

A primal-dual majorization-minimization method for large-scale linear programs

Xin-Wei Liu,^{*} Yu-Hong Dai,[†] and Ya-Kui Huang[‡]

Abstract. We present a primal-dual majorization-minimization method for solving large-scale linear programs. A smooth barrier augmented Lagrangian (SBAL) function with strict convexity for the dual linear program is derived. The majorization-minimization approach is naturally introduced to develop the smoothness and convexity of the SBAL function. Our method only depends on a factorization of the constant matrix independent of iterations and does not need any computation on step sizes, thus can be expected to be particularly appropriate for large-scale linear programs. The method shares some similar properties to the first-order methods for linear programs, but its convergence analysis is established on the differentiability and convexity of our SBAL function. The global convergence is analyzed without prior requiring either the primal or dual linear program to be feasible. Under the regular conditions, our method is proved to be globally linearly convergent, and a new iteration complexity result is given.

Key words: linear programming, majorization-minimization method, augmented Lagrangian, global convergence, linear convergence

AMS subject classifications. 90C05, 90C25

^{*}Institute of Mathematics, Hebei University of Technology, Tianjin 300401, China. E-mail: math-lxw@hebut.edu.cn. The research is supported by the Chinese NSF grants (nos. 12071108 and 11671116).

[†]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China & School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. This author is supported by the NSFC grants (nos. 12021001, 11991021, 11991020 and 11971372), the National Key R&D Program of China (nos. 2021YFA 1000300 & 2021YFA 1000301), the Strategic Priority Research Program of Chinese Academy of Sciences (no. XDA27000000).

[‡]Institute of Mathematics, Hebei University of Technology, Tianjin 300401, China. This author is supported by the Chinese NSF grant (no. 11701137).

1. Introduction

We consider to solve the linear program in the dual form

$$\min_y -b^T y \quad \text{s.t. } A^T y \leq c, \quad (1.1)$$

where $y \in \mathfrak{R}^m$ is the unknown, $b \in \mathfrak{R}^m$, $A \in \mathfrak{R}^{m \times n}$ and $c \in \mathfrak{R}^n$ are given data. Corresponding to the dual problem (1.1), the primal linear program has the form

$$\min_x c^T x \quad \text{s.t. } Ax = b, \quad x \geq 0, \quad (1.2)$$

where $x \in \mathfrak{R}^n$. Problem (1.2) is called the standard form of linear programming. In the literature, most of the methods and theories for linear programming are developed with the standard form (see, for example, [30, 35, 36, 37]). Moreover, it is often assumed that $m < n$, $\text{rank}(A) = m$.

The simplex methods are the most efficient and important methods for linear programming before 1980s. These methods search the optimal solution in vertices of a polyhedral set along the boundary of the feasible region of linear programming. The initial point should be a so-called basic feasible solution corresponding to a vertex of the polyhedron which may be obtained by solving some auxiliary linear programming problem with a built-in starting point. The main computation for a new iteration point is the solution of the linear systems

$$Bu = a, \quad B^T v = d, \quad (1.3)$$

where $u \in \mathfrak{R}^m$ and $v \in \mathfrak{R}^m$ are the unknowns, $B \in \mathfrak{R}^{m \times m}$ is a nonsingular sub-matrix of A and its one column is rotated in every iteration, $a \in \mathfrak{R}^m$ and $d \in \mathfrak{R}^m$ are some given vectors. The simplex methods are favorite since the systems in (1.3) are thought to be easily solved.

It was discovered in [20], however, that the simplex approach could be inefficient for certain pathological problems since the number of iterations (also known as the worst-case time complexity) was exponential in the sizes of problems. In contrast, the interior-point approach initiated in 1984 by Karmarkar [19] has been proved to be of the worst-case polynomial time complexity, a much better theoretical property than that for the simplex methods. Up to now, the best worst-case polynomial time complexity on interior-point methods is $\mathcal{O}(\sqrt{n} \log \frac{1}{\epsilon})$ (see, for example, [36, 37]).

In general, interior-point methods converge to the optimal solution along a central path of the feasible polytope. The central path is usually defined by a parameter-perturbed Karush-Kuhn-Tucker (KKT) system. The system can be induced by the KKT conditions of the logarithmic-barrier problem

$$\min c^T x - \mu \sum_{i=1}^n \ln x_i \quad \text{s.t. } Ax = b, \quad (1.4)$$

where $\mu > 0$ is the barrier parameter, $x_i > 0$ for $i = 1, \dots, n$ (that is, x should be an interior-point). It is known that the well-defined central path depends on the nonempty of the set of the primal-dual interior-points

$$\mathcal{F} := \{(x, y, s) | Ax = b, A^T y + s = c, x > 0, s > 0\}.$$

Although there are various interior-point methods, such as the affine-scaling methods, the logarithmic-barrier methods, the potential-reduction methods, the path-following methods, etc., all these methods share some common features that distinguish them from the simplex methods. Distinct from the simplex methods in starting from a feasible point, the interior-point methods require the initial point to be an interior-point which may not be feasible to the problem. While the simplex methods usually require a larger number of relatively inexpensive iterations, every interior-point iteration needs to solve a system with the form

$$AS^{-1}XA^T v = d, \quad (1.5)$$

where $S = \text{diag}(s)$ and $X = \text{diag}(x)$. This is generally more expensive to compute than (1.3) but can make significant progress towards the solution. In particular, as the primal and dual iterates tend to the solutions of the primal and dual problems, some components of x and s can be very close to zero, which can bring about both huge and tiny values of the elements of $S^{-1}X$ and an ill-conditioned Jacobian matrix of the system (1.5) (see [30]). Some advanced methods for improving classic interior-point methods have been proposed, including the sparse matrix factorization, the Krylov subspace method and the preconditioned conjugate gradient method (see, for example, [2, 6, 8, 10, 14, 15]).

Recently, some first-order methods for solving linear programs and linear semidefinite programming have been presented, see [22, 38] and the references therein. These methods are mainly the alternating direction augmented Lagrangian methods of multipliers (ADMM)-based methods, and can be free of solving systems (1.3) and (1.5). Since the solved problems may be reformulated in different ways which result in various augmented Lagrangian function, these methods may be distinct in the augmented Lagrangian subproblems. For example, Lin et al. [22] proposed their ADMM-based interior-point method based on the well-behaved homogeneous self-dual embedded linear programming model [37], while the method in [38] is established on using the classic augmented Lagrangian function and the projection on the cone of positive semidefinite matrices.

1.1. Our contributions. We present a primal-dual majorization-minimization method on basis of solving linear programs in dual form (1.1). In our method, y_i ($i = 1, \dots, m$) are the primal variables, and x_j ($j = 1, \dots, n$) the dual variables. The method is originated from a combination of the Fiacco-McCormick logarithmic-barrier method and the Hestenes-Powell augmented Lagrangian method (see [25] for more details on general nonlinear inequality-constrained optimization). A smooth barrier augmented Lagrangian (SBAL) function with strict convexity for the dual linear program is derived. Based on the smoothness and convexity of SBAL function, a majorization surrogate function is naturally designed to find the approximate minimizer of the augmented Lagrangian on primal variables, and the dual estimates are derived by a step for maximizing a minorization surrogate function of the augmented Lagrangian on dual variables. Our method can avoid the computation on the ill-conditioned Jacobian matrix like (1.5) and does not solve some iteration-varying system (1.3) or (1.5) like the simplex methods and interior-point methods.

Our method initiates from the logarithmic-barrier reformulation of problem (1.1), thus can be

thought of an interior-point majorization-minimization method, and shares some similar features as [22]. It can also be taken as a smooth version of [38] for linear programs, but it does not depend on any projection and computes more steps on primal iterates. Differing from the fixed-point framework for proving convergence in [38], based on the smoothness and convexity of our augmented Lagrangian, we can do the global convergence and prove the results on convergence rate and iteration complexity based on the well developed theories on convex optimization [29].

Our proposed method only needs the factorization of the constant matrix AA^T , which is distinguished from the existing simplex methods and interior-point methods for linear programs necessary to solve either (1.3) or (1.5) varied in every iteration. Since the factorization is independent of iterations and can be done in preprocessing, our method can be implemented easily with very cheap computations, thus is especially suitable for large-scale linear programs. In addition, our method does not need any computation on step sizes, which is the other outstanding feature of our method in contrast to the existing interior-point methods for linear programs. Similar to [22], the global convergence is analyzed without prior requiring either the primal or dual linear program to be feasible. Moreover, under the regular conditions, we prove that our method can be of globally linear convergence, and a new iteration complexity result is obtained.

1.2. Some related works. The augmented Lagrangian methods minimize an augmented Lagrangian function approximately and circularly with update of multipliers. The augmented Lagrangian function has been playing a very important role in the development of effective numerical methods and theories for convex and nonconvex optimization problems (see some recent references, such as [3, 4, 7, 11, 12, 13, 17, 18, 22, 23, 24, 38]). The augmented Lagrangian was initially proposed by Hestenes [16] and Powell [32] for solving optimization problems with only equality constraints. The Hestenes-Powell augmented Lagrangian method was then generalized by Rockafellar [34] to solve the optimization problems with inequality constraints. Since most of the augmented Lagrangian functions for inequality-constrained optimization depend on some kind of projection, the subproblems on the augmented Lagrangian minimization are generally solved by the first-order methods.

The majorization-minimization (MM) algorithm operates on a simpler surrogate function that majorizes the objective in minimization [21]. Majorization can be understood to be a combination of tangency and domination. Similarly, we have the minorization-maximization algorithm when we want to maximize an objective. The MM principle can be dated to Ortega and Rheinboldt [31] in 1970, where the majorization idea has been stated clearly in the context of line searches. The famed expectation-maximization (EM) principle [28] of computational statistics is a special case of the MM principle. So far, MM methods have been developed and applied efficiently for imaging and inverse problems, computer vision problems, and so on (for example, see [1, 5, 9, 21, 33]).

Recently, by combining the Hestenes-Powell augmented Lagrangian and the interior-point logarithmic-barrier technique ([26, 27, 30, 35]), the authors of [25] introduce a novel barrier augmented Lagrangian function for nonlinear optimization with general inequality constraints. Distinct from the classic augmented Lagrangian function for inequality constrained optimization only first-order differentiable, the newly proposed one shares the same-order differentiability with

the objective and constraint functions and is convex when the optimization is convex. In order to distinguish the new barrier augmented Lagrangian function to those proposed in [11, 13], we refer to it as the smooth barrier augmented Lagrangian (SBAL for short). For linear problems (1.1) and (1.2), the SBAL functions are strictly convex and concave, respectively, with respect to the primal and dual variables. In particular, the SBAL functions are well defined without requiring either primal or dual iterates to be interior-points. These outstanding features of the SBAL functions provide natural selections for the majorization-minimization methods.

1.3. Organization and notations. Our paper is organized as follows. In section 2, we describe the application of our augmented Lagrangian method in [25] to the linear programs and present the associated preliminary results. The majorized functions and our primal-dual majorization-minimization method are proposed in section 3. The analysis on the global convergence and the convergence rates is done, respectively, in sections 4 and 5. We conclude our paper in the last section.

Throughout the paper, all vectors are column vectors. We use capital letters to represent matrices, and a capital letter with a subscript such as A_i means the i th column of matrix A . The small letters are used to represent vectors, and a small letter with a subscript such as s_i means the i th component of vector s . The capital letter S means the diagonal matrix of which the components of vector s are the diagonal elements. In general, we use the subscripts k and ℓ to illustrate the letters to be related to the k th and ℓ th iterations, and i and j the i th and j th components of a vector or the i th and j th sub-vectors of a matrix. In other cases, it should be clear from the context. To quantify the convergence of sequences, we introduce the weighted norm $\|y\|_M = \sqrt{y^T M y}$, where y is a column vector, M is either a positive semi-definite or positive definite symmetric matrix with the same order as y . The symbol e is the all-one vector, for which the dimension may be varying and can be known by the context. For the symmetric positive definite matrix B , we use $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ to represent the minimum and maximum of eigenvalues of B , respectively. As usual, we use the capital letters in calligraphy to represent the index sets, $\|\cdot\|$ is the Euclidean norm, $x \circ s$ is the Hadamard product of vectors x and s , and $x \in \mathfrak{R}_{++}^n$ means $x \in \mathfrak{R}^n$ and $x > 0$ in componentwise.

2. The SBAL function and some preliminary results

Recently, the authors in [24, 25] presented a novel barrier augmented Lagrangian function for nonlinear optimization with general inequality constraints. For problem (1.1), we reformulate it as

$$\min_{y,s} -b^T y \quad \text{s.t.} \quad A^T y + s = c, \quad s \geq 0, \quad (2.1)$$

where $s \in \mathfrak{R}^n$ is a slack vector. The logarithmic-barrier problem associated with (2.1) has the form

$$\min_{y,s} -b^T y - \mu \sum_{i=1}^n \ln s_i \quad \text{s.t.} \quad s - c + A^T y = 0, \quad (2.2)$$

where $s = (s_i) > 0$, $\mu > 0$ is the barrier parameter. Noting that problem (2.2) is one with only equality constraints, we can use the Hestenes-Powell augmented Lagrangian function to reformulate it into a unconstrained optimization problem as follows,

$$\min_{y,s} F_{(\mu,\rho)}(y, s; x) := -\rho b^T y - \rho\mu \sum_{i=1}^n \ln s_i + \rho x^T (s - c + A^T y) + \frac{1}{2} \|s - c + A^T y\|^2, \quad (2.3)$$

where $\rho > 0$ is the penalty parameter which may be reduced adaptively if necessary, $x \in \mathfrak{R}^n$ is an estimate of the Lagrange multiplier vector.

Since $\frac{\partial^2 F_{(\mu,\rho)}(y,s;x)}{\partial s_i^2} = \frac{\rho\mu}{s_i^2} + 1 > 0$, no matter what are (y, s) and x , $F_{(\mu,\rho)}(y, s; x)$ is a strictly convex function with respect to s_i . Therefore, $F_{(\mu,\rho)}(y, s; x)$ will take the minimizer when

$$\frac{\partial F_{(\mu,\rho)}(y, s; x)}{\partial s_i} = -\frac{\rho\mu}{s_i} + \rho x_i + (s_i - c_i + A_i^T y) = 0,$$

where $A_i \in \mathfrak{R}^m$ is the i th column vector of A . Equivalently, one has

$$s_i = \frac{1}{2} (\sqrt{(\rho x_i - c_i + A_i^T y)^2 + 4\rho\mu} - (\rho x_i - c_i + A_i^T y)).$$

Based on the observation that s_i will be altered with y and x and is dependent on the parameters μ and ρ , and for simplicity of statement, we define $s = s(y, x; \mu, \rho)$ and $z = z(y, x; \mu, \rho)$ in componentwise as

$$s_i(y, x; \mu, \rho) = \frac{1}{2} (\sqrt{(\rho x_i - c_i + A_i^T y)^2 + 4\rho\mu} - (\rho x_i - c_i + A_i^T y)), \quad (2.4)$$

$$z_i(y, x; \mu, \rho) = \frac{1}{2} (\sqrt{(\rho x_i - c_i + A_i^T y)^2 + 4\rho\mu} + (\rho x_i - c_i + A_i^T y)), \quad (2.5)$$

where $i = 1, \dots, n$. By (2.4) and (2.5), $z = s - c + A^T y + \rho x$. Correspondingly, the objective function $F_{(\mu,\rho)}(y, s; x)$ of the unconstrained optimization problem (2.3) can be written as

$$L_B(y, x; \mu, \rho) = -\rho b^T y + \sum_{i=1}^n h_i(y, x; \mu, \rho), \quad (2.6)$$

where $y \in \mathfrak{R}^m$ and $x \in \mathfrak{R}^n$ are the primal and dual variables of problem (1.1), $\mu > 0$ and $\rho > 0$ are, respectively, the barrier parameter and the penalty parameter,

$$h_i(y, x; \mu, \rho) = -\rho\mu \ln s_i(y, x; \mu, \rho) + \frac{1}{2} z_i(y, x; \mu, \rho)^2 - \frac{1}{2} \rho^2 x_i^2. \quad (2.7)$$

We may write s and z for simplicity in the sequel when their dependence on (y, x) and (μ, ρ) is clear from the context.

Similar to [25], we can prove the differentiability of the functions s , z defined by (2.4), (2.5), and the barrier augmented Lagrangian function $L_B(y, x; \mu, \rho)$ defined by (2.6).

Lemma 2.1 *For given $\mu > 0$ and $\rho > 0$, let $L_B(y, x; \mu, \rho)$ be defined by (2.6), $s = (s_i(y, x; \mu, \rho)) \in \mathfrak{R}^n$ and $z = (z_i(y, x; \mu, \rho)) \in \mathfrak{R}^n$, $S = \text{diag}(s)$ and $Z = \text{diag}(z)$.*

(1) Both s and z are differentiable with respect to y and x , and

$$\nabla_y s = -A(S + Z)^{-1}S, \quad \nabla_y z = A(S + Z)^{-1}Z, \quad (2.8)$$

$$\nabla_x s = -\rho(S + Z)^{-1}S, \quad \nabla_x z = \rho(S + Z)^{-1}Z. \quad (2.9)$$

(2) The function $L_B(y, x; \mu, \rho)$ is twice continuously differentiable with respect to y , and

$$\nabla_y L_B(y, x; \mu, \rho) = Az(y, x; \mu, \rho) - \rho b,$$

$$\nabla_{yy}^2 L_B(y, x; \mu, \rho) = A(S + Z)^{-1}ZA^T.$$

Thus, $L_B(y, x; \mu, \rho)$ is strictly convex with respect to y .

(3) The function $L_B(y, x; \mu, \rho)$ is twice continuously differentiable and strictly concave with respect to x , and

$$\nabla_x L_B(y, x; \mu, \rho) = \rho(s(y, x; \mu, \rho) - c + A^T y),$$

$$\nabla_{xx}^2 L_B(y, x; \mu, \rho) = -\rho^2(S + Z)^{-1}S.$$

Proof. (1) By (2.4) and (2.5), $s - z = c - A^T y - \rho x$ and

$$s_i + z_i = \sqrt{(\rho x_i - c_i + A_i^T y)^2 + 4\rho\mu}.$$

Thus, one has

$$\nabla_y s - \nabla_y z = -A,$$

$$\nabla_y s + \nabla_y z = A(S + Z)^{-1} \text{diag}(\rho x - c + A^T y) = A(I - 2(S + Z)^{-1}S).$$

Thus, by doing summation and subtraction, respectively, on both sides of the preceding equations, we have

$$2\nabla_y s = -2A(S + Z)^{-1}S,$$

$$-2\nabla_y z = -2A(I - (S + Z)^{-1}S) = -2A(S + Z)^{-1}Z.$$

Therefore, (2.8) follows immediately. The results in (2.9) can be derived in the same way by differentiating with respect to x .

(2) Let $h(y, x; \mu, \rho) = (h_i(y, x; \mu, \rho)) \in \mathfrak{R}^n$. Due to (2.7) and noting that $SZ = \rho\mu I$,

$$\nabla_y h(y, x; \mu, \rho) = -\rho\mu\nabla_y s S^{-1} + \nabla_y z Z = A(S + Z)^{-1}(\rho\mu I + Z^2) = AZ.$$

Thus, $\nabla_y L_B(y, x; \mu, \rho) = -\rho b + \nabla_y h(y, x; \mu, \rho)e = Az - \rho b$. Furthermore, by (1),

$$\nabla_{yy}^2 L_B(y, x; \mu, \rho) = \nabla_y z A^T = A(S + Z)^{-1}ZA^T.$$

(3) Note that

$$\nabla_x h(y, x; \mu, \rho) = -\rho\mu S^{-1}\nabla_x s + Z\nabla_x z - \rho^2 X = \rho(Z - \rho X),$$

$$\nabla_{xx}^2 h(y, x; \mu, \rho) = \rho(\nabla_x Z - \rho\nabla_x X),$$

and $\nabla_x L_B(y, x; \mu, \rho) = \nabla_x h(y, x; \mu, \rho)e$, $\nabla_{xx}^2 L_B(y, x; \mu, \rho) = \rho(\nabla_x z - \rho \nabla_x x)$. The desired formulae in (3) can be derived immediately from the equation $s - c + A^T y = z - \rho x$ and the results of (1). \square

The next result gives the relation between the SBAL function and the logarithmic-barrier problem.

Theorem 2.2 *For given $\mu > 0$ and $\rho > 0$, let $L_B(y, x; \mu, \rho)$ be defined by (2.6). Then $((y^*, s^*), x^*)$ is a KKT pair of the logarithmic-barrier problem (2.2) if and only if $s^* - c + A^T y^* = 0$ and*

$$L_B(y^*, x; \mu, \rho) \leq L_B(y^*, x^*; \mu, \rho) \leq L_B(y, x^*; \mu, \rho), \quad (2.10)$$

i.e., (y^, x^*) is a saddle point of the SBAL function $L_B(y, x; \mu, \rho)$.*

Proof. Due to Lemma 2.1 (3), for any y such that $c_i - A_i^T y > 0$, $L_B(y, x; \mu, \rho)$ reaches its maximum with respect to x_i at $x_i^* = \frac{\mu}{c_i - A_i^T y}$ since $\frac{\partial L_B(y, x; \mu, \rho)}{\partial x_i} \Big|_{x_i = x_i^*} = 0$. If $c_i - A_i^T y \leq 0$, then $\frac{\partial L_B(y, x; \mu, \rho)}{\partial x_i} > 0$, which means that $L_B(y, x; \mu, \rho)$ is strictly monotonically increasing to ∞ as $x_i \rightarrow \infty$. Thus,

$$\operatorname{argmax}_{x_i \in \mathbb{R}} L_B(y, x; \mu, \rho) = \begin{cases} \frac{\mu}{c_i - A_i^T y}, & \text{if } c_i - A_i^T y > 0; \\ \infty, & \text{otherwise.} \end{cases} \quad (2.11)$$

If $((y^*, s^*), x^*)$ is a KKT pair of the logarithmic-barrier problem (2.2), then $s^* > 0$ and

$$Ax^* = b, \quad s^* - c + A^T y^* = 0, \quad \text{and } x_i^* s_i^* = \mu, \quad i = 1, \dots, n.$$

Thus, $s_i^* = c_i - A_i^T y^* > 0$ and $x_i^* = \frac{\mu}{c_i - A_i^T y^*}$, $i = 1, \dots, n$. Therefore, by (2.11),

$$L_B(y^*, x^*; \mu, \rho) = -\rho b^T y^* - \rho \mu \sum_{i=1}^n \ln(c_i - A_i^T y^*) \geq L_B(y^*, x; \mu, \rho).$$

Furthermore, the condition $x_i^* s_i^* = \mu$ implies $z_i(y^*, x^*; \mu, \rho) - \rho x_i^* = 0$. Thus, $Az(y^*, x^*; \mu, \rho) = \rho b$. It follows from Lemma 2.1 (2), y^* is the minimizer of $L_B(y, x^*; \mu, \rho)$. That is, the right-hand-side inequality in (2.10) holds.

In reverse, if (y^*, x^*) satisfies (2.10), then y^* is a minimizer of $L_B(y, x^*; \mu, \rho)$ and x^* is a maximizer of $L_B(y^*, x; \mu, \rho)$. Thus, due to Lemma 2.1 (2) and (3), one has

$$Az(y^*, x^*; \mu, \rho) = \rho b, \quad s(y^*, x^*; \mu, \rho) - c + A^T y^* = 0.$$

The second equation further implies $z(y^*, x^*; \mu, \rho) - \rho x^* = 0$ and $x_i^*(c_i - A_i^T y^*) = \mu$, $i = 1, \dots, n$. Let $s^* = s(y^*, x^*; \mu, \rho)$. Then $s^* = c - A^T y^*$, and $((y^*, s^*), x^*)$ is a KKT pair of the logarithmic-barrier problem (2.2). \square

The following result shows that, under suitable conditions, a minimizer of problem (1.1) is an approximate minimizer of the SBAL function.

Theorem 2.3 *Let y^* be a minimizer of the problem (1.1) and x^* is the associated Lagrange multiplier vector. If the Slater constraint qualification holds, then for $\mu > 0$ sufficiently small and for $\rho > 0$, there exists a neighborhood of x^* such that y^* is a $\sqrt{\rho\mu}$ -approximate strict global minimizer of the augmented Lagrangian $L_B(y, x; \mu, \rho)$ (that is, there a scalar $\delta > 0$ such that $\|\nabla_y L_B(y^*, x; \mu, \rho)\| \leq \delta\sqrt{\rho\mu}$).*

Proof. Under the conditions of the theorem, x^* is a KKT point of problem (1.1). Thus,

$$Ax^* = b, \quad A^T y^* \leq c, \quad x^* \geq 0, \quad (x^*)^T (c - A^T y^*) = 0. \quad (2.12)$$

Let $z_i^* = z_i(y^*, x^*; \mu, \rho)$. Note that $x_i^*(c_i - A_i^T y^*) = 0$ for $i = 1, \dots, n$. Then

$$z_i^* = \begin{cases} \frac{1}{2}(\sqrt{(\rho x_i^*)^2 + 4\rho\mu} + \rho x_i^*), & \text{if } c_i - A_i^T y^* = 0, \quad x_i^* > 0; \\ \frac{1}{2}(\sqrt{(c_i - A_i^T y^*)^2 + 4\rho\mu} - (c_i - A_i^T y^*)), & \text{if } c_i - A_i^T y^* > 0, \quad x_i^* = 0; \\ \sqrt{\rho\mu}, & \text{otherwise.} \end{cases}$$

Since $\sqrt{(\rho x_i^*)^2 + 4\rho\mu} \leq \rho x_i^* + 2\sqrt{\rho\mu}$ and $\sqrt{(c_i - A_i^T y^*)^2 + 4\rho\mu} \leq (c_i - A_i^T y^*) + 2\sqrt{\rho\mu}$, one has

$$\rho x^* \leq z^* \leq \rho x^* + \sqrt{\rho\mu}, \quad \|z^* - \rho x^*\|_\infty \leq \sqrt{\rho\mu}. \quad (2.13)$$

We will prove the result by showing $\|\nabla_y L_B(y^*, x^*; \mu, \rho)\| \leq \delta\sqrt{\rho\mu}$ for some scalar δ and $\nabla_{yy}^2 L_B(y^*, x^*; \mu, \rho)$ is positive definite for $\rho > 0$. By using Lemma 2.1 (2), and (2.12), (2.13), we have

$$\|\nabla_y L_B(y^*, x^*; \mu, \rho)\| = \|Az^* - \rho b\| = \|A(z^* - \rho x^*)\| \leq \sqrt{\rho\mu}\|A\|_1,$$

which verifies the first part of the result.

Now we prove the second part of the result by showing that $d^T \nabla_{yy}^2 L_B(y^*, x^*; \mu, \rho)d > 0$ for all nonzero $d \in \mathfrak{R}^n$ and $\rho > 0$. Let $s_i^* = s_i(y^*, x^*; \mu, \rho)$. Then

$$\frac{z_i^*}{s_i^* + z_i^*} = \begin{cases} \frac{1}{2}(1 + \frac{\rho x_i^*}{\sqrt{(\rho x_i^*)^2 + 4\rho\mu}}), & \text{if } c_i - A_i^T y^* = 0, \quad x_i^* > 0; \\ \frac{1}{2}(1 - \frac{(c_i - A_i^T y^*)}{\sqrt{(c_i - A_i^T y^*)^2 + 4\rho\mu}}), & \text{if } c_i - A_i^T y^* > 0, \quad x_i^* = 0; \\ \frac{1}{2}, & \text{otherwise.} \end{cases}$$

Therefore, by Lemma 2.1 (2),

$$\begin{aligned} & \nabla_{yy}^2 L_B(y^*, x^*; \mu, \rho) \\ &= \sum_{i=1}^n \frac{z_i^*}{s_i^* + z_i^*} A_i A_i^T \quad (A_i \text{ is the } i\text{th column of } A) \\ &= \frac{1}{2} \left(\sum_{i \in \mathcal{I}_1} (1 + \frac{\rho x_i^*}{\sqrt{(\rho x_i^*)^2 + 4\rho\mu}}) A_i A_i^T + \sum_{i \in \mathcal{I}_2} A_i A_i^T \right) \\ & \quad + \frac{1}{2} \sum_{i \in \mathcal{I}_3} (1 - \frac{(c_i - A_i^T y^*)}{\sqrt{(c_i - A_i^T y^*)^2 + 4\rho\mu}}) A_i A_i^T \\ &\geq \frac{1}{2} (1 - \max\{\frac{(c_i - A_i^T y^*)}{\sqrt{(c_i - A_i^T y^*)^2 + 4\rho\mu}}, i = 1, \dots, n\}) A A^T, \end{aligned} \quad (2.14)$$

where $\mathcal{I}_1 = \{i | c_i - A_i^T y^* = 0, x_i^* > 0\}$, $\mathcal{I}_2 = \{i | c_i - A_i^T y^* = 0, x_i^* = 0\}$, $\mathcal{I}_3 = \{i | c_i - A_i^T y^* > 0, x_i^* = 0\}$. The result follows easily because of the positive definiteness of AA^T . \square

Based on the newly proposed barrier augmented Lagrangian function, [25] presented a novel augmented Lagrangian method of multipliers for optimization with general inequality constraints. The method alternately updates the primal and dual iterates by

$$y_{k+1} = \operatorname{argmin}_y L_B(y, x_k; \mu_k, \rho_k), \quad (2.15)$$

$$x_{k+1} = \frac{1}{\rho_k} z(y_{k+1}, x_k; \mu_k, \rho_k). \quad (2.16)$$

The update of parameters μ_{k+1} and ρ_{k+1} depend on the residual $\|s(y_{k+1}, x_{k+1}; \mu_k, \rho_k) - c + A^T y_{k+1}\|$ and the norm $\|x_{k+1}\|$ of dual multiplier vector.

To end this section, we show some monotone properties of our defined functions $L_B(y, x; \mu, \rho)$, $s_i(y, x; \mu, \rho)$ and $z_i(y, x; \mu, \rho)$ with respect to the parameters.

Lemma 2.4 Denote $L_B(y, x; \mu, \rho) = \rho\phi(y, x; \mu, \rho) + \frac{1}{2}R^2(y, x; \mu, \rho)$, where

$$\begin{aligned} \phi(y, x; \mu, \rho) &= -b^T y - \mu \sum_{i=1}^n \ln s_i(y, x; \mu, \rho) + x^T (s(y, x; \mu, \rho) - c + A^T y), \\ R(y, x; \mu, \rho) &= \|s(y, x; \mu, \rho) - c + A^T y\|. \end{aligned}$$

Let $\hat{y}_{k+1} = \operatorname{argmin}_y L_B(y, x_k; \mu_k, \hat{\rho}_k)$ and $\tilde{y}_{k+1} = \operatorname{argmin}_y L_B(y, x_k; \mu_k, \tilde{\rho}_k)$ be attained. If $\hat{\rho}_k > \tilde{\rho}_k$, then

$$\phi(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k) < \phi(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k), \quad R(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k) > R(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k).$$

Proof. Let $\hat{s}_{k+1} = s(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k)$ and $\tilde{s}_{k+1} = s(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k)$. Then, by (2.3),

$$(\hat{y}_{k+1}, \hat{s}_{k+1}) = \operatorname{argmin}_{y,s} F_{(\mu_k, \hat{\rho}_k)}(y, s; x_k), \quad (\tilde{y}_{k+1}, \tilde{s}_{k+1}) = \operatorname{argmin}_{y,s} F_{(\mu_k, \tilde{\rho}_k)}(y, s; x_k).$$

Thus, if we denote $\psi_\mu(y, s; x) = -b^T y - \mu \sum_{i=1}^n \ln s_i + x^T (s - c + A^T y)$ and $W(y, s; x) = \|s - c + A^T y\|$, then $F_{(\mu, \rho)}(y, s; x) = \rho\psi_\mu(y, s; x) + \frac{1}{2}W^2(y, s; x)$, and

$$\begin{aligned} \phi(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k) &= \psi_{\mu_k}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k), \quad \phi(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k) = \psi_{\mu_k}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k), \\ R(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k) &= W(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k), \quad R(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k) = W(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k). \end{aligned}$$

Moreover,

$$\begin{aligned} F_{(\mu_k, \hat{\rho}_k)}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) &< F_{(\mu_k, \hat{\rho}_k)}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k), \\ F_{(\mu_k, \tilde{\rho}_k)}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) &< F_{(\mu_k, \tilde{\rho}_k)}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k). \end{aligned}$$

It follows that

$$\begin{aligned} &F_{(\mu_k, \tilde{\rho}_k)}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) - F_{(\mu_k, \hat{\rho}_k)}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) \\ &+ F_{(\mu_k, \hat{\rho}_k)}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) - F_{(\mu_k, \tilde{\rho}_k)}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) \\ &= (\hat{\rho}_k - \tilde{\rho}_k)(\psi_{\mu_k}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) - \psi_{\mu_k}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k)) > 0. \end{aligned}$$

That is, $\phi(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k) = \psi_{\mu_k}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) > \psi_{\mu_k}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) = \phi(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k)$. Therefore,

$$\begin{aligned} \frac{1}{2}R^2(\hat{y}_{k+1}, x_k; \mu_k, \hat{\rho}_k) &= \frac{1}{2}W^2(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) \\ &> \frac{1}{2}W^2(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) + \tilde{\rho}_k(\psi_{\mu_k}(\hat{y}_{k+1}, \hat{s}_{k+1}; x_k) - \psi_{\mu_k}(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k)) \\ &> \frac{1}{2}W^2(\tilde{y}_{k+1}, \tilde{s}_{k+1}; x_k) = \frac{1}{2}R^2(\tilde{y}_{k+1}, x_k; \mu_k, \tilde{\rho}_k), \end{aligned}$$

which completes our proof. \square

Lemma 2.5 For given parameters $\mu > 0$ and $\rho > 0$, the following results are true.

- (1) Both $s_i(y, x; \mu, \rho)$ and $z_i(y, x; \mu, \rho)$ are monotonically increasing with respect to μ .
- (2) If $(s_i(y, x; \mu, \rho) - c_i + A_i^T y) \neq 0$, then $(s_i(y, x; \mu, \rho) - c_i + A_i^T y)^2$ will be decreasing as ρ is decreasing.
- (3) If $\|s(y, x; \mu, \rho) - c + A^T y\| \neq 0$, then the function $\frac{1}{\rho}L_B(y, x; \mu, \rho)$ is monotonically decreasing with respect to ρ .
- (4) The function $\frac{1}{\rho}L_B(y, x; \mu, \rho)$ is strictly convex with respect to ρ .

Proof. It should be noted that all related functions are differentiable with respect to μ and ρ . In addition, due to

$$\frac{\partial s_i(y, x; \mu, \rho)}{\partial \mu} = \frac{\partial z_i(y, x; \mu, \rho)}{\partial \mu} = \frac{\rho}{s_i + z_i} > 0, \quad (2.17)$$

$$\frac{\partial ((s_i(y, x; \mu, \rho) - c_i + A_i^T y)^2)}{\partial \rho} = \frac{2}{\rho} \frac{s_i}{s_i + z_i} (s_i - c_i + A_i^T y)^2 > 0, \quad (2.18)$$

$$\frac{\partial \frac{1}{\rho}L_B(y, x; \mu, \rho)}{\partial \rho} = -\frac{1}{2} \frac{1}{\rho^2} \|s - c + A^T y\|^2 < 0, \quad (2.19)$$

$$\frac{\partial^2 (\frac{1}{\rho}L_B(y, x; \mu, \rho))}{\partial \rho^2} = \frac{1}{\rho^3} (s - c + A^T y)^T (S + Z)^{-1} Z (s - c + A^T y), \quad (2.20)$$

the desired results are obtained immediately. \square

3. Our primal-dual majorization-minimization method

Our method in this paper focuses on how to solve the subproblem (2.15) efficiently. Noting the strict convexity of the SBAL function $L_B(y, x; \mu, \rho)$ with respect to y and the special structure of the Hessian matrix $\nabla_{yy}^2 L(y, x; \mu, \rho)$, the introduction of the majorization-minimization method is a natural selection. In particular, we will see that the dual update is precisely a step which can be derived by the minorization-maximization.

Let (y_k, x_k) be the current iteration point, $\mu_k > 0$ and $\rho_k > 0$ are the current values of the parameters. For any given $x \in \mathfrak{R}^n$, we consider the quadratic surrogate function $Q_k(\cdot, x)$:

$\Re^m \rightarrow \Re$,

$$Q_k(y, x) = L_B(y_k, x; \mu_k, \rho_k) + (Az(y_k, x; \mu_k, \rho_k) - \rho_k b)^T (y - y_k) + \frac{1}{2}(y - y_k)^T AA^T (y - y_k), \quad (3.1)$$

which is an approximate function of the objective in (2.15) and majorizes the objective function with respect to y .

Lemma 3.1 *For any given $x = \hat{x}$ and the parameters $\mu_k > 0$ and $\rho_k > 0$, there holds $Q_k(y_k, \hat{x}) = L_B(y_k, \hat{x}; \mu_k, \rho_k)$ and $L_B(y, \hat{x}; \mu_k, \rho_k) \leq Q_k(y, \hat{x})$ for all $y \in \Re^m$.*

Proof. The equation $Q_k(y_k, \hat{x}) = L_B(y_k, \hat{x}; \mu_k, \rho_k)$ is obtained from (3.1).

By Taylor's theorem with remainder,

$$L_B(y, \hat{x}; \mu_k, \rho_k) = L_B(y_k, \hat{x}; \mu_k, \rho_k) + \nabla_y L_B(y_k, \hat{x}; \mu_k, \rho_k)^T (y - y_k) + \int_0^1 (\nabla_y L_B(y_k + \tau(y - y_k), \hat{x}; \mu_k, \rho_k) - \nabla_y L_B(y_k, \hat{x}; \mu_k, \rho_k))^T (y - y_k) d\tau. \quad (3.2)$$

Due to Lemma 2.1 (2), one has

$$\begin{aligned} & \nabla_y L_B(y_k + \tau(y - y_k), \hat{x}; \mu_k, \rho_k) - \nabla_y L_B(y_k, \hat{x}; \mu_k, \rho_k) \\ &= \int_0^1 \tau \nabla_{yy}^2 L_B(y_k + \alpha\tau(y - y_k), \hat{x}; \mu_k, \rho_k) (y - y_k) d\alpha \\ &= \int_0^1 \tau A(\hat{S}_k + \hat{Z}_k)^{-1} \hat{Z}_k A^T (y - y_k) d\alpha \\ &= \tau AA^T (y - y_k) - \int_0^1 \tau A(\hat{S}_k + \hat{Z}_k)^{-1} \hat{S}_k A^T (y - y_k) d\alpha, \end{aligned}$$

where $\hat{S}_k = \text{diag}(s(y_k + \tau(y - y_k), \hat{x}; \mu_k, \rho_k))$ and $\hat{Z}_k = \text{diag}(z(y_k + \tau(y - y_k), \hat{x}; \mu_k, \rho_k))$. Noting

$$\int_0^1 \int_0^1 \tau (y - y_k)^T A(\hat{S}_k + \hat{Z}_k)^{-1} \hat{S}_k A^T (y - y_k) d\alpha d\tau \geq 0,$$

the inequality $L_B(y, \hat{x}; \mu_k, \rho_k) \leq Q_k(y, \hat{x})$ follows from Lemma 2.1 (2) and (3.2) immediately. \square

In a similar way, if for given $y \in \Re^m$ and the parameters $\mu_k > 0$ and $\rho_k > 0$, we define $P_k(y, \cdot) : \Re^n \rightarrow \Re$ be the function

$$P_k(y, x) = L_B(y, x_k; \mu_k, \rho_k) + \rho_k (s(y, x_k; \mu_k, \rho_k) - c + A^T y)^T (x - x_k) - \frac{1}{2} \rho_k^2 (x - x_k)^T (x - x_k), \quad (3.3)$$

then $P_k(y, x_k) = L_B(y, x_k; \mu_k, \rho_k)$ and $L_B(y, x; \mu_k, \rho_k) \geq P_k(y, x)$ for all $x \in \Re^n$. That is, $P_k(y, x)$ is an approximate surrogate function of the objective in optimization

$$\max_x L_B(y, x; \mu_k, \rho_k)$$

and minorizes the objective function with respect to x (i.e., majorizes the negative objective function).

By the strict convexity of $Q_k(\cdot, x)$ and the strict concavity of $P_k(y, \cdot)$, there are a unique minimizer of $Q_k(y, \hat{x})$ and a unique maximizer of $P_k(\hat{y}, x)$, where $\hat{x} \in \mathfrak{R}^n$ and $\hat{y} \in \mathfrak{R}^m$ are any given vectors.

Lemma 3.2 *Given $\mu_k > 0$ and $\rho_k > 0$. Let $Q_k(\cdot, x) : \mathfrak{R}^m \rightarrow \mathfrak{R}$ and $P_k(y, \cdot) : \mathfrak{R}^n \rightarrow \mathfrak{R}$ be functions defined by (3.1) and (3.3), respectively.*

(1) *For any given \hat{x} , $Q_k(y, \hat{x})$ has a unique minimizer y_k^* . Moreover, y_k^* satisfies the equation*

$$AA^T(y - y_k) = -(Az(y_k, \hat{x}; \mu_k, \rho_k) - \rho_k b). \quad (3.4)$$

(2) *For any given \hat{y} , $P_k(\hat{y}, x)$ has a unique maximizer x_k^* , and*

$$x_k^* = x_k + \frac{1}{\rho_k}(s(\hat{y}, x_k; \mu_k, \rho_k) - c + A^T \hat{y}). \quad (3.5)$$

(3) *For any given \hat{x} and \hat{y} , one has*

$$L_B(y_k^*, \hat{x}; \mu_k, \rho_k) - L_B(y_k, \hat{x}; \mu_k, \rho_k) \leq -\frac{1}{2}\|Az(y_k, \hat{x}; \mu_k, \rho_k) - \rho_k b\|_{(AA^T)^{-1}}^2, \quad (3.6)$$

$$L_B(\hat{y}, x_k^*; \mu_k, \rho_k) - L_B(\hat{y}, x_k; \mu_k, \rho_k) \geq \frac{1}{2}\|s(\hat{y}, x_k; \mu_k, \rho_k) - c + A^T \hat{y}\|^2. \quad (3.7)$$

Proof. Since

$$\begin{aligned} \nabla_y Q_k(y, \hat{x}) &= AA^T(y - y_k) + (Az(y_k, \hat{x}; \mu_k, \rho_k) - \rho_k b), \\ \nabla_x P_k(\hat{y}, x) &= -\rho_k^2(x - x_k) + \rho_k(s(\hat{y}, x_k; \mu_k, \rho_k) - c + A^T \hat{y}), \end{aligned}$$

and noting the strict convexity of $Q_k(y, \hat{x})$ with respect to y , and the strict concavity of $P_k(\hat{y}, x)$ with respect to x , the results (1) and (2) are obtained immediately from the optimality conditions of general unconstrained optimization (see [30, 35]).

By the preceding results, one has

$$\begin{aligned} Q_k(y_k^*, \hat{x}) &= Q_k(y_k, \hat{x}) - \frac{1}{2}\|Az(y_k, \hat{x}; \mu_k, \rho_k) - \rho_k b\|_{(AA^T)^{-1}}^2, \\ P_k(\hat{y}, x_k^*) &= P_k(\hat{y}, x_k) + \frac{1}{2}\|s(\hat{y}, x_k; \mu_k, \rho_k) - c + A^T \hat{y}\|^2. \end{aligned}$$

Due to Lemma 3.1, there hold

$$\begin{aligned} L_B(y_k^*, \hat{x}; \mu_k, \rho_k) - L_B(y_k, \hat{x}; \mu_k, \rho_k) &\leq Q_k(y_k^*, \hat{x}) - Q_k(y_k, \hat{x}), \\ L_B(\hat{y}, x_k^*; \mu_k, \rho_k) - L_B(\hat{y}, x_k; \mu_k, \rho_k) &\geq P_k(\hat{y}, x_k^*) - P_k(\hat{y}, x_k), \end{aligned}$$

which complete our proof. \square

Because of (2.4) and (2.5), (3.5) is equivalent to $x_k^* = \frac{1}{\rho_k} z(\hat{y}, x_k; \mu_k, \rho_k)$, which is consistent with (2.16). This fact shows that the dual update x_{k+1} in (2.16) can be obtained from maximizing the minorized function $P_k(y_{k+1}, x)$. In the following, we describe our algorithm for linear programming.

Algorithm 3.3 (*A primal-dual majorization-minimization method for problem (1.1)*)

Step 0. Given $(y_0, x_0) \in \mathfrak{R}^m \times \mathfrak{R}^n$, $\mu_0 > 0$, $\rho_0 > 0$, $\delta > 0$, $\gamma \in (0, 1)$, $\epsilon > 0$. Set $k := 0$.

Step 1. Approximately minimize $L_B(y, x_k; \mu_k, \rho_k)$ by the majorization-minimization method starting from y_k .

Set $\hat{y}_0 = y_k$, $\hat{\rho}_0 = \rho_k$, $\ell := 0$.

Step 1.1. Solve the equation

$$AA^T(y - \hat{y}_\ell) = -(Az(\hat{y}_\ell, x_k; \mu_k, \hat{\rho}_\ell) - \hat{\rho}_\ell b) \quad (3.8)$$

to obtain the solution $\hat{y}_{\ell+1}$. Evaluate

$$E_{k+1}^{primal} = \|Az(\hat{y}_{\ell+1}, x_k; \mu_k, \hat{\rho}_\ell) - \hat{\rho}_\ell b\|.$$

If $E_{k+1}^{primal} > \mu_k$, set $\hat{\rho}_{\ell+1} = \hat{\rho}_\ell$, $\ell := \ell + 1$ and repeat Step 1.1. Otherwise, compute

$$E_{k+1}^{dual} = \|s(\hat{y}_{\ell+1}, x_k; \mu_k, \hat{\rho}_\ell) - c + A^T \hat{y}_{\ell+1}\|.$$

If $E_{k+1}^{dual} > \max\{\hat{\rho}_\ell, \mu_k\}$, set $\hat{\rho}_{\ell+1} \geq 0.5\hat{\rho}_\ell$, $\ell := \ell + 1$ and repeat Step 1.1; else set $y_{k+1} = \hat{y}_{\ell+1}$, $\rho_{k+1} = \hat{\rho}_\ell$, end.

Step 2. Update x_k to

$$x_{k+1} = x_k + \frac{1}{\rho_{k+1}} (s(y_{k+1}, x_k; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}). \quad (3.9)$$

Step 3. If $\mu_k < \epsilon$, stop the algorithm. Otherwise, set $\mu_{k+1} \leq \gamma\mu_k$, $\rho_{k+1} = \min\{\rho_{k+1}, \frac{\delta}{\|x_{k+1}\|_\infty}\}$, $k := k + 1$. End (while)

The initial point for our algorithm can be arbitrary, which is different from both the simplex methods and the interior-point methods starting from either a feasible point or an interior-point. Theoretically, since the augmented Lagrangian function is an exact penalty function, we can always select the initial penalty parameter ρ_0 sufficiently small such that, under desirable conditions, E_{k+1}^{dual} is sufficiently small. The initial barrier parameter μ_0 can be selected to be small without affecting the well-definedness of the algorithm, but it may impact the strict convexity of the SBAL function and bring about more iterations for solving the subproblem (2.15).

The Step 1 is the core and the main computation of our algorithm. For fixed x_k , μ_k and ρ_k , we attempt to find a new estimate y_{k+1} , which is an approximate minimizer of the SBAL function $L_B(y, x_k; \mu_k, \rho_k)$ with respect to y . The main computation is in solving the system

(3.8), which depends on the decomposition of AA^T . Since AA^T is independent of the iteration, its decomposition can be fulfilled in preprocessing. If $L_B(y, x_k; \mu_k, \rho_k)$ is lower bounded, then the Step 1 will terminate in a finite number of iterations.

By Step 2 of Algorithm 3.3, we have $x_{k+1} = \frac{1}{\rho_{k+1}} z(y_{k+1}, x_k; \mu_k, \rho_{k+1})$, thus $x_{k+1} > 0$ for all $k \geq 0$. Due to Lemma 3.2 (3) and the strict concavity, one has

$$\|s(y_{k+1}, x_{k+1}; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}\| < \|s(y_{k+1}, x_k; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}\|. \quad (3.10)$$

Due to the Step 3, $\mu_k \rightarrow 0$ as $k \rightarrow \infty$, $\rho_{k+1} \|x_{k+1}\|_\infty \leq \delta$ for all $k > 0$.

4. Global convergence

We analyze the convergence of Algorithm 3.3 in this section. Firstly, we prove that, if the original problem has a minimizer, then the Step 1 will always terminate in a finite number of iterations and $\{y_k\}$ will be obtained. After that, we prove that, without prior requiring either the primal or the dual linear problem to be feasible, our algorithm can recognize the KKT point of problem (1.1), or illustrate that either its dual problem (1.2) is unbounded as problem (1.1) is feasible, or a point with least violations of constraints is found as problem (1.1) is infeasible.

Lemma 4.1 *If problem (1.1) has a solution, then for any given $x_k \in \mathfrak{R}_{++}^n$ and any given parameters $\mu_k > 0$ and $\rho_k > 0$, the SBAL function $L_B(y, x_k; \mu_k, \rho_k)$ is lower bounded from $-\infty$, and the Step 1 will terminate in a finite number of iterations.*

Proof. If problem (1.1) has a solution, then the logarithmic-barrier problem (2.2) is feasible when the original problem is strictly feasible (that is, the Slater constraint qualification holds), otherwise problem (2.2) is infeasible. Correspondingly, the objective $-b^T y - \mu \sum_{i=1}^n \ln s_i$ of problem (2.2) either takes its minimizer at an interior-point of problem (1.1) (in this case the minimizer is attained) or is $+\infty$. It is noted that $A_i^T y \rightarrow -\infty$ for any $i = 1, \dots, n$ if and only if $A_i^T y < c_i$, the corresponding constraint of problem (1.1) is strictly feasible. The preceding result shows that no matter when y is such that $A_i^T y \rightarrow -\infty$ for any $i = 1, \dots, n$, the minimizer of $-b^T y - \mu \sum_{i=1}^n \ln s_i$ with $s_i = \max\{c_i - A_i^T y, 0\}$ will be lower bounded away from $-\infty$.

If $L_B(y, x_k; \mu_k, \rho_k)$ is not lower bounded, then $L_B(y, x_k; \mu_k, \rho_k) \rightarrow -\infty$ as $A_i^T y \rightarrow -\infty$ for some $i = 1, \dots, n$. Let $\mathcal{I}(y) = \{i | A_i^T y \rightarrow -\infty\}$. Since

$$\begin{aligned} L_B(y, x_k; \mu_k, \rho_k) &\geq -b^T y - \mu_k \sum_{i=1}^n \ln s_i(y, x_k; \mu_k, \rho_k) \\ &= -b^T y - \mu_k \sum_{i \in \mathcal{I}(y)} \ln(c_i - A_i^T y) - \mu_k \sum_{i \in \mathcal{I}(y)} \ln \frac{s_i(y, x_k; \mu_k, \rho_k)}{c_i - A_i^T y} - \mu_k \sum_{i \notin \mathcal{I}(y)} \ln s_i(y, x_k; \mu_k, \rho_k) \\ &> -\infty, \end{aligned}$$

it shows that $L_B(y, x_k; \mu_k, \rho_k)$ is lower bounded away from $-\infty$.

Now we prove that for any fixed $\hat{\rho}_\ell$, if the Step 1 of Algorithm 3.3 does not terminate finitely, then $E_{k+1}^{primal} \rightarrow 0$ as $\ell \rightarrow \infty$. By Lemma 3.2, $\{L_B(\hat{y}_\ell, x_k; \mu_k, \hat{\rho}_\ell)\}$ is monotonically non-increasing as $\ell \rightarrow \infty$. Thus either there is a finite limit for the sequence $\{L_B(\hat{y}_\ell, x_k; \mu_k, \hat{\rho}_\ell)\}$ or the whole sequence tends to $-\infty$. Since $L_B(y, x_k; \mu_k, \hat{\rho}_\ell)$ is bounded below, due to (3.6), one has

$$\lim_{\ell \rightarrow \infty} \|Az(\hat{y}_\ell, x_k; \mu_k, \hat{\rho}_\ell) - \hat{\rho}_\ell b\|_{(AA^T)^{-1}} = 0, \quad (4.1)$$

which shows that the condition $E_{k+1}^{primal} \leq \mu_k$ will be satisfied in a finite number of iterations.

Since problem (1.1) is supposed to be feasible, for every $s > 0$ one has $s - c + A^T y > 0$. It follows from Lemma 2.4 that there is a scalar $\rho_{k+1} > 0$ such that for given $\mu_k > 0$ and for all $\hat{\rho}_\ell \leq \rho_{k+1}$, $E_{k+1}^{dual} \leq \mu_k$ as ℓ is large enough. Thus, the Step 1 will terminate in a finite number of iterations. \square

The next result shows that, if the Step 1 does not terminate finitely, then either problem (1.1) is unbounded or a point with least constraint violations will be found.

Lemma 4.2 *For given $x_k \in \mathfrak{R}_{++}^n$ and parameters $\mu_k > 0$ and $\rho_k > 0$, if the Step 1 of Algorithm 3.3 does not terminate finitely and an infinite sequence $\{\hat{y}_\ell\}$ is generated, then either problem (1.1) is unbounded or any cluster point of $\{\hat{y}_\ell\}$ is an infeasible stationary point y^* satisfying*

$$A \max\{A^T y^* - c, 0\} = 0. \quad (4.2)$$

The point y^ is also a solution for minimizing the ℓ_2 -norm of constraint violations of problem (1.1), and shows that problem (1.2) is unbounded.*

Proof. If that the Step 1 of Algorithm 3.3 does not terminate finitely is resulted from E_{k+1}^{primal} not being small enough for given ρ_k , then $\{\hat{y}_\ell\}$ is unbounded and $L_B(\hat{y}_\ell, x_k; \mu_k, \hat{\rho}_\ell) \rightarrow -\infty$ as $\ell \rightarrow \infty$, which by the arguments in the proof of Lemma 4.1 implies that problem (1.1) is feasible and unbounded.

Now we consider the case that $\{\hat{y}_\ell\}$ is bounded for given ρ_k . Suppose that \mathcal{L} is a subset of indices such that $\hat{y}_\ell \rightarrow \hat{y}_\ell^*$ as $\ell \in \mathcal{L}$ and $\ell \rightarrow \infty$ for given ρ_k . Then

$$Az(\hat{y}_\ell^*, x_k; \mu_k, \rho_k) - \rho_k b = 0. \quad (4.3)$$

Due to $z(\hat{y}_\ell^*, x_k; \mu_k, \rho_k) > 0$, (4.3) shows that problem (1.2) is feasible. Furthermore, considering the fact that the Step 1 of Algorithm 3.3 does not terminate finitely, one has $\hat{\rho}_\ell \rightarrow 0$. Thus, the result (4.2) follows since $z(\hat{y}_\ell^*, x_k; \mu_k, \hat{\rho}_\ell) \rightarrow \max\{A^T y^* - c, 0\}$ as $\hat{\rho}_\ell \rightarrow 0$.

In addition, since $s(\hat{y}_\ell^*, x_k; \mu_k, \hat{\rho}_\ell) - c + A^T \hat{y}_\ell^* > \mu_k$ for given $\mu_k > 0$ and $\ell \in \{\ell | \hat{\rho}_{\ell+1} \leq 0.5\hat{\rho}_\ell\}$, and

$$s(\hat{y}_\ell^*, x_k; \mu_k, \hat{\rho}_\ell) - c + A^T \hat{y}_\ell^* \rightarrow \max\{A^T y^* - c, 0\} \text{ as } \hat{\rho}_\ell \rightarrow 0,$$

then $\max\{A^T y^* - c, 0\} \geq \mu_k > 0$. That is, y^* is infeasible to the problem (1.1), which by [30, 35, 36, 37] implies that problem (1.2) is unbounded. Noting that (4.2) suggests that y^* satisfies the stationary condition of the linear least square problem

$$\min_y \frac{1}{2} \|\max\{A^T y - c, 0\}\|^2,$$

y^* is a point with the least ℓ_2 -norm of constraint violations of problem (1.1). \square

In the following analysis of this section, we suppose that the Step 1 of Algorithm 3.3 terminates finitely for every k . In order to analyze the convergence of Algorithm 3.3, we also suppose that Algorithm 3.3 does not terminate finitely, and an infinite sequence $\{y_k\}$ is generated. Corresponding to the sequence $\{y_k\}$, we also have the sequence $\{\mu_k\}$ of barrier parameters, the sequence $\{\rho_k\}$ of penalty parameters, the sequence $\{x_k\}$ of the estimates of multipliers. In particular, $\{\mu_k\}$ is a monotonically decreasing sequence and tends to 0, $\{\rho_k\}$ is a monotonically non-increasing sequence which either keeps unchanged after a finite number of steps or tends to 0,

$$\begin{aligned} x_{k+1} &= x_k + \frac{1}{\rho_{k+1}}(s(y_{k+1}, x_k; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}) \\ &= x_{k-1} + \frac{1}{\rho_k}(s(y_k, x_{k-1}; \mu_{k-1}, \rho_k) - c + A^T y_k) + \frac{1}{\rho_{k+1}}(s(y_{k+1}, x_k; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}) \\ &= x_0 + \sum_{\ell=0}^k \frac{1}{\rho_{\ell+1}}(s(y_{\ell+1}, x_\ell; \mu_\ell, \rho_{\ell+1}) - c + A^T y_{\ell+1}). \end{aligned}$$

If the sequence $\{x_k\}$ is bounded, then $\frac{1}{\rho_{k+1}}(s(y_{k+1}, x_k; \mu_k, \rho_{k+1}) - c + A^T y_{k+1}) \rightarrow 0$ as $k \rightarrow \infty$, and $\{\rho_k\}$ is bounded away from zero.

Lemma 4.3 *If $\rho_k \rightarrow 0$, then any cluster point of $\{y_k\}$ is a Fritz-John point of problem (1.1). In particular, there exists an infinite subset \mathcal{K} of indices such that for $k \in \mathcal{K}$ and $k \rightarrow \infty$, $y_k \rightarrow y^*$, $z_k \rightarrow z^* \geq 0$, $s_k \rightarrow s^* \geq 0$ and*

$$s^* - c + A^T y^* = 0, \quad Az^* = 0, \quad z^* \circ s^* = 0, \quad (4.4)$$

which shows that problem (1.1) is feasible but problem (1.2) is unbounded.

Proof. Without loss of generality, we assume that $\{y_k\}$ is bounded. Because of the boundedness of $\{\rho_k x_k\}$, both $\{s_k\}$ and $\{z_k\}$ are bounded. Without loss of generality, we let $y_k \rightarrow y^*$, $z_k \rightarrow z^*$, $s_k \rightarrow s^*$ for $k \in \mathcal{K}$ and $k \rightarrow \infty$. Then $z^* \geq 0$ and $s^* \geq 0$. Therefore, (4.4) follows immediately from

$$\mu_k \rightarrow 0, \quad E_k^{primal} \leq \mu_k, \quad E_k^{dual} \leq \max\{\rho_k, \mu_k\}, \quad \text{and } z_k \circ s_k = \rho_k \mu_k e.$$

That is, y^* is a Fritz-John point of problem (1.1).

The equations in (4.4) show that, if $\rho_k \rightarrow 0$, Algorithm 3.3 will converge to a feasible point y^* of (1.1). By the first part of the proof of Lemma 4.2, the finite termination of Step 1 implies that problem (1.2) is strictly feasible. Thus, its set of solutions are unbounded since for any feasible point x of problem (1.2), due to (4.4), $x + \alpha z^*$ is feasible to problem (1.2) and $c^T(x + \alpha z^*) = c^T x$ for all $\alpha \geq 0$. \square

In what follows, we prove the convergence of Algorithm 3.3 to a KKT point.

Lemma 4.4 *If ρ_k is bounded away from zero, then $\{x_k\}$ is bounded, and every cluster point of $\{(y_k, x_k)\}$ is a KKT pair of problem (1.1).*

Proof. Suppose that $\rho_k \geq \rho^* > 0$ for all $k \geq 0$ and for some scalar ρ^* , then by Step 3 of Algorithm 3.3, $\frac{1}{\|x_k\|_\infty} \geq \rho^*$. Thus, $\|x_k\|_\infty \leq \frac{1}{\rho^*}$.

Since $\|x_k\|$ is bounded, $\lim_{k \rightarrow \infty} E_k^{dual} = 0$. Thus, $E_k^{dual} \leq \rho^*$ for all k sufficiently large. Together with the facts $\mu_k \rightarrow 0$ and

$$E_k^{primal} = \|\rho_k Ax_k - \rho_k b\|_{(AA^T)^{-1}} \leq \mu_k,$$

one has the result immediately. \square

In summary, we have the following global convergence results on Algorithm 3.3.

Theorem 4.5 *One of following three cases will arise when implementing Algorithm 3.3.*

(1) *The Step 1 does not terminate finitely for some $k \geq 0$, $\hat{\rho}_\ell \rightarrow 0$, either problem (1.1) is unbounded, or problem (1.1) is infeasible and problem (1.2) is unbounded, and a point for minimizing the ℓ_2 norm of constraint violations is found.*

(2) *The Step 1 terminate finitely for all $k \geq 0$, $\mu_k \rightarrow 0$ and $\rho_k \rightarrow 0$ as $k \rightarrow \infty$, problem (1.2) is unbounded, problem (1.1) is feasible and every cluster point of $\{y_k\}$ is a Fritz-John point of problem (1.1).*

(3) *The Step 1 terminate finitely for all $k \geq 0$, $\mu_k \rightarrow 0$ as $k \rightarrow \infty$, and ρ_k is bounded away from zero, both problems (1.1) and (1.2) are feasible and every cluster point of $\{y_k\}$ is a KKT point of problem (1.1).*

Proof. The results can be obtained straightforward from the preceding results Lemmas 4.1, 4.2, 4.3, and 4.4 in this section. \square

For reader's convenient, we summarize our global convergence results in Table 1.

Table 1: The overview on the global convergence results of Algorithm 3.3.

Algorithm 3.3	Results		
	Dual LP (1.1)	Primal LP (1.2)	Solution obtained
$\hat{\rho}_\ell \rightarrow 0, \mu_k > 0$	unbounded	-	-
	infeasible	unbounded	A point for minimizing constraint violations of LP (1.1)
$\mu_k \rightarrow 0, \rho_k \rightarrow 0$	feasible	unbounded	A Fritz-John point of LP (1.1)
$\mu_k \rightarrow 0, \rho_k > 0$	feasible	feasible	A KKT point

5. Convergence rates and the complexity

In this section, we concern about the convergence rate of Algorithm 3.3 under the situation

that both problems (1.1) and (1.2) are feasible, which corresponds to the result (3) of the preceding global convergence theorem. Firstly, without any additional assumption, based on theory on convex optimization [29], we prove that for given penalty parameter ρ_k , the convergence rate of the sequence of objective function values on the SBAL minimization subproblem is $O(\frac{1}{\ell})$, where $\ell > 0$ is a positive integer which is also the number of iterations of the Step 1. Secondly, under the regular conditions on the solution, we show that the iterative sequence $\{\hat{y}_\ell\}$ on the SBAL minimization subproblem is globally linearly convergent. Finally, without loss of generality, by assuming that ρ_k is small enough such that in Step 1, $E_{k+1}^{dual} \leq \max\{\rho_k, \mu_k\}$ for given ρ_k , and using the preceding global linear convergence result, we can establish the iteration complexity of our algorithm.

Theorem 5.1 *For given x_k and parameters μ_k and ρ_k , let $F_k(y) = L_B(y, x_k; \mu_k, \rho_k)$, $\{\hat{y}_\ell\}$ be a sequence generated by Step 1 of Algorithm 3.3 for minimizing $F_k(y)$, and $F_k^* = \inf_y F_k(y)$, $y_k^* = \operatorname{argmin}_y F_k(y)$. Then*

$$F_k(\hat{y}_\ell) - F_k^* \leq \frac{1}{2\ell} \|\hat{y}_0 - y_k^*\|_{AA^T}^2, \quad (5.1)$$

where \hat{y}_0 is an arbitrary starting point.

Proof. It follows from Lemma 3.2 that

$$\begin{aligned} F_k(\hat{y}_{\ell+1}) &\leq F_k(\hat{y}_\ell) - \frac{1}{2} \|Az(\hat{y}_\ell, x_k; \mu_k, \rho_k) - \rho_k b\|_{(AA^T)^{-1}}^2 \\ &\leq F_k^* + \nabla F_k(\hat{y}_\ell)^T (\hat{y}_\ell - y_k^*) - \frac{1}{2} \|Az(\hat{y}_\ell, x_k; \mu_k, \rho_k) - \rho_k b\|_{(AA^T)^{-1}}^2 \\ &= F_k^* + \frac{1}{2} (\|\hat{y}_\ell - y_k^*\|_{AA^T}^2 - \|\hat{y}_\ell - y_k^* - (AA^T)^{-1} \nabla F_k(\hat{y}_\ell)\|_{AA^T}^2) \\ &= F_k^* + \frac{1}{2} (\|\hat{y}_\ell - y_k^*\|_{AA^T}^2 - \|\hat{y}_{\ell+1} - y_k^*\|_{AA^T}^2), \end{aligned} \quad (5.2)$$

where the second inequality follows from the convexity of $F_k(y)$, and the last equality is obtained by (3.8). Thus,

$$\begin{aligned} \sum_{t=1}^{\ell} (F_k(\hat{y}_t) - F_k^*) &\leq \sum_{t=1}^{\ell} \frac{1}{2} (\|\hat{y}_{t-1} - y_k^*\|_{AA^T}^2 - \|\hat{y}_t - y_k^*\|_{AA^T}^2) \\ &= \frac{1}{2} (\|\hat{y}_0 - y_k^*\|_{AA^T}^2 - \|\hat{y}_\ell - y_k^*\|_{AA^T}^2), \end{aligned}$$

which implies $F_k(\hat{y}_\ell) - F_k^* \leq \frac{1}{2\ell} \|\hat{y}_0 - y_k^*\|_{AA^T}^2$. \square

In order to derive the convergence rate of the iterative sequence $\{\hat{y}_\ell\}$ of our method for the subproblem, we need to prove some lemmas.

Lemma 5.2 *For given x_k and parameters μ_k and ρ_k , let $F_k(y) = L_B(y, x_k; \mu_k, \rho_k)$. Then for any $u, v \in \mathfrak{R}^m$,*

$$(\nabla F_k(u) - \nabla F_k(v))^T (u - v) \geq \|\nabla F_k(u) - \nabla F_k(v)\|_{(AA^T)^{-1}}^2. \quad (5.3)$$

Proof. For proving (5.3), we consider the auxiliary function

$$G_u(v) = F_k(v) - \nabla F_k(u)^T v,$$

where v is the variable and u is any given vector. Then $\nabla G_u(u) = 0$ and $\nabla^2 G_u(v) = \nabla^2 F_k(v)$, which means that $G_u(v)$ is convex as $F_k(v)$ and u is precisely a global minimizer of $G_u(v)$. Therefore, we have a similar result to Lemma 3.1 (1), that is, for every $w, v \in \mathfrak{R}^m$,

$$G_u(w) \leq G_u(v) + \nabla G_u(v)^T (w - v) + \frac{1}{2} (w - v)^T A A^T (w - v),$$

which implies $G_u(v) - G_u(u) \geq \frac{1}{2} \|\nabla G_u(v)\|_{(A A^T)^{-1}}^2$ for every $v \in \mathfrak{R}^m$. Because of $\nabla G_u(v) = \nabla F_k(v) - \nabla F_k(u)$, the preceding inequality is equivalent to

$$F_k(v) - F_k(u) - \nabla F_k(u)^T (v - u) \geq \frac{1}{2} \|\nabla F_k(v) - \nabla F_k(u)\|_{(A A^T)^{-1}}^2. \quad (5.4)$$

Similarly, one can prove

$$F_k(u) - F_k(v) - \nabla F_k(v)^T (u - v) \geq \frac{1}{2} \|\nabla F_k(v) - \nabla F_k(u)\|_{(A A^T)^{-1}}^2. \quad (5.5)$$

Summarizing two sides of (5.4) and (5.5) brings about our desired result. \square

In the subsequent analysis, let y^* be the solution of problem (1.1) and x^* be the associated Lagrange multiplier vector, and $s^* = c - A^T y^*$. Thus, $x^* \circ s^* = 0$. We need the following blanket assumption.

Assumption 5.3 Denote $\mathcal{I} = \{i = 1, \dots, n \mid x_i^* > 0\}$. Suppose that the strict complementarity holds, and the columns of A corresponding to the positive components of x^* are linearly independent. That is, $x^* + s^* > 0$ and $|\mathcal{I}| = m$, $B = A_{\mathcal{I}} A_{\mathcal{I}}^T$ is positive definite, where $|\cdot|$ is the cardinality of the set, $A_{\mathcal{I}}$ is a submatrix of A consisting of A_i , $i \in \mathcal{I}$.

Under the Assumption 5.3, there exists a scalar $\delta > 0$ such that, for $i \in \mathcal{I}$ and for all $\ell \geq 0$, $(s_{\ell i} + z_{\ell i})^{-1} z_{\ell i} \geq \delta > 0$. Thus, for any $y \in \mathfrak{R}^m$,

$$\begin{aligned} y^T A(S + Z)^{-1} Z A^T y &\geq y^T (A_{\mathcal{I}}(S_{\mathcal{I}} + Z_{\mathcal{I}})^{-1} Z_{\mathcal{I}} A_{\mathcal{I}}^T) y \\ &\geq \delta y^T (A_{\mathcal{I}} A_{\mathcal{I}}^T) y \geq \delta' y^T y \geq \delta'' y^T A A^T y, \end{aligned}$$

where $\delta' \leq \delta \lambda_{\min}(A A^T)$ and $\delta'' \leq \frac{\delta'}{\lambda_{\max}(A A^T)} < 1$.

Lemma 5.4 For given x_k and parameters μ_k and ρ_k , let $F_k(y) = L_B(y, x_k; \mu_k, \rho_k)$. Under the Assumption 5.3, there exists a scalar $\delta'' \in (0, 1)$ such that, for any $u, v \in \mathfrak{R}^m$,

$$\begin{aligned} &(\nabla F_k(u) - \nabla F_k(v))^T (u - v) \\ &\geq \frac{1}{1 + \delta''} \|\nabla F_k(u) - \nabla F_k(v)\|_{(A A^T)^{-1}}^2 + \frac{\delta''}{1 + \delta''} \|u - v\|_{A A^T}^2. \end{aligned} \quad (5.6)$$

Proof. Let $G_k(y) = F_k(y) - \frac{1}{2}\delta'' y^T A A^T y$. Then $G_k(y)$ and $\frac{1}{2}(1 - \delta'')y^T A A^T y - G_k(y)$ are convex, which suggests that $G_k(y)$ shares the similar properties with $F_k(y)$. Thus, the result of Lemma 5.2 still holds for $G_k(y)$, i.e., for any $u, v \in \mathfrak{R}^m$,

$$(\nabla G_k(u) - \nabla G_k(v))^T(u - v) \geq \|\nabla G_k(u) - \nabla G_k(v)\|_{(A A^T)^{-1}}^2.$$

Due to $\nabla G_k(y) = \nabla F_k(y) - \delta'' A A^T y$, the preceding inequality can be rewritten as

$$\begin{aligned} & (\nabla F_k(u) - \nabla F_k(v))^T(u - v) \\ & \geq \frac{1}{1 - \delta''} \|\nabla F_k(u) - \nabla F_k(v) - \delta'' A A^T(u - v)\|_{(A A^T)^{-1}}^2 + \delta'' \|u - v\|_{A A^T}^2. \end{aligned}$$

Thus, one has

$$\begin{aligned} & (\nabla F_k(u) - \nabla F_k(v))^T(u - v) \\ & \geq \frac{1}{1 + \delta''} \|\nabla F_k(u) - \nabla F_k(v)\|_{(A A^T)^{-1}}^2 + \frac{\delta''}{1 + \delta''} \|u - v\|_{A A^T}^2, \end{aligned}$$

which completes our proof. \square

Set $u = \hat{y}_\ell$ and $v = y_k^*$. Due to $\nabla F_k(y_k^*) = 0$,

$$\nabla F_k(\hat{y}_\ell)^T(\hat{y}_\ell - y_k^*) \geq \frac{1}{1 + \delta''} \|\nabla F_k(\hat{y}_\ell)\|_{(A A^T)^{-1}}^2 + \frac{\delta''}{1 + \delta''} \|\hat{y}_\ell - y_k^*\|_{A A^T}^2.$$

The next result shows that sequence $\{\hat{y}_\ell\}$ can be of global linear convergence for the SBAL minimization subproblem.

Theorem 5.5 *Let $y_k^* = \operatorname{argmin} F_k(y)$. Under Assumption 5.3, there is a scalar $\tau \in (0, 1)$ such that*

$$\|\hat{y}_\ell - y_k^*\|_{A A^T}^2 \leq \tau^\ell \|\hat{y}_0 - y_k^*\|_{A A^T}^2.$$

That is, $\{\hat{y}_\ell\}$ is of global linear convergence to y_k^ .*

Proof. Note that

$$\begin{aligned} & \|\hat{y}_{\ell+1} - y_k^*\|_{A A^T}^2 \\ & = \|\hat{y}_\ell - (A A^T)^{-1} \nabla F_k(\hat{y}_\ell) - y_k^*\|_{A A^T}^2 \\ & = \|\hat{y}_\ell - y_k^*\|_{A A^T}^2 - 2 \nabla F_k(\hat{y}_\ell)^T(\hat{y}_\ell - y_k^*) + \|\nabla F_k(\hat{y}_\ell)\|_{(A A^T)^{-1}}^2 \\ & \leq \left(1 - \frac{2\delta''}{1 + \delta''}\right) \|\hat{y}_\ell - y_k^*\|_{A A^T}^2 + \left(1 - \frac{2}{1 + \delta''}\right) \|\nabla F_k(\hat{y}_\ell)\|_{(A A^T)^{-1}}^2 \\ & = \frac{1 - \delta''}{1 + \delta''} \|\hat{y}_\ell - y_k^*\|_{A A^T}^2 - \frac{1 - \delta''}{1 + \delta''} \|\nabla F_k(\hat{y}_\ell)\|_{(A A^T)^{-1}}^2 \\ & \leq \frac{1 - \delta''}{1 + \delta''} \|\hat{y}_\ell - y_k^*\|_{A A^T}^2. \end{aligned}$$

By setting $\tau = \frac{1 - \delta''}{1 + \delta''}$, the result follows immediately. \square

Finally, based on the preceding global linear convergence result, we can obtain a new iteration complexity result on the algorithms for linear programs.

Theorem 5.6 *Suppose that both problems (1.1) and (1.2) are feasible, and Assumption 5.3 holds. For ρ_0 sufficiently small, if Algorithm 3.3 is terminated when $\mu_k < \epsilon$, where $\epsilon > 0$ is a pre-given tolerance, then the iteration complexities of the MM methods for the subproblem and for problem (1.1) are respectively*

$$T_{\text{MM}} = O\left(\frac{1}{\ln \sqrt{\frac{\kappa_A + \delta}{\kappa_A - \delta}}} \ln\left(\frac{1}{\epsilon}\right)\right), \quad T_{\text{PDMM}} = O\left(\frac{1}{\ln \sqrt{\frac{\kappa_A + \delta}{\kappa_A - \delta}}} \left(\ln\left(\frac{1}{\epsilon}\right)\right)^2\right),$$

where $\delta \in (0, 1)$ is a scalar independent of k .

Proof. Due to Lemma 2.1 (2), one has

$$\begin{aligned} & \|Az(\hat{y}_{\ell+1}, x_k; \mu_k, \rho_0) - \rho_0 b\| \\ &= \|\nabla_y L(\hat{y}_{\ell+1}, x_k; \mu_k, \rho_0) - \nabla_y L(y_k^*, x_k; \mu_k, \rho_0)\| \\ &\leq \|\hat{y}_{\ell+1} - y_k^*\|_{AA^T}. \end{aligned}$$

In order to obtain $\|Az(\hat{y}_{\ell+1}, x_k; \mu_k, \rho_0) - \rho_0 b\| \leq \mu_k \leq \epsilon$, by Theorem 5.5, T_{MM} should satisfy

$$\sqrt{\tau}^{T_{\text{MM}}} \|y_k - y_k^*\|_{AA^T} \leq \epsilon,$$

where $y_k = \hat{y}_0$, τ is denoted in Theorem 5.5 and can be replaced by $\tau = \frac{\kappa_A - \delta}{\kappa_A + \delta}$ ($\kappa_A = \lambda_{\max}(AA^T)/\lambda_{\min}(AA^T)$, $\delta \in (0, 1)$ is denoted such that $(s_{\ell i} + z_{\ell i})^{-1} z_{\ell i} \geq \delta$ for all $i \in \mathcal{I}$ and for all $\ell > 0$ in Assumption 5.3). Thus,

$$T_{\text{MM}} \ln \frac{1}{\sqrt{\tau}} \geq \ln \frac{\|y_k - y_k^*\|_{AA^T}}{\epsilon}.$$

That is,

$$T_{\text{MM}} = O\left(\frac{1}{\ln \sqrt{\frac{\kappa_A + \delta}{\kappa_A - \delta}}} \ln\left(\frac{1}{\epsilon}\right)\right).$$

In addition, similarly, the number of iterations needed for driving $\mu_k < \epsilon$ is

$$T_{\text{out}} \geq \frac{1}{\ln \frac{1}{\gamma}} \ln \frac{\mu_0}{\epsilon}.$$

Thus, we have the estimate on the total number of iterations

$$T_{\text{PDMM}} = \sum_{k=1}^{T_{\text{out}}} T_{\text{MM}} = T_{\text{out}} T_{\text{MM}} = O\left(\frac{1}{\ln \sqrt{\frac{\kappa_A + \delta}{\kappa_A - \delta}}} \left(\ln\left(\frac{1}{\epsilon}\right)\right)^2\right), \quad (5.7)$$

which completes our proof. \square

6. Conclusion

The simplex methods and the interior-point methods are two kinds of main and effective methods for solving linear programs. Relatively, the former is more inexpensive for every iteration but may require more iterations to find the solution, while the latter is more expensive for

one iteration but the number of iterations may not be changed greatly with different problems. Theoretically, the iteration complexity of the simplex methods can be exponential on the sizes of linear programs, while the interior-point methods can be polynomial.

In this paper, we present a primal-dual majorization-minimization method for linear programs. The method is originated from the application of the Hestenes-Powell augmented Lagrangian method to the logarithmic-barrier problems. A novel barrier augmented Lagrangian (SBAL) function with second-order smoothness and strict convexity is proposed. Based the SBAL function, a majorization-minimization approach is introduced to solve the augmented Lagrangian subproblems. Distinct from the existing simplex methods and interior-point methods for linear programs, but similar to some alternate direction methods of multipliers (ADMM), the proposed method only depends on a factorization of the constant matrix independent of iterations which can be done in the preprocessing, and does not need any computation on step sizes, thus is much more inexpensive for iterations and can be expected to be particularly appropriate for large-scale linear programs. The global convergence is analyzed without prior assuming either primal or dual problem to be feasible. Under the regular conditions, based on theory on convex optimization, we prove that our method can be of globally linear convergence. The results show that the iteration complexity on our method is dependent on the conditioned number of the product matrix of the coefficient matrix and its transpose.

References

- [1] M. ALLAIN, J. IDIER, AND Y. GOUSSARD, *On global and local convergence of half-quadratic algorithms*, IEEE Trans. Im. Proc., 15(2006), 1130–1142.
- [2] S.J. BENSON, Y. YE, AND X. ZHANG, *Solving large-scale sparse semidefinite programs for combinatorial optimization*, SIAM J. Optim. 10(2000), 443–461.
- [3] E.G. BIRGIN AND J.M. MARTINEZ, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Book Series: Fundamentals of Algorithms, SIAM, PA 19104-2688 USA, 2014.
- [4] N. CHATZIPANAGIOTIS, D. DENTCHEVA AND M.M. ZAVLANOS, *An augmented Lagrangian method for distributed optimization*, Math. Program., 152(2015), 405–434.
- [5] L.D. COHEN, *Auxiliary variables and two-step iterative algorithms in computer vision problems*, J. Math. Im. Vision, 6(1996), 59–83.
- [6] Y. CUI, K. MORIKUNI, T. TSUCHIYA, AND K. HAYAMI, *Implementation of interior-point methods for LP based on Krylov subspace iterative solvers with inner-iteration preconditioning*, Comput. Optim. Appl. 74(2019), 143–176.
- [7] Y.-H. DAI, X.-W. LIU AND J. SUN, *A primal-dual interior-point method capable of rapidly detecting infeasibility for nonlinear programs*, J. Ind. Manag. Optim., 2018, doi: 10.3934/jimo.2018190.
- [8] T.A. DAVIS, *Direct Methods for Sparse Linear Systems*, SIAM Fundamentals of Algorithms, SIAM, Philadelphia, 2006.

- [9] M. FIGUEIREDO, J. BIOCAS-DIAS AND R. NOWAK, *Majorization-minimization algorithms for wavelet-based image restoration*, IEEE Trans. Image. Process., 16(2006), 2980–2991.
- [10] K. FOUNTOULAKIS, J. GONDZIO, AND P. ZHLOBICH, *Matrix-free interior point method for compressed sensing problems*, Math. Program. Comput. 6(2014), 1–31.
- [11] P.E. GILL, V. KUNGURTSEV AND D.P. ROBINSON, *A shifted primal-dual penalty-barrier method for nonlinear optimization*, SIAM J. Optim., 30 (2020), 1067–1093.
- [12] P.E. GILL AND D.P. ROBINSON, *A primal-dual augmented Lagrangian*, Comput. Optim. Appl., 51 (2012), 1–25.
- [13] D. GOLDFARB, R. A. POLYAK, K. SCHEINBERG AND I. YUZEFOVICH, *A modified barrier-augmented Lagrangian method for constrained minimization*, Comput. Optim. Appl., 14 (1999), 55–74.
- [14] J. GONDZIO, *Interior point methods 25 years later*, Eur. J. Oper. Res. 218(2012), 587–601.
- [15] J. GONDZIO, *Matrix-free interior point method*, Comput. Optim. Appl. 51(2012), 457–480.
- [16] M.R. HESTENES, *Multiplier and gradient methods*. J. Optim. Theory Appl. 4(1969), 303–320.
- [17] M.Y. HONG AND Z.Q. LUO, *On the linear convergence of the alternating direction method of multipliers*, Math. Program., 162 (2017), 165–199.
- [18] B. HOUSKA, J. FRASCH AND M. DIEHL, *An augmented Lagrangian based algorithm for distributed nonconvex optimization*, SIAM J. Optim., 26(2016), 1101–1127.
- [19] N. KARMARKAR, *A new polynomial-time algorithm for linear programming*, Combinatorics, 4 (1984), 373–395.
- [20] V. KLEE AND G.J. MINTY, *How good is the simplex algorithm?* in Inequalities, O. Shisha, ed., Academic Press, New York, 1972, 159–175.
- [21] K. LANGE, *MM Optimization Algorithms*, SIAM, 2016.
- [22] T. LIN, S. MA, Y. YE AND S. ZHANG, *An ADMM-based interior-point method for large-scale linear programming*, Optim. Methods Soft., 36 (2021), 389–424.
- [23] X.-W. LIU AND Y.-H. DAI, *A globally convergent primal-dual interior-point relaxation method for nonlinear programs*, Math. Comput., 89 (2020), 1301–1329.
- [24] X.-W. LIU, Y.-H. DAI AND Y.-K. HUANG, *A primal-dual interior-point relaxation method with global and rapidly local convergence for nonlinear programs*, Optimization Online, April 2020. To appear in Math. Meth. Oper. Res.
- [25] X.-W. LIU, Y.-H. DAI, Y.-K. HUANG AND J. SUN, *A novel augmented Lagrangian method of multipliers for optimization with general inequality constraints*, arXiv:2106.15044 [math.OC], June 2021. To appear in Math. Comput.
- [26] X.-W. LIU AND J. SUN, *A robust primal-dual interior point algorithm for nonlinear programs*, SIAM J. Optim., 14 (2004), 1163–1186.

- [27] X.-W. LIU AND Y.-X. YUAN, *A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties*, Math. Program., 125(2010), 163–193.
- [28] G.J. MCLACHLAN AND T. KRISHNAN, *The EM Algorithm and Extensions*, 2nd ed. Wiley, Hoboken, NJ, 2008.
- [29] Y. NESTEROV, *Lectures on Convex Optimization*, 2nd ed., Berlin: Springer, 2018.
- [30] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [31] J.M. ORTEGA AND W.C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
- [32] M.J.D. POWELL, *A method for nonlinear constraints in minimization problems*. In Fletcher, R. (ed.) *Optimization*, pp. 283–298. Academic, New York (1972).
- [33] T. QIU, P. BABU AND D.P. PALOMAR, *PRIME: phase retrieval via majorization-minimization*, arXiv preprint:1511.01669, 2015.
- [34] R.T. ROCKAFELLAR, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Math. Program. 5 (1973), 354–373.
- [35] W. SUN AND Y. YUAN, *Optimization Theory and Methods: Nonlinear Programming*, Springer, New York, 2006.
- [36] S.J. WRIGHT, *Primal-Dual Interior-Point Methods*, SIAM Publications, Philadelphia, Pa, 1997.
- [37] Y. YE, *Interior Point Algorithms: Theory and Analysis*, John Wiley & Sons, Inc., 1997.
- [38] Z. WEN, D. GOLDFARB AND W. YIN, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Math. Prog. Comp., 2 (2010), 203–230.