



A strategy of global convergence for the affine scaling algorithm for convex semidefinite programming

Xun Qian¹ · Li-Zhi Liao¹ · Jie Sun^{2,3}

Received: 1 March 2017 / Accepted: 19 July 2018 / Published online: 28 July 2018

© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2018

Abstract

The affine scaling algorithm is one of the earliest interior point methods developed for linear programming. This algorithm is simple and elegant in terms of its geometric interpretation, but it is notoriously difficult to prove its convergence. It often requires additional restrictive conditions such as nondegeneracy, specific initial solutions, and/or small step length to guarantee its global convergence. This situation is made worse when it comes to applying the affine scaling idea to the solution of semidefinite optimization problems or more general convex optimization problems. In (Math Program 83(1–3):393–406, 1998), Muramatsu presented an example of linear semidefinite programming, for which the affine scaling algorithm with either short or long step converges to a non-optimal point. This paper aims at developing a strategy that guarantees the global convergence for the affine scaling algorithm in the context of linearly constrained convex semidefinite optimization in a least restrictive manner. We propose a new rule of step size, which is similar to the Armijo rule, and prove that such an affine scaling algorithm is globally convergent in the sense that each accumulation point of the sequence generated by the algorithm is an optimal solution as long as the optimal solution set is nonempty and bounded. The algorithm is least restrictive in the sense that it allows the problem to be degenerate and it may start from any interior feasible point.

Keywords Affine scaling · Interior point method · Convex semidefinite programming

Mathematics Subject Classification 90C25 · 90C51 · 37C75 · 37N40

The work of L.-Z. Liao was supported in part by grants from Hong Kong Baptist University (FRG) and General Research Fund (GRF) of Hong Kong. The work of J. Sun was partially supported by Australia Research Council under grant DP160102819.

Extended author information available on the last page of the article

1 Introduction

Let \mathcal{S}^n denote the vector space of real symmetric $n \times n$ matrices. The standard inner product on \mathcal{S}^n is

$$A \bullet B = \text{tr}(AB) = \sum_{i,j} A_{ij} B_{ij} \quad \text{and} \quad \text{tr}(\cdot) = \text{trace}(\cdot).$$

By $X \geq 0$ ($X \succ 0$), where $X \in \mathcal{S}^n$, we mean that X is positive semidefinite (positive definite). Consider the following convex semidefinite programming (SDP) problem

$$\begin{aligned} \min_{X \in \mathcal{S}^n} \quad & f(X) \\ \text{s.t.} \quad & A_k \bullet X = b_k, \quad k = 1, \dots, m, \\ & X \geq 0, \end{aligned} \tag{P}$$

where $f : \mathcal{S}^n \rightarrow \mathbb{R}$ is convex and continuously differentiable, $b \in \mathbb{R}^m$, and $A_k \in \mathcal{S}^n$, $k = 1, \dots, m$. As a blanket assumption, we assume that the optimal value for problem (P) is finite and attainable, therefore, we use \min rather than \inf in problem (P).

The following notations will be used in our subsequent discussions

$$\begin{aligned} \mathcal{R}_+^n &= \{x \in \mathbb{R}^n \mid x \geq 0\}, & \mathcal{R}_{++}^n &= \{x \in \mathbb{R}^n \mid x \succ 0\}, \\ \mathcal{S}_+^n &= \{X \in \mathcal{S}^n \mid X \geq 0\}, & \mathcal{S}_{++}^n &= \{X \in \mathcal{S}^n \mid X \succ 0\}, \\ \mathcal{P}^+ &= \{X \in \mathcal{S}^n \mid A_k \bullet X = b_k, \quad k = 1, \dots, m, \quad X \geq 0\}, \\ \mathcal{P}^{++} &= \{X \in \mathcal{S}^n \mid A_k \bullet X = b_k, \quad k = 1, \dots, m, \quad X \succ 0\}. \end{aligned}$$

A comprehensive study of SDP with linear objective function can be found in [35].

There are many interior point algorithms for solving problem (P), for example, [1,16,17,19,25,28,38] for $f(X)$ being linear, [12,21,30,31] for $f(X)$ being convex quadratic, and [36,37] for general nonlinear semidefinite programming. Some related continuous trajectories were studied for semidefinite programming, for instance, [8,10,11,13,15]. Most of interior point algorithms are primal–dual path-following algorithms that are extensions of primal–dual path-following algorithm for linear programming (LP). The affine scaling algorithm for LP was originally proposed by Dikin [5], and further discussed by Barnes [2], and Vanderbei et al. [33]. For more details on the development of the affine scaling algorithm, see [7,20], and the references therein. Unfortunately, the affine scaling algorithm for linear SDP with either the short or the long step version may fail [20], even though the affine scaling continuous trajectory, which is contained in Cauchy trajectories, converges to the optimal solution set of problem (P) [13].

In [20], Muramatsu gave an example of linear SDP such that the affine scaling algorithm converges to a non-optimal point. In that example, for both the short and the long step version of the affine scaling algorithm, there exists a region of starting points such that the generated sequence converges to a non-optimal point. In the concluding remarks of [20], Muramatsu pointed that it may still be possible to prove the global convergence from well-chosen starting points, or by allowing variable step sizes. In

this paper, we focus on the second strategy—allowing variable step sizes, and propose a new step size rule which is similar to the Armijo-type rule [3]. Under this new rule of step size, we can prove that starting from any interior feasible point, every accumulation point of the affine scaling algorithm is an optimal solution of problem (P). It should be noted that Renegar and Sondjaja studied a polynomial-time affine scaling method for linear semidefinite and hyperbolic programming in [22], where the method is actually a variant of Dikin's affine scaling method. The algorithm in [22] shares the similar spirit as Dikin's method in that at each iteration, the cone in the original optimization problem is replaced by an ellipsoidal cone centered at the current iterate, rather than an ellipsoid as in Dikin's method. The algorithm differs from Dikin's method in that at each iteration, the ellipsoidal cone is chosen to contain the original cone rather than to be contained by it.

In Sect. 5, we will consider a special case where X and A_i ($i = 1, \dots, m$) are all diagonal in (P) and with a slightly different but less restrictive step size rule. In this special case, problem (P) becomes linearly constrained convex programming which has been studied in [7] in the context of affine scaling algorithms and in [32] through a first-order interior point method, which includes the affine scaling algorithm as a special case. However, to ensure optimality, they both require nondegeneracy assumptions, and the strong convergence of the (first-order) affine scaling algorithm is still open. The line search procedure in [7] is quite general, in order to guarantee optimality, the step size in [32] needs to be bounded. The second-order affine scaling algorithm for linearly constrained convex programming has been studied in [18,27]. In [27], Sun proved the global convergence at a local linear rate under a Hessian similarity condition of the objective function, and an ϵ -optimal solution can be obtained if the step size is in the order of $O(\epsilon)$ without nondegeneracy assumptions. In [18], Monteiro and Wang studied a version of the second-order affine scaling algorithm in which a fraction of the ellipsoid used at each iteration is selected according to a trust region strategy for linearly constrained convex and concave programs. In the convex case of [18], optimality is obtained under the primal nondegeneracy assumption, and it was shown that the sequences of iterates and objective function values generated by the algorithm converge R-linearly and Q-linearly, respectively, under certain invariance condition on the Hessian of the objective function.

For simplicity, in what follows, we use $\|\cdot\|$ to denote either the 2-norm of vector or the spectral norm of matrix, and use $\|\cdot\|_F$ to denote the Frobenius norm of matrix. x_i denotes the i th component of a vector x , and I stands for the identity matrix whose dimension will be clear from the context. $\text{diag}(x)$ for the vector x denotes the diagonal matrix whose diagonal entries are that of x . For any $Q \in \mathcal{S}^n$, $\lambda_{\min}(Q)$ denotes the smallest eigenvalue of Q .

The definition of the affine scaling direction in linear semidefinite programming can be found in [6] or [20]. In the convex case, the affine scaling direction can be defined similarly. Similar to [20], the affine scaling direction for problem (P) can be obtained by first defining the associated dual estimate. For a point $X \in \mathcal{P}^{++}$, we define the dual estimate $(u(X), S(X)) \in \mathcal{R}^m \times \mathcal{S}^n$ as the unique solution of the following optimization problem

$$\begin{aligned} \min_{S \in \mathcal{S}^n, u \in \mathbb{R}^m} \quad & \|X^{\frac{1}{2}} S X^{\frac{1}{2}}\|_F^2 \\ \text{s.t.} \quad & S + \sum_{i=1}^m u_i A_i = \frac{\partial f}{\partial X}, \end{aligned} \quad (1)$$

where $X^{\frac{1}{2}} \in \mathcal{S}_{++}^n$ is the unique square root matrix of X . Using the KKT condition of problem (1), $(u(X), S(X))$ can be solved as (see (4) in [20])

$$u(X) = G(X)^{-1} p(X), \quad S(X) = \frac{\partial f}{\partial X} - \sum_{i=1}^m u_i(X) A_i, \quad (2)$$

where $G(X) \in \mathcal{S}^m$ and $p(X) \in \mathbb{R}^m$ are such that $G_{ij}(X) = \text{tr}(A_i X A_j X)$ and $p_j(X) = \text{tr}(A_j X \frac{\partial f}{\partial X} X)$ for $i, j = 1, 2, \dots, m$, respectively. Assumption 2 below guarantees that $G(X)$ is invertible (see Proposition 1 in [20]).

Then the affine scaling direction $D(X)$ for problem (P) can be defined as

$$D(X) = -X S(X) X = -X \left(\frac{\partial f}{\partial X} - \sum_{i=1}^m u_i(X) A_i \right) X. \quad (3)$$

The rest of this paper is organized as follows. In Sect. 2, we study some properties of the affine scaling direction. In Sect. 3, we propose an affine scaling algorithm with a new step size rule, which is similar to the Armijo-type rule. In Sect. 4, we prove that any accumulation point of the affine scaling algorithm with the new step size rule is an optimal solution of problem (P) for any starting interior feasible point without nondegeneracy assumptions. In Sect. 5, we consider a special case of problem (P) where X and A_i 's are diagonal. A slightly different step size rule, which is less restrictive, is proposed for the affine scaling algorithm. For any accumulation point of the resulting algorithm, optimality is obtained as well. In Sect. 6, the convergence of our affine scaling algorithm on the counter example in [20] and on some randomly generated linear SDP problems are illustrated. Finally, some concluding remarks are drawn in Sect. 7.

2 Properties of the affine scaling direction

The following assumptions are made throughout this paper.

Assumption 1 \mathcal{P}^{++} is nonempty.

Assumption 2 The matrix \mathcal{A} has full row rank m .

Assumption 3 The optimal solution set of problem (P) is nonempty and bounded.

Now we introduce another form of the affine scaling direction $D(X)$ in (3). Firstly we need the following notations.

- We define the map $\text{svec} : \mathcal{S}^n \rightarrow \mathbb{R}^{n(n+1)/2}$ as

$$\text{svec}(U) := (u_{11}, \sqrt{2}u_{21}, \dots, \sqrt{2}u_{n1}, u_{22}, \sqrt{2}u_{32}, \dots, \sqrt{2}u_{n2}, \dots, u_{nn})^T,$$

where $U \in \mathcal{S}^n$ and “T” is the transpose.

- The symmetrized Kronecker product \otimes is defined as

$$(G \otimes K)\text{svec}(H) = \frac{1}{2}\text{svec}(KHG^T + GHK^T),$$

where $G, K \in \mathbb{R}^{n \times n}$ and $H \in \mathcal{S}^n$. The properties of operator \otimes can be found in the Appendix of [1,24].

- Let matrix \mathcal{A} be defined as follows

$$\mathcal{A} = \begin{pmatrix} \text{svec}(A_1)^T \\ \vdots \\ \text{svec}(A_m)^T \end{pmatrix} \in \mathbb{R}^{m \times n(n+1)/2}.$$

From (2), we can rewrite $G_{ij}(X)$ and $p_j(X)$ as $\text{svec}(A_i)(X \otimes X)\text{svec}(A_j)$ and $\text{svec}(A_j)(X \otimes X)\text{svec}\left(\frac{\partial f}{\partial X}\right)$, respectively. Therefore, $G(X)$ and $p(X)$ can be denoted as $\mathcal{A}(X \otimes X)\mathcal{A}^T$ and $\mathcal{A}(X \otimes X)\text{svec}\left(\frac{\partial f}{\partial X}\right)$, respectively, consequently

$$u(X) = (\mathcal{A}(X \otimes X)\mathcal{A}^T)^{-1}\mathcal{A}(X \otimes X)\text{svec}\left(\frac{\partial f}{\partial X}\right), \tag{4}$$

and

$$\text{svec}(S(X)) = (I - P_{\mathcal{A}X}(X \otimes X))\text{svec}\left(\frac{\partial f}{\partial X}\right), \tag{5}$$

where we denote $P_{\mathcal{A}X} = \mathcal{A}^T(\mathcal{A}(X \otimes X)\mathcal{A}^T)^{-1}\mathcal{A}$ for simplicity. Now we can present another form of the affine scaling direction $D(X)$ as

$$\text{svec}(D(X)) = - (I - (X \otimes X)P_{\mathcal{A}X})(X \otimes X)\text{svec}\left(\frac{\partial f}{\partial X}\right). \tag{6}$$

Using the above notations, the affine scaling direction in (3) can be also derived from the following optimization problem

$$\begin{aligned} \min_{D \in \mathbb{R}^{n \times n}} \quad & \frac{\partial f}{\partial X} \bullet D \\ \text{s.t.} \quad & A_k \bullet D = 0, \quad k = 1, \dots, m, \\ & \|X^{-\frac{1}{2}}DX^{-\frac{1}{2}}\|_F^2 \leq \tilde{\beta}^2 < 1, \end{aligned} \tag{7}$$

where $X^{-\frac{1}{2}}$ is the inverse of $X^{\frac{1}{2}}$. In fact, if the current point $X \in \mathcal{P}^{++}$ is not an optimal solution of problem (P), then by the KKT condition of problem (7), it is not difficult to obtain the solution of problem (7) as

$$\text{svec}(D) = - \frac{\tilde{\beta} (I - (X \otimes X) P_{AX}) (X \otimes X) \text{svec} \left(\frac{\partial f}{\partial X} \right)}{\left\| \left(I - \left(X^{\frac{1}{2}} \otimes X^{\frac{1}{2}} \right) P_{AX} \left(X^{\frac{1}{2}} \otimes X^{\frac{1}{2}} \right) \right) \left(X^{\frac{1}{2}} \otimes X^{\frac{1}{2}} \right) \text{svec} \left(\frac{\partial f}{\partial X} \right) \right\|},$$

or equivalently

$$D = - \frac{\tilde{\beta} X S(X) X}{\|X^{\frac{1}{2}} S(X) X^{\frac{1}{2}}\|_F}.$$

We can see this D and the $D(X)$ in (3) represent the same direction.

For linear SDP, the properties of the affine scaling direction have been discussed in [20]. For convex SDP, these properties can be obtained similarly. The proof of Theorem 1 below can be obtained identically from the proof of Proposition 2 in [20] except replacing matrix C with $\frac{\partial f}{\partial X}$, hence the proof is omitted.

Theorem 1 *We have*

$$A_i \bullet D(X) = 0, \quad (8)$$

for all $i = 1, \dots, m$, and

$$\frac{\partial f}{\partial X} \bullet D(X) = - \|X^{\frac{1}{2}} S(X) X^{\frac{1}{2}}\|_F^2. \quad (9)$$

3 A new step size rule

It has been shown in [20] that the affine scaling algorithm for linear SDP can fail no matter it uses a long step strategy or a short step strategy. Hence here we design a new step size strategy which is similar to the Armijo-type rule [3]. For any initial point $X_0 \in \mathcal{P}^{++}$, the iterations in the affine scaling algorithm have the form

$$X_{k+1} = X_k + \alpha^k D(X_k), \quad k = 0, 1, \dots, \quad (10)$$

where $\alpha^k > 0$ is the step size and $D(X)$ is given in (3). In order to state our step size strategy, we first introduce some notations. Let

$$\rho(X) = \sup\{\rho > 0 \mid X + \rho D(X) \succ 0\},$$

for any $X \in \mathcal{S}_{++}^n$ and select a positive sequence $\{a_i\}_{i=0}^{+\infty}$ such that $\lim_{i \rightarrow +\infty} a_i = 0$ and $\lim_{s \rightarrow +\infty} \sum_{i=0}^s a_i = +\infty$. For instance, the sequence can be $\{\frac{1}{(i+1)^\alpha}\}_{i=0}^{+\infty}$ with $0 < \alpha \leq 1$ or $\{\frac{1}{\ln(i+2)}\}_{i=0}^{+\infty}$. Then α^k in (10) can be defined from the following two steps:

Step 1:

$$\alpha_0^k = \min \left\{ \frac{a_k}{\|S_k X_k\|_C (\|S_k\|)}, \tau \rho(X_k) \right\} > 0, \quad (11)$$

where $S_k = S(X_k)$, $0 < \tau < 1$ is a constant, and $c(x)$ is a scalar function which satisfies $c_1x \leq c(x) \leq \max(c_2x, c_3)$, where $0 < c_1 \leq c_2, c_3 > 0$ are constants.

Step 2: α^k is the largest $\alpha \in \{\alpha_0^k \beta^l\}_{l=0,1,\dots}$ satisfying

$$f(X_k + \alpha D(X_k)) \leq f(X_k) + \sigma \alpha G_k \bullet D(X_k), \tag{12}$$

where $G_k = \frac{\partial f}{\partial X}|_{X=X_k}$, $0 < \beta, \sigma < 1$ are constants.

It should be noticed that in (11) S_k should be a nonzero matrix. In fact, if $S_k = 0$, then it is easy to verify that X_k is actually an optimal solution from the KKT condition, and the iteration should stop, hence we will not consider this trivial case for brevity.

4 Optimality of the affine scaling algorithm

In this section, we will show that the affine scaling algorithm with our step size strategy (12) will succeed without nondegeneracy assumptions. We begin our discussions with the following lemmas.

Lemma 1 (Section 3.1.3, [4]) *Suppose f is differentiable (i.e., its gradient ∇f exists at each point in $\text{dom} f$). Then f is convex if and only if $\text{dom} f$ is convex and*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \tag{13}$$

holds for all $x, y \in \text{dom} f$.

According to [9], a vector β is said to majorize a vector α if

$$\min \left\{ \sum_{j=1}^k \beta_{i_j} : 1 \leq i_1 < \dots < i_k \leq n \right\} \geq \min \left\{ \sum_{j=1}^k \alpha_{i_j} : 1 \leq i_1 < \dots < i_k \leq n \right\},$$

for any $k = 1, 2, \dots, n$ with equality for $k = n$.

Lemma 2 (Theorem 4.3.26, [9]) *Let A be Hermitian. Then the vector composed of the diagonal entries of A majorizes the vector composed of the eigenvalues of A .*

Lemma 3 *The level set $\mathcal{F} = \{X \in \mathcal{P}^+ | f(X) \leq f(X_0)\}$ is bounded.*

Proof Let $\delta_{\mathcal{P}^+}(X)$ be the indicator function of \mathcal{P}^+ which is defined by

$$\delta_{\mathcal{P}^+}(X) = \begin{cases} 0 & \text{if } X \in \mathcal{P}^+, \\ +\infty & \text{Otherwise.} \end{cases}$$

Then $f(X) + \delta_{\mathcal{P}^+}(X)$ is a closed proper convex function, and the optimal solution set of problem (P) can be expressed as

$$\{X \in \mathcal{S}^n | f(X) + \delta_{\mathcal{P}^+}(X) \leq f^*\},$$

where f^* is the optimal objective value for problem (P). According to Assumption 3, the optimal solution set of problem (P) is nonempty and bounded, hence Corollary 8.7.1 in [23] implies that

$$\{X \in \mathcal{S}^n | f(X) + \delta_{\mathcal{P}^+}(X) \leq f(X_0)\} = \mathcal{F} = \{X \in \mathcal{P}^+ | f(X) \leq f(X_0)\} \quad (14)$$

is also bounded. \square

Theorem 2 *Let $\{X_k\}$ be generated by the affine scaling algorithm (10) with the step size $\{\alpha^k\}$ chosen by (12). Then*

- (i) $X_k \in \mathcal{P}^{++}$, $\{f(X_k)\}$ is nonincreasing, $\{X_k\}$ and $\{D(X_k)\}$ are bounded;
- (ii) every accumulation point of $\{X_k\}$ is an optimal solution for problem (P).

Proof Proof of (i). Since $X_0 \in \mathcal{P}^{++}$ and $\alpha^k \leq \tau\rho(X_k)$ at each step, by using an induction argument on k , we have that $X_k \in \mathcal{S}_{++}^n$ for all k . Moreover, from (8) in Theorem 1 and $X_0 \in \mathcal{P}^{++}$, we have $X_k \in \mathcal{P}^{++}$ for all k . Also, since $f(X)$ is continuously differentiable and $\alpha_0^k > 0$, $0 < \sigma < 1$ in (12), we know $\alpha^k > 0$ for all k . Combining (9) and (12), we have for all k

$$f(X_{k+1}) - f(X_k) \leq \sigma\alpha^k G_k \bullet D(X_k) = -\sigma\alpha^k \left\| X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right\|_F^2 \leq 0, \quad (15)$$

thus $\{f(X_k)\}$ is nonincreasing. Then $X_k \in \mathcal{F}$ for all k . Since the level set \mathcal{F} in (14) is bounded from Lemma 3, we know $\{X_k\}$ is bounded as well. For $D(X_k)$, since

$$\text{svec}(D(X_k)) = \left(X_k^{\frac{1}{2}} \otimes X_k^{\frac{1}{2}} \right) \mathcal{P}_k \left(X_k^{\frac{1}{2}} \otimes X_k^{\frac{1}{2}} \right) G_k,$$

where $\mathcal{P}_k = I - (X_k^{\frac{1}{2}} \otimes X_k^{\frac{1}{2}}) P_{AX_k} (X_k^{\frac{1}{2}} \otimes X_k^{\frac{1}{2}})$ is an idempotent matrix which implies $\|\mathcal{P}_k\| \leq 1$ for all k . Then along with the facts that $\{X_k\}$ is bounded and $f(X)$ is continuously differentiable, we have $\{D(X_k)\}$ is also bounded.

Proof of (ii). From (i), we have $\{X_k\}$ is bounded, hence $\{X_k\}$ must have at least one accumulation point. Let \bar{X} be any accumulation point of $\{X_k\}$, we will show it is actually an optimal solution for problem (P) by contradiction.

Assume \bar{X} is not an optimal solution of problem (P). First, since $X_k \in \mathcal{P}^{++}$, we have $\bar{X} \in \mathcal{P}^+$. From Assumption 3, we can choose a point $X^* \in \mathcal{P}^+$ such that X^* is an optimal solution for problem (P). According to the hypothesis and the fact that $\{f(X_k)\}$ is nonincreasing from (i), we have

$$f(X_0) \geq f(\bar{X}) = \lim_{k \rightarrow +\infty} f(X_k) > f(X^*).$$

Let us define

$$\bar{Y} = \frac{f(\bar{X}) - f(X^*)}{2(f(X_0) - f(X^*))} X_0 + \left[1 - \frac{f(\bar{X}) - f(X^*)}{2(f(X_0) - f(X^*))} \right] X^*,$$

then $\bar{Y} \in \mathcal{P}^{++}$. Since $f(X)$ is convex, we have

$$\begin{aligned} f(\bar{Y}) &\leq \frac{f(\bar{X}) - f(X^*)}{2(f(X_0) - f(X^*))} f(X_0) + \left[1 - \frac{f(\bar{X}) - f(X^*)}{2(f(X_0) - f(X^*))} \right] f(X^*) \\ &= \frac{f(\bar{X}) + f(X^*)}{2}. \end{aligned}$$

Let us define

$$V(X) = \ln \det X + \text{tr}(X^{-1}\bar{Y}),$$

where $X \in \mathcal{S}_{++}^n$. (Remark: The definition of $V(x)$ is inspired by the potential function in Losert and Akin [14].) Then at X_k , we can define a scalar function $V_k(\alpha)$ as

$$V_k(\alpha) = V(X_k + \alpha D(X_k)),$$

where $0 \leq \alpha \leq \alpha_0^k$. Obviously, $V_k(0) = V(X_k)$ and $V_k(\alpha^k) = V(X_{k+1})$. Moreover,

$$\begin{aligned} \frac{dV_k(\alpha)}{d\alpha} &= \text{tr}(X(\alpha)^{-1} D(X_k)) - \text{tr}(X(\alpha)^{-1} \bar{Y} X(\alpha)^{-1} D(X_k)) \\ &= \text{tr} \left[(X(\alpha) - \bar{Y}) X(\alpha)^{-1} D(X_k) X(\alpha)^{-1} \right] \\ &= \text{tr} \left[(\bar{Y} - X(\alpha)) X(\alpha)^{-1} X_k S_k X_k X(\alpha)^{-1} \right], \end{aligned}$$

where $X(\alpha) = X_k + \alpha D(X_k) = X_k - \alpha X_k S_k X_k$. Notice $X_k^{-1} X(\alpha) = I - \alpha S_k X_k$, which implies

$$X(\alpha)^{-1} X_k = I + \alpha S_k X_k (I - \alpha S_k X_k)^{-1}, \tag{16}$$

hence

$$\begin{aligned} \frac{dV_k(\alpha)}{d\alpha} &= \text{tr} \left[(\bar{Y} - X(\alpha)) X(\alpha)^{-1} X_k S_k X_k X(\alpha)^{-1} \right] \\ &= \text{tr} \left\{ (\bar{Y} - X(\alpha)) [I + \alpha S_k X_k W] S_k [I + \alpha W^T X_k S_k] \right\} \\ &= \text{tr} \left\{ (\bar{Y} - X(\alpha)) [S_k + \alpha S_k X_k W S_k] \right\} \\ &\quad + \text{tr} \left\{ (\bar{Y} - X(\alpha)) \left[\alpha S_k W^T X_k S_k + \alpha^2 S_k X_k W S_k W^T X_k S_k \right] \right\}, \end{aligned}$$

where $W = X(\alpha)^{-1} X_k = (I - \alpha S_k X_k)^{-1}$. Next we show that when $0 \leq \alpha \leq \alpha^k$, $\frac{dV_k(\alpha)}{d\alpha}$ is always negative if k is large enough. Since the level set \mathcal{F} in (14) is bounded and $f(X)$ is continuously differentiable, there exist $M_1, M_2 > 0$ such that $\|X\| \leq M_1, \|\frac{\partial f}{\partial X}\| \leq M_2$ if $X \in \mathcal{F}$. From Theorem 1 and Lemma 1, we have

$$\begin{aligned} \text{tr}((\bar{Y} - X_k) S_k) &= \text{tr}((\bar{Y} - X_k) G_k) \leq f(\bar{Y}) - f(X_k) \\ &\leq f(\bar{Y}) - f(\bar{X}) \leq \frac{f(X^*) - f(\bar{X})}{2} < 0, \end{aligned}$$

which implies

$$\frac{f(\bar{X}) - f(X^*)}{2} \leq \text{tr}((X_k - \bar{Y})S_k) \leq \|X_k - \bar{Y}\| \cdot \|S_k\| \leq 2M_1 \|S_k\|,$$

thus

$$\|S_k\| \geq \frac{f(\bar{X}) - f(X^*)}{4M_1}, \quad (17)$$

for all k . Then when $0 \leq \alpha \leq \alpha^k$ we have

$$\begin{aligned} \text{tr}((\bar{Y} - X(\alpha))S_k) &= \text{tr}[(\bar{Y} - X_k + \alpha X_k S_k X_k)G_k] \\ &\leq \frac{f(X^*) - f(\bar{X})}{2} + \alpha M_1 M_2 \|S_k X_k\| \\ &\leq \frac{f(X^*) - f(\bar{X})}{2} + M_1 M_2 \frac{a_k}{c_1 \|S_k\|} \\ &\leq \frac{f(X^*) - f(\bar{X})}{2} + \frac{4M_1^2 M_2 a_k}{c_1 (f(\bar{X}) - f(X^*))}. \end{aligned}$$

From (16), if $0 < a_k < \frac{c_1(f(\bar{X})-f(X^*))}{4M_1}$ and $0 \leq \alpha \leq \alpha^k$, we have

$$\|W\| \leq 1 + \alpha \|S_k X_k\| \cdot \|W\| \leq 1 + \frac{4M_1 a_k}{c_1 (f(\bar{X}) - f(X^*))} \|W\|,$$

which indicates that

$$\|W^T\| = \|W\| \leq \frac{1}{1 - \frac{4M_1 a_k}{c_1 (f(\bar{X}) - f(X^*))}}. \quad (18)$$

Therefore if $0 < a_k < \frac{c_1(f(\bar{X})-f(X^*))}{4M_1}$ and $0 \leq \alpha \leq \alpha^k$, we have

$$\begin{aligned} \frac{dV_k(\alpha)}{d\alpha} &\leq \frac{f(X^*) - f(\bar{X})}{2} + \frac{4M_1^2 M_2 a_k}{c_1 (f(\bar{X}) - f(X^*))} + 4\alpha M_1 \|S_k X_k\| \cdot \|W\| \cdot \|S_k\| \\ &\quad + 2\alpha^2 M_1 \|S_k X_k\|^2 \cdot \|W\|^2 \cdot \|S_k\| \\ &\leq \frac{f(X^*) - f(\bar{X})}{2} + \frac{4M_1^2 M_2 a_k}{c_1 (f(\bar{X}) - f(X^*))} + \frac{4M_1 a_k \|W\|}{c_1} + \frac{2M_1 a_k^2 \|W\|^2}{c_1^2 \|S_k\|}, \end{aligned}$$

then from (17), (18), and the fact that $\lim_{k \rightarrow +\infty} a_k = 0$, we know there exists a $K > 0$ such that for all $k \geq K$, if $0 \leq \alpha \leq \alpha^k$, then $\frac{dV_k(\alpha)}{d\alpha} < 0$. Especially, we have

$$V(X_{k+1}) = V_k(\alpha^k) < V_k(0) = V(X_k),$$

for all $k \geq K$. Hence there exists an $M_3 \in R$ such that $V(X_k) \leq M_3$ for all k . When $X \in S_{++}^n$, let $X = Q\Lambda Q^T$ be an eigenvalue decomposition of X , and $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of X . Then

$$V(X) = \ln \det X + \text{tr}(Q\Lambda^{-1}Q^T\bar{Y}) = \sum_{i=1}^n \ln \lambda_i + \text{tr}(\Lambda^{-1}Q^T\bar{Y}Q),$$

since $\bar{Y} \in \mathcal{P}^{++}$, we have

$$\lambda_{\min}(Q^T\bar{Y}Q) = \lambda_{\min}(\bar{Y}) > 0.$$

Therefore from Lemma 2, we have

$$\begin{aligned} V(X) &= \sum_{i=1}^n \ln \lambda_i + \text{tr}(\Lambda^{-1}Q^T\bar{Y}Q) \\ &\geq \sum_{i=1}^n \ln \lambda_i + \sum_{i=1}^n \lambda_i^{-1} \lambda_{\min}(\bar{Y}) \\ &= \sum_{i=1}^n (\ln \lambda_i + \lambda_i^{-1} \lambda_{\min}(\bar{Y})). \end{aligned}$$

For each i , $\ln \lambda_i + \lambda_i^{-1} \lambda_{\min}(\bar{Y}) \geq \ln \lambda_{\min}(\bar{Y}) + 1$ and $\lim_{\lambda_i \rightarrow 0} [\ln \lambda_i + \lambda_i^{-1} \lambda_{\min}(\bar{Y})] = +\infty$. Thus, by $V(X_k) \leq M_3$ for all k , we know there exists an $M_4 > 0$ such that for all k , we have

$$\lambda_{\min}(X_k) \geq M_4 > 0.$$

Let us define

$$\mathcal{H} = \{X \in S_{++}^n \mid \lambda_{\min}(X) \geq M_4\}.$$

Then for all k , $X_k \in \mathcal{H} \cap \mathcal{F}$ which implies $\|X_k^{-\frac{1}{2}}\| \leq \frac{1}{\sqrt{M_4}}$. Along with (17), we get for all k ,

$$\left\| X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right\|_F \geq \left\| X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right\| \geq \frac{\|S_k\|}{\|X_k^{-\frac{1}{2}}\|^2} \geq \frac{M_4(f(\bar{X}) - f(X^*))}{4M_1} > 0. \tag{19}$$

From (15), (19), and $\lim_{k \rightarrow +\infty} f(X_{k+1}) - f(X_k) = 0$, we know $\lim_{k \rightarrow +\infty} \alpha^k = 0$. Next we show that the index set

$$\mathcal{I} = \{k \mid \alpha^k = \alpha_0^k \beta^l, l \geq 1 \text{ in (12)}\}$$

is finite. If not, then we can choose a subsequence $\{X_k\}_{k \in \mathcal{K}}$ ($\mathcal{K} \subseteq \{0, 1, \dots\}$) such that $\mathcal{K} \subseteq \mathcal{I}$ and $\lim_{k \in \mathcal{K}, k \rightarrow +\infty} X_k = \tilde{X} \in \mathcal{H} \cap \mathcal{F}$. Then for $k \in \mathcal{K}$, the condition (12) is violated by $\alpha = \alpha^k/\beta$, i.e.,

$$\frac{f\left(X_k + \frac{\alpha^k}{\beta} D(X_k)\right) - f(X_k)}{\frac{\alpha^k}{\beta}} > \sigma G_k \bullet D(X_k). \tag{20}$$

Since $\lim_{k \rightarrow +\infty} \alpha^k = 0$ and $f(X)$ is continuously differentiable, from (20) we have $\tilde{G} \bullet D(\tilde{X}) \geq \sigma \tilde{G} \bullet D(\tilde{X})$ where $\tilde{G} = \frac{\partial f}{\partial X}|_{X=\tilde{X}}$, which implies

$$- \left\| \tilde{X}^{\frac{1}{2}} S(\tilde{X}) \tilde{X}^{\frac{1}{2}} \right\|_F^2 = \tilde{G} \bullet D(\tilde{X}) \geq 0,$$

but this contradicts with (19). Thus the index set \mathcal{I} is finite and there must exist an $N_1 > 0$ such that for all $k \geq N_1$, $\alpha^k = \alpha_0^k$ in (12). From (4), we know $\|u(X)\|$ is a continuous function on $\mathcal{H} \cap \mathcal{F}$ which is closed and bounded, thus $\|u(X)\|$ will be bounded on $\mathcal{H} \cap \mathcal{F}$. Along with (2) and (17), there exists an $M_5 > 0$ such that for all k , $\|S_k\| \leq M_5$ and

$$\frac{c_1(f(\tilde{X}) - f(X^*))}{4M_1} \leq c(\|S_k\|) \leq M_5. \tag{21}$$

Since $\lim_{k \rightarrow +\infty} a_k = 0$, there exists an $N_2 > 0$ such that for all $k \geq N_2$, we have $a_k < \frac{\tau c_1(f(\tilde{X}) - f(X^*))}{4M_1}$. Then for $k \geq N_2$, by using Theorem 1.3.20 in [9] and (21), for $\alpha = \frac{a_k}{\tau \|S_k X_k\| c(\|S_k\|)}$, we have

$$\begin{aligned} \left\| \alpha X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right\| &= \alpha r\left(X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\right) = \alpha r(S_k X_k) \\ &\leq \alpha \|S_k X_k\| = \frac{a_k}{\tau c(\|S_k\|)} \leq \frac{4M_1 a_k}{\tau c_1(f(\tilde{X}) - f(X^*))} < 1, \end{aligned}$$

where $r(A)$ denotes the spectral radius of matrix A , this implies

$$X_k + \alpha D(X_k) = X_k^{\frac{1}{2}} \left(I - \alpha X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right) X_k^{\frac{1}{2}} \in \mathcal{S}_{++}^n,$$

therefore $\alpha \leq \rho(X_k)$ and then $\frac{a_k}{\|S_k X_k\| c(\|S_k\|)} \leq \tau \rho(X_k)$. Thus for all $k \geq N_2$, by (21) we have

$$\alpha_0^k = \frac{a_k}{\|S_k X_k\| c(\|S_k\|)} \geq \frac{a_k}{\|S_k\| \cdot \|X_k\| c(\|S_k\|)} \geq \frac{a_k}{M_1 M_5^2}. \tag{22}$$

Combining (15), (19), and (22), we know for all $k \geq N_3 = \max(N_1, N_2)$,

$$f(X_{k+1}) - f(X_k) \leq -\sigma \alpha_0^k \frac{M_4^2 (f(\tilde{X}) - f(X^*))^2}{16M_1^2} \leq -\frac{\sigma M_4^2 (f(\tilde{X}) - f(X^*))^2}{16M_1^3 M_5^2} a_k,$$

this implies

$$f(\bar{X}) - f(X_{N_3}) \leq -\frac{\sigma M_4^2 (f(\bar{X}) - f(X^*))^2}{16M_1^3 M_5^2} \sum_{k \geq N_3}^{+\infty} a_k = -\infty,$$

which is a contradiction. Therefore any accumulation point of $\{X_k\}$ is an optimal solution for problem (P). □

5 A special case of problem (P)

In this section, we consider a special case of problem (P) where X and A_i ($i = 1, \dots, m$) are all diagonal. The results of this section are therefore applicable to the linearly constrained convex programming. We will show that in this special case, the step size can be larger in (12) in the sense that the positive sequence $\{a_i\}_{i=0}^{+\infty}$ is not required. If $X = \text{diag}(x)$, where $x \in R_{++}^n$, then $S(X)$, $D(X)$, and $\frac{\partial f}{\partial X}$ are all diagonal matrices. Let $A \in R^{m \times n}$ such that $A_{ij} = (A_i)_{jj}$ and $\nabla f \in R^n$ such that $(\nabla f)_i = (\frac{\partial f}{\partial X})_{ii}$. Then from (4), $u(X)$ can be denoted as

$$u(X) = (AX^2A^T)^{-1}AX^2\nabla f. \tag{23}$$

In this special case, the step size α^k is also chosen by (12) but α_0^k is defined as

$$0 < \alpha_0^k = \min \left\{ \frac{c_5}{\|S_k X_k\|^{c_4}}, \tau \rho(X_k) \right\}, \tag{24}$$

where $0 \leq c_4 < 1, c_5 > 0$ are constants.

Theorem 3 *Let $\{X_k\}$ be generated by the affine scaling algorithm (10) with the step size $\{\alpha^k\}$ chosen by (12) and (24). Then*

- (i) $X_k \in \mathcal{P}^{++}$, $\{f(X_k)\}$ is nonincreasing, $\{X_k\}$ and $\{D(X_k)\}$ are bounded;
- (ii) $\lim_{k \rightarrow +\infty} \|X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\|_F = 0$;
- (iii) every accumulation point of $\{X_k\}$ is an optimal solution for problem (P).

Proof Proof of (i). Similar to the proof of (i) in Theorem 2.

Proof of (ii). We prove this by contradiction. Assume $\lim_{k \rightarrow +\infty} \|X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\|_F = 0$ is not true, then since $\{X_k\}, \{D(X_k)\}$ are both bounded, there must exist a subsequence $\{X_k\}_{k \in \mathcal{K}}$ ($\mathcal{K} \subseteq \{0, 1, \dots\}$) and an $\bar{M}_1 > 0$ such that $\lim_{k \in \mathcal{K}, k \rightarrow +\infty} \|X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\|_F = \bar{M}_1, \lim_{k \in \mathcal{K}, k \rightarrow +\infty} X_k = \hat{X}$, and $\lim_{k \in \mathcal{K}, k \rightarrow +\infty} D(X_k) = \hat{D}$. Then from (15) and $\lim_{k \rightarrow +\infty} f(X_{k+1}) - f(X_k) = 0$, we know

$$\lim_{k \in \mathcal{K}, k \rightarrow +\infty} \alpha^k = 0. \tag{25}$$

From Lemma 3 and the Remark in Sun [26], we know that if $x > 0$, then every entry of $(AX^2A^T)^{-1}AX^2$ is bounded, and the bound depends only on A and n . Thus from (23), we know that $u(X_k)$ is bounded which implies that $S(X_k)$ is also bounded. Hence there exists an $\bar{M}_2 > 0$ such that $\|S_k X_k\| \leq \bar{M}_2$ for all k . If $\alpha = \frac{1}{2\bar{M}_2}$, then

$$\left\| \alpha X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right\| = \alpha \|S_k X_k\| < 1,$$

which implies

$$X_k + \alpha D(X_k) = X_k^{\frac{1}{2}} \left(I - \alpha X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}} \right) X_k^{\frac{1}{2}} \in \mathcal{S}_{++}^n,$$

therefore $\rho(X_k) \geq \frac{1}{2\bar{M}_2}$ for all k . Let $\bar{M}_3 = \min(\frac{c_5}{\bar{M}_2^{c_4}}, \tau \frac{1}{2\bar{M}_2})$. Then from (24), we have $\alpha_0^k \geq \bar{M}_3 > 0$. Hence from (25), we know for all $k \in \mathcal{K}$ sufficiently large, $\alpha^k < \alpha_0^k$, which implies that condition (12) is violated by $\alpha = \alpha^k/\beta$, i.e.,

$$\frac{f\left(X_k + \frac{\alpha^k}{\beta} D(X_k)\right) - f(X_k)}{\frac{\alpha^k}{\beta}} > \sigma G_k \bullet D(X_k). \tag{26}$$

Since $\lim_{k \in \mathcal{K}, k \rightarrow +\infty} \alpha^k = 0$ and $f(X)$ is continuously differentiable, from (26) we have $\hat{G} \bullet \hat{D} \geq \sigma \hat{G} \bullet \hat{D}$ where $\hat{G} = \frac{\partial f}{\partial X}|_{X=\hat{X}}$, which implies

$$\hat{G} \bullet \hat{D} = 0,$$

but this contradicts with $\lim_{k \in \mathcal{K}, k \rightarrow +\infty} \|X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\|_F = \bar{M}_1 > 0$. Hence the hypothesis is not true, and $\lim_{k \rightarrow +\infty} \|X_k^{\frac{1}{2}} S_k X_k^{\frac{1}{2}}\|_F = 0$.

Proof of (iii). Similar to Theorem 2, we also prove this by contradiction. From (ii), we have

$$\begin{aligned} 0 \leq \liminf_{k \rightarrow +\infty} \alpha^k \|S_k X_k\| &\leq \limsup_{k \rightarrow +\infty} \alpha^k \|S_k X_k\| \\ &\leq \limsup_{k \rightarrow +\infty} \alpha_0^k \|S_k X_k\| \leq \limsup_{k \rightarrow +\infty} c_5 \|S_k X_k\|^{1-c_4} = 0, \end{aligned}$$

which implies $\lim_{k \rightarrow +\infty} \alpha^k \|S_k X_k\| = 0$. Moreover, $\|S_k\|$ is also bounded, hence similar to the proof of Theorem 2, we can also get (19) ($M_1, M_4 > 0$ may be different), but this contradicts with (ii). Hence every accumulation point of $\{X_k\}$ is an optimal solution for problem (P). Thus the proof is complete. \square

6 Numerical experiments

In this section, we provide computational results of our affine scaling algorithm on the counter example in [20] and on some randomly generated linear SDP problems. The algorithm is implemented in Matlab (version 2017a) on a PC, and it should be mentioned that the main goal of this paper is on the theoretical aspect, hence the program is coded only for the demonstration purpose. In our algorithm, for linear SDP, $\alpha^k = \alpha_0^k$, and for simplicity, we set $c_1 = c_2 = c_3 = 1$ and $c(\|S_k\|) = \|S_k\|$ in the following numerical experiments.

The example in [20] is the following SDP problem:

$$\min_{X \in \mathcal{S}^n} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bullet X \quad \text{s.t.} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bullet X = 2, \quad X \succeq 0. \tag{27}$$

The equality constraint in problem (27) implies that $X = \begin{pmatrix} x & 1 \\ 1 & y \end{pmatrix}$ for $x, y \in \mathbb{R}$. In fact, problem (27) has the following equivalent form (see (18) in [20]):

$$\min_{x, y \in \mathbb{R}} x + y \quad \text{s.t.} \quad x \geq 0, \quad y \geq 0, \quad xy \geq 1. \tag{28}$$

More details related to problem (27) can be found in Sect. 3 in [20]. From (24)-(26) in [20], we have

$$u(X) = \frac{x + y}{xy + 1}, \quad S(X) = \begin{pmatrix} 1 & -u(X) \\ -u(X) & 1 \end{pmatrix},$$

and

$$D(X) = -\frac{xy - 1}{xy + 1} \begin{pmatrix} x^2 - 1 & 0 \\ 0 & y^2 - 1 \end{pmatrix}.$$

The initial point (x^0, y^0) in [20] is chosen from the set

$$\mathcal{L} = \{(x, y) | xy > 1, \quad x < 1, \quad y > 1\},$$

and in this situation, from Proposition 3 and (32) in [20], we can obtain

$$\rho(X) = \frac{xy + 1}{xy - 1} \cdot \frac{-(x + y) + \sqrt{(x + y)^2 + 4R(x, y)}}{2R(x, y)},$$

where $R(x, y) = \frac{(1-x^2)(y^2-1)}{xy-1}$. In the numerical tests, we also choose the initial point (x^0, y^0) from \mathcal{L} . In particular, we choose an x^0 in $(0, 1)$ and set $\epsilon(x^0) = \frac{1-x^0}{2}$. Finally, we choose a $y^0 > 1$ such that

$$\sqrt{x^0 y^0 - 1} < \frac{\tau^2 \epsilon(x^0) (1 - \epsilon(x^0) - x^0)}{2}.$$

Table 1 Numerical results for problem (27) with $\tau = 0.5$

x^0	$\epsilon(x^0)$	$x^0 y^0 - 1$	<i>Iter_{es}</i>	<i>Iter</i>	x_f
0.2	0.4	3.9×10^{-4}	11,865	144,367	0.215
0.4	0.3	1.2×10^{-4}	8672	215,459	0.416
0.6	0.2	2.4×10^{-5}	13,429	550,369	0.611
0.8	0.1	1.5×10^{-6}	57,513	5,078,004	0.804

The values of $|x + y - 2|$ are not listed since they are all less than 10^{-6}

Table 2 Numerical results for problem (27) with $\tau = 0.9$

x^0	$\epsilon(x^0)$	$x^0 y^0 - 1$	<i>Iter_{es}</i>	<i>Iter</i>	x_f
0.2	0.4	4.0×10^{-3}	2298	18,145	0.220
0.4	0.3	1.3×10^{-3}	1156	16,234	0.423
0.6	0.2	2.5×10^{-4}	1361	26,059	0.615
0.8	0.1	1.5×10^{-5}	3484	92,600	0.805

The values of $|x + y - 2|$ are not listed since they are all less than 10^{-6}

From Theorem 7 in [20], the limit point of the iterations $\{(x^k, y^k)\}$ will be contained in the closure of $\mathcal{L}_{\epsilon(x^0)} = \{(x, y) \in \mathcal{L} | x < 1 - \epsilon(x^0)\}$ if $\alpha^k = \tau\rho(X_k)$ at each iteration. In the numerical tests, we set $a_k = \frac{1}{\sqrt{x^0 y^0 - 1 \cdot \ln(2+k)}}$ to make a_0 and $\rho(X_0)$ have the same magnitude.

The following two tables list the numerical results of our algorithm for problem (27) with different τ values and initial points. The unique solution of problem (28) is (1, 1), and the stopping criterion of the program is $|x + y - 2| < 10^{-6}$. *Iter_{es}* denotes the iteration number when (x^k, y^k) escaped from $\mathcal{L}_{\epsilon(x^0)}$ the first time, and *Iter* denotes the total iteration number when the algorithm stopped. We also report the behaviors of the long step affine scaling algorithm in [20]. We set $\alpha^k = \tau\rho(X_k)$ at each iteration and use x_f to denote the value of x after 10^7 iterations.

From Tables 1 and 2, we can see for the long step affine scaling algorithm where $\alpha^k = \tau\rho(X_k)$ at each iteration, the numerical results confirm the findings of Theorem 7 in [20], and even after 1.0×10^7 iterations, (x^k, y^k) is still contained in $\mathcal{L}_{\epsilon(x^0)}$. However, by adopting our new step size rule, we can obtain the optimal solution within 10^{-6} accuracy successfully in all cases.

Next we report the numerical results of our affine scaling algorithm on some randomly generated linear SDP problems. Let $C \bullet X$ be the objective function of randomly generated linear SDP. Then each element of C and A_i ($i = 2, \dots, m$) is chosen randomly and uniformly in $[-1, 1]$, while A_1 is set at the identity matrix I . Finally, we set $b_i = A_i \bullet I$ ($i = 1, \dots, m$). In this way, I will be an interior feasible point for problem (P), and we use I as the initial point in our numerical experiments. These randomly generated SDP problems are actually the same as the ones in Section 4.1 of [34].

In our tests, after a linear SDP problem is generated, we firstly use SDPT3 [29] to obtain the optimal objective function value f^* with the default accuracy tolerance of 10^{-8} . Then we run our program from the initial point $X_0 = I$, and the stopping

Table 3 Numerical results for randomly generated linear SDPs

m	n	CPU (s)	$Iter$	$\frac{ f(X_k) - f^* }{ f^* + 1}$	$\ Asvec(X_k) - b\ $
20	20	0.030	32.25	8.00×10^{-7}	4.11×10^{-8}
40	20	0.032	24.35	7.72×10^{-7}	2.95×10^{-8}
80	20	0.046	23.00	7.20×10^{-7}	3.04×10^{-8}
200	40	0.234	23.40	7.25×10^{-7}	1.26×10^{-7}
200	100	1.334	24.85	6.84×10^{-7}	1.06×10^{-6}
400	40	0.513	23.00	8.18×10^{-7}	1.07×10^{-7}
400	100	2.726	24.05	7.90×10^{-7}	8.77×10^{-7}
800	40	1.336	22.80	7.50×10^{-7}	4.17×10^{-7}
800	100	6.254	23.95	6.32×10^{-7}	1.07×10^{-6}
800	200	28.045	25.05	6.39×10^{-7}	5.52×10^{-6}

criterion is set as $\frac{|f(X_k) - f^*|}{|f^*| + 1} < 10^{-6}$. For each (m, n) pair, we run the program for 20 times, and each number reported in Table 3 is the average of 20 runs. In all of our tests, we set $a_k = \frac{mn}{\ln(k+2)}$ and $\tau = 0.5$.

7 Concluding remarks

In this paper, an affine scaling algorithm with a new step size rule for linearly constrained convex semidefinite programming is proposed. It is proven that, starting from any feasible interior point, the accumulation points of the sequence generated by this algorithm are optimal solutions if the optimal solution set is nonempty and bounded. This global convergence does not depend on nondegeneracy assumptions. This confirms our conjecture that by just selecting a suitable step size at each iteration, the affine scaling algorithm can be made convergent even if it is applied to semidefinite programming, which includes linearly constrained convex programming as a special case.

A possible future topic of research is the convergence of the generated sequence, which we call strong convergence. Even in the special case where X is diagonal, as far as we know, the strong convergence of the affine scaling algorithm for convex programming has not been resolved. Tseng et al. [32] showed that for some first-order interior point method, the strong convergence can be obtained for the quadratic case, but not for the affine scaling algorithm. It would be of theoretical interest to know if this is true if we allow more flexibility in selecting the step size or initial points.

Acknowledgements The authors would like to thank the two anonymous referees for their constructive comments and suggestions on the earlier version of this paper. We really appreciate their valuable inputs.

References

1. Alizadeh, F., Haeberly, J.P.A., Overton, M.L.: Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J. Optim.* **8**, 746–768 (1998)

2. Barnes, E.R.: A variation on Karmarkar's algorithm for solving linear programming problems. *Math. Program.* **36**, 174–182 (1986)
3. Bertsekas, D.P.: *Nonlinear Programming*, 2nd edn. Athena Scientific, Belmont (1999)
4. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
5. Dikin, I.I.: Iterative solution of problems of linear and quadratic programming. *Sov. Math. Dokl.* **8**, 674–675 (1967). (in Russian)
6. Faybusovich, L.: Dikin's algorithm for matrix linear programming problems. In: *System Modelling and Optimization*, Springer, pp. 237–247 (1994)
7. Gonzaga, C.C., Carlos, L.A.: A primal affine scaling algorithm for linearly constrained convex programs. Tech. Report ES-238/90, Department of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, Rio de Janeiro (1990)
8. Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F.: On the central path for nonlinear semidefinite programming. *RAIRO Oper. Res.* **34**, 331–345 (2000)
9. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (2012)
10. Halická, M., de Klerk, E., Roos, C.: On the convergence of the central path in semidefinite optimization. *SIAM J. Optim.* **12**, 1090–1099 (2002)
11. Halická, M., Trnovská, M.: Limiting behavior and analyticity of weighted central paths in semidefinite programming. *Optim. Methods Softw.* **25**, 247–262 (2010)
12. Li, L., Toh, K.C.: A polynomial-time inexact primal-dual infeasible path-following algorithm for convex quadratic SDP. *Pac. J. Optim.* **7**(1), 43–61 (2011)
13. López, J., Ramírez, H.C.: On the central paths and cauchy trajectories in semidefinite programming. *Kybernetika* **46**, 524–535 (2010)
14. Losert, V., Akin, E.: Dynamics of games and genes: discrete versus continuous time. *J. Math. Biol.* **17**, 241–251 (1983)
15. Lu, Z., Monteiro, R.D.C.: Limiting behavior of the Alizadeh–Haerberly–Overton weighted paths in semidefinite programming. *Optim. Methods Softw.* **22**, 849–870 (2007)
16. Luo, Z.Q., Sturm, J.F., Zhang, S.Z.: Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. *SIAM J. Optim.* **8**(1), 59–81 (1998)
17. Monteiro, R.D.C.: Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.* **7**, 663–678 (1997)
18. Monteiro, R.D.C., Wang, Y.: Trust region affine scaling algorithms for linearly constrained convex and concave programs. *Math. Program.* **80**(3), 283–313 (1998)
19. Monteiro, R.D.C., Zhang, Y.: A unified analysis for a class of long-step primal–dual path-following interior-point algorithms for semidefinite programming. *Math. Program.* **81**(3), 281–299 (1998)
20. Muramatsu, M.: Affine scaling algorithm fails for semidefinite programming. *Math. Program.* **83**(1–3), 393–406 (1998)
21. Nie, J.W., Yuan, Y.X.: A predictor-corrector algorithm for QSDP combining Dikin-type and Newton centering steps. *Ann. Oper. Res.* **103**, 115–133 (2001)
22. Renegar, J., Sondjaja, M.: A polynomial-time affine-scaling method for semidefinite and hyperbolic programming [arXiv: 1410.6734](https://arxiv.org/abs/1410.6734) (2014)
23. Rockafellar, R.T.: *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton (1970)
24. Sim, C.-K., Zhao, G.: Asymptotic behavior of HKM paths in interior point methods for monotone semidefinite linear complementarity problems: general theory. *J. Optim. Theory Appl.* **137**, 11–25 (2008)
25. Sturm, J.F., Zhang, S.: Symmetric primal–dual path-following algorithms for semidefinite programming. *Appl. Numer. Math.* **29**(3), 301–315 (1999)
26. Sun, J.: A convergence proof for an affine scaling algorithm for convex quadratic programming without nondegeneracy assumptions. *Math. Program.* **60**, 69–79 (1993)
27. Sun, J.: A convergence analysis for a convex version of Dikin's algorithm. *Ann. Oper. Res.* **62**(1), 357–374 (1996)
28. Todd, M.J., Toh, K.C., Tütüncü, R.H.: On the Nesterov–Todd direction in semidefinite programming. *SIAM J. Optim.* **8**, 769–796 (1998)
29. Toh, K.C., Todd, M.J., Tütüncü, R.H.: SDPT3a MATLAB software package for semidefinite programming, version 1.3. *Optim. Methods Softw.* **11**(1–4), 545–581 (1999)
30. Toh, K.C.: An inexact primal–dual path following algorithm for convex quadratic SDP. *Math. Program.* **112**(1), 221–254 (2008)

31. Toh, K.C., Tütüncü, R.H., Todd, M.J.: Inexact primal–dual path-following algorithms for a special class of convex quadratic SDP and related problems. *Pac. J. Optim.* **3**(1), 135–164 (2007)
32. Tseng, P., Bomze, I.M., Schachinger, W.: A first-order interior point method for linearly constrained smooth optimization. *Math. Program.* **127**, 399–424 (2011)
33. Vanderbei, R.J., Meketon, M.S., Freedman, B.A.: A modification of Karmarkar’s linear programming algorithm. *Algorithmica* **1**, 395–407 (1986)
34. Yamashita, M., Fujisawa, K., Kojima, M.: Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). *Optim. Methods Softw.* **18**(4), 491–505 (2003)
35. Wolkowicz, H., Saigal, R., Vandenberghe, L.: *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, vol. 27. Springer, Berlin (2012)
36. Yamashita, H., Yabe, H.: Local and superlinear convergence of a primal–dual interior point method for nonlinear semidefinite programming. *Math. Program.* **132**(1–2), 1–30 (2012)
37. Yamashita, H., Yabe, H., Harada, K.: A primalCdual interior point method for nonlinear semidefinite programming. *Math. Program.* **135**(1–2), 89–121 (2012)
38. Zhang, Y.: On extending some primal–dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. Optim.* **8**, 365–386 (1998)

Affiliations

Xun Qian¹ · Li-Zhi Liao¹  · Jie Sun^{2,3}

✉ Li-Zhi Liao
liliao@hkbu.edu.hk

Xun Qian
13479784@life.hkbu.edu.hk

Jie Sun
jie.sun@curtin.edu.au

¹ Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, People’s Republic of China

² Department of Mathematics and Statistics, Curtin University, Perth, Australia

³ School of Business, National University of Singapore, Singapore, Singapore