_+)]\\ &\le q_k(0, \lambda_k, \hat{\sigma}) - q_k(\hat{d}, \lambda_k, \hat{\sigma}).\end{aligned}$$

So (2.8) holds. It follows from Lemma 3.3, $\hat{\sigma} \ge \sigma_+$ and $m_k(\tilde{d}_k^\infty) \le m_k(d_k^\infty)$ that

$$m_k(0) - m_k(\hat{d}) \ge m_k(0) - m_k(d_k(\sigma_+)).$$

Thus (2.7) also holds. □

Lemma 3.5. ([16]) *The sequence $\{v_k^{\max}\}$ is monotonically nonincreasing. Further, we have that $v_k^{\max} > 0$ and $0 \le v_k \le v_k^{\max}$ for all k.*

Lemma 3.6. *If Algorithm 2.1 does not terminate at x_k, the algorithm can get an $f-$type or $c-$type successful step after reducing the trust region radius in finite times.*

Proof. By Algorithm 2.1, if it does not terminate at x_k, we can without loss of generality assume that

$$\Delta_k^{(j)} = \gamma_1^j \Delta_k,\ \sigma_k^{(j)} = \sigma_k, \quad \text{for } j = 0, 1, 2, \cdots,$$

where the index j denotes the number of inner cycles. Let $d_k^{(j)} = d_k(\sigma_k^{(j)}, \Delta_k^{(j)})$ be a solution to the subproblem (2.4). We consider the following two cases.

(1) $v_k = 0$. In this case, $m_k(0) = m_k(d_k(\sigma_\infty, \Delta)) = 0$ for any $\Delta > 0$. By (3.2), we know that

$$\begin{aligned}pred_k^f(d_k^{(j)}) &\ge \frac{\sigma_k}{2}\|A_k^T d_k^{(j)}\|^2 + \lambda_k^T A_k^T d_k^{(j)} + \frac{1}{2}\|\tilde{g}_k\| \min\left\{\Delta_k^{(j)}, \frac{\|\tilde{g}_k\|}{\|B_k + \sigma_k A_k A_k^T\|}\right\}\\ &\ge \frac{1}{2}\|\tilde{g}_k\|\Delta_k^{(j)} - \|A_k^T d_k^{(j)}\|\|\lambda_k\|\end{aligned} \tag{3.5}$$

holds for sufficiently large j. By Lemma 3.3, we assume without loss of generality that σ_k is large enough such that

$$pred_k^f(d_k^{(j)}) \ge \frac{1}{4}\|\tilde{g}_k\|\Delta_k^{(j)}.$$

So (2.9) always holds. Furthermore, by Lemma 3.1, we have that

$$|\rho_k^f(d_k^{(j)}) - 1| \le \frac{2(M_B + M_{ff})}{\|\tilde{g}_k\|}\Delta_k\gamma_1^j \to 0 \quad (j \to +\infty),$$

which implies that $\rho_k^f(d_k^{(j)}) \geq \eta_1^f$ holds as long as j large enough. Moreover,

$$v(x_k^{(j)} + d_k^{(j)}) = \frac{1}{2}\|c(x_k + d_k^{(j)})\|^2$$
$$= \frac{1}{2}\|c_k + A_k^T d_k^{(j)} + O(\|d_k^{(j)}\|^2)\|^2, \tag{3.6}$$

which with $c_k = 0$ and Assumption A indicates that there exists a constant $M_c > 0$ such that $v(x_k^{(j)} + d_k^{(j)}) \leq M_c\|d_k^{(j)}\|^2$. Thus $v(x_k^{(j)} + d_k^{(j)}) \leq v_k^{max}$ holds as long as j large enough. Therefore $d_k^{(j)}$ is an $f-$type successful iterate step if j is large enough.

(2) $v_k > 0$. Since $\|d_k^{(j)}\| \leq \Delta_k^{(j)} = \gamma_1^j \Delta_k$, we have that

$$pred_k^f(d_k^{(j)}) = -g_k^T d_k^{(j)} - \frac{1}{2}(d_k^{(j)})^T B_k d_k^{(j)} \to 0, \quad \text{as } j \to \infty.$$

So (2.9) does not hold for all large j. By Algorithm 2.1, all inner cycles can only be $c-$type iterations for large j and the trial step $d_k^{(j)} = d_k(\sigma_\infty, \Delta_k^{(j)})$ in the $j-$th inner cycle.

Since the algorithm does not terminate at x_k and $v_k > 0$, it is easy to know that $A_k c_k \neq 0$. It follows from Lemmas 3.1 and 3.3 that

$$|\rho_k^c(d_k^{(j)}) - 1| = \frac{|ared_k^c(d_k^{(j)}) - pred_k^c(d_k^{(j)})|}{pred_k^c(d_k^{(j)})}$$
$$\leq \frac{O(\|d_k^{(j)}\|^2)}{0.5\|A_k c_k\| \min\left\{\Delta_k^{(j)}, \dfrac{\|A_k c_k\|}{1 + \|A_k A_k^T\|}\right\}} \leq \frac{O\left(\Delta_k^{(j)}\right)}{0.5\|A_k c_k\|} \to 0, \quad \text{as } j \to \infty.$$

So $\rho_k^c(d_k^{(j)}) \geq \eta_1^c$ holds for j large enough, which implies that $d_k^{(j)}$ is a $c-$type successful iterate step. This completes our proof. □

4. Global Convergence

In order to analyze the global convergence properties of Algorithm 2.1, we define

$$\mathcal{C} = \{k | v_{k+1}^{max} \neq v_k^{max}\}.$$

In other words, the set $\mathcal{C}$ is the indices set of all $c-$type successful iterates. Let $|\mathcal{C}|$ denote the cardinal number of the set $\mathcal{C}$. If Algorithm 2.1 does not terminate finitely, we have the following results under Assumption A.

Lemma 4.1. *If $|\mathcal{C}| < +\infty$, every limit point of $\{x_k\}$ is a feasible point of problem* (2.1).

Proof. If $|\mathcal{C}| < +\infty$, there exists an index k_0 such that the iterations are all $f-$type iterations for $k \geq k_0$. Therefore the relations (2.9), (2.10) and (2.11) hold for $k \geq k_0$. Thus,

$$f(x_{k_0}) - f(x_{k+1}) = \sum_{i=k_0}^{k} ared_i^f(d_i) \geq \eta_1^f \sum_{i=k_0}^{k} pred_i^f(d_i).$$

By Assumption A, $\{f(x_k)\}_{k \geq k_0}$ is monotonically decreasing and is bounded below. This with the above relation implies that $\lim_{k\to\infty} pred_k^f(d_k) = 0$. By (2.9), $\lim_{k\to\infty} v_k = 0$. This completes our proof. □

Theorem 4.1. *If* $|\mathcal{C}| < +\infty$*, every limit point of* $\{x_k\}$ *is a stationary point of* (2.1) *or the linear independence constraint qualification fails at the limit point.*

Proof. If $|\mathcal{C}| < +\infty$, there exists an index k_0 such that all iterations are $f-$type iterations for $k \geq k_0$. Therefore, (2.9), (2.10) and (2.11) all hold and $v_k^{max} \equiv v_{k_0}^{max} > 0$ for $k \geq k_0$.

Let $\bar{x}$ be any limit point of $\{x_k\}$. There exists an infinite index set $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} x_k = \bar{x}$. By Lemma 4.1, $v(\bar{x}) = 0$. Assume without loss of generality that $v(x_k) > 0$ for all $k \in \mathcal{K}$.

We now proceed by contradiction and assume that $\bar{x}$ is not a stationary point of (2.1) and that the linear independence constraint qualification holds at $\bar{x}$ (*i.e.*, $\nabla c_1(\bar{x}), \nabla c_2(\bar{x}), \cdots, \nabla c_m(\bar{x})$ are linearly independent). Then there exists a unit vector $\bar{d} \in R^n$ such that

$$g(\bar{x})^T\bar{d} < 0, \quad A(\bar{x})^T\bar{d} = 0, \quad Z(\bar{x})^T g(\bar{x}) \neq 0, \tag{4.1}$$

where $Z(x)$ is a matrix whose columns form an orthogonal basis for the null space $A(x)^T$. If $k_1 \geq k_0$ is large enough, the matrix A_k has full column rank for $k \in \mathcal{K}, k \geq k_1$. Denote

$$\begin{aligned} &P_k = P(x_k) = I - A_k(A_k^TA_k)^{-1}A_k^T = Z_kZ_k^T, \quad Z_k = Z(x_k), \quad k \in \mathcal{K},\ k \geq k_1. \\ &t_k = -A_k(A_k^TA_k)^{-1}c_k, \quad s_k = s(x_k) = P(x_k)\bar{d}/\|P(x_k)\bar{d}\|. \end{aligned} \tag{4.2}$$

Since $s(\bar{x}) = \bar{d}$, it follows from the continuity that there exist $\bar{\epsilon} > 0$ and $k_2 \geq k_1$ such that

$$g_k^Ts_k \leq -\bar{\epsilon},\ A_k^Ts_k = 0,\ \|Z_k^Tg_k\| \geq \bar{\epsilon} \tag{4.3}$$

holds for $k \in \mathcal{K}$, $k \geq k_2$. By Assumption A, there exist $b_1 > 0$ and $b_2 > 0$ such that

$$\|t_k\| = \| - A_k(A_k^TA_k)^{-1}c_k\| \leq b_1\|c_k\| = b_1\sqrt{2v_k}, \tag{4.4}$$

$$v(x_k + d_k) = \frac{1}{2}\|c_k + A_k^Td_k + O(\|d_k\|^2)\|^2 \leq \frac{1}{2}(\|c_k + A_k^Td_k\| + b_2\|d_k\|^2)^2. \tag{4.5}$$

Let $j^{'}$ be the smallest positive integer such that

$$\begin{aligned} (\gamma_1)^j \leq \frac{1}{\Delta_{\max}} \min \Bigg\{ &1, \left(\frac{\bar{\epsilon}}{2\kappa_nM_B}\right)^{\frac{1}{1+\mu}}, \frac{\bar{\epsilon}}{2(1+M_B)}, \frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}(1-\eta_1^f)}{16(M_{ff}+M_B)}, \\ &\left(\frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}}{16M_g + 8M_B}\right)^{\frac{1}{\mu}}, \left(\frac{v_{k_0}^{\max}}{2b_2^2}\right)^{1/4} \Bigg\}. \end{aligned} \tag{4.6}$$

Suppose that $x_{k+1} = x_k + d_k^{(j_k)} (k \geq k_0)$, where $d_k^{(j_k)}$ is an $f-$type successful iterate step, j_k is the number of the corresponding inner cycle iteration. Then, we can prove that $j_k \leq j^{'}$ for all $k \in \mathcal{K}$ large enough.

Since A_k has full column rank for $k \in \mathcal{K}, k \geq k_2$, it follows from (2.14) that $\lambda_k = (A_k^TA_k)^{-1}A_k^Tg_k$, $k \in \mathcal{K}$, $k \geq k_2$, which implies that there exists $M_\lambda > 0$ such that $\|\lambda_k\| \leq M_\lambda$ holds for all $k \in \mathcal{K}, k \geq k_2$. By $\lim_{k\to\infty} v_k = 0$, there exists $k_3 \geq k_2$ such that, for $k \in \mathcal{K}$, $k \geq k_3$,

$$\begin{aligned} &\|c_k\| = \sqrt{2v_k} \\ \leq &\min\left\{\frac{\kappa_n}{b_1}\left((\gamma_1)^{j^{'}}\Delta_{min}\right)^{1+\mu}, \frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}}{16}\Delta_{min}(\gamma_1)^{j^{'}}\min\left\{\frac{\sqrt{2}}{\delta}, \frac{1}{M_\lambda}\right\}\right\}. \end{aligned} \tag{4.7}$$

For $k \in \mathcal{K}, k \geq k_3$, we claim that if at least one of (2.9), (2.10) and (2.11) does not hold for $j = 0, 1, 2, \cdots, j' - 1$, then $d_k^{(j')} = d_k(\sigma_k, \Delta_k^{(j')})$, which is a solution to (2.4) corresponding to $\Delta_k^{(j')} = (\gamma_1)^{j'}\Delta_k$, must be an $f-$type successful step. In fact, let

$$\tilde{d_k}^{(j')} = t_k + (\Delta_{k,j'}^c - \|t_k\|)s_k,$$

where $\Delta_{k,j'}^c = \kappa_n(\Delta_k^{(j')})^{1+\mu}$. Then,

$$c_k + A_k^T \tilde{d_k}^{(j')} = c_k + A_k^T t_k + (\Delta_k^{(j')} - \|t_k\|)A_k^T s_k = 0.$$

Moreover,

$$\begin{aligned}
&\|t_k\| \leq b_1\sqrt{2v_k} \overset{(4.7)}{\leq} \kappa_n\left((\gamma_1)^{j'}\Delta_{min}\right)^{1+\mu} \leq \kappa_n\left(\Delta_k^{(j')}\right)^{1+\mu} = \Delta_{k,j'}^c,\\
&\|\tilde{d_k}^{(j')}\| = \sqrt{\|t_k\|^2 + (\Delta_{k,j'}^c - \|t_k\|)^2\|s_k\|^2} = \sqrt{(\Delta_{k,j'}^c)^2 - 2\|t_k\|(\Delta_{k,j'}^c - \|t_k\|)} \leq \Delta_{k,j'}^c,
\end{aligned}$$

which implies that $\tilde{d_k}^{(j')}$ is a feasible solution to (2.5) and $m_k(\tilde{d_k}^{(j')}) = 0$. Here notice by the choice of j' that $\Delta_k^{(j')} \leq 1$. Thus,

$$m_k(d_k(\sigma_\infty, \Delta_k^{(j')})) = 0.$$

Let u_k be a solution to the following subproblem

$$\begin{array}{ll}\min & \bar{g}_k^T Z_k u + \frac{1}{2}u^T Z_k^T B_k Z_k u\\ \text{s.t.} & \|u\| \leq \bar{\Delta}_{k,j'},\end{array} \tag{4.8}$$

where $\bar{g}_k = g_k + B_k t_k$ and

$$\bar{\Delta}_{k,j'} = \sqrt{(\Delta_k^{(j')})^2 - \|t_k\|^2} \geq \sqrt{1-\kappa_n^2}\Delta_k^{(j')}.$$

Denote $\bar{d}_k^{(j')} = t_k + Z_k u_k$. Then $\|\bar{d}_k^{(j')}\| \leq \Delta_k^{(j')}$,

$$\|Z_k^T\bar{g}_k\| \geq \|Z_k^T g_k\| - \|Z_k^T B_k\|\|t_k\| \geq \bar{\epsilon} - M_B\kappa_n\left(\Delta_k^{(j')}\right)^{1+\mu} \overset{(4.6)}{\geq} \frac{\bar{\epsilon}}{2}. \tag{4.9}$$

$$\begin{aligned}
pred_k^f(\bar{d}_k^{(j')}) &= -\bar{g}_k^T Z_k u_k - \frac{1}{2}u_k^T Z_k^T B_k Z_k u_k - g_k^T t_k - \frac{1}{2}{t_k}^T B_k t_k\\
&\geq \frac{1}{2}\|Z_k^T\bar{g}_k\| \min\left\{\sqrt{1-\kappa_n^2}\Delta_k^{(j')}, \frac{\|Z_k^T\bar{g}_k\|}{1+\|Z_k^T B_k Z_k\|}\right\} - (M_g + 0.5M_B)\left(\Delta_k^{(j')}\right)^{1+\mu}\\
&\geq \frac{\bar{\epsilon}}{4}\sqrt{1-\kappa_n^2}\Delta_k^{(j')} - (M_g + 0.5M_B)\left(\Delta_k^{(j')}\right)^{1+\mu}\\
&\overset{(4.6)}{\geq} \frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2}\Delta_k^{(j')}.
\end{aligned} \tag{4.10}$$

$$q_k(0, \lambda_k, \sigma_k) - q_k(d_k^{(j')}, \lambda_k, \sigma_k) = pred_k^f(d_k^{(j')}) + \sigma_k(m_k(0) - m_k(d_k^{(j')})) + \lambda_k^T A_k^T d_k^{(j')}.$$

Since $\bar{d}_k^{(j')}$ is a feasible solution to (2.4), we have that $q_k(d_k^{(j')}, \lambda_k, \sigma_k) \leq q_k(\bar{d}_k^{(j')}, \lambda_k, \sigma_k)$. Thus,

$$q_k(0, \lambda_k, \sigma_k) - q_k(d_k^{(j')}, \lambda_k, \sigma_k) \geq q_k(0, \lambda_k, \sigma_k) - q_k(\bar{d}_k^{(j')}, \lambda_k, \sigma_k).$$

By (2.7), we have $\|c_k + A_k^T d_k^{(j')}\| \le \|c_k\|$. Consequently,

$$\begin{aligned} pred_k^f(d_k^{(j')}) &\ge pred_k^f(\bar{d}_k^{(j')}) + \sigma_k m_k(d_k^{(j')}) - \lambda_k^T(c_k + A_k^T d_k^{(j')}) \qquad (4.11) \\ &\ge pred_k^f(\bar{d}_k^{(j')}) - \|\lambda_k\|\|c_k + A_k^T d_k^{(j')}\| \\ &\overset{(4.10)}{\ge} \frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2}\Delta_k^{(j')} - M_\lambda\|c_k\| \overset{(4.7)}{\ge} \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\Delta_k^{(j')} \qquad (4.12) \\ &\ge \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\Delta_{min}(\gamma_1)^{j'} \overset{(4.7)}{\ge} \delta\sqrt{v_k} \ge \delta(v_k)^\gamma \end{aligned}$$

holds for $k \in \mathcal{K}, k \ge k_3$, *i.e.*, (2.9) holds for $j = j'$.

It follows from Lemma 3.1 and (4.12) that

$$|\rho_k^f(d_k^{(j')}) - 1| \le \frac{16(M_{ff} + M_B)(\Delta_k^{(j')})^2}{\bar{\epsilon}\sqrt{1-\kappa_n^2}\Delta_k^{(j')}} \overset{(4.6)}{\le} 1 - \eta_1^f, \quad k \in \mathcal{K},\ k \ge k_3.$$

So (2.10) holds for $j = j'$.

By (2.7), $\|c_k + A_k^T d_k^{(j')}\| \le \sqrt{1-\epsilon_1}\|c_k\|$. By (4.5), we have that

$$\begin{aligned} v(x_k + d_k^{(j')}) &\le m_k(d_k^{(j')}) + b_2\|c_k + A_k^T d_k^{(j')}\|\|d_k^{(j')}\|^2 + b_2^2\|d_k^{(j')}\|^4 \\ &\le \frac{1}{2}(1-\epsilon_1)\|c_k\|^2 + b_2\sqrt{1-\epsilon_1}\|c_k\|\|d_k^{(j')}\|^2 + b_2^2\|d_k^{(j')}\|^4, \end{aligned}$$

which with $\lim\limits_{k\to+\infty} c_k = 0$ implies that there exists an index $k_4 \ge k_3$ such that for all $k \ge k_4$,

$$v(x_k + d_k^{(j')}) \le 2b_2^2\|d_k^{(j')}\|^4 \le 2b_2^2\left(\Delta_{\max}\gamma_1^{j'}\right)^4 \overset{(4.6)}{\le} v_{k_0}^{max}.$$

So (2.11) also holds for $j = j'$.

Summarizing the discussion above, we know that $j_k \le j'$, which implies that

$$\Delta_k = \Delta_k^{(j_k)} \ge \Delta_k^{(j')} = \tilde{\Delta} \quad \forall\ k \in \mathcal{K}_1, k \ge k_4.$$

Similarly, $\bar{d}_k^{(j_k)}$ is defined as $\bar{d}_k^{(j')}$, where the corresponding trust region radius is $\Delta_k^{(j_k)}$. It follows from $\|Z_k^T\bar{g}_k\| \ge 0.5\bar{\epsilon}$, $\lim\limits_{k\to\infty} c_k = 0$ and (4.9) that there exist $b_3 > 0$ and $k_5 \ge k_4$ such that $pred_k^f(\bar{d}_k^{(j_k)}) \ge b_3$ holds for all $k \in \mathcal{K}_1, k \ge k_5$. By (4.11), $pred_k^f(d_k^{(j_k)}) \ge 0.5b_3$ holds for all $k \in \mathcal{K}_1, k \ge k_6$, where $k_6 \ge k_5$ is large enough. Since all the iterations after k_0 are $f-$type successful iterations,

$$f(x_k) - f(x_{k+1}) \ge \eta_1^f pred_k^f(d_k) \ge 0.5\eta_1^f b_3 > 0, \quad k \in \mathcal{K},\ k \ge k_6,$$

which contradicts the property of the sequence $\{f(x_k)\}$. This completes our proof. □

Theorem 4.2. *If* $|\mathcal{C}| = +\infty$, $\lim_{k\to\infty} v_k^{max} = \bar{v} > 0$, *then any limit point of* $\{x_k\}_{k\in\mathcal{C}}$ *is an infeasible stationary point of problem* (2.1).

Proof. Let $\bar{x}$ be any limit point of $\{x_k\}_{k\in\mathcal{C}}$. Then there exists an infinite index set $\mathcal{K} \subseteq \mathcal{C}$ such that $\lim_{k\in\mathcal{K}} x_k = \bar{x}$. We can prove that $v(\bar{x}) = \bar{v} > 0$.

In fact, it follows from $\lim_{k\to\infty} v_k^{\max} = \bar{v} > 0$ that there exists an index k_7 such that

$$v_{k+1}^{\max} = \beta_2 v_k + (1-\beta_2) v_{k+1}$$

holds for $k \in \mathcal{C}$, $k \geq k_7$. By Algorithm 2.1 and $k \in \mathcal{C}$,

$$v_k - v_{k+1} \geq \eta_1^c pred_k^c(d_k^\infty), \qquad k \in \mathcal{C}.$$

Thus,

$$v_{k+1}^{\max} \leq \beta_2 v_k + (1-\beta_2) v_k - \eta_1^c (1-\beta_2) pred_k^c(d_k^\infty), \qquad k \in \mathcal{C}, k \geq k_7,$$

which follows that

$$0 \leq \eta_1^c (1-\beta_2) pred_k^c(d_k^\infty) \leq v_k - v_{k+1}^{\max}, \qquad k \in \mathcal{C},\ k \geq k_1. \tag{4.13}$$

Let $k \in \mathcal{K}$, $k \to +\infty$, then $\bar{v} \leq v(\bar{x})$. On the other hand, by $v_k \leq v_k^{\max}$, $v(\bar{x}) \leq \bar{v}$. Therefore, $v(\bar{x}) = \bar{v}$.

Now we prove by contradiction that $\lim \inf_{k\in\mathcal{K}} \|A_k c_k\| = 0$. Suppose that $\lim \inf_{k\in\mathcal{K}} \|A_k c_k\| > 0$. Then there exists $\bar{\epsilon} > 0$ such that

$$\|A_k c_k\| \geq \bar{\epsilon} > 0,\ \forall k \in \mathcal{K}.$$

By (4.13) and $v(\bar{x}) = \bar{v}$, $\lim_{k\in\mathcal{K}} pred_k^c(d_k^\infty) = 0$, which follows from (3.1) that $\lim_{k\in\mathcal{K}} \Delta_k = 0$. For $k \in \mathcal{K}$ large enough, we have that

$$v_k - v(x_k + d_k(\sigma_\infty, \Delta_k)) \geq \eta_1^c pred_k^c(d_k(\sigma_\infty, \Delta_k)), \tag{4.14a}$$
$$v_k - v(x_k + d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)) < \eta_1^c pred_k^c(d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)). \tag{4.14b}$$

By Lemma 3.2,

$$\begin{aligned} &pred_k^c(d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)) \geq \frac{\bar{\epsilon}}{2} \min\left\{\kappa_n(\gamma_1)^{-(1+\mu)}(\Delta_k)^{1+\mu}, \frac{\bar{\epsilon}}{1+M_{AA}}\right\} \\ =\ & \frac{\bar{\epsilon}}{2}\kappa_n(\gamma_1)^{-(1+\mu)}(\Delta_k)^{1+\mu} \end{aligned}$$

holds for $k \in \mathcal{K}$ large enough. It follows from Lemma 3.1 that

$$\begin{aligned} &|\rho_k^c(d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)) - 1| \\ \leq\ & \frac{O(\Delta_k^2)}{0.5\bar{\epsilon}\kappa_n(\gamma_1)^{-(1+\mu)}(\Delta_k)^{1+\mu}} = O((\Delta_k)^{1-\mu}) \to 0, \qquad \text{as } k \to \infty,\ k \in \mathcal{K}, \end{aligned}$$

i.e.,

$$v_k - v(x_k + d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)) \geq \eta_1^c pred_k^c(d_k(\sigma_\infty, \gamma_1^{-1}\Delta_k)),$$

which is a contradiction to (4.14b). This completes our proof. □

Theorem 4.3. *If* $|\mathcal{C}| = +\infty$, $\lim_{k\to\infty} v_k^{max} = 0$, *then any limit point of* $\{x_k\}_{k\in\mathcal{C}}$ *is a stationary point of problem* (2.1) *or the linear independence constraint qualification does not hold at the limit point.*

Proof. Let $x_{k+1} = x_k + d_k^{(j_k)}$, where $d_k^{(j_k)}$ is an $f-$type or $c-$type successful iterate step, j_k is the number of the corresponding inner cycle iteration.

Let $\bar{x}$ be any limit point of $\{x_k\}_{k\in\mathcal{C}}$. Then there exists an infinite index set $\mathcal{K} \subseteq \mathcal{C}$ such that $\lim_{k\in\mathcal{K}} x_k = \bar{x}$. Thus $v(\bar{x}) = 0$.

Suppose, by contradiction, that $\bar{x}$ is not a stationary point of (2.1) and the linear independence constraint qualification holds at $\bar{x}$, that is, $\nabla c_1(\bar{x}), \nabla c_2(\bar{x}), \cdots, \nabla c_m(\bar{x})$ are linearly independent and there exists a unit vector $\bar{d} \in R^n$ such that (4.1) holds. It suffices to prove that $d_k^{(j_k)}$ must be an $f-$type successful step for $k \in \mathcal{K}$ large enough. This will give a contradiction to the fact that $k \in \mathcal{K} \subseteq \mathcal{C}$ and complete our proof.

In fact, we see that there exists an index k_8 such that A_k has full column rank for $k \in \mathcal{K}$, $k \geq k_8$. Thus $s(\bar{x}) = \bar{d}$, where $s(x)$ is defined as (4.2). It follows from the continuity that there exist $\bar{\epsilon} > 0$ and $k_9 \geq k_8$ such that (4.3) holds for $k \in \mathcal{K}$, $k \geq k_9$.

Now we prove that $\lim\limits_{k\in\mathcal{K}} \Delta_k = 0$. Suppose, by contradiction, that there exists $\tilde{\Delta} > 0$ and $\mathcal{K}_1 \subseteq \mathcal{K}$ such that

$$\Delta_k \geq \tilde{\Delta}, \quad \forall k \in \mathcal{K}_1.$$

Similar to the proof of (4.12), we have that

$$pred_k^f(d_k) \geq \frac{\bar{\epsilon}}{16} \min\left\{\sqrt{1-k_n^2}\tilde{\Delta}, \frac{0.5\bar{\epsilon}}{1+M_B}\right\} \geq \delta(v_k)^\gamma$$

holds for all sufficiently large $k \in \mathcal{K}$, which contradicts with $k \in \mathcal{K}_1 \subseteq \mathcal{C}$. So $\lim_{k\in\mathcal{K}} \Delta_k = 0$.

Let $j^{'}$ be the smallest index which satisfies

$$\gamma_1^j \leq \frac{1}{\Delta_{\max}} \min\left\{\left(\frac{\bar{\epsilon}}{2\kappa_n M_B}\right)^{\frac{1}{1+\mu}}, \quad \frac{\bar{\epsilon}}{2(1+M_B)}, \left(\frac{b_1\bar{\epsilon}\sqrt{1-\kappa_n^2}}{16\kappa_n M_\lambda}\right)^{\frac{1}{\mu}}, \right.$$
$$\left. \left(\frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}}{8\kappa_n(M_g+0.5M_B)}\right)^{\frac{1}{\mu}}, \left(\frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}(1-\eta_1^f)}{16(M_{ff}+M_B)}\right)^{\frac{1}{1-\mu}}\right\}, \tag{4.15}$$

where M_λ is defined as the proof in Theorem 4.1. It follows from $\lim\limits_{k\to\infty} v_k^{max} = 0$ that there exists an index $k_{10} \geq k_9$ such that

$$\sqrt{2v_k^{max}} \leq \kappa_n \left(\Delta_{min}(\gamma_1)^{j^{'}}\right)^{1+\mu} \min\left\{\frac{1}{b_1}, \frac{\sqrt{2}\bar{\epsilon}}{16\delta\kappa_n}\sqrt{1-\kappa_n^2}\right\} \tag{4.16}$$

holds for $k \geq k_{10}$. Thus,

$$\|t_k\| \overset{(4.4)}{\leq} b_1\sqrt{2v_k} \leq b_1\sqrt{2v_k^{max}} \overset{(4.16)}{\leq} \kappa_n\left(\Delta_{min}(\gamma_1)^{j^{'}}\right)^{1+\mu} \leq \kappa_n\left(\Delta_k^{(j^{'})}\right)^{1+\mu} \leq \Delta_{k,j}^c \tag{4.17}$$

holds for $j = 0, 1, 2, \cdots, j^{'}$ and $k \in \mathcal{K}$, $k \geq k_{10}$. Denote

$$\tilde{d_k}^{(j)} = t_k + (\Delta_{k,j}^c - \|t_k\|)s_k, \quad j = 0, 1, \cdots, j',$$

where $\Delta_{k,j}^c = \kappa_n \min\{1, (\Delta_k^{(j)})^\mu\}\Delta_k^{(j)}$. By $\|t_k\| \leq \Delta_{k,j}^c$, $\tilde{d_k}^{(j)}$ is a feasible solution to (2.5) and $m_k(\tilde{d_k}^{(j)}) = 0$. Let $\bar{d}_k^{(j)} = t_k + Z_k u_k$, where u_k is a solution to (4.8) corresponding to $\bar{\Delta}_k^{(j)}$ and $\bar{\Delta}_k^{(j)} \geq \sqrt{1-\kappa_n^2}\Delta_k^{(j)}$. For $k \in \mathcal{K}, k \geq k_{10}$ large enough,

$$\|Z_k^T\bar{g}_k\| \geq \|Z_k^T g_k\| - \|Z_k^T B_k\|\|t_k\| \geq \frac{\bar{\epsilon}}{2}.$$

Therefore,

$$\begin{aligned}
pred_k^f(\bar{d}_k^{(j)}) &= -\bar{g}_k^T Z_k u_k - \frac{1}{2} u_k^T Z_k^T B_k Z_k u_k - g_k^T t_k - \frac{1}{2} t_k{}^T B_k t_k \\
&\geq \frac{\bar{\epsilon}}{4} \min\left\{ \sqrt{1-\kappa_n^2}\Delta_k^{(j)}, \frac{0.5\bar{\epsilon}}{1+M_B} \right\} - (M_g + 0.5M_B)\kappa_n \left(\Delta_k^{(j)}\right)^{1+\mu} \\
&= \left(\frac{\bar{\epsilon}}{4}\sqrt{1-\kappa_n^2} - (M_g+0.5M_B)\kappa_n \left(\Delta_k^{(j)}\right)^{\mu} \right) \Delta_k^{(j)} \\
&\overset{(4.15)}{\geq} \frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2}\Delta_k^{(j)} \qquad (4.18)
\end{aligned}$$

holds for $k \in \mathcal{K}, k \geq k_{10}$ and $j = 0, 1, \cdots, j'$. Similar to (4.12),

$$pred_k^f(d_k^{(j)}) \geq \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\Delta_k^{(j)} \qquad (4.19)$$

holds for $k \in \mathcal{K}, k \geq k_{10}$ and $j = 0, 1, \cdots, j'$. Let

$$b_4 = \min\left\{ 1, \frac{0.5\bar{\epsilon}}{(1+M_B)\sqrt{1-\kappa_n^2}}, \left(\frac{\bar{\epsilon}\sqrt{1-\kappa_n^2}}{16(M_g+0.5M_B)} \right)^{\frac{1}{\mu}} \right\}.$$

By (4.19),

$$\begin{aligned}
pred_k^f(d_k^{(j)}) &\geq \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\left(\Delta_k^{(j)}\right)^{1+\mu} \\
&\geq \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\left(\Delta_{min}(\gamma_1)^{j'}\right)^{1+\mu} \\
&\overset{(4.16)}{\geq} \delta\sqrt{v_k^{\max}} \geq \delta v_k \geq \delta(v_k)^{\gamma}.
\end{aligned}$$

Therefore, (2.9) holds for $k \in \mathcal{K}, k \geq k_{10}$ and $j = 0, 1, \cdots, j^{'}$.

If there exists some $j = 0, 1, 2, \cdots, j^{'}$ such that (2.10) and (2.11) all hold, then the current iteration is $f-$type successful iteration, which contradicts with $k \in \mathcal{K} \subseteq \mathcal{C}$.

Now we suppose that either of (2.10) and (2.11) does not hold for $j = 0, 1, \cdots, j^{'}$.

By $\lim_{k\to+\infty} v_k^{\max} = 0$, for $k \in \mathcal{K}$, $k \geq k_3$, there exists an index j such that

$$\begin{aligned}
&\left(\max\left\{ \frac{16\delta}{\bar{\epsilon}\sqrt{1-\kappa_n^2}}, \frac{b_1}{\kappa_n} \right\} \sqrt{2v_k^{max}} \right)^{\frac{1}{1+\mu}} \\
&\leq \Delta_k^{(j)} \leq \min\left\{ \sqrt{\frac{\epsilon_1}{2b_2\sqrt{2(1-\epsilon_1)}}}, \left(\frac{\epsilon_1}{2b_2^2}\right)^{\frac{1}{4}} \right\} (v_k^{max})^{\frac{1}{4}}. \qquad (4.20)
\end{aligned}$$

Let $\tilde{j_k}$ be the smallest index which satisfies (4.20).

If $\tilde{j_k} \leq j'$, we have shown that, for the trial step $d_k^{(\tilde{j_k})}$, which is a solution to (2.4) corresponding to $\Delta_k^{(\tilde{j_k})}$, (2.9) holds.

If $\tilde{j_k} > j'$, we have that

$$\|t_k\| \leq b_1\sqrt{2v_k} \leq b_1\sqrt{2v_k^{max}} \overset{(4.20)}{\leq} \kappa_n \left(\Delta_k^{(\tilde{j_k})}\right)^{1+\mu} \leq \Delta_{k,j}^c$$

holds for $j = j^{'}+1, j^{'}+2, \cdots, \tilde{j_k}$ and $k \in \mathcal{K}$, $k \geq k_{10}$. Let $\bar{d}_k^{(j)} = t_k + Z_k u_k$, u_k is a solution to (4.8) corresponding to $\bar{\Delta}_k^{(j)}$. Noting that $d_k^{(j)}$ $(j = j^{'}+1, \cdots, \tilde{j_k})$ is a solution to (2.4) corresponding to $\Delta_k^{(j)}$, it follows from (4.11), (4.18) and $\|c_k + A_k^T d_k^{(j)}\| \leq \|c_k\|$ that

$$
\begin{aligned}
pred_k^f(d_k^{(j)}) &\geq \frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2}\Delta_k^{(j)} - M_\lambda \|c_k\| \\
&\overset{(4.17)}{\geq} \left(\frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2} - \frac{\kappa_n M_\lambda}{b_1}\left(\Delta_k^{(j)}\right)^\mu \right)\Delta_k^{(j)} \\
&\geq \left(\frac{\bar{\epsilon}}{8}\sqrt{1-\kappa_n^2} - \frac{\kappa_n M_\lambda}{b_1}\left(\Delta_k^{(j')}\right)^\mu \right)\Delta_k^{(j)} \qquad (j' < j) \\
&\overset{(4.15)}{\geq} \frac{\bar{\epsilon}}{16}\sqrt{1-\kappa_n^2}\left(\Delta_k^{(\tilde{j_k})}\right)^{1+\mu} \qquad (j \leq \tilde{j_k}) \\
&\overset{(4.20)}{\geq} \delta\sqrt{2v_k^{max}} \geq \delta\sqrt{2v_k} \geq \delta(v_k)^\gamma
\end{aligned}
$$

holds for $k \in \mathcal{K}$, $k \geq k_{10}, j = j^{'}+1, \cdots, \tilde{j_k}$. In a word, (2.9) holds for $j = 0, 1, \cdots, \tilde{j_k}$.

Suppose that neither (2.10) nor (2.11) holds for $j = j^{'}+1, \cdots, \tilde{j_k}-1$. Then for $j = \tilde{j_k}$, $k \in \mathcal{K}$, $k \geq k_{10}$, we have that

$$
\left| \frac{ared_k^f(d_k^{(\tilde{j_k})})}{pred_k^f(d_k^{(\tilde{j_k})})} - 1 \right| \leq \frac{8(M_{ff}+M_B)(\Delta_k^{(\tilde{j_k})})^2}{\bar{\epsilon}\sqrt{1-\kappa_n^2}\left(\Delta_k^{(\tilde{j_k})}\right)^{1+\mu}} \overset{(4.15)}{\leq} 1 - \eta_1^f,
$$

which implies the truth of (2.10). Moreover, by (4.5), (2.7) and $\|c_k\| = \sqrt{2v_k} \leq \sqrt{2v_k^{max}}$,

$$
\begin{aligned}
v(x_k + d_k^{(\tilde{j_k})}) &\leq m_k(d_k^{(\tilde{j_k})}) + b_2\|c_k + A_k^T d_k^{(\tilde{j_k})}\|\|d_k^{(\tilde{j_k})}\|^2 + b_2^2\|d_k^{(\tilde{j_k})}\|^4 \\
&\leq \frac{1}{2}(1-\epsilon_1)\|c_k\|^2 + b_2\sqrt{1-\epsilon_1}\|c_k\|\|d_k^{(\tilde{j_k})}\|^2 + b_2^2\|d_k^{(\tilde{j_k})}\|^4 \\
&\leq (1-\epsilon_1)v_k^{max} + b_2\sqrt{1-\epsilon_1}\sqrt{2v_k^{max}}\|d_k^{(\tilde{j_k})}\|^2 + b_2^2\|d_k^{(\tilde{j_k})}\|^4 \\
&\overset{(4.20)}{\leq} (1-\epsilon_1)v_k^{max} + \frac{\epsilon_1}{2}v_k^{max} + \frac{\epsilon_1}{2}v_k^{max} = v_k^{max},
\end{aligned}
$$

which implies the truth of (2.11).

Summarizing the discussion above, we know that $d_k^{(\tilde{j_k})}$ is an $f-$type successful iterate step for $k \in \mathcal{K}, k \geq k_5$. This contradicts the fact that $k \in \mathcal{K} \subseteq \mathcal{C}$ and completes our proof. □

5. Numerical Results

In order to test the efficiency of Algorithm 2.1, we wrote a MATLAB code for the algorithm. The experiments were performed on PC Thinkpad S3 Yoga I7-4510U Laptop with an Intel dual core CPU at 2.00GHz, 2.59GHz and 8GB of RAM. The parameters in Algorithm 2.1 are set as follows.

$$
\begin{aligned}
&\delta = 1,\ \gamma = 2.1,\ \eta_1^f = \eta_1^c = 0.001,\ \epsilon_1 = 0.02,\ \epsilon_2 = 0.01,\ \kappa_n = 0.8,\ \mu = 0.5, \\
&\beta_1 = 0.9,\ \beta_2 = 0.75,\ \Delta_{\min} = 0.001,\ \Delta_{\max} = 100,\ \Delta_0 = 1,\ \gamma_1 = 0.25,\ \sigma_0 = 1.
\end{aligned}
$$

The matrix B_k is chosen to be the exact Hessian of the Lagrangian function. At the beginning of each iteration, we set $\sigma_{k+1} = \|\lambda_k\|_\infty + 1$, where λ_k is the Lagrangian multiplier estimation. If

$\max\{\|c_k\|, \|d_k\|\} \le 10^{-6}$ or an infeasible stationary point is obtained, Algorithm 2.1 terminates. Five difficult problems were used for testing, which exhibit inconsistent constraint linearization at some iterate, the gradients of the constraint functions being linearly dependent, infeasible problem, nonuniqueness of Lagrange multipliers at the solution.

Example 5.1. The problem

$$\begin{array}{ll} \min & (x_2-1)^2 \\ \text{s.t.} & x_1^2 = 0, \\ & x_1^3 = 0 \end{array}$$

is presented in Fletcher *et al.* [14] and is also discussed by Chen and Goldfarb [10]. MFCQ is violated at the solution $x^* = (0,1)^T$ and the linearized constraints are inconsistent at every infeasible point.

Example 5.2. $x^* = (1,0,0)^T$ is the solution to the following problem

$$\begin{array}{ll} \min & 0.5(x_1-1)^2 + 0.5(x_2-1)^2 + 0.5x_3^2 \\ \text{s.t.} & x_1x_2 = 0, \\ & x_2 = 0. \end{array}$$

The gradients of the constrained functions are linearly dependent at any point of the form $x_k = (a,0,b)^T$ with the constants a,b. Moreover, the Lagrangian multipliers at the solution is not unique and unbounded.

Example 5.3. $x^* = (0, \sqrt[3]{1.5}, \sqrt[3]{1.5})^T$ is the solution to the following problem [18]

$$\begin{array}{ll} \min & x_1 + \frac{1}{2}x_2^2 + \frac{1}{2}(x_2-x_3)^2 \\ \text{s.t.} & \frac{1}{2}x_1^2 = 0, \\ & x_1 + x_2^3 - \frac{3}{2} = 0. \end{array}$$

The gradients of the constrained functions are linearly dependent at any point of the form $(a,0,b)^T$ with $a,b \in R$. Moreover, there are no Lagrange multipliers at the solution x^*.

Example 5.4. The problem

$$\begin{array}{ll} \min & -5x_1^2 + x_2^2 \\ \text{s.t.} & x_1 - 1 = 0 \end{array}$$

is introduced by Nocedal and Wright [21] to show that the quadratic penalty function is unbounded by any penalty factor $\sigma < 10$. For such values of σ, the iterates generated by an unconstrained minimization method would usually diverge.

Example 5.5. The following problem (see also problem (1.1))

$$\begin{array}{ll} \min & (x_1-1)^2 \\ \text{s.t.} & x_2^2 + 1 = 0 \end{array}$$

is infeasible. $x^* = (1,0)^T$ is an infeasible stationary point.

The numerical results are listed in Table 5.1, where x_0 is the initial point, *iter* means the number of iterations, $f(x_k)$ and $||c(x_k)||_\infty$ stand for the value of the objective function and the measure of the constraint violation, respectively. σ_k is the final value of the penalty penalty. *Res* means the KKT error, *i.e.*, $Res = \|g_k - A_k\lambda_k\|$.

Table 5.1: Numerical Results of Algorithm 2.1 for five difficult problems.

Example	x_0^T	iter	Res	$\|\|c(x_k)\|\|_\infty$	$f(x_k)$	σ_k
1	$(1,0)$	11	0.0000e+00	5.9298e-07	0.0000e+00	1.0000e+00
2	(1,0,10)	4	3.3688e-16	1.0105e-16	5.0000e-01	1.5000e+00
3	(10,0,10)	6	5.7651e-08	5.1062e-07	6.5447e-01	7.0252e+02
4	(10,10)	5	0.0000e+00	0.0000e+00	-5.0000e+00	1.1000e+01
5	(10,10)	15	Infeasible	1.0000e+00	0.0000e+00	1.0000e+00

From Table 5.1, we can see that Algorithm 2.1 successfully solved the five difficult problems, which shows the efficiency of the algorithm.

Acknowledgments. The authors are very grateful to the two anonymous referees whose comments improved the quality of this paper greatly. The research of Kou is supported by Chinese NSF grant (Nos. 11401038, 11471052); the research of Chen is supported by Chinese NSF grant (No. 11371273), and the research of Dai is supported by the Chinese NSF (Nos. 11631013, 11331012, 71331001) and the National 973 Program of China (No. 2015CB856000).

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez and M.L. Schuverdt, Augmented Lagrangian methods under the constant positive linear dependence constraint qualification, *Math. Program.*, **111** (2008), 5-32.

[2] N.S. Aybat and G. Iyengar, A first-order augmented Lagrangian method for compressed sensing, *SIAM J. Optim.*, **22** (2012), 429-459.

[3] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Athena Scientific, Belmont, Massachusetts, 1996.

[4] E.G. Birgin and J.M. Martínez, Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization, *Comput. Optim. Appl.*, **51** (2012), 941-965.

[5] P.T. Boggs and J.W. Tolle, Sequential quadratic programming, *Acta Numer.*, **4** (1995), 1-51.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, **3**:1 (2011), 1-122.

[7] R.H. Byrd, G. Lopez-Calva and J. Nocedal, A line search exact penalty method using steering rules, *Math. Program.*, **133** (2012), 39-73.

[8] R.H. Byrd, J. Nocedal and R.A. Waltz, Steering exact penalty methods, *Optim. Methods Softw.*, **23**:2 (2008), 197-213.

[9] Z.W. Chen and Y.H. Dai, A line search exact penalty method with bi-object strategy for nonlinear constrained optimization, *J. Comput. Appl. Math.*, **300** (2016), 245-258.

[10] L. Chen and D. Goldfarb, Interior-point l_2 penalty methods for nonlinear programming with strong global convergence properties, *Math. Program.*, **108**:1, (2006), 1-36.

[11] A.R. Conn, N.I.M. Gould and Ph. L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM J. Numer. Anal.*, **28** (1991), 545-572.

[12] A.R. Conn, N.I.M. Gould and Ph. L. Toint, Lancelot: A Fortran Package for Large Scale Nonlinear Optimization (Released A), Springer, New York, 1992.

[13] R. Fletcher, Practical Methods of Optimization, 2nd ed., John Wiley and Sons, Chichester, UK, 1987.

[14] R. Fletcher, S. Leyffer and Ph. L. Toint, On the global convergence of a filter-SQP algorithm, *SIAM J. Optim.*, **13**:1 (2002), 44-59.

[15] H.W. Ge and Z.W. Chen, A penalty-free method with line search for nonlinear equality constrained optimization, *Appl. Math. Model.*, **37** (2013), 9934-9949.

[16] N.I.M. Gould and P.L. Toint, Nonlinear programming without a penalty function or a filter, *Math. Program.*, **122** (2010), 155-196.

[17] M.R. Hestense, Multiplier and gradient method, J. Optim. *Theory Appl.*, **4** (1969), 303-320.

[18] X.W. Liu and Y.X. Yuan, A robust algorithm for optimization with general equality and inequality constraints, *SIAM J. Sci. Comput.*, **22** (2000), 517-534.

[19] X W. Liu and Y.X. Yuan, A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization, *SIAM J. Optim.*, **21** (2011), 545-571.

[20] L.F. Niu and Y.X. Yuan, A trust-region algorithm for nonlinear constrained optimization, *J. Comput. Math.*, **1** (2010), 72-86.

[21] J. Nocedal and S.J. Wright, Numerical Optimization, Springer, 2nd ed., 2006.

[22] M.J.D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, Fletcher R., ed., Academic Press, London, 1969, 283-298.

[23] S.Q. Qiu and Z.W. Chen, A globally convergent penalty-free method for optimization with general constraints and simple bounds, *Acta Appl. Math.*, **142**, (2016), 39-60.

[24] X. Wang and Y. X. Yuan, An augmented Lagrangian trust region method for equality constrained optimization, *Optim. methods softw.*, **30**, (2015), 559-582.

[25] J. Yang, Y. Zhang and W. Yin, A fast alternating direction method for tv11-12 signal reconstruction from partial fourier data, *IEEE J. Sel. Top. Signal Proces.*, **4** (2010), 288-297.

[26] Y.X. Yuan, On the convergence of a new trust region algorithm, *Numer. Math.*, **70** (1995), 515-539.

[27] Y.X. Yuan, Recent advances in trust region algorithms, *Math. Program.*, **151** (2015), 249-281.