AN AUGMENTED LAGRANGIAN AFFINE SCALING METHOD FOR NONLINEAR PROGRAMMING *

XIAO WANG † AND HONGCHAO ZHANG ‡

Abstract. In this paper, we propose an Augmented Lagrangian Affine Scaling (ALAS) algorithm for general nonlinear programming, for which a quadratic approximation to the augmented Lagrangian is minimized at each iteration. Different from the classical sequential quadratic programming (SQP), the linearization of nonlinear constraints is put into the penalty term of this quadratic approximation, which results smooth objective of the subproblem and avoids possible inconsistency among the linearized constraints and trust region constraint. By applying affine scaling techniques to handle the strict bound constraints, an active set type affine scaling trust region subproblem is proposed. Through special strategies of updating Lagrange multipliers and adjusting penalty parameters, this subproblem is able to reduce the objective function value and feasibility errors in an adaptive well-balanced way. Global convergence of the ALAS algorithm is established under mild assumptions. Furthermore, boundedness of the penalty parameter is analyzed under certain conditions. Preliminary numerical experiments of ALAS are performed for solving a set of general constrained nonlinear optimization problems in CUTEr problem library.

Key words. augmented Lagrangian, affine scaling, equality and inequality constraints, bound constraints, trust region, global convergence, boundedness of penalty parameter

AMS subject classifications. 90C30, 65K05

1. Introduction.

In this paper, we consider the following nonlinear programming problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & \mathbf{c}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{c} = (c_1, \ldots, c_m)^\mathsf{T}$ with $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are Lipschitz continuously differentiable. Here, to simplify the exposition, we only assume nonnegative lower bounds on the variables. However, we emphasize that the analysis in this paper can be easily extended to the general case with bounds $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$.

The affine scaling method is first proposed by Dikin [18] for linear programming. Then, affine scaling techniques have been widely used to solve bound constrained optimization quite effectively [4, 10, 11, 25, 27, 28, 40]. Affine scaling methods for solving polyhedral constrained optimization are also intensively studied. Interested readers are referred to [6, 12, 22, 26, 30, 31, 37, 38]. However, for optimization problems with nonlinear constraints, there are only relatively few works that have applied affine scaling techniques. In [17], a special class of nonlinear programming problems arising from optimal control is studied. These problems have a special structure that the variables can be separated into two independent parts, with one part satisfying bound constraints. Then an affine scaling method is proposed to keep the bound constraints strictly feasible. Some other affine scaling methods are introduced from the view of equivalent KKT equations of nonlinear programming [8, 9, 21]. In these methods, affine scaling techniques are often combined with primal and dual interior point approaches, where the iterates are scaled to avoid converging prematurely to boundaries.

To solve (1.1), standard sequential quadratic programming (SQP) trust-region methods [33, 23] need to solve a constrained quadratic subproblem at each iteration. In this subproblem, the Hessian is often constructed by a quasi-Newton technique and the constraints are local linearizations of original nonlinear constraints in a region around the current iterate, called trust region. One fundamental issue with this standard approach is that because of this newly introduced trust region constraint, the subproblems are often infeasible. There are two commonly used approaches for dealing with this issue. One is to relax

* This material is based upon work supported by the Postdoc Grant of China 119103S175, UCAS President grant Y35101AY00, NSFC grant of China 11301505 and the National Science Foundation grant of USA 1016204.

† wangxiao@ucas.ac.cn, School of Mathematical Sciences, University of Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China.

‡ hozhang@math.lsu.edu, http://www.math.lsu.edu/∼hozhang, 303 Lockett Hall, Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918. Phone (225) 578-1982. Fax (225) 578-4276.

the linear constraints. The other commonly used approach is to split the trust region into two separate parts. One part is used to reduce the feasibility error in the range space of the linear constraints, while the other is to reduce the objective in its null space. To avoid the possible infeasibility, some modern SQP methods are developed based on penalty functions. Two representative approaches are sequential l_∞ quadratic programming (Sl_∞QP) [20] and sequential l_1 quadratic programming (Sl_1QP) [42]. In these two methods, the linearization of nonlinear constraints in standard SQP subproblems are moved into the objective by adding an l_∞ and an l_1 penalty term, respectively. Then, the l_∞ penalty function and l_1 penalty function are used as merit functions in these two methods accordingly.

Recently, Wang and Yuan [41] propose an Augmented Lagrangian Trust Region (ALTR) method for the equality constrained nonlinear optimization:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & \mathbf{c}(\mathbf{x}) = \mathbf{0}. \end{aligned}$$

To handle the aforementioned infeasibility issues, at the k-th iteration ALTR solves the following trust region subproblem:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & (\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}\mathbf{B}_k\mathbf{d} + \frac{\sigma_k}{2}\|\mathbf{c}_k + \mathbf{A}_k\mathbf{d}\|_2^2 \\ \text{s. t.} \quad & \|\mathbf{d}\|_2 \le \Delta_k, \end{aligned} \tag{1.2}$$

where $\boldsymbol{\lambda}_k$ and σ_k are the Lagrange multipliers and the penalty parameter, respectively, and $\mathbf{B}_k$ is the Hessian of the Lagrange function $f(\mathbf{x}) - \boldsymbol{\lambda}_k^\mathsf{T}\mathbf{c}(\mathbf{x})$ at $\mathbf{x}_k$ or its approximation. Here, $\mathbf{A}_k := (\nabla c_1(\mathbf{x}_k), \ldots, \nabla c_m(\mathbf{x}_k))^\mathsf{T}$ is the Jacobian matrix of equality constraints at $\mathbf{x}_k$. Notice that the objective function in (1.2) is a local approximation of the augmented Lagrangian function $L(\mathbf{x}; \lambda_k, \sigma_k) := f(\mathbf{x}) - \boldsymbol{\lambda}_k^\mathsf{T}\mathbf{c}(\mathbf{x}) + \frac{\sigma_k}{2}\|\mathbf{c}(\mathbf{x})\|_2^2$, which is actually obtained by applying the second-order Taylor expansion at $\mathbf{x}_k$ to approximate the Lagrange function, and using the linearization of $\mathbf{c}(\mathbf{x}_k)$ at $\mathbf{x}_k$ in the augmented quadratic penalty term. In ALTR, it is natural to adopt the augmented Lagrangian function as the merit function. One major advantage of (1.2) compared with Sl_∞QP and Sl_1QP subproblems is that its objective is a smooth quadratic function. Therefore, the subproblem (1.2) is a standard trust region subproblem, for solving which many efficient algorithms exist, such as the *gqtpar* subroutine [32] and the *SMM* algorithm [24]. Promising numerical results of ALTR have been reported in [41]. More recently, we noticed that Curtis *et al.* [16] also developed a similar trust region approach based on approximations to the augmented Lagrangian functions for solving constrained nonlinear optimization. In this paper, motivated by nice properties of the affine scaling technique and the augmented Lagrangian function, we extend the method developed in [41] to solve the general constrained nonlinear optimization (1.1), in which the inequality constraints are also allowed.

Our major contributions in this paper are the followings. Firstly, to overcome possible infeasibility of SQP subproblems, we move the linearized equality constraints into a penalty term of a quadratic approximation to the augmented Lagrangian. This quadratic approximation can be viewed as a local approximation of the augmented Lagrangian function at fixed Lagrange multipliers and a fixed penalty parameter. In addition, we incorporate the affine scaling technique to deal with simple bound constraints. Secondly, we propose a new strategy to update penalty parameters with the purpose of adaptively reducing the objective function value and constraint violations in a well-balanced way. Thirdly, we establish the global convergence of our new method under the Constant Positive Linear Dependence (CPLD) constraint qualification. We also give a condition under which the penalty parameters in our method will be bounded.

The remainder of this paper is organized as follows. In Section 2, we describe in detail the ALAS algorithm. Section 3 studies the global convergence properties of the proposed algorithm. In Section 4, we analyze the boundedness of penalty parameters. The preliminary numerical results are given in Section 5. Finally, we draw some conclusions in Section 6.

Notation. In this paper, $\mathbb{R}^n$ is denoted as the n dimensional real vector space and $\mathbb{R}^n_+$ is the nonnegative orthant of $\mathbb{R}^n$. Let $\mathbb{N}$ be the set of all nonnegative integers. For any vector $\mathbf{z}$ in $\mathbb{R}^n$, z_i is the i-th component of $\mathbf{z}$ and $\text{supp}(\mathbf{z}) = \{i : z_i \neq 0\}$. We denote $\mathbf{e}_i$ as the i-th coordinate vector. If not specified, $\|\cdot\|$ refers to the

Euclidean norm and $B_\rho(\mathbf{z})$ is the Euclidean ball centered at $\mathbf{z}$ with radius $\rho > 0$. For a closed and convex set $\Omega \in \mathbb{R}^n$, the operator $P_\Omega(\cdot)$ denotes the Euclidean projection onto Ω, and when $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}\}$, we simply let $(\mathbf{z})^+ = P_\Omega(\mathbf{z})$. Given any matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ and an index set $\mathcal{F}$, $[\mathbf{H}]_\mathcal{F}$ denotes the submatrix of $\mathbf{H}$ with rows indexed by $\mathcal{F}$, while $[\mathbf{H}]_{\mathcal{F},\mathcal{I}}$ stands for the submatrix of $\mathbf{H}$ with rows indexed by $\mathcal{F}$ and columns indexed by $\mathcal{I}$. In addition, $\mathbf{H}^\dagger$ denotes the general inverse of $\mathbf{H}$.

We denote the gradient of f at $\mathbf{x}$ as $\mathbf{g}(\mathbf{x}) := \nabla f(\mathbf{x})$, a column vector, and the Jacobian matrix of $\mathbf{c}$ at $\mathbf{x}$ as $\mathbf{A}(\mathbf{x})$, where $\mathbf{A}(\mathbf{x}) = (\nabla c_1(\mathbf{x}), \ldots, \nabla c_m(\mathbf{x}))^\mathsf{T}$. The subscript $_k$ refers to the iteration number. For convenience, we abbreviate $\mathbf{g}(\mathbf{x}_k)$ as $\mathbf{g}_k$, and similarly f_k, $\mathbf{c}_k$ and $\mathbf{A}_k$ are also used. The i-th component of $\mathbf{x}_k$ is denoted as x_{ki}.

2. An augmented Lagrangian affine scaling method. In this section, we propose a new algorithm for nonlinear programming (1.1). The goal of this algorithm is to generate a sequence of iterates $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$, converging to a KKT point of (1.1), which is defined as follows (for instance, see [33]).

DEFINITION 2.1. *A point $\mathbf{x}^*$ is called a KKT point of* (1.1)*, if there exists a vector $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_m^*)^\mathsf{T}$ such that the following KKT conditions are satisfied at $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$:*

$$\begin{cases} (\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*)_i \geq 0, & \text{if} \quad x_i^* = 0, \\ (\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*)_i = 0, & \text{if} \quad x_i^* > 0, \\ \mathbf{c}^* = 0, \quad \mathbf{x}^* \geq \mathbf{0}. \end{cases} \tag{2.1}$$

where $\mathbf{c}^*$, $\mathbf{g}^*$ and $\mathbf{A}^*$ denote $\mathbf{c}(\mathbf{x}^*)$, $\mathbf{g}(\mathbf{x}^*)$ and $\mathbf{A}(\mathbf{x}^*)$ respectively.

Before presenting the ALAS algorithm, we would like to discuss several important issues first, such as the construction of the subproblem, the update of Lagrange multipliers and penalty parameters.

2.1. Affine scaling trust region subproblem. The augmented Lagrangian function (see, e.g., [14]) associated with the objective function $f(\mathbf{x})$ and equality constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ is defined as

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}, \sigma) = f(\mathbf{x}) - \boldsymbol{\lambda}^\mathsf{T}\mathbf{c}(\mathbf{x}) + \frac{\sigma}{2}\|\mathbf{c}(\mathbf{x})\|^2, \tag{2.2}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$ denotes the vector of Lagrange multipliers and $\sigma \in \mathbb{R}_+$ is the penalty parameter. For (1.1), at the k-th iteration with fixed $\boldsymbol{\lambda}_k$ and σ_k the classical augmented Lagrangian methods solve the following subproblem :

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}_k, \sigma_k) \\ \text{s. t.} \quad & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{2.3}$$

Normally (2.3) is solved to find $\mathbf{x}_{k+1}$ such that

$$\|(\mathbf{x}_k - \nabla_x \mathcal{L}(\mathbf{x}_{k+1}, \boldsymbol{\lambda}_k, \sigma_k))^+ - \mathbf{x}_k\| \ \leq \ w_k,$$

where w_k, $k = 1, 2, \ldots$, is a sequence of preset tolerances gradually converging to zero. For more details, one could refer to [33]. However, different from classical augmented Lagrangian methods, a new subproblem is proposed in this paper.

Instead of minimizing $\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}, \sigma)$ at the k-th iteration as in (2.3), we work with its second order approximation. In order to make this approximation adequate, the approximation is restricted in a trust region. So, at the current iterate $\mathbf{x}_k$, $\mathbf{x}_k \geq 0$, a trust region subproblem can be defined with explicit bound constraints as:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & q_k(\mathbf{d}) := (\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}\mathbf{B}_k\mathbf{d} + \frac{\sigma_k}{2}\|\mathbf{c}_k + \mathbf{A}_k\mathbf{d}\|^2 \\ & \qquad = \bar{\mathbf{g}}_k^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}(\mathbf{B}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k)\mathbf{d} + \frac{\sigma_k}{2}\|\mathbf{c}_k\|^2 \\ \text{s. t.} \quad & \mathbf{x}_k + \mathbf{d} \geq \mathbf{0}, \quad \|\mathbf{d}\| \leq \Delta_k, \end{aligned} \tag{2.4}$$

where $\Delta_k > 0$ is the trust region radius, $\mathbf{B}_k$ is the Hessian of the Lagrange function $f(\mathbf{x}) - \boldsymbol{\lambda}_k^\mathsf{T}\mathbf{c}(\mathbf{x})$ or its approximation and $\bar{\mathbf{g}}_k = \nabla_x \mathcal{L}(\mathbf{x}_k; \boldsymbol{\lambda}_k, \sigma_k)$, i.e.,

$$\bar{\mathbf{g}}_k = \mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k \mathbf{A}_k^\mathsf{T}\mathbf{c}_k. \tag{2.5}$$

Instead of solving (2.4) directly, we would like to apply affine scaling techniques to deal with the explicit bound constraints effectively. Before giving our new subproblem, let us first recall a most commonly used affine scaling technique proposed by Coleman and Li [11] for the simple bound constrained optimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ f(\mathbf{x}) \quad \text{s. t.} \quad \mathbf{l} \le \mathbf{x} \le \mathbf{u}.$$

At the current strict interior iterate $\mathbf{x}_k$, i.e., $\mathbf{l} < \mathbf{x}_k < \mathbf{u}$, the affine scaling trust region subproblem is defined as

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & \nabla f(\mathbf{x}_k)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}\mathbf{W}_k\mathbf{d} \\ \text{s. t.} \quad & \mathbf{l} \le \mathbf{x}_k + \mathbf{d} \le \mathbf{u}, \quad \|(\mathbf{D}_k^c)^{-1}\mathbf{d}\| \le \Delta_k, \end{aligned}$$

where $\mathbf{W}_k$ is equal to $\nabla^2 f(\mathbf{x_k})$ or its approximation, $\mathbf{D}_k^c$ is an affine scaling diagonal matrix defined by $[\mathbf{D}_k^c]_{ii} = \sqrt{|v_{ki}|}$, $i = 1, \ldots, n$, and $\mathbf{v}_k$ is given by

$$v_{ki} = \begin{cases} x_{ki} - u_i, & \text{if} \quad g_{ki} < 0 \text{ and } u_i < \infty, \\ x_{ki} - l_i, & \text{if} \quad g_{ki} \ge 0 \text{ and } l_i > -\infty, \\ -1, & \text{if} \quad g_{ki} < 0 \text{ and } u_i = \infty, \\ 1, & \text{if} \quad g_{ki} \ge 0 \text{ and } l_i = -\infty. \end{cases}$$

Therefore, by applying similar affine scaling technique to (2.4), we obtain the following affine scaling trust region subproblem:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & q_k(\mathbf{d}) \\ \text{s. t.} \quad & \|\mathbf{D}_k^{-1}\mathbf{d}\| \le \Delta_k, \quad \mathbf{x}_k + \mathbf{d} \ge \mathbf{0}, \end{aligned} \tag{2.6}$$

where $\mathbf{D}_k$ is the scaling matrix given by

$$[\mathbf{D}_k]_{ii} = \begin{cases} \sqrt{x_{ki}}, & \text{if } \bar{g}_{ki} > 0, \\ 1, & \text{otherwise}, \end{cases} \tag{2.7}$$

and $\bar{\mathbf{g}}_k$ is defined in (2.5). By Definition 2.1, we can see that $\mathbf{x}_k$ is a KKT point of (1.1) if and only if

$$\mathbf{x}_k \ge \mathbf{0}, \qquad \mathbf{c}_k = \mathbf{0} \qquad \text{and} \qquad \mathbf{D}_k\bar{\mathbf{g}}_k = \mathbf{0}. \tag{2.8}$$

In subproblem (2.6), the scaled trust region $\|\mathbf{D}_k^{-1}\mathbf{d}\| \le \Delta_k$ requires $[\mathbf{D}_k]_{ii} > 0$ for each i, which can be guaranteed if $\mathbf{x}_k$ is maintained in the strict interior of the feasible region, namely, $\mathbf{x}_k > \mathbf{0}$. However, in our approach such requirement is actually not needed. In other words, we allow some components of $\mathbf{x}_k$ to be zero. For those i with $[\mathbf{D}_k]_{ii} = 0$, we simply solve the subproblem in the reduced subspace by fixing $d_i = 0$. So we translate the subproblem (2.6) to the following active set type affine scaling trust region subproblem:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & \tilde{q}_k(\mathbf{d}) := (\mathbf{D}_k\bar{\mathbf{g}}_k)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}[\mathbf{D}_k(\mathbf{B}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k)\mathbf{D}_k]\mathbf{d} + \frac{\sigma_k}{2}\|\mathbf{c}_k\|^2 \\ \text{s. t.} \quad & \|\mathbf{d}\| \le \Delta_k, \quad \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \ge \mathbf{0}, \\ & d_i = 0, \quad \text{if } [\mathbf{D}_k]_{ii} = 0. \end{aligned} \tag{2.9}$$

Obviously, $\tilde{q}_k(\mathbf{d}) = q_k(\mathbf{D}_k\mathbf{d})$. Hence, letting $\mathbf{s}_k = \mathbf{D}_k\tilde{\mathbf{s}}_k$, where $\tilde{\mathbf{s}}_k$ denotes the solution of (2.9), we define the predicted reduction of the augmented Lagrangian function (2.2) at $\mathbf{x}_k$ with step $\mathbf{s}_k$ as

$$\mathrm{Pred}_k := q_k(\mathbf{0}) - q_k(\mathbf{s}_k) = \tilde{q}_k(\mathbf{0}) - \tilde{q}_k(\tilde{\mathbf{s}}_k). \tag{2.10}$$

Note that (2.9) ensures that $\mathbf{s}_k$ is a feasible trial step for the bound constraints, i.e., $\mathbf{x}_k + \mathbf{s}_k \geq \mathbf{0}$. Of course, it is impractical and inefficient to solve (2.9) exactly. To guarantee the global convergence, theoretically we only need to the trial step $\mathbf{s}_k$ satisfies the following sufficient model reduction condition:

$$\mathrm{Pred}_k \geq \bar{\beta}\ [\tilde{q}_k(\mathbf{0}) - \min\{\tilde{q}_k(\mathbf{d}) : \|\mathbf{d}\| \leq \Delta_k,\ \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \geq \mathbf{0}\}], \tag{2.11}$$

where $\bar{\beta}$ is some constant in (0,1).

2.2. Update of Lagrange multipliers. We now discuss how to update the Lagrange multiplier $\boldsymbol{\lambda}_k$. In our method, we do not necessarily update $\boldsymbol{\lambda}_k$ at each iterate. Its update depends on the performance of the algorithm. If the iterates are still far away from the feasible region, the algorithm will keep $\boldsymbol{\lambda}_k$ unchanged and focuses on reducing the constraint violation $\|\mathbf{c}_k\|$ by increasing the penalty parameter σ_k. However, if the constraint violation $\|\mathbf{c}_k\|$ is reduced, we think that the algorithm is performing well as the iterates approaching the feasible region. In this situation, the algorithm will update $\boldsymbol{\lambda}_k$ according to the latest information. In particular, we set a switch condition:

$$\|\mathbf{c}_k\| \ \leq\ R_{k-1},$$

where R_k, $k = 1, 2, \ldots$, is a sequence of positive controlling factors gradually converging to zero as iterates approach the feasible region. In classical augmented Lagrangian methods, solving (2.3) yields a new iterate $\mathbf{x}_{k+1}$, then it follows from the KKT conditions that

$$\mathbf{g}_{k+1} - \mathbf{A}_{k+1}^{\mathsf{T}}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k = \boldsymbol{\mu}_k, \tag{2.12}$$

where $\boldsymbol{\mu}_k \geq \mathbf{0}$ is the Lagrange multiplier associated with $\mathbf{x} \geq \mathbf{0}$. Therefore, by Definition 2.1, a good estimate of $\boldsymbol{\lambda}_{k+1}$ would be $\boldsymbol{\lambda}_k - \sigma_k\mathbf{c}_{k+1}$. However, this estimate is not so well suited for our method, because our new iterate is only obtained by approximately solving the subproblem (2.9). To ensure global convergence, theoretically Lagrange multipliers are normally required to be bounded or grow not too fast compared with penalty parameters (see, e.g., Lemma 4.1-4.2 in [13]). More precisely, the ratio $\|\boldsymbol{\lambda}_k\|/\sigma_k$ needs to approach zero, when σ_k increases to infinity. The easiest way to realize this is to restrict $\boldsymbol{\lambda}_k$ in a bounded region, e.g. $[\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}]$. Therefore, in our method, the following scheme is proposed to update $\boldsymbol{\lambda}_k$. We calculate

$$\tilde{\boldsymbol{\lambda}}_k = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \psi_k(\boldsymbol{\lambda}) := \|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda})^+ - \mathbf{x}_k\|^2 \tag{2.13}$$

and let $\boldsymbol{\lambda}_k = P_{[\boldsymbol{\lambda}_{\min},\boldsymbol{\lambda}_{\max}]}\tilde{\boldsymbol{\lambda}}_k$. Here, $\boldsymbol{\lambda}_{\min}$ and $\boldsymbol{\lambda}_{\max}$ are preset safeguards of upper and lower bounds of Lagrange multipliers, and in practice they are often set to be very small and large, respectively. Note that $\psi_k(\boldsymbol{\lambda})$ defined in (2.13) is a continuous, piecewise quadratic function with respect to $\boldsymbol{\lambda}$. Effective techniques could be applied to identify its different quadratic components and hence be able to solve (2.13) quite efficiently. In addition, it can be shown (see Section 4) that under certain nondegeneracy assumptions, when $\mathbf{x}_k$ is close to a KKT point $\mathbf{x}^*$ of (1.1), (2.13) is equivalent to the following smooth least squares problem:

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad \|[\mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}]_{\mathcal{I}_*}\|^2,$$

where $\mathcal{I}_*$ consists of all indices of inactive bound constraints at $\mathbf{x}^*$. This property provides us a practical way of obtaining a good initial guess for computing $\tilde{\boldsymbol{\lambda}}_k$. Suppose that at the k-th iteration $\mathcal{I}_k$ is a good estimate of $\mathcal{I}_*$. Then, a good initial guess of $\tilde{\boldsymbol{\lambda}}_k$ could be the solution of the following smooth least squares problem:

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \quad \|[\mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}]_{\mathcal{I}_k}\|^2.$$

REMARK 2.2. *Although in our method* (2.13) *is adopted to compute* $\tilde{\boldsymbol{\lambda}}_k$*, actually an other practical easiest way could be simply setting*

$$\tilde{\boldsymbol{\lambda}}_k = \boldsymbol{\lambda}_{k-1} - \sigma_{k-1}\mathbf{c}_k. \tag{2.14}$$

However, the theoretical advantage of using (2.13) *is that a stronger global convergence property can be achieved, as shown in the following Theorem 3.5.*

2.3. Update of penalty parameters. We now discuss the strategy of updating the penalty parameter σ_k. Note that in subproblem (2.9), it may happen that $\mathbf{D}_k\bar{\mathbf{g}}_k = \mathbf{0}$, which by the definition of $\bar{\mathbf{g}}_k$ in (2.5) indicates that $\mathbf{d} = \mathbf{0}$ is a KKT point of the following problem:

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & \mathcal{L}(\mathbf{x}_k + \mathbf{D}_k\mathbf{d}; \boldsymbol{\lambda}_k, \sigma_k) \\ \text{s. t.} \quad & \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \geq \mathbf{0}. \end{aligned}$$

In this case, if $\mathbf{x}_k$ is feasible, by (2.8) we know that it is a KKT solution of (1.1). Hence, we can terminate the algorithm. Otherwise, if $\mathbf{x}_k$ is infeasible (note that $\mathbf{x}_k$ is always kept feasible to the bound constraints), that is

$$\mathbf{D}_k\bar{\mathbf{g}}_k = \mathbf{0} \quad \text{and} \quad \mathbf{c}_k \neq \mathbf{0}, \tag{2.15}$$

we break the first equality by repeatedly increasing σ_k. If for all sufficiently large σ, $\mathbf{D}_k\nabla_x\mathcal{L}(\mathbf{x}_k; \boldsymbol{\lambda}_k, \sigma) = \mathbf{0}$ and $\mathbf{c}_k \neq \mathbf{0}$, it can be shown by Lemma 2.3 that $\mathbf{x}_k$ is a stationary point of minimizing $\|\mathbf{c}(\mathbf{x})\|^2$ with the bound constraints. So, our algorithm detects whether (2.15) holds at $\mathbf{x}_k$ or not. If (2.15) holds, we increase the penalty parameter σ_k to break the the first equality of (2.15). Notice that increasing σ_k also helps reducing the constraint violation.

Now assume that the subproblem (2.9) returns a nonzero trial step $\mathbf{s}_k$. Then, our updating strategies for σ_k mainly depend on the improvement of constraint violations. In many approaches, the penalty parameter σ_k increases if sufficient improvement on the feasibility is not obtained, that is σ_k will be increased, if $\|\mathbf{c}_{k+1}\| \geq \tilde{\tau}\|\mathbf{c}_k\|$ for some constant $0 < \tilde{\tau} < 1$. An adaptive way of updating penalty parameters for augmented Lagrangian methods can also be found in [16]. Inspired by [41], we propose a different strategy to update the penalty parameter. In [41] for equality constrained optimization, the authors propose to test the condition $\text{Pred}_k < \delta_k\sigma_k \min\{\Delta_k\|\mathbf{c}_k\|, \|\mathbf{c}_k\|^2\}$ to decide whether to update σ_k or not, where δ_k, $k = 1, 2, \ldots$, is a prespecified sequence of parameters converging to zero. In this paper, however, with respect to the predicted reduction Pred_k defined in (2.10), we propose the following condition:

$$\text{Pred}_k < \frac{\delta}{\sigma_k}\min\{\Delta_k\|\mathbf{c}_k\|, \|\mathbf{c}_k\|^2\}. \tag{2.16}$$

where $\delta > 0$ is a constant. If (2.16) holds, we believe that the constraint violation $\|\mathbf{c}_k\|$ is still relatively large. So we increase σ_k, which to some extent can help reducing constraint violations in future iterations. We believe that the switching condition (2.16) will provide a more adaptive way to reduce the objective function value and the constraint violation simultaneously. Besides, this new condition (2.16) would increase σ_k less frequently compared with the strategy in [41], therefore the subproblem would be more well-conditioned for numerical solving.

2.4. Algorithm description. We now summarize the above discussions into the following algorithm.

Algorithm 1 Augmented Lagrangian Affine Scaling (ALAS) Algorithm

Step 0: Initialization. Given initial guess $\mathbf{x}_0$, compute f_0, $\mathbf{g}_0$, $\mathbf{c}_0$ and $\mathbf{A}_0$. Given initial guesses $\mathbf{B}_0$, σ_0, Δ_0 and parameters $\beta \in (0,1)$, $\delta > 0$, $\Delta_{\max} > 0$, $\boldsymbol{\lambda}_{\min} < \mathbf{0} < \boldsymbol{\lambda}_{\max}$,

$$\theta_1,\ \theta_2 > 1, \quad 0 < \eta < \eta_1 < \frac{1}{2}, \quad R > 0,$$

set $R_0 = \max\{\|\mathbf{c}_0\|, R\}$. Set $k := 0$. Compute $\tilde{\boldsymbol{\lambda}}_0$ by (2.13) and set $\boldsymbol{\lambda}_0 = P_{[\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}]}\tilde{\boldsymbol{\lambda}}_0$.

Step 1: Computing Scaling Matrices. Compute $\bar{\mathbf{g}}_k$ by (2.5) and $\mathbf{D}_k$ by (2.7).

Step 2: Termination Test. If $\mathbf{c}_k = \mathbf{0}$ and $\mathbf{D}_k\bar{\mathbf{g}}_k = \mathbf{0}$, stop and return $\mathbf{x}_k$ as the solution.

Step 3: Computing Trial Steps. Set $\sigma_k^{(0)} = \sigma_k$, $\mathbf{D}_k^{(0)} = \mathbf{D}_k$ and $j := 0$.
While $\mathbf{D}_k^{(j)}\nabla_x\mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k, \sigma_k^{(j)}) = \mathbf{0}$ and $\mathbf{c}_k \neq 0$,

$$\sigma_k^{(j+1)} = \theta_1\sigma_k^{(j)}; \tag{2.17}$$

$$[\mathbf{D}_k^{(j+1)}]_{ii} = \begin{cases} \sqrt{x_{ki}}, & \text{if } [\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}_k, \boldsymbol{\lambda}_k, \sigma_k^{(j+1)})]_i > 0, \\ 1, & \text{otherwise;} \end{cases} \tag{2.18}$$

$$j := j + 1;$$

Endwhile
Set $\sigma_k = \sigma_k^{(j)}$ and $\mathbf{D}_k = \mathbf{D}_k^{(j)}$. Solve the subproblem (2.9) to yield a trial step $\tilde{\mathbf{s}}_k$ satisfying (2.11) and set $\mathbf{s}_k = \mathbf{D}_k\tilde{\mathbf{s}}_k$.

Step 4: Updating Iterates. Let

$$\text{Ared}_k = \mathcal{L}(\mathbf{x}_k; \boldsymbol{\lambda}_k, \sigma_k) - \mathcal{L}(\mathbf{x}_k + \mathbf{s}_k; \boldsymbol{\lambda}_k, \sigma_k) \quad \text{and} \quad \rho_k = \frac{\text{Ared}_k}{\text{Pred}_k},$$

where Pred_k is defined by (2.10).
If $\rho_k < \eta$, $\Delta_{k+1} = \|\tilde{\mathbf{s}}_k\|/4$, $\mathbf{x}_{k+1} = \mathbf{x}_k$, $k := k+1$, go to Step 3; otherwise, set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$. Calculate f_{k+1}, $\mathbf{g}_{k+1}$, $\mathbf{c}_{k+1}$ and $\mathbf{A}_{k+1}$.

Step 5: Updating Multipliers and Penalty Parameters.
If $\|\mathbf{c}_{k+1}\| \leq R_k$, then compute

$$\begin{aligned} &\tilde{\boldsymbol{\lambda}}_{k+1} \text{ as the minimizer of } \psi_{k+1}(\boldsymbol{\lambda}) \text{ defined by (2.13) or through (2.14),} \\ &\text{set} \quad \boldsymbol{\lambda}_{k+1} = P_{[\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}]}\tilde{\boldsymbol{\lambda}}_{k+1} \quad \text{and} \quad R_{k+1} = \beta R_k; \end{aligned} \tag{2.19}$$

otherwise, $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k$, $R_{k+1} = R_k$.
If (2.16) is satisfied,

$$\sigma_{k+1} = \theta_2\sigma_k; \tag{2.20}$$

otherwise, $\sigma_{k+1} = \sigma_k$.
Compute $\mathbf{B}_{k+1}$, which is (or some approximation to) the Hessian of $f(\mathbf{x}) - \boldsymbol{\lambda}_{k+1}\mathbf{c}(\mathbf{x})$ at $\mathbf{x}_{k+1}$.

Step 6: Updating Trust Region Radii.
Set

$$\Delta_{k+1} = \begin{cases} \min\{\max\{\Delta_k, 1.5\|\tilde{\mathbf{s}}_k\|\}, \Delta_{\max}\}, & \text{if } \rho_k \in [1-\eta_1, \infty), \\ \Delta_k, & \text{if } \rho_k \in [\eta_1, 1-\eta_1), \\ \max\{0.5\Delta_k, 0.75\|\tilde{\mathbf{s}}_k\|\}, & \text{if } \rho_k \in [\eta, \eta_1). \end{cases} \tag{2.21}$$

Let $k := k+1$ and go to Step 1.

In ALAS, B_k is updated as the Hessian of the Lagrange function $f(\mathbf{x}) - \boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x})$ or its approximation. But from global convergence point of view, we only need $\{\mathbf{B}_k\}$ is uniformly bounded. We update the penalty parameters in the loop (2.17)-(2.18) to move the iterates away from infeasible local minimizers of the augmented Lagrangian function. The following lemma shows that if it is an infinite loop, then $\mathbf{x}_k$ would

be a stationary point of minimizing $\|\mathbf{c}(\mathbf{x})\|^2$ subject to bound constraints.

LEMMA 2.3. *Suppose that* $\mathbf{x}_k$ *is an iterate satisfying* $\mathbf{D}_k^{(j)}\nabla_x\mathcal{L}(\mathbf{x}_k;\boldsymbol{\lambda}_k,\sigma_k^{(j)}) = \mathbf{0}$ *with* $\sigma_k^{(j)}$ *and* $\mathbf{D}_k^{(j)}$ *updated by* (2.17)-(2.18) *in an infinite loop. Then* $\mathbf{x}_k$ *is a KKT point of*

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \|\mathbf{c}(\mathbf{x})\|^2 \\ \text{s. t.} \quad & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{2.22}$$

Proof. If $\sigma_k^{(j)}$ is updated by (2.17) in an infinite loop, then $\lim_{j\to\infty}\sigma_k^{(j)} = \infty$ and for all j

$$\mathbf{D}_k^{(j)}\nabla_x\mathcal{L}(\mathbf{x}_k;\boldsymbol{\lambda}_k,\sigma_k^{(j)}) \;=\; \mathbf{D}_k^{(j)}(\mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}_k + \sigma_k^{(j)}\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k) \;=\; \mathbf{0}.$$

Since $\{\mathbf{D}_k^{(j)}\}$ is equal to 1 or $\sqrt{x_{ki}} \geq 0$, it has a subsequence, still denoted as $\mathbf{D}_k^{(j)}$, converging to a diagonal matrix, say $\bar{\mathbf{D}}_k$, with $[\bar{\mathbf{D}}_k]_{ii} \geq 0$. Then, for those i with $[\bar{\mathbf{D}}_k]_{ii} = 0$, we have $(\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k)_i \geq 0$ and $x_{ki} = 0$ due to the definition of $\mathbf{D}_k$. And for those i with $[\bar{\mathbf{D}}_k]_{ii} > 0$, we have $[\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k]_i = 0$. Hence, we have

$$\begin{cases} (\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k)_i \geq 0, & \text{if } x_{ki} = 0, \\ (\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k)_i = 0, & \text{if } x_{ki} > 0, \end{cases}$$

which indicates that $\mathbf{x}_k$ is a KKT point of (2.22). □

Theoretically, the loop (2.17)-(2.18) can be infinite. However, in practical computation, once $\sigma_k^{(j)}$ reaches a preset large number, we simply terminate this algorithm and return $\mathbf{x}_k$ as an approximate infeasible stationary point of (1.1). Thus, in the following context we assume that the loop (2.17)-(2.18) finishes finitely at each iteration. Once the loop terminates, we have $\mathbf{D}_k\bar{\mathbf{g}}_k \neq \mathbf{0}$ and the subproblem (2.9) is solved. Then, it is not difficult to see that (2.9) always generates a trial step $\tilde{\mathbf{s}}_k$ such that $\rho_k \geq \eta$, when the trust region radius Δ_k is sufficiently small (see, e.g., [15]). Therefore, the loop between Step 3 and Step 4 in ALAS will stop in finite number of inner iterations.

3. Global convergence.

In this section, we assume that an infinite sequence of iterates $\{\mathbf{x}_k\}$ is generated by ALAS. The following assumptions are required throughout the analysis of this paper.

AS.1 $f:\mathbb{R}^n \to \mathbb{R}$ *and* $\mathbf{c}:\mathbb{R}^n \to \mathbb{R}^m$ *are Lipschitz continuously differentiable.*

AS.2 $\{\mathbf{x}_k\}$ *and* $\{\mathbf{B}_k\}$ *are bounded.*

Without the assumption **AS.2**, ALAS may allow unbounded minimizers. There are many problem-dependent sufficient conditions to support **AS.2**. When the optimization problem has both finite lower and upper bound constraints, all the iterates $\{\mathbf{x}_k\}$ will be certainly bounded. Another sufficient condition could be the existence of $M, \epsilon > 0$ such that the set $\{x \in \mathbb{R}^n : |f(\mathbf{x})| < M,\ \|\mathbf{c}(\mathbf{x})\| < \epsilon,\ \mathbf{x} \geq \mathbf{0}\}$ is bounded. We shall start with the following important property of ALAS.

LEMMA 3.1. *Under assumptions* ***AS.1***-***AS.2****, given any two integers* p *and* q *with* $0 \leq p \leq q$*, we have*

$$\|\mathbf{c}(\mathbf{x}_{q+1})\|^2 \;\leq\; \|\mathbf{c}(\mathbf{x}_p)\|^2 + 2\,M_0\,\sigma_p^{-1}, \tag{3.1}$$

where $M_0 = 2\,f_{\max} + 2\,\pi_\lambda\,(c_{\max} + R_0\,/\,(1-\beta))$*,* $f_{\max}$ *is an upper bound of* $\{|f(\mathbf{x}_k)|\}$*,* $c_{\max}$ *is an upper bound of* $\{\|\mathbf{c}_k\|\}$ *and* π_λ *is an upper bound of* $\{\|\boldsymbol{\lambda}_k\|\}$*.*

Proof. As the proof is almost identical to that of Lemma 3.1 in [41], we give its proof in the appendix. □

The following lemma provides a lower bound on the predicted objective function value reduction obtained by the subproblem (2.9).

LEMMA 3.2. *The predicted objective function value reduction defined by* (2.10) *satisfies*

$$\text{Pred}_k \;\geq\; \frac{\bar{\beta}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\mathbf{D}_k\bar{\mathbf{B}}_k\mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|}\right\}, \tag{3.2}$$

where $\bar{\mathbf{B}}_k = \mathbf{B}_k + \sigma_k\mathbf{A}_k^{\mathsf{T}}\mathbf{A}_k$*.*

Proof. Define $\mathbf{d}_k(\tau) = -\tau \frac{\mathbf{D}_k \bar{\mathbf{g}}_k}{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}$ with $\tau \geq 0$. Then, due to the definition of $\mathbf{D}_k$ in (2.7), $\mathbf{d}(\tau)$ is feasible for (2.9) if

$$\tau \in \left[0, \min_{\bar{g}_{ki}>0} \left\{ \Delta_k, \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{\bar{g}_{ki}} \right\} \right], \tag{3.3}$$

which implies that $\mathbf{d}(\tau)$ is feasible if $\tau \in [0, \min\{\Delta_k, \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|}\}]$. So considering the largest reduction of $\tilde{q}_k$ along $\mathbf{d}(\tau)$, we have

$$\begin{aligned} &\tilde{q}_k(\mathbf{0}) - \min\left\{ \tilde{q}_k(\mathbf{d}(\tau)) : 0 \leq \tau \leq \min\left\{ \Delta_k, \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|} \right\} \right\} \\ \geq\ & \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{2} \min\left\{ \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{\|\mathbf{D}_k \bar{\mathbf{B}}_k \mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k \bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|} \right\}. \end{aligned}$$

Then, (2.11) indicates that (3.2) holds. □

3.1. Global convergence with bounded penalty parameters. From the construction of ALAS, the penalty parameters $\{\sigma_k\}$ is a monotone nondecreasing sequence. Hence, when k goes to infinity, σ_k has a limit, either infinite or finite. So our following analysis is separated into two parts. We first study the global convergence of ALAS when $\{\sigma_k\}$ is bounded. And the case when $\{\sigma_k\}$ is unbounded will be addressed in Section 3.2. The following lemma shows that any accumulation point of $\{\mathbf{x}_k\}$ is feasible if $\{\sigma_k\}$ is bounded.

LEMMA 3.3. *Under assumptions **AS.1-AS.2**, assuming that* $\lim_{k\to\infty} \sigma_k = \sigma < \infty$*, we have*

$$\lim_{k\to\infty} \mathbf{c}_k = \mathbf{0}, \tag{3.4}$$

which implies that all the accumulation points of $\{\mathbf{x}_k\}$ *are feasible.*

Proof. If $\lim_{k\to\infty} \sigma_k = \sigma < \infty$, by (2.20) in ALAS, the predicted reduction Pred_k satisfies

$$\mathrm{Pred}_k \geq \frac{\delta}{\sigma} \min\{\Delta_k \|\mathbf{c}_k\|, \|\mathbf{c}_k\|^2\}, \tag{3.5}$$

for all sufficiently large k. Without loss of generality, we assume that $\sigma_k = \sigma$ for all k. The update scheme of $\boldsymbol{\lambda}_k$ in Step 4 together with **AS.2** implies that the sum of all $-\boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_k + \boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_{k+1}$, $k = 0, 1, \ldots$, is bounded, which is shown in (A.4) in Appendix:

$$\sum_{k=0}^{\infty} (-\boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_k + \boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_{k+1}) \leq 2\pi_\lambda \left(c_{\max} + \frac{1}{1-\beta} R_0 \right) < \infty,$$

where $c_{\max}$ is an upper bound of $\{\|\mathbf{c}_k\|\}$ and π_λ is an upper bound of $\{\|\boldsymbol{\lambda}_k\|\}$. Thus, the sum of Ared_k is bounded from above:

$$\begin{aligned} \sum_{k=0}^{\infty} \mathrm{Ared}_k &= \sum_{k=0}^{\infty} (f_k - f_{k+1}) + \sum_{k=0}^{\infty} (-\boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_k + \boldsymbol{\lambda}_k^\mathsf{T} \mathbf{c}_{k+1}) + \frac{\sigma}{2} \sum_{k=0}^{\infty} (\|\mathbf{c}_k\|^2 - \|\mathbf{c}_{k+1}\|^2) \\ &\leq 2 f_{\max} + 2\pi_\lambda \left(c_{\max} + \frac{1}{1-\beta} R_0 \right) + \sigma c_{\max}^2 < \infty, \end{aligned} \tag{3.6}$$

where $f_{\max}$ is an upper bound of $\{|f(\mathbf{x}_k)|\}$.

To prove (3.4), we first show

$$\liminf_{k\to\infty} \|\mathbf{c}_k\| = 0 \tag{3.7}$$

by contradiction. Suppose there exists a constant $\bar{\tau} > 0$ such that $\|\mathbf{c}_k\| \geq \bar{\tau}$ for all k. Then (3.5) gives

$$\mathrm{Pred}_k \geq \frac{\delta}{\sigma} \min\{\Delta_k \bar{\tau}, \bar{\tau}^2\} \quad \text{for all large } k.$$

Let $\mathcal{S}$ be the set of iteration numbers corresponding to successful iterations, i.e.,

$$\mathcal{S} = \{k \in \mathbb{N} : \rho_k \geq \eta\}. \tag{3.8}$$

Since $\{\mathbf{x}_k\}$ is an infinite sequence, $\mathcal{S}$ is an infinite set. Then (3.6) indicates that

$$\frac{\delta}{\sigma}\sum_{k\in\mathcal{S}} \min\{\Delta_k\bar{\tau}, \bar{\tau}^2\} \;\leq\; \sum_{k\in\mathcal{S}} \mathrm{Pred}_k \;\leq \frac{1}{\eta}\sum_{k\in\mathcal{S}} \mathrm{Ared}_k \;=\; \frac{1}{\eta}\sum_{k=0}^{\infty} \mathrm{Ared}_k \;<\; \infty, \tag{3.9}$$

which yields $\{\Delta_k\}_{\mathcal{S}} \to 0$ as $k \to \infty$. Then, by update rules (2.21) of trust region radii, we have

$$\lim_{k\to\infty} \Delta_k = 0. \tag{3.10}$$

By **AS.1-AS.2** and the boundedness of $\boldsymbol{\lambda}_k$, we have

$$|\mathrm{Ared}_k - \mathrm{Pred}_k| \leq \left| f(\mathbf{x}_k) - \boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}_k + (\mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}_k)^{\mathsf{T}}\mathbf{s}_k + \frac{1}{2}\mathbf{s}^{\mathsf{T}}\mathbf{B}_k\mathbf{s}_k - (f(\mathbf{x}_k+\mathbf{s}_k) - \boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_k+\mathbf{s}_k))\right| +$$
$$\frac{\sigma_k}{2}\left|\|\mathbf{c}_k + \mathbf{A}_k\mathbf{s}_k\|^2 - \|\mathbf{c}(\mathbf{x}_k+\mathbf{s}_k)\|^2\right| \leq M\|\mathbf{s}_k\|^2 \leq M\|\mathbf{D}_k\|^2\Delta_k^2,$$

where M is a positive constant. By the definition of $\mathbf{D}_k$ and **AS.2**, $\{\mathbf{D}_k\}$ is bounded, i.e., $\|\mathbf{D}_k\| \leq D_{\max}$ for some $D_{\max} > 0$. So, (3.10) implies that

$$|\rho_k - 1| = \frac{|\mathrm{Ared}_k - \mathrm{Pred}_k|}{\mathrm{Pred}_k} \leq \frac{MD_{\max}^2\Delta_k^2}{\frac{\delta}{\sigma}\min\{\Delta_k\bar{\tau}, \bar{\tau}^2\}} \to 0, \quad \text{as } k \to \infty,$$

Hence, again by the rules of updating trust region radii, we have $\Delta_{k+1} \geq \Delta_k$ for all large k. So, Δ_k is bounded away from 0. However, this contradicts (3.10). Therefore (3.7) holds.

We next prove the stronger result that $\lim_{k\to\infty}\|\mathbf{c}_k\| = 0$ by contradiction. Since $\{\mathbf{x}_k\}$ is an infinite sequence, $\mathcal{S}$ is infinite. So we assume that there exists an infinite set $\{m_i\} \subset \mathcal{S}$ and $\nu > 0$ such that

$$\|\mathbf{c}_{m_i}\| \geq 2\nu. \tag{3.11}$$

Because of (3.7), there exists a sequence $\{n_i\}$ such that

$$\|\mathbf{c}_k\| \geq \nu \;(m_i \leq k < n_i) \quad \text{and} \quad \|\mathbf{c}_{n_i}\| < \nu. \tag{3.12}$$

Now, let us consider the set

$$\mathcal{K} = \bigcup_i \{k \in \mathcal{S} : m_i \leq k < n_i\}.$$

By (3.6), we have $\mathrm{Ared}_k \to 0$, which implies that for all sufficiently large $k \in \mathcal{K}$,

$$\mathrm{Ared}_k \geq \eta\frac{\delta}{\sigma}\min[\Delta_k\nu, \nu^2] \geq \xi\Delta_k,$$

where $\xi = \eta\delta\nu/\sigma$. Hence, when i is sufficiently large we have

$$\|\mathbf{x}_{m_i} - \mathbf{x}_{n_i}\| \leq \sum_{k=m_i}^{n_i-1}\|\mathbf{x}_k - \mathbf{x}_{k+1}\| \leq \sum_{k=m_i,k\in\mathcal{S}}^{n_i-1}\|\mathbf{D}_k\|\Delta_k \leq \frac{D_{\max}}{\xi}\sum_{k=m_i,k\in\mathcal{S}}^{n_i-1}\mathrm{Ared}_k. \tag{3.13}$$

The boundedness of $\sum_{k=0}^{\infty}\mathrm{Ared}_k$ shown in (3.6) also implies

$$\lim_{i\to\infty}\sum_{k=m_i,k\in\mathcal{S}}^{n_i-1}\mathrm{Ared}_k = 0.$$

Hence, $\lim_{i\to\infty}\|\mathbf{x}_{m_i}-\mathbf{x}_{n_i}\|=0$. Therefore, from (3.12) we have $\|\mathbf{c}_{m_i}\|<2\nu$ for large i, which contradicts (3.11). Consequently, (3.4) holds. Since $\mathbf{x}_k\geq\mathbf{0}$ for all k, all the accumulation points of $\{\mathbf{x}_k\}$ are feasible. □

By applying Lemma 3.2, we obtain the following lower bound on Pred_k when $\{\sigma_k\}$ is bounded.

LEMMA 3.4. *Under assumptions **AS.1-AS.2**, assume that* $\lim_{k\to\infty}\sigma_k=\sigma<\infty$. *Then the predicted objective function value reduction defined by* (2.10) *satisfies*

$$\mathrm{Pred}_k \;\geq\; \frac{\bar\beta}{\bar M}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|\min\{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|,\Delta_k\}, \tag{3.14}$$

where $\bar M$ is a positive constant.

Proof. By Lemma 3.2, we have

$$\mathrm{Pred}_k \;\geq\; \frac{\bar\beta\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\mathbf{D}_k\bar{\mathbf{B}}_k\mathbf{D}_k\|},\Delta_k,\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|}\right\}, \tag{3.15}$$

where $\bar{\mathbf{B}}_k=\mathbf{B}_k+\sigma_k\mathbf{A}_k^{\mathsf T}\mathbf{A}_k$. Since $\{\sigma_k\}$ is bounded, $\{\bar{\mathbf{B}}_k\}$ and $\{\bar{\mathbf{g}}_k\}$ are all bounded. Then the definition of $\{\mathbf{D}_k\}$ in (2.7) and **AS.2** indicate that $\mathbf{D}_k$ is bounded as well. So there exists a positive number $\bar M$ such that

$$2\max\{\|\bar{\mathbf{g}}_k\|,\|\mathbf{D}_k\bar{\mathbf{B}}_k\mathbf{D}_k\|,1\}\leq\bar M.$$

Then (3.15) yields that

$$\mathrm{Pred}_k\geq\frac{\bar\beta\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\max\{\|\bar{\mathbf{g}}_k\|,\|\mathbf{D}_k\bar{\mathbf{B}}_k\mathbf{D}_k\|\}},\Delta_k\right\},$$

which further yields (3.14). □

We now give the main convergence result for the case that $\{\sigma_k\}$ is bounded.

THEOREM 3.5. *Under assumptions **AS.1-AS.2**, assuming that* $\lim_{k\to\infty}\sigma_k=\sigma<\infty$ *, we have*

$$\liminf_{k\to\infty}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|=0, \tag{3.16}$$

which implies that at least one accumulation point of $\{\mathbf{x}_k\}$ is a KKT point of (1.1). *Furthermore, if $\boldsymbol{\lambda}_k=\arg\min\psi_k(\boldsymbol{\lambda})$ for all large k, where ψ_k is defined by* (2.13), *we have*

$$\lim_{k\to\infty}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|=0, \tag{3.17}$$

which implies that all accumulation points of $\{\mathbf{x}_k\}$ are KKT points of (1.1).

Proof. By contradiction we can prove that

$$\liminf_{k\to\infty}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|=0. \tag{3.18}$$

Suppose that there exists $\epsilon>0$ such that $\|\mathbf{D}_k\bar{\mathbf{g}}_k\|\geq\epsilon$ for all k. Then (3.14) indicates that

$$\mathrm{Pred}_k \;\geq\; \frac{\bar\beta}{\bar M}\epsilon\min\{\epsilon,\Delta_k\}, \tag{3.19}$$

Then by mimicking the analysis after (3.7), we can derive a contradiction, so we obtain (3.18). Hence, due to $\lim_{k\to\infty}\mathbf{c}_k=\mathbf{0}$ by Lemma 3.3, we have

$$\liminf_{k\to\infty}\|\mathbf{D}_k(\mathbf{g}_k-\mathbf{A}_k^{\mathsf T}\boldsymbol{\lambda}_k)\|=0.$$

Then, as $\{\boldsymbol{\lambda}_k\}$ and $\{\mathbf{x}_k\}$ are bounded, there exist subsequences $\{\mathbf{x}_{k_i}\}$, $\{\boldsymbol{\lambda}_{k_i}\}$, $\mathbf{x}^*$ and $\boldsymbol{\lambda}^*$ such that $\mathbf{x}_{k_i}\to\mathbf{x}^*$, $\boldsymbol{\lambda}_{k_i}\to\boldsymbol{\lambda}^*$ and

$$\lim_{i\to\infty}\mathbf{D}_{k_i}(\mathbf{g}_{k_i}-\mathbf{A}_{k_i}^{\mathsf T}\boldsymbol{\lambda}_{k_i})=\mathbf{0}.$$

Therefore, from the definition of $\mathbf{D}_k$ in Step 3 of ALAS and $\mathbf{0} = \mathbf{c}(\mathbf{x}^*) = \lim_{i\to\infty} \mathbf{c}(\mathbf{x}_{k_i})$, we can derive

$$\begin{cases} [\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*]_i = 0, & \text{if} \quad x_i^* > 0, \\ [\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*]_i \geq 0, & \text{if} \quad x_i^* = 0. \end{cases} \tag{3.20}$$

This together with $\mathbf{x}_k \geq \mathbf{0}$ gives that $\mathbf{x}^*$ is a KKT point of (1.1). Hence, (3.16) holds.

Now, if $\boldsymbol{\lambda}_k = \arg\min \psi_k$ for all large k, we now prove (3.17) in three steps. Without loss of generality, we assume $\boldsymbol{\lambda}_k = \arg\min \psi_k$ for all k. Firstly, we prove that for any two iterates $\mathbf{x}_k$ and $\mathbf{x}_l$ there exists a constant $C > 0$, independent of k and l, such that

$$\big|\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| - \|(\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_l\|\big| \leq C\|\mathbf{x}_k - \mathbf{x}_l\|. \tag{3.21}$$

Since $\boldsymbol{\lambda}_k = \arg\min \psi_k$ with ψ_k defined in (2.13), we have

$$\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| \ \leq \ \|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_k\|.$$

Hence,

$$\begin{aligned} &\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| - \|(\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_l\| \\ \leq\ &\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_k\| - \|(\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_l\| \\ \leq\ &\|((\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_k) - ((\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)^+ - \mathbf{x}_l)\| \\ \leq\ &\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_l)^+ - (\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)^+\| + \|\mathbf{x}_k - \mathbf{x}_l\| \\ \leq\ &\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_l) - (\mathbf{x}_l - \mathbf{g}_l + \mathbf{A}_l^\mathsf{T}\boldsymbol{\lambda}_l)\| + \|\mathbf{x}_k - \mathbf{x}_l\| \\ \leq\ &C\|\mathbf{x}_k - \mathbf{x}_l\|, \end{aligned}$$

where $C > 0$ is a constant. Here, the last inequality follows from the assumptions **AS.1**-**AS.2** and the boundedness of $\{\boldsymbol{\lambda}_k\}$. Then, by the arbitrary of $\mathbf{x}_k$ and $\mathbf{x}_l$, we have that (3.21) holds.

Secondly, we show that

$$\lim_{k\in\mathcal{K}, k\to\infty} \mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k) = \mathbf{0} \tag{3.22}$$

is equivalent to

$$\lim_{k\in\mathcal{K}, k\to\infty} (\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k = \mathbf{0}, \tag{3.23}$$

where $\mathcal{K}$ is any infinite subset of $\mathbb{N}$. If (3.22) is satisfied, it follows from the same argument as showing (3.16) that any accumulation point $\mathbf{x}^*$ of $\{\mathbf{x}_k : k \in \mathcal{K}\}$ is a KKT point of (1.1). Hence, from KKT conditions (2.1), there exists $\boldsymbol{\lambda}^*$ such that $(\mathbf{x}^* - \mathbf{g}(\mathbf{x}^*) + (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*)^+ - \mathbf{x}^* = \mathbf{0}$. Then, by (3.21), we obtain

$$\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| \leq C\|\mathbf{x}_k - \mathbf{x}^*\|, \quad \forall k \in \mathcal{K}.$$

Since $\mathbf{x}^*$ can be any accumulation point of $\{\mathbf{x}_k : k \in \mathcal{K}\}$, the above inequality implies that (3.23) holds. We now assume that (3.23) is satisfied. Assume that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is any accumulation point of $\{(\mathbf{x}_k, \boldsymbol{\lambda}_k), k \in \mathcal{K}\}$. This implies that there exists a $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\mathbf{x}_k \to \mathbf{x}^*$ and $\boldsymbol{\lambda}_k \to \boldsymbol{\lambda}^*$ as $k \in \bar{\mathcal{K}}$, $k \to \infty$. Hence, we have

$$\lim_{k\in\bar{\mathcal{K}}, k\to\infty} \mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k = \mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*. \tag{3.24}$$

From (3.23) and (3.24) and $\mathbf{x}_k \geq \mathbf{0}$ for all k, we have $\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^* \geq \mathbf{0}$. For those i with $[\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*]_i = 0$, we have by (3.24) and the boundedness of $\{\mathbf{D}_k\}$ that $[\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)]_i \to 0$ as $k \in \bar{\mathcal{K}}, k \to \infty$. For those i with $[\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*]_i > 0$, on one hand, we have from (3.23) and (3.24) that $x_i^* = 0$; on the other hand, we have from (3.24) and $\lim_{k\to\infty} \|\mathbf{c}_k\| = 0$ (Lemma 3.3) that

$$\lim_{k\in\bar{\mathcal{K}}, k\to\infty} \bar{g}_{ki} = [\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*]_i > 0,$$

which implies $[\mathbf{D}_k]_{ii} = \sqrt{x_{ki}}$ for all large $k \in \bar{\mathcal{K}}$. Then it follows from $x_{ki} \to x_i^* = 0$ that $[\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)]_i \to 0$ as $k \in \bar{\mathcal{K}}, k \to \infty$. Hence, $\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)] \to \mathbf{0}$ as $k \in \bar{\mathcal{K}}, k \to \infty$ Then, since $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ can be any accumulation point of $\{(\mathbf{x}_k, \boldsymbol{\lambda}_k), k \in \mathcal{K}\}$, it implies (3.22). Therefore, we obtain the equivalence between (3.22) and (3.23).

Thirdly, we now prove $\lim_{k\to\infty} \mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k) = \mathbf{0}$ by contradiction. If this does not hold, then there exists an infinite indices set $\{m_i\} \subset \mathcal{S}$ and a constant $\bar{\nu} > 0$ such that $\|\mathbf{D}_{m_i}(\mathbf{g}_{m_i} - \mathbf{A}_{m_i}^\mathsf{T}\boldsymbol{\lambda}_{m_i})\| \geq \bar{\nu}$, where $\mathcal{S}$ is defined in (3.8). So, from the equivalence between (3.22) and (3.23), there exists a constant $\bar{\epsilon} > 0$ such that $\|(\mathbf{x}_{m_i} - \mathbf{g}_{m_i} + \mathbf{A}_{m_i}^\mathsf{T}\boldsymbol{\lambda}_{m_i})^+ - \mathbf{x}_{m_i}\| \geq \bar{\epsilon}$ for all m_i. Let $\bar{\delta} = \bar{\epsilon}/(3C)$. Then, it follows from (3.21) that

$$\|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| \geq 2\bar{\epsilon}/3, \qquad \text{if} \quad \mathbf{x}_k \in B_{\bar{\delta}}(\mathbf{x}_{m_i}). \tag{3.25}$$

Now, let $\mathcal{S}(\epsilon) = \{k : \|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)^+ - \mathbf{x}_k\| \geq \epsilon\}$. Then it again follows from (3.22) and (3.23) that there exists a $\hat{\nu} \in (0, \bar{\nu}]$ such that

$$\|\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)\| \geq \hat{\nu}, \quad \text{for all } k \in \mathcal{S}(2\bar{\epsilon}/3). \tag{3.26}$$

By (3.16) and $\hat{\nu} \leq \bar{\nu}$, we can find a subsequence $\{n_i\}$ such that $\|\mathbf{D}_{n_i}(\mathbf{g}_{n_i} - \mathbf{A}_{n_i}^\mathsf{T}\boldsymbol{\lambda}_{n_i})\| < \hat{\nu}/2$ and $\|\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)\| \geq \hat{\nu}/2$ for any $k \in [m_i, n_i)$. Following the same arguments as showing (3.13) in Lemma 3.3 and the boundedness of $\sum_{k=0}^{\infty} \text{Pred}_k$, we can obtain

$$\lim_{i\to\infty} (\mathbf{x}_{m_i} - \mathbf{x}_{n_i}) = \mathbf{0}.$$

Hence, for all large n_i, we have $\mathbf{x}_{n_i} \in B_{\bar{\delta}}(\mathbf{x}_{m_i})$. Thus by (3.25) we obtain $\|(\mathbf{x}_{n_i} - \mathbf{g}_{n_i} + \mathbf{A}_{n_i}^\mathsf{T}\boldsymbol{\lambda}_{n_i})^+ - \mathbf{x}_{n_i}\| \geq 2\bar{\epsilon}/3$. Then, it follows from (3.26) that $\|\mathbf{D}_{n_i}(\mathbf{g}_{n_i} - \mathbf{A}_{n_i}^\mathsf{T}\boldsymbol{\lambda}_{n_i})\| \geq \hat{\nu}$. However, this contradicts the way of choosing $\{n_i\}$ such that $\|\mathbf{D}_{n_i}(\mathbf{g}_{n_i} - \mathbf{A}_{n_i}^\mathsf{T}\boldsymbol{\lambda}_{n_i})\| < \hat{\nu}/2$. Hence, (3.17) holds, which implies that any accumulation point of $\{\mathbf{x}_k\}$ is a KKT point of (1.1). □

REMARK 3.6. *In Remark 2.2, we have mentioned that the computation of* $\tilde{\boldsymbol{\lambda}}$ *in* (2.13) *can be replaced with* (2.14)*. If* (2.14) *is adopted, the first part of Theorem 3.5 still holds, while the second part will not be guaranteed.*

3.2. Global convergence with unbounded penalty parameters. In this subsection, we study the behavior of ALAS when the penalty parameters $\{\sigma_k\}$ are unbounded. The following lemma shows that in this case the limit of $\{\|\mathbf{c}_k\|\}$ always exists .

LEMMA 3.7. *Under assumptions* ***AS.1-AS.2****, assume that* $\lim_{k\to\infty} \sigma_k = \infty$*, then* $\lim_{k\to\infty} \|\mathbf{c}_k\|$ *exists.*

Proof. Let $c_{\min} = \liminf_{k\to\infty} \|\mathbf{c}(\mathbf{x}_k)\|$ and $\{\mathbf{x}_{k_i}\}$ be a subsequence of $\{\mathbf{x}_k\}$ such that $c_{\min} = \lim_{i\to\infty} \|\mathbf{c}(\mathbf{x}_{k_i})\|$. By Lemma 3.1, substituting $\mathbf{x}_p$ in the right side of (3.1) by $\mathbf{x}_{k_i}$, for all $q > k_i$ we have that

$$\|\mathbf{c}(\mathbf{x}_{q+1})\|^2 \;\leq\; \|\mathbf{c}(\mathbf{x}_{k_i})\|^2 + 2M_0\sigma_p^{-1}.$$

Since $\lim_{p\to\infty} \sigma_p = \infty$, it follows from above inequality that $c_{\min} = \lim_{k\to\infty} \|\mathbf{c}(\mathbf{x}_k)\|$. □

Lemma 3.7 shows that the unboundedness of $\{\sigma_k\}$ ensures convergence of equality constraint violations $\{\|\mathbf{c}_k\|\}$. Then two cases might happen: $\lim_{k\to\infty} \|\mathbf{c}_k\| = 0$ or $\lim_{k\to\infty} \|\mathbf{c}_k\| > 0$. It is desirable if the iterates generated by ALAS converge to a feasible point. However, the constraints $\mathbf{c}(\mathbf{x}) = 0$ in (1.1) may not be feasible for any $\mathbf{x} \geq \mathbf{0}$. Hence, in the following we further divide our analysis into two parts. We first study the case that $\lim_{k\to\infty} \|\mathbf{c}_k\| > 0$. The following theorem shows that in this case any infeasible accumulation point is a stationary point of minimizing $\|\mathbf{c}(\mathbf{x})\|^2$ subject to the bound constraints.

THEOREM 3.8. *Under assumptions* ***AS.1-AS.2****, assume that* $\lim_{k\to\infty} \sigma_k = \infty$ *and* $\lim_{k\to\infty} \|\mathbf{c}_k\| > 0$*. Then all accumulation points of* $\{\mathbf{x}_k\}$ *are KKT points of* (2.22)*.*

Proof. We divide the proof into three major steps.

Step 1: We prove by contradiction that

$$\liminf_{k\to\infty} \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\sigma_k} = 0. \tag{3.27}$$

Assume that (3.27) does not hold. Then there exists a constant $\xi > 0$ such that

$$\|\mathbf{D}_k\bar{\mathbf{g}}_k\| \geq \xi\sigma_k \tag{3.28}$$

for all large k. Hence, when k is sufficiently large, (2.17) will never happen and σ_k will be updated only through (2.20). Then, by Lemma 3.2, we have

$$\begin{aligned}\mathrm{Pred}_k &\geq \frac{\bar{\beta}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\mathbf{D}_k(\mathbf{B}_k+\sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k)\mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|}\right\}\\ &= \frac{\bar{\beta}\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|/\sigma_k}{\|\mathbf{D}_k(\mathbf{B}_k/\sigma_k+\mathbf{A}_k^\mathsf{T}\mathbf{A}_k)\mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|/\sigma_k}{\|\bar{\mathbf{g}}_k\|/\sigma_k}\right\}.\end{aligned} \tag{3.29}$$

As $\{\mathbf{B}_k/\sigma_k + \mathbf{A}_k^\mathsf{T}\mathbf{A}_k\}$ and $\{\bar{\mathbf{g}}_k/\sigma_k\}$ are all bounded, (3.28) and (3.29) imply that there exists a constant $\zeta > 0$ such that

$$\mathrm{Pred}_k \ \geq\ \zeta\sigma_k\min\{\zeta, \Delta_k\}, \quad \text{for all large } k.$$

However, this contradicts with $\sigma_k \to \infty$, because the above inequality implies that σ_k will not be increased for all large k. Therefore, (3.27) holds. Then, by $\sigma_k \to \infty$ and (3.27), we have

$$\liminf_{k\to\infty}\ \|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\| = \liminf_{k\to\infty}\ \left\|\mathbf{D}_k\left(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)/\sigma_k + \mathbf{A}_k^\mathsf{T}\mathbf{c}_k\right)\right\| = \liminf_{k\to\infty}\ \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\sigma_k} = 0. \tag{3.30}$$

Step 2: We now prove by contradiction that

$$\lim_{k\in\mathcal{S},k\to\infty}\ \mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k = \mathbf{0}, \tag{3.31}$$

where $\mathcal{S}$ is defined in (3.8). Assume that (3.31) does not hold. Then, there exist $\bar{\mathbf{x}} \geq \mathbf{0}$, $\bar{\epsilon} > 0$ and a subset $\bar{\mathcal{K}} \subseteq \mathcal{S}$ such that

$$\lim_{k\in\bar{\mathcal{K}},k\to\infty}\ \mathbf{x}_k = \bar{\mathbf{x}} \quad \text{and} \quad \|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\| \geq \bar{\epsilon}\ \text{ for all } k \in \bar{\mathcal{K}}. \tag{3.32}$$

Let $\bar{\mathcal{F}} = \bar{\mathcal{F}}^+ \cup \bar{\mathcal{F}}^-$, where

$$\bar{\mathcal{F}}^+ := \{i : [\mathbf{A}(\bar{\mathbf{x}})^\mathsf{T}\mathbf{c}(\bar{\mathbf{x}})]_i > 0\} \quad \text{and} \quad \bar{\mathcal{F}}^- := \{i : [\mathbf{A}(\bar{\mathbf{x}})^\mathsf{T}\mathbf{c}(\bar{\mathbf{x}})]_i < 0\}.$$

Then, it follows from the boundedness of $\{\mathbf{D}_k\}$ and (3.32) that $\bar{\mathcal{F}}$ is not empty. Since $\sigma_k \to \infty$, by the definition of $\bar{\mathbf{g}}_k$, there exists a constant $\bar{\delta} > 0$ such that

$$\bar{g}_{ki}\begin{cases} > 0, & \text{if} \quad \mathbf{x}_k \in B_{\bar{\delta}}(\bar{\mathbf{x}}) \quad \text{and} \quad i \in \bar{\mathcal{F}}^+;\\ < 0, & \text{if} \quad \mathbf{x}_k \in B_{\bar{\delta}}(\bar{\mathbf{x}}) \quad \text{and} \quad i \in \bar{\mathcal{F}}^-.\end{cases}$$

Accordingly, we have

$$[\mathbf{D}_k]_{ii} = \begin{cases} \sqrt{x_{ki}}, & \text{if} \quad \mathbf{x}_k \in B_{\bar{\delta}}(\bar{\mathbf{x}}) \quad \text{and} \quad i \in \bar{\mathcal{F}}^+;\\ 1, & \text{if} \quad \mathbf{x}_k \in B_{\bar{\delta}}(\bar{\mathbf{x}}) \quad \text{and} \quad i \in \bar{\mathcal{F}}^-.\end{cases}$$

This together with (3.32) and continuity indicates that when $\bar{\delta}$ is sufficiently small

$$\|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\| \geq \bar{\epsilon}/2, \quad \text{if } \mathbf{x}_k \in B_{\bar{\delta}}(\bar{\mathbf{x}}). \tag{3.33}$$

Let $\varphi_k(\mathbf{d}) := \|\mathbf{c}_k + \mathbf{A}_k\mathbf{D}_k\mathbf{d}\|^2/2$. We consider the following problem:

$$\begin{aligned}\min_{\mathbf{d}\in\mathbb{R}^n} \quad & \varphi_k(\mathbf{d}) = (\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k)^\mathsf{T}\mathbf{d} + \frac{1}{2}\mathbf{d}^\mathsf{T}(\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k\mathbf{D}_k)\mathbf{d} + \frac{1}{2}\|\mathbf{c}_k\|^2\\ \text{s. t.} \quad & \|\mathbf{d}\| \leq \Delta_k, \quad \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \geq \mathbf{0},\end{aligned}$$

with its solution denoted by $\bar{\mathbf{d}}_k$. Then, similar to the proof of Lemma 3.2, we can show that for any k

$$\begin{aligned}&\frac{1}{2}(\|\mathbf{c}_k\|^2 - \|\mathbf{c}_k + \mathbf{A}_k\mathbf{D}_k\bar{\mathbf{d}}_k\|^2)\\ \geq\ &\frac{\|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\|}{2}\min\left\{\frac{\|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\|}{\|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k\mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\|}{\|\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\|}\right\}. \end{aligned} \tag{3.34}$$

Let $\mathcal{K} := \{k \in \mathcal{S} : \|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\| \geq \bar{\epsilon}/2\}$. Since $\bar{\mathcal{K}} \subseteq \mathcal{K}$, $\mathcal{K}$ is an infinite set. And for any $k \in \mathcal{K}$, it follows from (3.34) that there exists $\bar{\eta} > 0$ such that

$$\|\mathbf{c}_k\|^2 - \|\mathbf{c}_k + \mathbf{A}_k\mathbf{D}_k\bar{\mathbf{d}}_k\|^2 \geq \bar{\eta}\min\{1, \Delta_k\}.$$

Since $\mathbf{D}_k\bar{\mathbf{d}}_k$ is a feasible point of subproblem (2.6), the predicted reduction obtained by solving (2.9) satisfies the following relations:

$$\begin{aligned}\mathrm{Pred}_k &= q_k(\mathbf{0}) - q_k(\mathbf{s}_k) \geq \bar{\beta}\,[q_k(\mathbf{0}) - q_k(\mathbf{D}_k\bar{\mathbf{d}}_k)]\\ &= \bar{\beta}\,[-(\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k))^\mathsf{T}\bar{\mathbf{d}}_k - \frac{1}{2}\bar{\mathbf{d}}_k^\mathsf{T}\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\bar{\mathbf{d}}_k\\ &\quad + \frac{\sigma_k}{2}(\|\mathbf{c}_k\|^2 - \|\mathbf{c}_k + \mathbf{A}_k\mathbf{D}_k\bar{\mathbf{d}}_k\|^2)].\end{aligned}$$

As $\|\bar{\mathbf{d}}_k\| \leq \Delta_k \leq \Delta_{\max}$, we have

$$\begin{aligned}&\left|(\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k))^\mathsf{T}\bar{\mathbf{d}}_k - \frac{1}{2}\bar{\mathbf{d}}_k^\mathsf{T}\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\bar{\mathbf{d}}_k\right|\\ \leq\ &\left(\|\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)\| + \frac{1}{2}\|\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\|\Delta_{\max}\right)\Delta_k\\ \leq\ &\left(\|\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)\| + \frac{1}{2}\|\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\|\Delta_{\max}\right)\max\{1, \Delta_{\max}\}\min\{1, \Delta_k\}.\end{aligned}$$

Therefore, for any $k \in \mathcal{K}$ and k sufficiently large, by (2.11) we have

$$\begin{aligned}\mathrm{Pred}_k &\geq \bar{\beta}\left[\frac{\bar{\eta}\sigma_k}{2} - \left(\|\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k)\| + \frac{1}{2}\|\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\|\Delta_{\max}\right)\max\{1, \Delta_{\max}\}\right]\min\{1, \Delta_k\}\\ &\geq \frac{\bar{\beta}\bar{\eta}\sigma_k}{4}\min\{1, \Delta_k\},\end{aligned} \tag{3.35}$$

where the second inequality follows from $\sigma_k \to \infty$. From (3.30) and (3.32), there exist infinite sets $\{m_i\} \in \bar{\mathcal{K}}$ and $\{n_i\}$ such that

$$\mathbf{x}_{m_i} \in B_{\bar{\delta}/2}(\bar{\mathbf{x}}),\ \|\mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k\| \geq \frac{\bar{\epsilon}}{2} \text{ for } m_i \leq k < n_i \ \text{ and }\ \|\mathbf{D}_{n_i}\mathbf{A}_{n_i}^\mathsf{T}\mathbf{c}_{n_i}\| < \frac{\bar{\epsilon}}{2}. \tag{3.36}$$

Hence, we have

$$\{k \in \mathcal{S} : m_i \leq k < n_i\} = \{k \in \mathcal{K} : m_i \leq k < n_i\}.$$

Therefore, by (3.35), when m_i is sufficiently large

$$\begin{aligned}\frac{\eta\bar{\beta}\bar{\eta}}{4}\sum_{k=m_i, k\in\mathcal{S}}^{n_i-1}\min\{1, \Delta_k\} &\leq \sum_{k=m_i, k\in\mathcal{S}}^{n_i-1}\frac{\eta}{\sigma_k}\,\mathrm{Pred}_k \leq \sum_{k=m_i, k\in\mathcal{S}}^{n_i-1}\frac{1}{\sigma_k}\,\mathrm{Ared}_k\\ &\leq \frac{1}{2}(\|\mathbf{c}_{m_i}\|^2 - \|\mathbf{c}_{n_i}\|^2) + \frac{M_0}{\sigma_{m_i}},\end{aligned} \tag{3.37}$$

where the last inequality follows from (A.5) with

$$M_0 \;=\; 2\, f_{\max} + 2\, \pi_\lambda\, (c_{\max} + R_0 \,/\, (1-\beta)).$$

By Lemma 3.7, we have $\lim_{k\to\infty} \|\mathbf{c}_k\|$ exists. It together with $\lim_{k\to\infty} \sigma_k \to \infty$ and (3.37) yields

$$\sum_{k=m_i, k\in\mathcal{S}}^{n_i-1} \min\{1, \Delta_k\} \to 0, \quad \text{as} \quad m_i \to \infty,$$

which obviously implies $\sum_{k=m_i,k\in\mathcal{S}}^{n_i-1} \Delta_k \to 0$, as $m_i \to \infty$. Hence, we have

$$\begin{aligned}\|\mathbf{x}_{m_i} - \mathbf{x}_{n_i}\| &\le \sum_{k=m_i}^{n_i-1} \|\mathbf{x}_k - \mathbf{x}_{k+1}\| \le \sum_{k=m_i,k\in\mathcal{S}}^{n_i-1} \|\mathbf{D}_k\|\Delta_k \\ &\le D_{\max} \sum_{k=m_i,k\in\mathcal{S}}^{n_i-1} \Delta_k \to 0, \quad \text{as} \quad m_i \to \infty,\end{aligned}$$

where $D_{\max}$ is an upper bound of $\{\|\mathbf{D}_k\|\}$. So, $\lim_{i\to\infty} \|\mathbf{x}_{m_i} - \mathbf{x}_{n_i}\| = 0$. Then, it follows from $\mathbf{x}_{m_i} \in B_{\bar\delta/2}(\bar{\mathbf{x}})$ by (3.36) that $\mathbf{x}_{n_i} \in B_{\bar\delta}(\bar{\mathbf{x}})$ when m_i is large. And therefore by (3.33) we have $\|\mathbf{D}_{n_i}\mathbf{A}_{n_i}^\mathsf{T}\mathbf{c}_{n_i}\| \ge \bar\epsilon/2$. However, this constricts $\|\mathbf{D}_{n_i}\mathbf{A}_{n_i}^\mathsf{T}\mathbf{c}_{n_i}\| < \bar\epsilon/2$ shown in (3.36). Hence, (3.31) holds.

Step 3: We now prove that any accumulation point $\mathbf{x}^*$ of $\{\mathbf{x}_k\}$ is a KKT point of (2.22). Since $\mathbf{x}^*$ is also an accumulation point of $\{\mathbf{x}_k : k \in \mathcal{S}\}$, by (3.31) there exists a subset $\bar{\mathcal{S}} \subset \mathcal{S}$ such that

$$\lim_{k\in\bar{\mathcal{S}}, k\to\infty} \mathbf{x}_k = \mathbf{x}^* \quad \text{and} \quad \lim_{k\in\bar{\mathcal{S}}, k\to\infty} \mathbf{D}_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k = \mathbf{0}. \tag{3.38}$$

For the components of $\mathbf{x}^*$, there are two cases to happen.

Case I. If $x_i^* > 0$, due to the definition of $\mathbf{D}_k$, we know that $[\mathbf{D}_k]_{ii} > 0$ for all large $k \in \bar{\mathcal{S}}$. Hence, it follows from (3.38) that $[(\mathbf{A}^*)^\mathsf{T}\mathbf{c}^*]_i = 0$.

Case II. If $x_i^* = 0$, then $[(\mathbf{A}^*)^\mathsf{T}\mathbf{c}^*]_i \ge 0$ must hold. We prove it by contradiction. Assume that there exists a positive constant $\bar\epsilon$ such that

$$[(\mathbf{A}^*)^\mathsf{T}\mathbf{c}^*]_i \;\le\; -\bar\epsilon < 0. \tag{3.39}$$

Then since $\lim_{k\to\infty} \sigma_k = \infty$, for sufficiently large $k \in \bar{\mathcal{S}}$ we have

$$[\bar{\mathbf{g}}_k]_i = [\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k]_i < 0,$$

which implies that $[\mathbf{D}_k]_{ii} = 1$. Then, (3.38) indicates that $[(\mathbf{A}^*)^\mathsf{T}\mathbf{c}^*]_i = 0$, which contradicts (3.39).

Since $\mathbf{x}_k \ge \mathbf{0}$ for all k, we have from **Case I** and **Case II** that $\mathbf{x}^*$ is a KKT point of (2.22). □

In the following of this section, we prove the global convergence of ALAS to KKT points under the case that $\lim_{k\to\infty} \|\mathbf{c}_k\| = 0$. We now start with the following lemma.

LEMMA 3.9. *Under assumptions **AS.1-AS.2**, assume that* $\lim_{k\to\infty} \sigma_k = \infty$ *and* $\lim_{k\to\infty} \|\mathbf{c}_k\| = 0$. *Then the iterates* $\{\mathbf{x}_k\}$ *generated by ALAS satisfy*

$$\liminf_{k\to\infty} \|\mathbf{D}_k\bar{\mathbf{g}}_k\| = 0. \tag{3.40}$$

Proof. If (2.17) happens at infinite number of iterates, then (3.40) obviously holds. So, we assume that (2.17) does not happen for all large k. By Lemma 3.2, we have

$$\begin{aligned}\text{Pred}_k &\ge \bar\beta \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{2} \min\left\{ \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\mathbf{D}_k(\mathbf{B}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{A}_k)\mathbf{D}_k\|}, \Delta_k, \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{\|\bar{\mathbf{g}}_k\|} \right\} \\ &\ge \frac{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|}{M_1} \frac{1}{\sigma_k} \min\{\|\mathbf{D}_k\bar{\mathbf{g}}_k\|, \Delta_k\}\end{aligned}$$

for some constant $M_1 > 0$, since $\mathbf{D}_k$, $\mathbf{B}_k$, $\boldsymbol{\lambda}_k$ and $\mathbf{A}_k$ are all bounded. This lower bound on Pred_k indicates that (3.40) holds. Otherwise, there exists a constant $\zeta > 0$ such that

$$\text{Pred}_k \ \geq \ \frac{\zeta}{M_1}\frac{1}{\sigma_k}\min\{\zeta, \Delta_k\},$$

which indicates that (2.16) does not hold for all large k as $\|\mathbf{c}_k\| \to 0$. Hence, σ_k will be a constant for all large k, which contradicts $\lim_{k\to\infty}\sigma_k = \infty$. □

The constraint qualification, *Linear Independence Constraint Qualification* (LICQ) [33], is widely used in analyzing the optimality of feasible accumulation points. With respect to the problem (1.1), the definition of LICQ is given as follows.

DEFINITION 3.10. *We say that LICQ condition holds at a feasible point* $\mathbf{x}^*$ *of problem* (1.1)*, if the gradients of active constraints (all the equality and active bound inequality constraints) are linearly independent at* $\mathbf{x}^*$*.*

Recently, a constraint qualification condition called *Constant Positive Linear Dependence* (CPLD) condition is proposed in [36]. The CPLD condition has been shown weaker than LICQ and has been widely studied [1, 2, 3]. In this paper, with respect to feasible accumulation points, we analyze their optimality properties under CPLD condition. We first give the following definition.

DEFINITION 3.11. *We say that a set of constraints of problem* (1.1) *with indices* $\bar{\mathcal{A}} = \mathcal{I} \cup \mathcal{J}$ *are Positive Linearly Dependent at* $\mathbf{x}$*, where* $\mathcal{I} \subseteq \{i = 1, \ldots, m\}$ *and* $\mathcal{J} \subseteq \{j = 1, \ldots, n\}$*, if there exist* $\boldsymbol{\lambda} \in \mathbb{R}^m$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n_+$ *with* $\sum_{i\in\mathcal{I}}|\lambda_i| + \sum_{j\in\mathcal{J}}\mu_j \neq 0$ *such that*

$$\sum_{i\in\mathcal{I}}\lambda_i\nabla\mathbf{c}_i(\mathbf{x}) + \sum_{j\in\mathcal{J}}\mu_j\mathbf{e}_j = \mathbf{0}.$$

Now, we give the definition of CPLD.

DEFINITION 3.12. *Given a feasible point* $\mathbf{x}^*$ *of* (1.1)*, we say that the CPLD condition holds at* $\mathbf{x}^*$ *if the positive linear dependence of any subset of the active constraints at* $\mathbf{x}^*$ *implies the linear dependence of their gradients at some neighborhood of* $\mathbf{x}^*$*.*

By Lemma 3.9, we can see there exist an accumulation point $\mathbf{x}^*$ and a subsequence $\{\mathbf{x}_{k_i}\}$ such that

$$\lim_{i\to\infty}\mathbf{x}_{k_i} = \mathbf{x}^* \quad \text{and} \quad \lim_{i\to\infty}\mathbf{D}_{k_i}\bar{\mathbf{g}}_{k_i} = \mathbf{0}. \tag{3.41}$$

The following theorem shows under certain conditions, this $\mathbf{x}^*$ will be a KKT point of (1.1).

THEOREM 3.13. *Under assumptions* ***AS.1-AS.2****, assume that* $\lim_{k\to\infty}\sigma_k = \infty$, $\lim_{k\to\infty}\|\mathbf{c}_k\| = 0$ *and* $\mathbf{x}^*$ *is an accumulation point at which the CPLD condition holds. Then* $\mathbf{x}^*$ *is a KKT point of* (1.1)*.*

Proof. By Lemma 3.9, (3.41). Therefore, there exist an accumulation point $\mathbf{x}^*$ and a subsequence $\{\mathbf{x}_{k_i}\}$, still denoted as $\{\mathbf{x}_k\}$ for notation simplicity, such that

$$\lim_{k\to\infty}\mathbf{x}_k = \mathbf{x}^*, \quad \lim_{k\to\infty}\mathbf{D}_k\bar{\mathbf{g}}_k = \lim_{k\to\infty}\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k) = \mathbf{0}. \tag{3.42}$$

So, by the definition of $\mathbf{D}_k$, there exists $\boldsymbol{\mu}_k \in \mathbb{R}^n_+$ such that

$$\lim_{k\to\infty}\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}(\boldsymbol{\lambda}_k - \sigma_k\mathbf{c}_k) - \sum_{j\in\mathcal{A}_*}\mu_{kj}\mathbf{e}_j = \mathbf{0}, \tag{3.43}$$

where $\mathcal{A}_* = \{i : x_i^* = 0\}$. Then, by (3.43) and the Caratheodory's Theorem for Cones, for any k there exist $\hat{\boldsymbol{\lambda}}_k \in \mathbb{R}^m$ and $\hat{\boldsymbol{\mu}}_k \in \mathbb{R}^n_+$ such that

$$\lim_{k\to\infty}\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\hat{\boldsymbol{\lambda}}_k - \sum_{j\in\mathcal{A}_*}\hat{\mu}_{kj}\mathbf{e}_j = \mathbf{0}, \tag{3.44}$$

and

$$\{\nabla c_i(\mathbf{x}_k) : i \in \text{supp}(\hat{\boldsymbol{\lambda}}_k)\} \cup \{\mathbf{e}_j : j \in \text{supp}(\hat{\boldsymbol{\mu}}_k) \subseteq \mathcal{A}_*\} \text{ are linearly independent.} \tag{3.45}$$

Let $\mathcal{I}_k = \text{supp}(\hat{\boldsymbol{\lambda}}_k)$ and $\mathcal{J}_k = \text{supp}(\hat{\boldsymbol{\mu}}_k)$. Since $\mathbb{R}^n$ is a finite dimensional space, without loss of generality, by taking subsequence if necessary we can assume that $\mathcal{I}_k = \mathcal{I}$ and $\mathcal{J}_k = \mathcal{J} \subseteq \mathcal{A}_*$ for all large k.

Let $\mathbf{y}_k = (\hat{\boldsymbol{\lambda}}_k, \hat{\boldsymbol{\mu}}_k)$. We now show that $\{\mathbf{y}_k\} = \{(\hat{\boldsymbol{\lambda}}_k, \hat{\boldsymbol{\mu}}_k)\}$ is bounded. Suppose $\|\mathbf{y}_k\| \to \infty$ as $k \to \infty$. From (3.44), we have

$$\lim_{k\to\infty} \sum_{i\in\mathcal{I}} \nabla \mathbf{c}_i(\mathbf{x}_k)\hat{\lambda}_{ki}/\|\mathbf{y}_k\| + \sum_{j\in\mathcal{J}} \mathbf{e}_j \hat{\mu}_{kj}/\|\mathbf{y}_k\| = \lim_{k\to\infty} \mathbf{g}_k/\|\mathbf{y}_k\| = \mathbf{0}.$$

Since $\mathbf{z}_k := (\hat{\boldsymbol{\lambda}}_k, \hat{\boldsymbol{\mu}}_k)/\|\mathbf{y}_k\|$ has unit norm, it has a subsequence converging to a limit $\mathbf{z}^* := (\mathbf{w}^*, \mathbf{v}^*)$. Without loss of generality, we assume that $\hat{\boldsymbol{\lambda}}_k/\|\mathbf{y}_k\| \to \mathbf{w}^*$ and $\hat{\boldsymbol{\mu}}_k/\|\mathbf{y}_k\| \to \mathbf{v}^*$. Hence we have

$$\sum_{i\in\mathcal{I}} w_i^* \nabla \mathbf{c}_i(\mathbf{x}^*) + \sum_{j\in\mathcal{J}} v_j^* \mathbf{e}_j = \mathbf{0}. \tag{3.46}$$

Since $\hat{\boldsymbol{\mu}}_k \geq 0$, we have $\mathbf{v}^* \geq 0$. So, (3.46) and $\|\mathbf{z}^*\| = 1$ imply that the constraints with indices $\bar{\mathcal{A}} = \mathcal{I} \cup \mathcal{J}$ are positive linearly dependent at $\mathbf{x}^*$. Then, since CPLD holds at $\mathbf{x}^*$ and $\mathbf{x}_k \to \mathbf{x}^*$, we have that $\{\nabla c_i(\mathbf{x}_k) : i \in \mathcal{I}\} \cup \{\mathbf{e}_j : j \in \mathcal{J}\}$ are linearly dependent for all large k. However, this contradicts (3.45). Hence $\{\mathbf{y}_k\} = \{(\hat{\boldsymbol{\lambda}}_k, \hat{\boldsymbol{\mu}}_k)\}$ is bounded. So, by (3.44), there exist $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^* \geq \mathbf{0}$ such that

$$\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^* = \sum_{j\in\mathcal{J}} \mu_j^* \mathbf{e}_j.$$

which by Definition 2.1 indicates that $\mathbf{x}^*$ is a KKT point of (1.1). □

4. Boundedness of penalty parameters. We have previously shown that the boundedness of penalty parameters plays a very important role in forcing $\mathbf{x}^*$ to be a KKT point of (1.1). In addition, large penalty parameters might lead to ill-conditioned Hessian of the augmented Lagrangian function (2.2), which normally brings numerical difficulties for solving trust region subproblems. Hence, in this section we study the behavior of penalty parameters and investigate the conditions under which the penalty parameters are bounded. Throughout this section, we assume that $\mathbf{x}^*$ is a KKT point of (1.1). For the analysis in this section, we need some additional assumptions.

AS.3 *The LICQ condition, defined in Definition 3.10*, holds at the KKT point $\mathbf{x}^*$.

The assumption **AS.3** implies that $[(\mathbf{A}^*)^\mathsf{T}]_{\mathcal{I}_*}$ has full column rank, i.e.,

$$\text{Rank}([(\mathbf{A}^*)^\mathsf{T}]_{\mathcal{I}_*}) = m, \tag{4.1}$$

where $\mathcal{I}_*$ is the index set of inactive bound constraints at $\mathbf{x}^*$, i.e., $\mathcal{I}_* = \{i : x_i^* > 0\}$.

AS.4 *The strict complementarity conditions hold at the KKT point $\mathbf{x}^*$, that is*

$$[\mathbf{g}(\mathbf{x}^*) - (\mathbf{A}(\mathbf{x}^*))^\mathsf{T}\boldsymbol{\lambda}^*]_i > 0, \quad \textit{if } i \in \mathcal{A}_* = \{i : x_i^* = 0\},$$

where $\boldsymbol{\lambda}^*$ is the Lagrange multiplier associated with the equality constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$ at $\mathbf{x}^*$.

We now give a property of $\tilde{\boldsymbol{\lambda}}_k$, defined in (2.13), in the following lemma.

LEMMA 4.1. *Assume that **AS.1-AS.4** hold at a KKT point $\mathbf{x}^*$ of* (1.1)*. Then, there exists a constant $\rho > 0$, such that if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$, $\tilde{\boldsymbol{\lambda}}_k$ given by* (2.13) *is unique and*

$$\tilde{\boldsymbol{\lambda}}_k = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}]_{\mathcal{I}_*}\|^2. \tag{4.2}$$

Proof. The strict complementarity conditions imply that

$$[\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*]_{\mathcal{I}_*} = \mathbf{0} \quad \text{and} \quad [\mathbf{g}^* - (\mathbf{A}^*)^\mathsf{T}\boldsymbol{\lambda}^*]_{\mathcal{A}_*} > \mathbf{0}. \tag{4.3}$$

Then, (4.1) indicates that $\boldsymbol{\lambda}^*$ is uniquely given by $\boldsymbol{\lambda}^* = [(\mathbf{A}(\mathbf{x}^*))^\mathsf{T}]^\dagger_{\mathcal{I}_*}[\mathbf{g}(\mathbf{x}^*)]_{\mathcal{I}_*}$.

We now denote $\boldsymbol{\mu}_k$ as

$$\boldsymbol{\mu}_k = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}]_{\mathcal{I}_*}\|^2. \tag{4.4}$$

AS.1 and **AS.3** indicate that there exists a $\rho > 0$ such that if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$, $\mathrm{Rank}([(\mathbf{A}_k)^\mathsf{T}]_{\mathcal{I}_*}) = m$, and thus $\boldsymbol{\mu}_k$ is a strict unique minimizer of (4.4), given by $\boldsymbol{\mu}_k = [\mathbf{A}_k^\mathsf{T}]^\dagger_{\mathcal{I}_*}[\mathbf{g}_k]_{\mathcal{I}_*}$, Moreover, by (4.3), reducing ρ if necessary, we have if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$,

$$\|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\| < \frac{x_i^*}{2} < x_{ki}, \quad \text{for any } i \in \mathcal{I}_*, \tag{4.5}$$

and there exists a constant $\bar{\xi} > 0$ independent of k such that

$$[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_i \geq x_{ki} + \bar{\xi}, \quad \text{for any } i \in \mathcal{A}_*. \tag{4.6}$$

Then (4.5) and (4.6) imply that if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$ the following equality holds:

$$\psi_k(\boldsymbol{\mu}_k) = \|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k)^+ - \mathbf{x}_k\|^2 = \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\|^2 + \|[\mathbf{x}_k]_{\mathcal{A}_*}\|^2. \tag{4.7}$$

In the following supposing $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$, we show $\tilde{\boldsymbol{\lambda}}_k = \boldsymbol{\mu}_k$ by way of contradiction. Assume that $\tilde{\boldsymbol{\lambda}}_k \neq \boldsymbol{\mu}_k$. Then, since $\boldsymbol{\mu}_k$ is the *unique* minimizer of (4.4), we have

$$\|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k]_{\mathcal{I}_*}\|^2 > \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\|^2.$$

This together with (4.5)-(4.6) implies that

$$\begin{aligned}
\|[(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k)^+ - \mathbf{x}_k]_{\mathcal{I}_*}\|^2 &\geq \sum_{i\in\mathcal{I}_*} \min\{[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k]_i^2, x_{ki}^2\} \\
&\geq \min\{\|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k]_{\mathcal{I}_*}\|^2, (\gamma^*)^2/4\} \\
&> \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\|^2 \\
&= \|[(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k)^+ - \mathbf{x}_k]_{\mathcal{I}_*}\|^2,
\end{aligned} \tag{4.8}$$

where $\gamma^* = \min_{i\in\mathcal{I}_*}\{x_i^*\}$. Therefore, it follows from $\psi_k(\tilde{\boldsymbol{\lambda}}_k) \leq \psi_k(\boldsymbol{\mu}_k)$ by the definition of $\tilde{\boldsymbol{\lambda}}_k$ and (4.7) that

$$\|[(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k)^+ - \mathbf{x}_k]_{\mathcal{A}_*}\|^2 < \|[(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k)^+ - \mathbf{x}_k]_{\mathcal{A}_*}\|^2 = \|[\mathbf{x}_k]_{\mathcal{A}_*}\|^2.$$

Consequently, there exists an index $j \in \mathcal{A}_*$, depending on k, such that

$$\begin{aligned}
x_{kj} &> \left|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k]_j\right| \\
&= \left|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_j + [\mathbf{A}_k^\mathsf{T}(\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k)]_j\right| \\
&\geq \left|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_j\right| - \|\mathbf{A}_k^\mathsf{T}\|\|\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k\|
\end{aligned}$$

which yields

$$\|\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k\| \;>\; \frac{\left|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_j\right| - x_{kj}}{\|\mathbf{A}_k^\mathsf{T}\|} \;\geq\; \frac{\bar{\xi}}{M}, \tag{4.9}$$

where $M = \max\{\|\mathbf{A}_k\|\} < \infty$ and the second inequality follows from (4.6). Recall that $\mathrm{Rank}([(\mathbf{A}_k)^\mathsf{T}]_{\mathcal{I}_*}) = m$ if $\mathbf{x}_k \in B_\rho(x^*)$. Hence, there exists $\hat{\xi} > 0$ such that $\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}(\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k)\| \geq \hat{\xi}\bar{\xi}/M$. Reducing ρ if necessary, since $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$, it again follows from (4.3) and $\boldsymbol{\mu}_k = [\mathbf{A}_k^\mathsf{T}]^\dagger_{\mathcal{I}_*}[\mathbf{g}_k]_{\mathcal{I}_*}$ as the unique minimizer of (4.4) that

$$\|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\| \leq \frac{\bar{\xi}\hat{\xi}}{2M},$$

which further gives that

$$\begin{aligned}\|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k]_{\mathcal{I}_*}\| &= \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*} + [\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}(\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k)\| \\ &\geq \|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}(\boldsymbol{\mu}_k - \tilde{\boldsymbol{\lambda}}_k)\| - \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\| \\ &> \frac{\bar{\xi}\hat{\xi}}{M} - \|[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\mu}_k]_{\mathcal{I}_*}\| \geq \frac{\bar{\xi}\hat{\xi}}{2M}. \end{aligned} \tag{4.10}$$

Hence, by (4.8) and (4.10)

$$\begin{aligned}\psi_k(\tilde{\boldsymbol{\lambda}}_k) &= \|(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k)^+ - \mathbf{x}_k\|^2 \\ &\geq \|[(\mathbf{x}_k - \mathbf{g}_k + \mathbf{A}_k^\mathsf{T}\tilde{\boldsymbol{\lambda}}_k)^+ - \mathbf{x}_k]_{\mathcal{I}_*}\|^2 \\ &\geq \min\left\{\left(\frac{\bar{\xi}\tilde{\xi}}{2M}\right)^2, \frac{(\gamma^*)^2}{4}\right\} =: \bar{\theta} > 0. \end{aligned} \tag{4.11}$$

However, by (4.3) and (4.7), we can choose ρ sufficiently small such that if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$, then $\psi_k(\boldsymbol{\mu}_k) \leq \bar{\theta}/2$, and therefore by (4.11), $\psi_k(\boldsymbol{\mu}_k) < \psi_k(\tilde{\boldsymbol{\lambda}}_k)$. This contradicts the definition of $\tilde{\boldsymbol{\lambda}}_k$ such that $\psi_k(\tilde{\boldsymbol{\lambda}}_k) \leq \psi_k(\boldsymbol{\mu}_k)$. Hence, if $\mathbf{x}_k \in B_\rho(\mathbf{x}^*)$ with ρ sufficiently small, then $\tilde{\boldsymbol{\lambda}}_k = \boldsymbol{\mu}_k$, which is uniquely given by (4.2). □

Lemma 4.1 shows that if the iterates $\{\mathbf{x}_k\}$ generated by ALAS converge to a KKT point satisfying **AS.1**-**AS.4**, then $\{\tilde{\boldsymbol{\lambda}}_k\}$ will converge to the unique Lagrange multiplier $\boldsymbol{\lambda}^*$ associated with the equality constraints at $\mathbf{x}^*$. Hence, if we choose components of $\boldsymbol{\lambda}_{\min}$ and $\boldsymbol{\lambda}_{\max}$ sufficient small and large respectively such that $\boldsymbol{\lambda}_{\min} < \boldsymbol{\lambda}^* < \boldsymbol{\lambda}_{\max}$, then we have $\boldsymbol{\lambda}_{\min} < \tilde{\boldsymbol{\lambda}}_k < \boldsymbol{\lambda}_{\max}$ for all large k. Then, according to the computation of $\boldsymbol{\lambda}_k$ in (2.19), $\boldsymbol{\lambda}_k = \tilde{\boldsymbol{\lambda}}_k$ for all large k. To show the boundedness of penalty parameters $\{\sigma_k\}$, we also need the following assumption.

AS.5 *Assume that for all large k, $\boldsymbol{\lambda}_k$ is updated by* (2.19) *with* $\boldsymbol{\lambda}_{\min} \leq \tilde{\boldsymbol{\lambda}}_k \leq \boldsymbol{\lambda}_{\max}$.

This assumption assumes that $\boldsymbol{\lambda}_k$ is updated for all later iterations and $\boldsymbol{\lambda}_k = \tilde{\boldsymbol{\lambda}}_k$. Recall that to decide whether to update $\boldsymbol{\lambda}_k$ or not, we set a test condition $\|\mathbf{c}_{k+1}\| \leq R_k$ in ALAS. **AS.5** essentially assumes that this condition is satisfied for all large k. Actually, this assumption is more practical than theoretical, because when $\mathbf{x}_k$ is close to the feasible region and σ_k is relatively large, normally the constraint violation will be decreasing. Then if the parameter β in ALAS is very close to 1, the condition $\|\mathbf{c}_{k+1}\| \leq R_k$ will usually be satisfied in later iterations. We now conclude this section with the following theorem.

THEOREM 4.2. *Suppose the iterates $\{\mathbf{x}_k\}$ generated by ALAS converge to a KKT point $\mathbf{x}^*$ satisfying assumptions **AS.1**-**AS.5**. Then*

$$\mathrm{Pred}_k \;\geq\; \frac{\delta}{\sigma_k}\min\{\Delta_k\|\mathbf{c}_k\|, \|\mathbf{c}_k\|^2\}, \quad \textit{for all large } k. \tag{4.12}$$

Furthermore, if ALAS does not terminate in finite number of iterations, the penalty parameters $\{\sigma_k\}$ are bounded.

Proof. We first prove (4.12). We consider the case $\|\mathbf{c}_k\| \neq 0$, otherwise (4.12) obviously holds. Adding constraints $\mathbf{d}_{\mathcal{A}_*} = \mathbf{0}$ to the subproblem (2.9), we obtain the following problem:

$$\begin{aligned} &\min \quad \tilde{q}_k(\mathbf{d}) \\ &\text{s. t.} \quad \|\mathbf{d}\| \leq \Delta_k, \quad \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \geq \mathbf{0}, \quad \mathbf{d}_{\mathcal{A}_*} = \mathbf{0}. \end{aligned} \tag{4.13}$$

Since $\mathbf{x}_k \to \mathbf{x}^*$, $\mathcal{I}_* \subseteq \mathcal{I}_k$ for large k. Hence, when k is sufficiently large, $[\mathbf{D}_k]_{ii} > 0$ for all $i \in \mathcal{I}_*$. Denote $\bar{\mathbf{d}} = \mathbf{D}_k\mathbf{d}$. Then, for k sufficiently large, (4.13) is equivalent to

$$\begin{aligned} \min \quad & \bar{q}_k(\bar{\mathbf{d}}) = ([\mathbf{g}_k]_{\mathcal{I}_*} - [\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\boldsymbol{\lambda}_k)^\mathsf{T}\bar{\mathbf{d}}_{\mathcal{I}_*} + \frac{1}{2}\bar{\mathbf{d}}_{\mathcal{I}_*}^\mathsf{T}[\mathbf{B}_k]_{\mathcal{I}_*,\mathcal{I}_*}\bar{\mathbf{d}}_{\mathcal{I}_*} + \frac{\sigma_k}{2}\|\mathbf{c}_k + [\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T}\bar{\mathbf{d}}_{\mathcal{I}_*}\|^2 \\ \text{s. t.} \quad & \|[\mathbf{D}_k]_{\mathcal{I}_*,\mathcal{I}_*}^{-1}\bar{\mathbf{d}}_{\mathcal{I}_*}\| \le \Delta_k, \quad [\mathbf{x}_k]_{\mathcal{I}_*} + \bar{\mathbf{d}}_{\mathcal{I}_*} \ge \mathbf{0}, \quad \bar{\mathbf{d}}_{\mathcal{A}_*} = \mathbf{0}. \end{aligned} \tag{4.14}$$

By **AS.5** and Lemma 4.1, we obtain that $[\mathbf{g}_k]_{\mathcal{I}_*} - [\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\boldsymbol{\lambda}_k \in \mathrm{Null}([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T})$ when k is sufficiently large. Here, $\mathrm{Null}([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T})$ denotes the null space of $[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T}$. Hence, if we restrict $\bar{\mathbf{d}}_{\mathcal{I}_*} \in \mathrm{Range}([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*})$, (4.14) turns into

$$\begin{aligned} \min \quad & \hat{q}_k(\bar{\mathbf{d}}) = \frac{1}{2}\bar{\mathbf{d}}_{\mathcal{I}_*}^\mathsf{T}[\mathbf{B}_k]_{\mathcal{I}_*,\mathcal{I}_*}\bar{\mathbf{d}}_{\mathcal{I}_*} + \frac{\sigma_k}{2}\|\mathbf{c}_k + [\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T}\bar{\mathbf{d}}_{\mathcal{I}_*}\|^2 \\ & = \sigma_k\left[([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k)^\mathsf{T}\bar{\mathbf{d}}_{\mathcal{I}_*} + \frac{1}{2}\bar{\mathbf{d}}_{\mathcal{I}_*}^\mathsf{T}\left([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}^\mathsf{T} + \frac{[\mathbf{B}_k]_{\mathcal{I}_*,\mathcal{I}_*}}{\sigma_k}\right)\bar{\mathbf{d}}_{\mathcal{I}_*} + \frac{1}{2}\|\mathbf{c}_k\|^2\right] \\ \text{s. t.} \quad & \|[\mathbf{D}_k^{-1}]_{\mathcal{I}_*,\mathcal{I}_*}\bar{\mathbf{d}}_{\mathcal{I}_*}\| \le \Delta_k, \quad [\mathbf{x}_k]_{\mathcal{I}_*} + \bar{\mathbf{d}}_{\mathcal{I}_*} \ge \mathbf{0}, \\ & \bar{\mathbf{d}}_{\mathcal{I}_*} \in \mathrm{Range}([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}), \quad \bar{\mathbf{d}}_{\mathcal{A}_*} = \mathbf{0}. \end{aligned} \tag{4.15}$$

Here, $\mathrm{Range}([\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*})$ denotes the range space of $[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}$. Denote the solution of (4.15) as $\bar{\mathbf{s}}_k$. By **AS.3** and $\|\mathbf{c}_k\| \neq 0$, we have $[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k \neq \mathbf{0}$ for k sufficiently large. Now, note that

$$\bar{\mathbf{d}}_k(\tau) = -\tau\frac{[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k}{\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|}, \quad \text{for any } \tau \in \left[0, \min_{i\in\mathcal{I}_*}\left\{\frac{\Delta_k\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|}{\|[\mathbf{D}_k^{-1}]_{\mathcal{I}_*,\mathcal{I}_*}[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|}, x_{ki}\right\}\right], \tag{4.16}$$

is a feasible point of (4.15). Since for all sufficiently large k, $[\mathbf{D}_k]_{ii}$ and x_{ki} are bounded from below for each $i \in \mathcal{I}_*$, by (4.16) there exists a constant $\gamma > 0$ such that

$$\bar{\mathbf{d}}_k(\tau) = -\tau\frac{[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k}{\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|}, \quad \text{for any } \tau \in [0, \gamma\Delta_k]$$

is feasible. Then direct calculations show that there exists a constant $M > 0$ such that

$$\tilde{q}_k(\mathbf{0}) - \tilde{q}_k(\bar{\mathbf{s}}_k) = \hat{q}_k(\mathbf{0}) - \hat{q}_k(\bar{\mathbf{s}}_k) \ge \hat{q}_k(\mathbf{0}) - \min_{\tau\in[0,\gamma\Delta_k]}\hat{q}_k(\bar{\mathbf{s}}_k) \ge \sigma_k\frac{\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|}{M}\min\{\Delta_k, \|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\|\},$$

where the existence of M is due to the boundedness of $\mathbf{A}_k$, $\mathbf{c}_k$, $\mathbf{B}_k$ and σ_k^{-1}. Since **AS.3** indicates that there exists a constant $\hat{\xi} > 0$ such that $\|[\mathbf{A}_k^\mathsf{T}]_{\mathcal{I}_*}\mathbf{c}_k\| \ \ge\ \hat{\xi}\|\mathbf{c}_k\|$ for all large k, the above inequality gives

$$\mathrm{Pred}_k \ge \bar{\beta}[\tilde{q}_k(\mathbf{0}) - \tilde{q}_k(\bar{\mathbf{s}}_k)] \ge \sigma_k\frac{\bar{\beta}\hat{\xi}}{M}\ \min\{\Delta_k\|\mathbf{c}_k\|, \hat{\xi}\|\mathbf{c}_k\|^2\}.$$

Therefore, (4.12) holds as long as $\sigma_k \ge \sqrt{(M\delta)/(\bar{\beta}\min\{\xi, \xi^2\})}$. Since whenever (4.12) does not hold, σ will be increased by a factor of $\theta_2 > 1$. So (4.12) holds for all large k.

Hence, for the second part of the theorem, to show the boundedness of $\{\sigma_k\}$, we only need to prove that (2.17) does not happen at infinite number of iterations. First, by the definition of $\mathbf{D}_k$, if $\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k) = \mathbf{0}$, we have

$$[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k]_{\mathcal{I}_k} = \mathbf{0} \quad \text{and} \quad [\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k]_{\mathcal{A}_k} \ge \mathbf{0}, \tag{4.17}$$

where $\mathcal{I}_k = \{i : x_{ki} > 0\}$ and $\mathcal{A}_k = \{i : x_{ki} = 0\}$. If (2.17) happens at infinite number of iterations, then $\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k) = \mathbf{0}$ at infinite number of iterations, which by (4.17) and $\mathcal{I}_* \subseteq \mathcal{I}_k$ for large k indicates that

$$[\mathbf{g}_k - \mathbf{A}_k^\mathsf{T}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^\mathsf{T}\mathbf{c}_k]_{\mathcal{I}_*} = \mathbf{0} \tag{4.18}$$

at infinite number of iterations. Under our assumptions, by Lemma 4.1, (4.2) holds and $\boldsymbol{\lambda}_k = \tilde{\boldsymbol{\lambda}}_{\bar{k}}$ for all sufficiently large k. Hence, by (4.18) there exists a $\bar{k}$ large enough such that

$$[\mathbf{g}_{\bar{k}} - \mathbf{A}_{\bar{k}}^{\mathsf{T}}(\boldsymbol{\lambda}_{\bar{k}} + \sigma_{\bar{k}}\mathbf{c}_{\bar{k}})]_{\mathcal{I}_*} = \mathbf{0} \quad \text{and} \quad \boldsymbol{\lambda}_{\bar{k}} = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^m} \|[\mathbf{g}_{\bar{k}} - \mathbf{A}_{\bar{k}}^{\mathsf{T}}\boldsymbol{\lambda}]_{\mathcal{I}_*}\|^2. \tag{4.19}$$

By **AS.3**, $\mathrm{Rank}([\mathbf{A}_{\bar{k}}^{\mathsf{T}}]_{\mathcal{I}_*}) = m$ when $\bar{k}$ is sufficiently large. This together with (4.19) implies $\boldsymbol{\lambda}_{\bar{k}} + \sigma_{\bar{k}}\mathbf{c}_{\bar{k}} = \boldsymbol{\lambda}_{\bar{k}}$, thus $\mathbf{c}_{\bar{k}} = \mathbf{0}$. This and (4.17) show that $\mathbf{x}_{\bar{k}}$ is a KKT point of (1.1). Then ALAS terminates finitely and returns $\mathbf{x}_{\bar{k}}$ as the solution. Therefore, if ALAS does not terminate finitely, (2.17) only happens at finite number of iterations, which shows that penalty parameters $\{\sigma_k\}$ are bounded. □

5. Numerical experiments. In this section, we present some preliminary numerical results on a set of general constrained nonlinear optimization problems in the CUTEr [5] library. Since the development of an efficient and reliable software for general nonlinear programming would be a fairly large, complicated project and would depend on many factors, our first step now is to implement ALAS in MATLAB in a very conservative way to show a general pattern of the performance of ALAS on a set of relatively smaller size problems. In the following Table 5.1, we list the tested 78 problems with inequality constraints. Both the number of variables and the number of constraints of these testing problems are less than or very close to 100. The numerical implementation in this section is executed in Matlab 7.6.0 (R2008a) on a PC with a 1.86 GHz Pentium Dual-Core microprocessor and 1GB of memory running Fedora 8.0.

In previous sections, our presentation of ALAS is based on nonlinear programming problems with equality and nonnegative bound constraints. However, as mentioned at the beginning of the paper, all the theoretical results developed previously can be easily applied to the following problem with general bound constraints:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & \mathbf{c}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{l} \le \mathbf{x} \le \mathbf{u}. \end{aligned} \tag{5.1}$$

For (5.1), the corresponding affine scaling trust region subproblem would be

$$\begin{aligned} \min_{\mathbf{d}\in\mathbb{R}^n} \quad & Q_k(\mathbf{x}_k + \mathbf{d}) = [\mathbf{D}_k(\mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}_k)]^{\mathsf{T}}\mathbf{d} + \frac{1}{2}\mathbf{d}^{\mathsf{T}}\mathbf{D}_k\mathbf{B}_k\mathbf{D}_k\mathbf{d} + \frac{\sigma_k}{2}\|\mathbf{c}_k + \mathbf{A}_k\mathbf{D}_k\mathbf{d}\|^2 \\ \text{s. t.} \quad & \mathbf{l} \le \mathbf{x}_k + \mathbf{D}_k\mathbf{d} \le \mathbf{u}, \quad \|\mathbf{d}\| \le \Delta_k, \end{aligned} \tag{5.2}$$

where the scaling matrix $\mathbf{D}_k$ is given by

$$[\mathbf{D}_k]_{ii} = \begin{cases} \sqrt{x_{ki} - l_i}, & \text{if } \bar{g}_{ki} > 0 \text{ and } l_i > -\infty; \\ \sqrt{u_i - x_{ki}}, & \text{if } \bar{g}_{ki} < 0 \text{ and } u_i < \infty; \\ 1, & \text{otherwise}, \end{cases}$$

and $\bar{\mathbf{g}}_k = \mathbf{g}_k - \mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}_k + \sigma_k\mathbf{A}_k^{\mathsf{T}}\mathbf{c}_k$. The general constrained nonlinear optimization problems in CUTEr library are given in the following standard format:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & \mathbf{cl} \le \mathbf{c}(\mathbf{x}) \le \mathbf{cu}, \quad \mathbf{bl} \le \mathbf{x} \le \mathbf{bu}. \end{aligned}$$

Slack variables are introduced in our implementation to formulate such problems into the format as (5.1). Since CUTEr can provide the exact Hessian of the objective function and we are still in a very early stage of code development on smaller size testing problems, we simply set $\mathbf{B}_k = \nabla_{xx}^2\mathcal{L}(\mathbf{x}_k; \boldsymbol{\lambda}_k, \sigma_k)$ in (5.2) in our current numerical experiments. Our future implementation would also include various options for defining $\mathbf{B}_k$ such as the BFGS, limited BFGS or Barzilai-Borwein quasi-Newton approximations of $\nabla_{xx}^2\mathcal{L}(\mathbf{x}_k; \boldsymbol{\lambda}_k, \sigma_k)$.

Another important issue related to the efficient implementation of ALAS is how to solve the subproblem (5.2) effectively. Note that (5.2) has a special structure that its feasible region is the intersection of bound

and trust region constraints. Therefore, it is possible to take advantage of this problem structure. In our current numerical implementation, we apply the idea from Powell [35] to solve (5.2). To decrease the objective function value of (5.2), starting with an interior feasible point, our algorithm executes the conjugate gradient iteration until the iterate reaches the boundary of the box constrained region or the boundary of the ball trust region. If the boundary of the box is met, the algorithm will decide whether or not to fix the active components of the variable to the box constraints and then restart the iteration with free components of the variable. On the other hand, if the iterate reaches the boundary of the ball trust region, the algorithm will truncate it on the boundary and keep searching a new point with lower objective function value along the ball boundary. These strategies would greedily reduce the objective function value as much as possible without increasing much computational cost. The practical termination criterion for solving (5.2) is set as

$$\|\mathrm{P}(\nabla Q_k(\mathbf{x}_k+\mathbf{d}))\|\Delta_k \le 0.01[Q_k(\mathbf{x}_k)-Q_k(\mathbf{x}_k+\mathbf{d})]$$

or

$$\|\mathrm{P}(\mathbf{d})\|^2\|\mathrm{P}(\nabla Q_k(\mathbf{x}_k+\mathbf{d}))\|^2-[\mathrm{P}(\mathbf{d})^{\mathsf{T}}\mathrm{P}(\nabla Q_k(\mathbf{x}_k+\mathbf{d}))]^2 \le 10^{-4}[Q_k(\mathbf{x}_k)-Q_k(\mathbf{x}_k+\mathbf{d})]^2,$$

where P is the projection on the normal cone of active bound constraints. Interested readers could referred to [35] for more details.

For comparison purpose, we also apply the code LANCELOT [14] to solve our testing problems. In each iteration, after introducing slack variables, LANCELOT minimizes the augmented Lagrangian function associated with the equality constraints and keeps explicit bound constraints. Therefore, the feasibility of bound constraints is always guaranteed in LANCELOT as well as in ALAS. In the implementation, to make the comparison between ALAS and LANCELOT as fair as possible, in both algorithms we use the following main termination criteria:

$$\|\mathbf{c}(\mathbf{x}_k)\|_\infty < 10^{-5} \quad \text{and} \quad \|\mathbf{x}_k - P_{[l,u]}(\mathbf{x}_k-(\mathbf{g}_k-\mathbf{A}_k^{\mathsf{T}}\boldsymbol{\lambda}_k))\|_\infty < 10^{-5}. \tag{5.3}$$

In ALAS, we also terminate the algorithm as long as $\|\mathbf{s}_k\|_\infty < 10^{-12}$, which indicates that no more progress could be obtained.

For both algorithms, we set the maximum number of objective function evaluations as 1000. In ALAS, we use the following initialization settings:

$$\Delta_0=1;\quad \theta_1=2;\quad \theta_2=10;\quad \eta=10^{-8},\quad \eta_1=0.1;\quad R=10^3,\quad \sigma_0=10;\quad \delta=10,\quad \beta=0.6.$$

In LANCELOT, we set $\Delta_0=1$, and other parameters are used as the default settings.

We present the numerical results of LANCELOT and ALAS in Table 5.1. For each problem, "n" and "m" denote the number of variables and the number of nonlinear constraints, including both equality and inequality constraints. Since the objective and constraint function values are calculated simultaneously by calling the subroutine "cfn" in CUTEr, "n_f" denotes the number of times of calling the subroutine "cfn". Similarly, "n_g" denotes the number of times of calling the subroutine "cgr" to compute both the gradients of objective function and the constraints simultaneously. Moreover, in Table 5.1, we record the objective function value and constraint violation $(f, \|c\|_\infty)$ at returned solutions.

From the Table 5, we can see that for both algorithms, there are two test problems which could not be solved within 1000 objective function evaluations. For LANCELOT, they are POLAK6 and ROSENMMX, while for ALAS they are SPIRAL and UBH5. It should be noted that for UBH5, ALAS returns an iterate with lower f and lower $\|c\|_\infty$, although the maximum number of objective function evaluations is reached. For majority of the test problems, both ALAS and LANCELOT return solutions with $\|c\|_\infty < 10^{-5}$. However, this feasibility criterion is not reached by ALAS for 6 problems: HALDMADS, PT, RES, SPIRAL, VANDERM1 and WOMFLET, and by LANCELOT for 7 problems: PFIT1, POLAK6, ROSENMMX, VANDERM1, VANDERM2, VANDERM3 and VANDERM4. These cases normally happen when the algorithms indicate that no further progress could be possible. For example, in ALAS the step length is too small. Besides, note that for most of the test problems, objective function values at returned solutions by both

TABLE 5.1
Number of function and gradient evaluations of ALAS and LANCELOT on CUTEr test problems

Problem	Prob. Dim.		LANCELOT			ALAS		
	n	m	n_f	n_g	$(f,\|c\|_\infty)$	n_f	n_g	$(f,\|c\|_\infty)$
ALSOTAME	1	2	6	7	(8.20850126D-02, 1.7083D-07)	7	7	(8.20849989e-02,4.0968e-09)
CANTILVR	1	5	22	23	(1.33995654D+00,4.1561D-07)	14	14	(1.33995523e+00,2.5169e-06)
CB2	3	3	14	15	(1.95222474D+00,3.8774D-07)	9	9	(1.95222449e+00,2.0041e-09)
CB3	3	3	16	16	(1.99999980D+00,1.5692D-06)	9	9	(1.99999968e+00,1.4071e-06)
CHACONN1	3	3	10	11	(1.95222448D+00, 7.6149D-08)	8	8	(1.95222450e+00,4.4503e-08)
CHACONN2	3	3	15	15	(2.00000171D+00,3.3514D-06)	9	9	(1.99999958e+00,1.7768e-06)
CHARDIS1	10	4	23	22	(6.49999880D+00,3.3648D-07)	26	25	(4.17721084e-11,1.6562e-06)
CONGIGMZ	10	4	47	46	(2.79999388D+01,7.6460D-06)	18	18	(2.79999925e+01,9.2878e-07)
CSFI1	5	4	175	151	(-4.90751986D+01,4.1058D-09)	29	27	(-4.90750459e+01,6.2039e-06)
CSFI2	5	4	96	87	(5.50176016D+01,2.0205D-07)	43	38	(5.50173905e+01,7.8503e-06)
DEMYMALO	3	3	23	22	(-3.00000621D+00,8.2886D-06)	16	14	(-3.00000000e+00,1.4563e-08)
DIPIGRI	7	4	313	313	(6.80630057D+02,8.4694D-09)	845	739	(6.80630057e+02,5.2012e-12)
DNIEPER	61	24	56	53	(1.87438976D+04,5.7774D-06)	108	57	(1.87440142e+04,2.1151e-06)
FLETCHER	4	4	20	20	(1.95253660D+01,1.3890D-08)	11	11	(1.95255284e+01,6.1029e-06)
GIGOMEZ1	3	3	26	25	(-2.99999996D+00,5.1353D-07)	16	14	(-3.00000000e+00,2.3249e-07)
GIGOMEZ2	3	3	13	14	(1.95222464D+00, 1.6299D-07)	9	9	(1.95222463e+00,3.2798e-07)
GIGOMEZ3	3	3	16	15	(2.00000236D+00,9.4271D-06)	9	9	(1.99999952e+00,2.0438e-06)
GOFFIN	51	50	110	111	(-1.34009432D-08,1.4452D-07)	14	14	(-6.65429083e-18,2.3479e-17)
HAIFAS	13	9	24	23	(-4.49996545D-01,3.7767D-06)	19	16	(-4.50003091e-01,3.3924e-06)
HALDMADS	6	42	53	51	(1.20668307D-04, 4.3992D-06)	213	64	(9.12965956e-04,9.8944e-03)
HS100	7	4	248	248	(6.80630057D+02,8.4542D-09)	905	750	(6.80630057e+02,4.9866e-10)
HS10	2	1	19	20	(-1.00000075D+00,1.5078D-06)	13	13	(-1.00000080e+00,1.6092e-06)
HS104	8	5	56	53	(3.95116530D+00,4.6342D-07)	96	79	(4.28675830e+00,3.5062e-08)
HS107	9	6	31	32	(5.05496073D+03,5.0121D-06)	96	79	(5.05501179e+03,9.3476e-10)
HS111	10	3	48	46	(-4.77610917D+01,8.5764D-08)	16	13	(-4.77611057e+01,5.8688e-06)
HS113	10	8	288	289	(2.43062090D+01,2.5848D-08)	64	55	(2.43062091e+01,2.9534e-08)
HS11	2	1	15	16	(-8.49846907D+00,1.5896D-06)	9	9	(-8.49846932e+00,1.6726e-06)
HS12	2	1	101	101	(-2.99999999D+01,1.4606D-08)	11	10	(-3.00000000e+01,2.4076e-08)
HS13	2	1	57	57	(9.58710521D-01,9.0801D-06)	74	73	(9.58684154e-01,9.0977e-06)
HS14	2	3	12	13	(1.39346318D+00,6.1204D-07)	10	10	(1.39344079e+00,8.2745e-06)
HS22	2	2	9	10	(9.99997213D-01,4.6053D-06)	8	8	(9.99999221e-01,1.2879e-06)
HS268	5	5	48	49	(3.63797880D-12,1.6962D-07)	129	80	(9.93532012e-09,2.0137e-07)
HS29	3	1	66	65	(-2.26274169D+01,8.6427D-08)	12	11	(-2.26274170e+01,9.1338e-10)
HS43	4	3	157	156	(-4.40000115D+01,8.5801D-06)	16	16	(-4.39999985e+01,1.2949e-06)
HS88	2	1	54	53	(1.36009987D+00,2.4365D-06)	64	62	(1.36247761e+00,1.6921e-07)
HS89	3	1	118	106	(1.36264613D+00, 1.0081D-08)	64	59	(1.36247734e+00,1.6946e-07)
HS90	4	1	58	56	(1.36010106D+00, 2.4354D-06)	64	62	(1.36247763e+00,1.6919e-07)
HS91	5	1	69	66	(1.36160172D+00,9.9962D-07)	180	133	(1.36237011e+00,2.7083e-07)
HS92	6	1	55	54	(1.36009987D+00,2.4365D-06)	64	62	(1.36247766e+00,1.6916e-07)
KISSING	43	105	136	118	(5.63950250D-01, 1.3307D-06)	14	13	(9.99997474e-01,4.1310e-06)
KIWCRESC	3	2	17	18	(1.04802163D-06, 2.0960D-06)	9	9	(-5.54405695e-07,1.1088e-06)
LUKVLI10	100	98	140	139	(3.49243841D+01,9.7776D-06)	26	22	(3.49244033e+01,2.4767e-07)
MADSEN	3	6	33	32	(6.16432452D-01,1.4456D-07)	14	12	(6.16431908e-01,1.1961e-06)
MADSSCHJ	5	6	59	50	(-2.61211036D+00,4.6572D-07)	24	21	(-2.61211170e+00,2.7363e-06)
MAKELA1	3	2	17	17	(-1.41421292D+00, 1.5359D-06)	10	9	(-1.41421376e+00,7.4268e-06)
MAKELA2	3	3	30	30	(7.20000065D+00, 7.3504D-07)	14	12	(7.19999980e+00,1.4658e-06)
MAKELA3	21	20	93	82	(3.72655031D-06,3.7266D-06)	91	87	(9.99094844e-06,9.9909e-06)
MAKELA4	21	40	28	29	(1.23158590D-18,1.5573D-18)	43	43	(4.88134746e-07,1.5632e-06)
MIFFLIN1	3	2	10	11	(-9.99999966D-01,1.3755D-06)	7	7	(-1.00000006e+00,3.2109e-06)
MIFFLIN2	3	2	44	40	(-9.99999950D-01,1.1461D-07)	16	13	(-9.99994086e-01,8.2751e-06)
MINMAXRB	3	4	71	69	(4.22500034D-19,3.2961D-06)	37	31	(2.69749500e-16,1.6727e-06)
MISTAKE	9	13	37	35	(-9.99999839D-01,5.1108D-06)	15	14	(-9.99999282e-01,1.4536e-06)

TABLE 5.1
Number of function and gradient evaluations of ALAS and LANCELOT on CUTEr test problems

Problem	Prob. Dim.		LANCELOT			ALAS		
	n	m	n_f	n_g	$(f,\|c\|_\infty)$	n_f	n_g	$(f,\|c\|_\infty)$
NGONE	12	19	40	37	(-6.20365844D-01,5.1612D-07)	104	69	(-6.20365661e-01,4.2345e-09)
OPTCDEG2	32	20	39	40	(3.40605292D+02, 2.0776D-07)	11	11	(8.34418294e-11,1.9435e-06)
PENTAGON	6	15	10	11	(1.50922801D-04,6.2111D-07)	25	24	(1.36568606e-04,7.5738e-06)
PFIT1	3	3	366	293	(8.16678140D-07,7.3320D-04)	255	241	(0.00000000e+00,9.5376e-06)
POLAK1	3	2	36	34	(2.71828538D+00,3.6116D-06)	33	33	(2.71828183e+00,5.8633e-09)
POLAK3	12	10	140	119	(5.93300226D+00, 2.3498D-06)	84	80	(5.93300330e+00,1.0269e-07)
POLAK5	3	2	11	12	(5.00000000D+01, 5.1490D-08)	13	13	(4.99999990e+01,1.4493e-06)
POLAK6	5	4	1000	981	(-3.45298208D+01,2.6952e-02)	733	643	(-4.40000000e+01,1.7619e-09)
PT	2	101	28	29	(1.78384401D-01,1.2525D-15)	246	246	(1.78360995e-01,3.6832e-05)
POWELL20	100	100	126	127	(5.27030499D+04,4.0240D-06)	49	49	(5.27030115e+04,6.3353e-06)
RES	22	14	8	9	(1.17055013D-29,3.0254D-15)	21	21	(0.00000000e+00,6.7626e-02)
ROBOT	14	2	29	26	(5.46284071D+00,2.1830D-07)	79	30	(5.73171891e+00,9.2683e-10)
ROSENMMX	5	4	1000	998	(-3.62341720D+01,6.1317e-02)	239	210	(-4.40000000e+01,1.3580e-08)
SPIRAL	3	2	107	92	(8.57261671D-07,8.5726D-07)	1000	961	(1.86291229e-02,5.5085e-03)
SYNTHES1	6	6	12	13	(7.59280429D-01,2.3793D-06)	102	53	(7.59284399e-01,3.6697e-09)
TRY-B	2	1	13	14	(9.99998098D-01,1.9018D-06)	8	8	(1.00000000e+00,9.6549e-10)
UBH5	110	70	127	121	(1.14735203D+00,1.4627E-06)	1000	1000	(-4.53695500e-08,7.8130e-09)
VANDERM1	5	9	19	18	(1.45210124D-08,8.6578D-05)	78	26	(0.00000000e+00,1.8986e-05)
VANDERM2	5	9	19	18	(1.45210124D-08,8.6578D-05)	22	22	(0.00000000e+00,9.7613e-06)
VANDERM3	4	7	22	21	(8.32165815D-09,6.6002D-05)	15	15	(0.00000000e+00,6.2734e-06)
VANDERM4	3	5	46	45	(4.40017829D-09,3.8774D-05)	24	20	(0.00000000e+00,5.3224e-06)
WACHBIEG	3	2	18	17	(1.00000398D+00,7.9752D-06)	13	13	(9.99999184e-01,1.6310e-06)
WOMFLET	3	3	56	54	(1.14364087D-07, 1.5236D-07)	70	17	(-7.84664492e-05,1.5983e-03)
ZECEVIC3	2	2	17	18	(9.73094011D+01,6.1048D-06)	9	9	(9.73095197e+01,8.6804e-06)
ZIGZAG	64	50	40	38	(1.79999810D+00,7.6984D-07)	257	207	(1.96363638e+00,5.1479e-09)
ZY2	3	2	10	11	(2.00000006D+00,2.4020D-07)	9	9	(2.00000241e+00,9.6715e-06)

algorithms are quite similar when the termination condition (5.3) are satisfied, except for the problems CHARDIS1, HS104, KISSING, OPTCDEG2, UBH5 and ZIGZAG. One possible reason is that for these problems, the algorithms may step into a different region of local minimizers. In general, out of 78 test problems, we can see ALAS uses fewer function evaluations for 50 problems and uses fewer gradient evaluations for 49 problems. Hence, ALAS performs very promising compared with LANCELOT for solving this set of testing problems with relatively smaller problem size.

6. Conclusion. In this paper, we propose an Augmented Lagrangian Affine Scaling (ALAS) method for general constrained nonlinear optimization. In this method, to avoid possibly infeasible subproblem arising in classical SQP methods and keep the smoothness of the subproblem, we move the linearized equality constraints into the l_2 penalty term of a local quadratic approximation to the augmented Lagrangian function. Then, an active set type affine scaling trust region subproblem is proposed in which the affine scaling techniques are applied to handle the explicit strict bound constraints. New subproblems, combined with special strategies for updating penalty parameters and Lagrange multipliers, are designed to reduce both the objective function value and the equality constraint violation in an adaptive well-balanced way. The information obtained through solving the subproblem is used to update penalty parameters and Lagrangian multipliers. Under some mild conditions, we have proved the global convergence of ALAS and studied the conditions under which penalty parameters are bounded. This paper has done comprehensive theoretical studies of ALAS. For the numerical performance, we have tested ALAS on a set of smaller sized problems in the CUTEr library. Our preliminary numerical experiments show that ALAS has potentials to be developed into an effective software for general nonlinear optimization.

Acknowledgements. We would like to thank the anonymous reviewer and the associate editor for their constructive comments and suggestions, which helped to improve the quality of the paper.

Appendix. The proof of Lemma 3.1.

Proof. The penalty parameters $\{\sigma_k\}$ are positive and monotonically nondecreasing. Therefore $\sigma_k^{-1}-\sigma_{k+1}^{-1}$ is nonnegative for all k. And for any two integers p and q with $0\le p<q$, we have

$$\textstyle\sum_{k=p}^{q-1}\,(\sigma_k^{-1}-\sigma_{k+1}^{-1})\;=\;\sigma_p^{-1}-\sigma_q^{-1}. \tag{A.1}$$

Thus, denoting $f_{\max}$ as an upper bound of $\{|f(\mathbf{x}_k)|\}$, we deduce the bound

$$\begin{aligned}&\textstyle\sum_{k=p}^{q}\,\sigma_k^{-1}\,[\,f(\mathbf{x}_k)-f(\mathbf{x}_{k+1})\,]\\ &=\sigma_p^{-1}\,f(\mathbf{x}_p)+\textstyle\sum_{k=p}^{q-1}\,(-\sigma_k^{-1}+\sigma_{k+1}^{-1})\,f(\mathbf{x}_{k+1})-\sigma_q^{-1}f(\mathbf{x}_{q+1})\\ &\le\sigma_p^{-1}\,f_{\max}+(\sigma_p^{-1}-\sigma_q^{-1})\,f_{\max}+\sigma_q^{-1}\,f_{\max}\;=\;2\,\sigma_p^{-1}\,f_{\max}.\end{aligned} \tag{A.2}$$

We also have the following identity

$$\begin{aligned}&\textstyle\sum_{k=p}^{q}\,\sigma_k^{-1}\,[\,\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})-\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_k)\,]=-\sigma_p^{-1}\,\boldsymbol{\lambda}_p^{\mathsf{T}}\mathbf{c}(\mathbf{x}_p)+\sigma_q^{-1}\,\boldsymbol{\lambda}_q^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{q+1})\\ &+\textstyle\sum_{k=p}^{q-1}\,\{\,\sigma_k^{-1}\,\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})-\sigma_{k+1}^{-1}\,\boldsymbol{\lambda}_{k+1}^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})\,\}.\end{aligned} \tag{A.3}$$

Let $\mathcal{I}(p,q)=\{k\in[p,q-1]:\boldsymbol{\lambda}_{k+1}\neq\boldsymbol{\lambda}_k\}$. ($\mathcal{I}(p,q)$ is possibly empty.) Then, from (2.19) of ALAS, we have

$$\begin{aligned}&\textstyle\sum_{k\in\mathcal{I}(p,q)}\,\{\,\sigma_k^{-1}\,\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})-\sigma_{k+1}^{-1}\,\boldsymbol{\lambda}_{k+1}^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})\,\}\\ &\le\textstyle\sum_{k\in\mathcal{I}(p,q)}\,\{\,\sigma_k^{-1}\,\|\boldsymbol{\lambda}_k\|+\sigma_{k+1}^{-1}\,\|\boldsymbol{\lambda}_{k+1}\|\,\}\,\|c(\mathbf{x}_{k+1})\|\\ &\le 2\,\sigma_p^{-1}\,\pi_\lambda\,\textstyle\sum_{k\in\mathcal{I}(p,q)}\,\|\mathbf{c}(\mathbf{x}_{k+1})\|\;\le\;2\,\sigma_p^{-1}\,\pi_\lambda\,R_0\,/\,(1-\beta),\end{aligned}$$

where π_λ is an upper bound of $\{\|\boldsymbol{\lambda}_k\|\}$, for example, π_λ can be chosen as

$$\sqrt{n}\,\max\{\|\boldsymbol{\lambda}_{\max}\|_\infty,\|\boldsymbol{\lambda}_{\min}\|_\infty\}.$$

In all other terms in the second line of (A.3), $\boldsymbol{\lambda}_k$ and $\boldsymbol{\lambda}_{k+1}$ are same, which gives the inequality

$$\sigma_k^{-1}\,\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})-\sigma_{k+1}^{-1}\boldsymbol{\lambda}_{k+1}^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})\;\le\;(\sigma_k^{-1}-\sigma_{k+1}^{-1})\,\pi_\lambda\,c_{\max},$$

where $c_{\max}$ is an upper bound of $\{\|\mathbf{c}_k\|\}$. Therefore, by (A.1) and (A.3), we find the bound

$$\begin{aligned}&\textstyle\sum_{k=p}^{q}\,\sigma_k^{-1}\,[\,\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_{k+1})-\boldsymbol{\lambda}_k^{\mathsf{T}}\mathbf{c}(\mathbf{x}_k)\,]\le\sigma_p^{-1}\,\pi_\lambda\,c_{\max}+\sigma_q^{-1}\,\pi_\lambda\,c_{\max}\\ &+2\,\sigma_p^{-1}\,\pi_\lambda\,R_0\,/\,(1-\beta)+(\sigma_p^{-1}-\sigma_q^{-1})\,\pi_\lambda\,c_{\max}\\ &=2\,\sigma_p^{-1}\,\pi_\lambda\,(c_{\max}+R_0\,/\,(1-\beta)).\end{aligned} \tag{A.4}$$

Since $\mathbf{x}_{k+1}\neq\mathbf{x}_k$ only if $\mathrm{Ared}_k>0$, for all k we have

$$L(\mathbf{x}_{k+1};\boldsymbol{\lambda}_k,\sigma_k)\;\le\;L(\mathbf{x}_k;\boldsymbol{\lambda}_k,\sigma_k).$$

Then, it follows from inequalities (A.2) and (A.4) that

$$\textstyle\sum_{k=p}^{q}\,\sigma_k^{-1}\,\{\,L(\mathbf{x}_k;\boldsymbol{\lambda}_k,\sigma_k)-L(\mathbf{x}_{k+1};\boldsymbol{\lambda}_k,\sigma_k)\,\}\le\;\tfrac{1}{2}\,\|\mathbf{c}(\mathbf{x}_p)\|^2-\tfrac{1}{2}\,\|\mathbf{c}(\mathbf{x}_{q+1})\|^2+M_0\,\sigma_p^{-1}, \tag{A.5}$$

where M_0 is the constant

$$M_0\;=\;2\,f_{\max}+2\,\pi_\lambda\,(c_{\max}+R_0\,/\,(1-\beta)). \tag{A.6}$$

The nonnegativity of the left side of inequality (A.5) implies that (3.1) holds. □

REFERENCES

[1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, *On Augmented Lagrangian Methods with General Lower-Level Constraints*, SIAM J. Optim., 18 (2007), pp. 1286–1309.

[2] 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), pp. 5–32.

[3] R. Andreani, J. M. Martínez, and M. L. Schuverdt,*On the Relation between Constant Positive Linear Dependence Condition and Quasinormality Constraint Qualification*, J. Optim. Theory and Appl., 125 (2005), pp. 473-485.

[4] S. Bellavia, M. Macconi and B. Morini, *An affine scaling trust-region approach to bound-constrained nonlinear systems*, Applied Numerical Mathematics, 44 (2003), pp. 257–280.

[5] I. Bongartz, A. R. Conn, N. I. M. Gould, and P. L. Toint, *CUTE: constrained and unconstrained testing environments*, ACM Trans. Math. Softw., 21 (1995), pp. 123–160.

[6] J. Bonnans and M. Bouhtou, *The trust region affine interior point algorithm for convex and nonconvex quadratic programming*, RAIRO-Recherche Opérationnelle, 29 (1995), pp. 195–217.

[7] 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 Mach. Learn., 3(2010): 1–122.

[8] R. H. Byrd, J. C. Gilbert and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Program., 89 (2000), pp. 149–185.

[9] R. H. Byrd, M. E. Hribar and J. Nocedal, *An interior point method for large scale nonlinear programming*, SIAM J. Optim., 9 (4), pp. 877–900.

[10] T. F. Coleman and Y. Li, *On the convergence of interior-reflection Newton methods for nonlinear minimization subject to bounds*, Math. Program., 67 (1994), pp. 189–224.

[11] T. F. Coleman and Y. Li, *An interior trust region approach for nonlinear minimization subject to bounds*, SIAM J. Optim. 6 (1996), pp. 418–445.

[12] T. F. Coleman and Y. Li, *A trust region and affine scaling interior point methods for nonconvex minimization with linear inequality constraints*, Math. Program., 88 (2000), pp. 1–31.

[13] 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), pp. 545–572.

[14] A. R. Conn, N. I. M. Gould and Ph. L. Toint, LANCELOT: A Fortran Package for Large-scale Nonlinear Optimization(Release A), Springer, Heidelberg, New York, USA, 1992.

[15] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, USA, 2000.

[16] F. E. Curtis, H. Jiang and D. P. Robinson, *An adaptive augmented Lagrangian method for large-scale constrained optimization*, Math. Program., Ser. A, DOI 10.1007/s10107-014-0784-y, 2014.

[17] J. E. Dennis, M. Heinkenschloss and L. N. Vicente, *Trust-region interior-point SQP algorithms for a class of nonlinear programming problems*, SIAM J. Control Optim., 36 (1998), pp. 1750–1794.

[18] I. I. Dikin, *Iterative solution of problems of linear and quadratic programming*, Sov. Math. Dokl., 8 (1967), pp. 674–675.

[19] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program., 91 (2002), pp. 201–213.

[20] R. Flether, *Numerical experiments with an l1 exact penalty function method*, in Nonlinear Programming, O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., Vol. 4, Academic Press, New York, 1981, pp. 99-C129.

[21] A. Forsgren, P. E. Gill and M. H. Wright, *Interior methods for nonlinear optimizaiton*, SIAM Review, 44 (2003), pp. 525–597.

[22] M. D. Gonzalez-Lima, W. W. Hager and H. Zhang, *An affine-scaling interior-point method for continuous knapsack constraints with application to support vector machines*, SIAM J. Optim, 21 (2011), pp. 361–390.

[23] N. I. M. Gould, D. Orban and P. L. Toint, *Numerical methods for large-scale nonlinear optimization*, Acta Numerica, 14 (2005), pp. 299–361.

[24] W. W. Hager, *Minimizing a quadratic over a sphere*, SIAM J. Optim., 12 (2001), pp. 188–208.

[25] W. W. Hager and H. Zhang, *An affine-scaling interior-point CBB methods for box constrained optimization*, Math. Program., 119 (2009), pp. 1–32.

[26] W. W. Hager and H. Zhang, *An affine scaling method for optimization problems with polyhedral constraints*, Comput. Optim. Appl., DOI 10.1007/s10589-013-9535-x, 2013.

[27] M. Heinkenschloss, M. Ulbrich and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Math. Program., 86 (1999), pp. 615–635.

[28] C. Kanzow and A. Klug, *On affine scaling interior-point newton methods for nonlinear minimization with bound constraints*, Comput. Optim. Appl., 35 (2006), pp. 177–197.

[29] M. Kraning, E. Chu, J. Lavaei and S. Boyd, *Dynamic network energy management via proximal message passing*, Foundations and Trends® in Optim., 1(2013): 70–122.

[30] R. D. C. Monteiro and T. Tsuchiya,*Global convergence of the affine scaling algorithm for convex quadratic programmming*, SIAM J. Optim., 8 (1998), pp. 26–58.

[31] R. D. C. Monteiro and Y. Wang,*Trust region affine scaling algorithms for linearly constrained convex and concave programs*, Math. Program., 80 (1998), pp. 283–313.

[32] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Comput., 4 (1983), pp. 553C-572.
[33] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, USA 2006.
[34] B. O'Donoghue, G. Stathopoulos and S. Boyd, *A splitting method for optimal control*, IEEE. Trans. Contr. Sys. Techn., 21(2013): 2432–2442.
[35] M. J. D. Powell, *The BOBYQA algorithm for bound constrained optimization without derivatives*, Report DAMTP 2009/NA06, University of Cambridge, 2009.
[36] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP emthods*, SIAM J. Optim., 10 (2000), pp. 963–981.
[37] J. Sun, *A convergence analysis for a convex version of Dikin's algorithm*, Ann. Oper. Res., 62 (1996), pp. 357–374.
[38] P. Tseng, *Convergence Properties of Dikin's Affine Scaling Algorithm for Nonconvex Quadratic Minimization*, J. Glob. Optim., 30 (2004), pp. 285–300.
[39] L. N. Vicente, *Local convergence of the affine-scaling interior-point algorithm for nonlinear programming*, Comput. Optim. Appl., 17 (2000), pp 23–35.
[40] X. Wang and Y. Yuan, *A trust region method for simple bounded optimization based on a new affine scaling technique*, Optim. Methods and Software, 28(2013), pp. 871–888.
[41] X. Wang and Y. Yuan, *An augmented Lagrangian trust region method for equality constrained optimization*, Optim. Methods and Sotware, DOI:10.1080/10556788.2014.940947, 2014.
[42] Y. Yuan, *On the convergence of a new trust region algorithm*, Numer. Math. 70 (1995), pp. 515C539.