Mathematical Programming 60 (1993) 69–79
North-Holland

A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions

Jie Sun

Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA

Received 15 August 1990
Revised manuscript received 28 November 1991

This paper presents a theoretical result on convergence of a primal affine-scaling method for convex quadratic programs. It is shown that, as long as the stepsize is less than a threshold value which depends on the input data only, Ye and Tse's interior ellipsoid algorithm for convex quadratic programming is globally convergent without nondegeneracy assumptions. In addition, its local convergence rate is at least linear and the dual iterates have an ergodically convergent property.

Key words: Quadratic programming, interior point methods, affine-scaling methods.

1. Introduction

A long-standing question about affine-scaling methods for linear programming (LP for short) was their convergence in the presence of degeneracy. Recently, Tseng and Luo [5] and Tsuchiya [6, 7] have resolved the problem. In papers [5] and [6], it is shown that the affine-scaling methods for LP are globally convergent without nondegeneracy assumptions of either primal or dual program, if some restrictions are imposed on the stepsize. Tsuchiya's analysis requires that the stepsize is $\frac{1}{8}$ of the maximum allowable distance, which is certainly important in practical computations. He also demonstrates in [7] that the convergence results of primal and dual affine-scaling methods are mathematically equivalent. However, his proof is quite complex and is based on the potential function from projective-scaling. On the other hand, Tseng and Luo's proof is straightforward but assumes a tiny step size of order $2^{-O(L)}$, where L is the input length. In achieving their goal, Tseng and Luo discover an interesting property of the methods. They show that the dual iterates generated by the methods are ergodically convergent, which means that, despite the seemingly irregular behavior of the dual iterates in the presence of degeneracy, certain weighted average of the dual iterates has a subsequence converging to a dual solution.

Correspondence to: Dr. Jie Sun, Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119, USA.
Research supported in part by the NSF under grant DDM-8721709.

Much less work has been done in the convergence analysis of affine-scaling methods for convex quadratic programming (CQP for short) although the idea of affine-scaling was extended to this case soon after Karmarkar's projective-scaling method for LP had been proposed. As a matter of fact, Ye and Tse suggested an "interior ellipsoid method" [11] in 1986, which is probably the first affine-scaling algorithm proposed for CQP. However, convergence (with or without degeneracy) of the method has been an open problem since that time. The interior ellipsoid method repeatedly minimizes the quadratic objective function over a series of ellipsoids of a fixed radius in the scaled space. Given the conceptual simplicity of the method and the potential of utilizing efficient implementation techniques developed for LP affine-scaling methods, two theoretical questions should be answered. First, is there a simple proof for its convergence? Second, how does the algorithm behave? In this paper we offer our answers to these questions by extending Tseng and Luo's analysis to the CQP case. It is shown that, as long as the stepsize is less than a threshold value which depends on the input data only, the algorithm generates iterates that converge globally without nondegeneracy assumptions; its local convergent rate is at least linear; and the dual iterates are also ergodically convergent. It should be noted that these results are only of theoretical interest because the analyzed algorithm takes the stepsize whose scaled norm equals $(\tau/\gamma)/(1+\tau/\gamma)$, where $\tau=2^{-O(L)}$ and γ is of the order of the condition number of the constraint matrix.

We consider the CQP problem

$$\text{(QP)}\qquad \text{minimize}\quad q(x)=\tfrac{1}{2}x^{\mathrm{T}}Qx+c^{\mathrm{T}}x$$
$$\text{subject to}\quad Ax=b,\quad x\geqslant 0,$$

where $x,\ c\in\mathbb{R}^n$, $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$, and Q is a symmetric positive semidefinite matrix. All input data are integers. The superscript "T" means the transpose. The following assumptions, which are typical for analysis of interior point methods, are made for (QP):

1. A has full rank.
2. The relative interior of the feasible set $F\equiv\{x\in\mathbb{R}^n \mid Ax=b,\ x\geqslant 0\}$ is nonempty.
3. Let x^0 be the initial interior solution. Then the set

$$K\equiv F\cap\{x\in\mathbb{R}^n \mid q(x^0)\geqslant q(x)\}$$

is included in a box $B\equiv\{x\in\mathbb{R}^n \mid \|x\|_\infty\leqslant 2^L\}$, where $L\equiv n^2+mn+\log_2|P|$ is the input length of (QP) (P is the product of the nonzero coefficients appearing in Q, c, A, and b).

The assumption 3 is reasonable as Ye and Tse argued in [11] that there is an optimal solution of (QP) whose ∞-norm is bounded by 2^L if (QP) has an optimal solution at all. In particular, if $Q=0$, this assumption corresponds to the assumption of boundedness of the optimal solution set in the case of LP. Notice also that the assumption implies that, for all $x\in B$, $\|x\|_1\leqslant 2^{2L}$ and the gradient of $q(x)$ is bounded by

$$\|\nabla q(x)\|_\infty \leq \|Q\|_\infty \|x\|_\infty + \|c\|_\infty \leq 2^L \cdot 2^L + 2^L \leq 2^{3L}. \tag{1.1}$$

For simplicity of notation, here and below, $\|\cdot\|$ denotes the Euclidean norm. Unless otherwise specified, we use capital letters to represent matrices or sets, lower-case letters to designate vectors, and Greek letters to stand for scalars.

The affine-scaling algorithm analyzed in this paper is as follows.

Algorithm 1.1. Let $x^k = (x_1^k, \ldots, x_n^k)^{\mathrm{T}} > 0$ be the current interior feasible solution to (QP). Let $X_k = \operatorname{diag}(x_1^k, \ldots, x_n^k)$. Solve

$$\text{(SP)} \quad \text{minimize} \quad \bar{q}(w) \equiv \tfrac{1}{2} w^{\mathrm{T}} Q w + \nabla q(x^k)^{\mathrm{T}} w$$

$$\text{subject to} \quad Aw = 0, \quad \|X_k^{-1} w\|^2 \leq \beta^2 < 1.$$

Let the solution be w^k. Set $x^{k+1} = x^k + w^k$. If x^{k+1} is optimal to (QP), stop; otherwise update k and iterate.

The solution w^k to the subproblem (SP) and the Kuhn–Tucker vector $(p^k, \mu_k)^{\mathrm{T}} \in \mathbb{R}^m \times \mathbb{R}_+$ satisfy

$$\begin{aligned} &Q_k w^k - A^{\mathrm{T}} p^k = -\nabla q(x^k), \quad Aw^k = 0, \quad \|X_k^{-1} w^k\|^2 \leq \beta^2, \\ &\mu_k(\|X_k^{-1} w^k\|^2 - \beta^2) = 0, \end{aligned} \tag{1.2}$$

where $Q_k = Q + \mu_k X_k^{-2}$. If $\mu_k = 0$, then (1.2) implies that w^k is the optimum of $\bar{q}(w)$ on $Aw = 0$. Thus x^{k+1} is an optimal solution to (QP); otherwise $\mu_k > 0$, from (1.2) one has

$$\begin{aligned} &w^k = -\frac{\beta X_k^2 r^k}{\|X_k r^k\|}, \quad r^k = \nabla q(x^{k+1}) - A^{\mathrm{T}} p^k, \\ &p^k = (A X_k^2 A^{\mathrm{T}})^{-1} A X_k^2 \nabla q(x^{k+1}), \end{aligned} \tag{1.3}$$

and $\mu_k = \beta^{-1} \|X_k r^k\|$.

In the next section we will show that if β is suitably chosen, then the sequence generated by the algorithm will converge to a limit point x^* such that $q(x^*)$ is within $\bar{\varepsilon}$ accuracy of the optimal value, where $\bar{\varepsilon}$ is a given tolerance. Consequently, by setting $\bar{\varepsilon} = 2^{-4L-1}$, in a finite number of iterations we will get an x^k that is within 2^{-4L} accuracy of the optimal value of (QP). An optimal solution to (QP) can then be obtained by rounding the error from this point, as discussed by Ye and Tse [11].

2. Convergence analysis of the algorithm

It is not hard to see that the algorithm generates a sequence of feasible solutions and we have $q(x^0) \geq q(x^1) \geq q(x^2) \geq \cdots$. Hence there is a limit value $\lim_{k\to\infty} q(x^k) \equiv \sigma_\infty$ because $q(x)$ is lower bounded on the set K. We will characterize convergence properties of the algorithm by showing three theorems:

(Theorem 1) $q(x^{k+1}) - \sigma_\infty \leq (1-\beta/\sqrt{n})[q(x^k) - \sigma_\infty]$, for large k.

(Theorem 2) The sequence $\{x^k\}$ converges to a limit point x^*.

(Theorem 3) $q(x^*)$ is within $2\beta\gamma 2^{5L}/(1-\beta)$ accuracy of the optimal value of (QP), where γ is a constant depending on A.

Theorem 1. *Let $\{x^k\}$ be the sequence generated by Algorithm* 1.1. *Then for k sufficiently large one has*

$$q(x^{k+1}) - \sigma_\infty \leq (1-\beta/\sqrt{n})[q(x^k) - \sigma_\infty].$$

Proof. The convex compact set $Y \equiv K \cap \{y \mid q(y) \leq \sigma_\infty\}$ is nonempty because any cluster point of $\{x^k\}$ belongs to it. Now we claim that there exists a positive integer $\bar{k}$, such that for all $k > \bar{k}$, one has

$$\min_{y\in Y} \|X_k^{-1}(y - x^k)\| \leq \sqrt{n}. \tag{2.1}$$

To show this, suppose that there is a subsequence $\{x^k\}_{k\in S}$ such that

$$\min_{y\in Y} \|X_k^{-1}(y - x^k)\| > \sqrt{n} \quad \forall k \in S. \tag{2.2}$$

We may further assume that $\{x^k\}_{k\in S} \to y^*$ because K is bounded. Note that $y^* \in Y$. Thus we have

$$\begin{aligned}\|X_k^{-1}(y^* - x^k)\|^2 &= \sum_{j=1}^{n} (1 - y_j^*/x_j^k)^2 \\ &= \sum_{y_j^*>0} (1 - y_j^*/x_j^k)^2 + \sum_{y_j^*=0} 1 \leq n \quad \text{(for large } k \in S),\end{aligned}$$

a contradiction to (2.2). Hence (2.1) is valid. Now consider $k > \bar{k}$ and $y^k = x^k + \beta[y^*(x^k) - x^k]/\sqrt{n}$, where $y^*(x^k)$ attains the minimum in (2.1). Then, from (2.1), $y^k - x^k$ is a feasible solution to (SP). Compared with the optimal solution w^k to (SP), we have

$$\begin{aligned}q(x^{k+1}) &= q(x^k + w^k) \leq q(x^k + y^k - x^k) \\ &= q((1-\beta/\sqrt{n})x^k + \beta y^*(x^k)/\sqrt{n}) \\ &\leq (1-\beta/\sqrt{n})q(x^k) + \beta q(y^*(x^k))/\sqrt{n} \\ &\leq (1-\beta/\sqrt{n})q(x^k) + \beta\sigma_\infty/\sqrt{n}.\end{aligned}$$

Therefore

$$q(x^{k+1}) - \sigma_\infty \leq (1-\beta/\sqrt{n})[q(x^k) - \sigma_\infty]. \qquad \square$$

In order to show Theorem 2, we need the following lemma. Its proof is obvious.

Lemma 1. *For any $k\times n$ matrix B and any k-vector d, if the linear system $Bx=d$ has a solution, then it has a solution whose norm is at most $\lambda\|d\|$, where λ is a constant that depends on B only.* □

Theorem 2. *The sequence $\{x^k\}$ converges to a limit point x^*.*

Proof. There exists an orthogonal mattrix P, depending on Q only, such that

$$P^{\mathrm{T}}QP=\begin{pmatrix} Q_1 & 0 \\ 0 & Q_2 \end{pmatrix},$$

where Q_1 is an $n_1\times n_1$ diagonal positive definite matrix, $Q_2=0\in\mathbb{R}^{n_2\times n_2}$, and $n_1+n_2=n$. Correspondingly, we express $w^k=v^k+u^k$, where

$$v^k=P\begin{pmatrix} y^k \\ 0 \end{pmatrix},\qquad u^k=P\begin{pmatrix} 0 \\ z^k \end{pmatrix},$$

$y^k\in\mathbb{R}^{n_1}$, and $z^k\in\mathbb{R}^{n_2}$.

Claim 1. $\|v^k\|\leqslant\sqrt{2\tau^{-1}}|q(x^k)-q(x^{k+1})|^{1/2}$ for all k, where τ is the smallest eigenvalue of Q_1 (hence the smallest positive eigenvalue of Q).

Proof of Claim 1. From $w^k=x^{k+1}-x^k$ and

$$q(x^k)-q(x^{k+1})=-\nabla q(x^{k+1})^{\mathrm{T}}w^k+\tfrac{1}{2}(w^k)^{\mathrm{T}}Qw^k$$

we have

$$q(x^k)-q(x^{k+1})\geqslant\tfrac{1}{2}(w^k)^{\mathrm{T}}Qw^k=\tfrac{1}{2}(y^k)^{\mathrm{T}}Q_1y^k \tag{2.3}$$

because $-\nabla q(x^{k+1})^{\mathrm{T}}w^k\geqslant 0$, which can be seen from the optimality of w^k to (SP). Therefore

$$\|y^k\|\leqslant\sqrt{\tau^{-1}}\|y^k\|_{Q_1}\leqslant\sqrt{2\tau^{-1}}|q(x^k)-q(x^{k+1})|^{1/2},$$

where $\|\cdot\|_{Q_1}$ is the Q_1-norm of the vector. Since $\|v^k\|=\|y^k\|$, the claim is proven.

Claim 2. There holds that

$$\|u^k\|\leqslant\alpha|q(x^k)-q(x^{k+1})|^{1/2} \tag{2.4}$$

for large k, where α is a constant.

Proof of Claim 2. Suppose that (2.4) is not valid. Then, because $\{u^k\}$ is a bounded sequence, there is a subsequence S of $\{1,2,\ldots\}$ and a nonempty subset J of $\{1,\ldots,n\}$ so that

$$\left\{\frac{|q(x^k)-q(x^{k+1})|^{1/2}}{\|u^k\|}\right\}_S\to 0, \tag{2.5}$$

$$\lim_{k\to\infty,k\in S}\frac{|u_j^k|}{\|u^k\|}>0\quad\forall j\in J$$

and

$$\lim_{k\to\infty, k\in S} \frac{u_j^k}{\|u^k\|} = 0 \quad \forall j \in \bar{J} \equiv \{1, \ldots, n\} - J. \tag{2.6}$$

From (2.6) and Claim 1, it is implied that

$$\{\|v^k\|/u_j^k\}_S \to 0 \quad \forall j \in J. \tag{2.7}$$

Consider the system

$$Au = -Av^k, \quad c^T u = c^T u^k, \quad u_j = u_j^k \ \ \forall j \in \bar{J}, \quad [P^T u]_i = 0 \ \ \forall i = 1, \ldots, n_1.$$

Here $[\cdot]_i$ denotes the ith component of the vector $[\cdot]$. This system of equations is solvable because at least u^k is a trivial solution. By Lemma 1 there is a solution $\bar{u}^k$ such that $\|\bar{u}^k\| \leq \lambda(\|Av^k\| + |c^T u^k| + \sum_{j\in\bar{J}} |u_j^k|)$, where λ depends on A, c, and Q only. Now we argue that $\{\bar{u}^k/\|u^k\|\}_S \to 0$. Since both $\|Av^k\|$ and $\sum_{j\in\bar{J}} |u_j^k|$ are in the order of $o(\|u^k\|)$ (see Claim 1 and (2.6)), it suffices to show that $|c^T u^k| = O(|q(x^k) - q(x^{k+1})|^{1/2})$. This is true because of the fact that

$$\begin{aligned} |-c^T u^k| &= |q(x^k) - q(x^{k+1}) + \nabla q(x^k)^T v^k + (x^k)^T Q u^k + \tfrac{1}{2}(w^k)^T Q w^k| \\ &\leq q(x^k) - q(x^{k+1}) + \|\nabla q(x^k)\| \cdot \|v^k\| \\ &\quad + \tfrac{1}{2}(y^k)^T Q_1 y^k \ (\text{using } Qu^k = 0) \\ &\leq (1 + 2^{3L}\sqrt{2\tau^{-1}} + 1)|q(x^k) - q(x^{k+1})|^{1/2} \\ &\quad (\text{using (2.3), (1.1), and Claim 1, for large } k). \end{aligned}$$

Let $d^k = u^k - \bar{u}^k \ \forall k \in S$. Then there hold that $Ad^k = 0$, $c^T d^k = 0$, and

$$Qd^k = P\begin{pmatrix} Q_1 & 0 \\ 0 & 0 \end{pmatrix} P^T(u^k - \bar{u}^k) = 0.$$

Thus for every $k \in S$ we have that

$$\begin{aligned} &\bar{q}(w^k - d^k) = \tfrac{1}{2}(w^k - d^k)^T Q (w^k - d^k) + (Qx^k + c)^T (w^k - d^k) = \bar{q}(w^k), \\ &A(w^k - d^k) = 0, \end{aligned} \tag{2.8}$$

and that

$$\begin{aligned} \|(X_k)^{-1}(w^k - d^k)\|^2 &= \sum_{j\in\bar{J}} \left(\frac{w_j^k}{x_j^k}\right)^2 + \sum_{j\in J} \left(\frac{w_j^k - u_j^k + \bar{u}_j^k}{x_j^k}\right)^2 \\ &= \sum_{j\in\bar{J}} \left(\frac{w_j^k}{x_j^k}\right)^2 + \sum_{j\in J} \left(\frac{w_j^k}{x_j^k}\right)^2 \left(\frac{v_j^k + \bar{u}_j^k}{w_j^k}\right)^2. \end{aligned} \tag{2.9}$$

Now, since $\{\bar{u}_j^k/\|u^k\|\}_S \to 0$, we have for all $j \in J$, $\{\bar{u}_j^k/u_j^k\}_S \to 0$, which together with (2.7) yields $\{(v_j^k + \bar{u}_j^k)/u_j^k\}_S \to 0 \ \forall j \in J$. Then for all $j \in J$, we have that

$$\{(v_j^k + \bar{u}_j^k)/w_j^k\}_S = \{(v_j^k + \bar{u}_j^k)/(v_j^k + u_j^k)\}_S \to 0,$$

so each term in the second sum of (2.9) is strictly less than the square of w_j^k/x_j^k for sufficiently large $k \in S$. Because J is not empty, we have for those k that

$$\|X_k^{-1}(w^k - d^k)\|^2 < \|X_k^{-1} w^k\|^2.$$

This, together with (2.8), means that $w^k - d^k$ is an interior optimal solution to (SP), so $x^k + w^k - d^k$ is an optimal solution to (QP). Since $q(x^k + w^k - d^k) = q(x^k + w^k)$, $x^k + w^k$ is optimal to (QP), too. Hence the algorithm will stop. In other words, the sequence $\{x^k\}$ will either stop at an optimum or have that $\|u^k\| \leqslant \alpha(|q(x^k) - q(x^{k+1})|^{1/2})$, where α is a constant. This completes the proof of Claim 2.

Now we prove the theorem. Using Theorem 1, Claim 1, and Claim 2, we obtain that for any large $k < r$,

$$\begin{aligned}\|x^k - x^r\| &\leqslant (\alpha + \sqrt{2\tau^{-1}}) \sum_{s=k}^{r-1} |q(x^s) - q(x^{s+1})|^{1/2} \\ &\leqslant (\alpha + \sqrt{2\tau^{-1}}) \sum_{s=k}^{r-1} |q(x^s) - \sigma_\infty|^{1/2} \\ &\leqslant (\alpha + \sqrt{2\tau^{-1}})\, |q(x^k) - \sigma_\infty|^{1/2} \sum_{s=0}^{\infty} (1 - \beta/\sqrt{n})^{s/2}.\end{aligned}$$

Hence $\{x^k\}$ is a Cauchy sequence and it must converge to a limit x^*. □

Remark. In a separate paper of Ye [10], a statement is given (without proof) that if Q is positive definite then $\{x^k\}$ converges. This can be seen clearly from the proof of Theorem 2. Theorem 2 offers a good supplement to Ye's observation.

The following three lemmas help to establish Theorem 3.

Lemma 2. *Suppose that $x \in K$ and $p \in \mathbb{R}^m$ satisfy*

$$\begin{aligned} &x_j = 0 \Rightarrow [\nabla q(x) - A^{\mathrm{T}} p]_j \geqslant -\varepsilon, \\ &x_j > 0 \Rightarrow \varepsilon \geqslant [\nabla q(x) - A^{\mathrm{T}} p]_j \geqslant -\varepsilon. \end{aligned} \tag{2.10}$$

Then $q(x)$ is within $2\varepsilon 2^{2L}$ of the optimal cost of (QP).

Proof. Consider the parameterized form of (QP):

(PQP) minimize $\tilde{q}(x, \varepsilon) = \frac{1}{2} x^{\mathrm{T}} Q x + (c + \varepsilon e)^{\mathrm{T}} x$

subject to $Ax = b, \quad x \geqslant 0,$

where $e = (1, \ldots, 1)^{\mathrm{T}}$. Its dual program is

(DPQP) maximize $d(y, p) = -\frac{1}{2} y^{\mathrm{T}} Q y + b^{\mathrm{T}} p$

subject to $\nabla q(y) - A^{\mathrm{T}} p \geqslant -\varepsilon e, \quad y \geqslant 0.$

Note that x is feasible to (QP). In addition, if (2.10) is satisfied by x and p, then (x, p) is feasible to (DPQP). By duality we note that

$$\begin{aligned} 0 \leqslant \tilde{q}(x, \varepsilon) - d(x, p) &= x^{\mathrm{T}} \nabla q(x) - b^{\mathrm{T}} p + \varepsilon e^{\mathrm{T}} x \\ &= x^{\mathrm{T}} (\nabla q(x) - A^{\mathrm{T}} p) + \varepsilon e^{\mathrm{T}} x \leqslant 2\varepsilon e^{\mathrm{T}} x. \end{aligned}$$

The inequality above says that the $q(x)$ is within $\varepsilon e^{\mathrm{T}}x$ of the optimal cost of (PQP). However, the optimal values of (QP) and (PQP) can differ at most by $\varepsilon 2^{2L}$. Thus $q(x)$ is within $2\varepsilon 2^{2L}$ of the optimal cost of (QP). □

Lemma 3. *For any $h \in D$, where D is a bounded set in $\mathbb{R}^n$, the range of the function $p(x^k) = (AX_k^2A^{\mathrm{T}})^{-1}AX_k^2h$ is a bounded set whose bound depends only on A, n, and D.*

Proof. Vanderbei and Lagarias [8, Section 4] have proven that $p(x^k)$ is bounded for any positive diagonal matrix X_k and for fixed A and h. Now let $h = e^j$, the jth unit vector in $\mathbb{R}^n$. We conclude that the jth column, hence each entry of the matrix $(AX_k^2A^{\mathrm{T}})^{-1}AX_k^2$, has a bound that depends on A only for any X_k. Therefore the range of $p(x^k)$ is contained in a bounded set if $h \in D$ and D is bounded. Moreover, the bound of the range can depend on A, n, and D only. □

Remark. The bound obtained in [8] for the ith component of $p(x^k)$ is

$$\max_{\substack{1 \le j_1 < \cdots < j_m \le n \\ \text{denominator} \ne 0}} \left| \frac{\det_{j_1,\ldots,j_m}(a^1, \ldots, h, \ldots, a^m)}{\det_{j_1,\ldots,j_m}(a^1, \ldots, a^m)} \right|,$$

where $\det_{j_1,\ldots,j_m}(a^1, \ldots, a^m)$ denotes the $m \times m$ subdeterminant of A by selecting columns $j_1, \ldots, j_m$ from A, where $a^1, \ldots, a^m$ are the rows of A. The numerator is defined similarly but a^j is replaced by h^{T}. Thus by Cramer's rule, if $h = e^j$, the entries of $(AX_k^2A^{\mathrm{T}})^{-1}AX_k^2$ are bounded by the largest absolute value of the entries of T^{-1} among all $m \times m$ invertible submatrices M of A. Let this value be η (which depends on A only). It is then implied that

$$\|I - A^{\mathrm{T}}(AX_k^2A^{\mathrm{T}})^{-1}AX_k^2\|_\infty \le 1 + \|A^{\mathrm{T}}\|_\infty(n\eta) \equiv \gamma,$$

where γ is of the order of the condition number of the constraint matrix A. Hence by (1.1) and (1.3) one has

$$\|r^k\|_\infty = \|[I - A^{\mathrm{T}}(AX_k^2A^{\mathrm{T}})^{-1}AX_k^2]\nabla q(x^{k+1})\| \le \gamma 2^{3L}.$$

The lemma below is a well-known result in mathematical analysis.

Lemma 4. *Suppose that $0 < \alpha_k \to +\infty$ and $h^k \to h^*$ as $k \to \infty$. Then $\sum_{i=0}^k \alpha_i h^i / \sum_{i=0}^k \alpha_i \to h^*$.* □

Now we prove the third theorem.

Theorem 3. *The sequence $\{x^k\}$ generated by Algorithm 1.1 converges to a limit point x^* with $q(x^*)$ being within $2\beta\gamma 2^{5L}/(1-\beta)$ accuracy of the optimal value of (QP).*

Proof. The proof is similar to the one of Tseng and Luo for linear programming except for the processing of some limits. For the reader's convenience, we write it in detail here.

Let $g^k = \nabla q(x^{k+1})$. For each k denote

$$\bar{p}^k = \frac{p^0/\|X_0 r^0\| + \cdots + p^k/\|X_k r^k\|}{1/\|X_0 r^0\| + \cdots + 1/\|X_k r^k\|}.$$

Note that from formulae (1.3), Lemma 3, and the remark after Lemma 3, the sequence $\{p^k\}$ is bounded, so are the sequences $\{\bar{p}^k\}$ and $\{r^k\}$.

Note that $\|X_k r^k\| \to 0$, which can be derived from

$$\begin{aligned} q(x^k) - q(x^{k+1}) &\geq -\nabla q(x^{k+1})^{\mathrm{T}} w^k \quad \text{(by convexity of } q(x)) \\ &= \beta \|X_k r^k\| \quad \text{(by formulae (1.3)).} \end{aligned}$$

Consider any convergent subsequence of $\{\bar{p}^k\}_{k\in S} \to \bar{p}^*$. We now show that x^* and $\bar{p}^*$ satisfy (2.10) with $\varepsilon = \beta\gamma 2^{3L}/(1-\beta)$. Because $\{g^k\} \to g^* = \nabla q(x^*)$, by Lemma 4 and the fact that $\|X_k r^k\| \to 0$, we have

$$\frac{\sum_{i=0}^k g_j^i/\|X_i r^i\|}{\sum_{i=0}^k 1/\|X_i r^i\|} \to g^*.$$

If $x_j^* > 0$, then from $X_k r^k \to 0$ we know $r_j^k \to 0$. Lemma 4 then implies that

$$[A^{\mathrm{T}}\bar{p}^k]_j - \frac{\sum_{i=0}^k g_j^i/\|X_i r^i\|}{\sum_{i=0}^k 1/\|X_i r^i\|} = -\frac{\sum_{i=0}^k r_j^i/\|X_i r^i\|}{\sum_{i=0}^k 1/\|X_i r^i\|} \to 0,$$

and thus for any $\varepsilon > 0$ we have that

$$|[A^{\mathrm{T}}\bar{p}^* - \nabla q(x^*)]_j| < \varepsilon. \tag{2.11}$$

If $x_j^* = 0$, we first prove the following result:

$$-\frac{\sum_{i=0}^k r_j^i/\|X_i r^i\|}{\sum_{i=0}^k 1/\|X_i r^i\|} \leq \frac{\theta_j^k}{\sum_{i=0}^k 1/\|X_i r^i\|} + \frac{\beta}{1-\beta}\frac{\sum_{i=0}^k |r_j^i|/\|X_i r^i\|}{\sum_{i=0}^k 1/\|X_i r^i\|}, \tag{2.12}$$

where

$$\theta_j^k = \frac{x_j^k \delta_j^k}{x_j^{k+1}} + \frac{x_j^{k-1}\delta_j^{k-1}}{x_j^k} + \cdots + \frac{x_j^0 \delta_j^0}{x_j^1} \quad \text{and} \quad \delta_j^k = -\frac{r_j^k}{\|X_k r^k\|}. \tag{2.13}$$

From

$$\begin{aligned} x_j^{k+1} &= x_j^k - \beta (x_j^k)^2 r_j^k/\|X_k r^k\| \\ &= x_j^k(1 - \beta x_j^k r_j^k/\|X_k r^k\|) = x_j^k(1 + \beta x_j^k \delta_j^k) \end{aligned} \tag{2.14}$$

we have that

$$\begin{aligned} (1+\beta)x_j^k &\geq x_j^{k+1} \quad \text{if } \delta_j^k > 0, \\ (1-\beta)x_j^k &\leq x_j^{k+1} \quad \text{if } \delta_j^k < 0, \end{aligned}$$

so that

$$\begin{aligned} \delta_j^k/(1+\beta) &\leq x_j^k \delta_j^k / x_j^{k+1} \quad \text{if } \delta_j^k > 0, \\ \delta_j^k/(1-\beta) &\leq x_j^k \delta_j^k / x_j^{k+1} \quad \text{if } \delta_j^k < 0. \end{aligned}$$

Since $1/(1+\beta)=1-\beta/(1+\beta)$ and $1/(1-\beta)=1+\beta/(1-\beta)$, this implies that

$$\delta_j^k+\cdots+\delta_j^0-\frac{\beta}{1+\beta}\sum_{\delta_j^l>0}\delta_j^l+\frac{\beta}{1-\beta}\sum_{\delta_j^l<0}\delta_j^l\leq\theta_j^k.$$

Hence by using the fact $\beta/(1+\beta)<\beta/(1-\beta)$,

$$\delta_j^k+\cdots+\delta_j^0\leq\theta_j^k+\frac{\beta}{1-\beta}(|\delta_j^k|+\cdots+|\delta_j^0|).$$

Dividing both sides by $1/\|X_k r^k\|+\cdots+1/\|X_0 r^0\|$ and using (2.13) gives (2.12).

Because $\|r^k\|_\infty\leq\gamma 2^{3L}$ (see the remark after Lemma 3), we have from (2.12) that

$$(A^{\mathrm{T}}\bar{p}^k)_j-\frac{\sum_{i=0}^{i} g_j^i/\|X_i r^i\|}{\sum_{i=0}^{k}1/\|X_i r^i\|}=-\frac{\sum_{i=0}^{k} r_j^i/\|X_i r^i\|}{\sum_{i=0}^{k}1/\|X_i r^i\|}\leq\frac{\theta_j^k}{\sum_{i=0}^{k}1/\|X_i r^i\|}+\frac{\beta\gamma 2^{3L}}{1-\beta}.$$

Thus by Lemma 4, there holds that

$$(A^{\mathrm{T}}\bar{p}^*-g^*)_j\leq\limsup_{k\to\infty,k\in S}\left\{\frac{\theta_j^k}{\sum_{i=0}^{k}1/\|X_i r^i\|}\right\}+\frac{\beta\gamma 2^{3L}}{1-\beta}. \tag{2.15}$$

Now since $x_j^*=0$, we obtain from (2.13) and (2.14) that

$$\begin{aligned}\theta_j^k&=\beta^{-1}\left[\frac{\beta x_j^k\delta_j^k}{x_j^{k+1}}+\frac{\beta x_j^{k-1}\delta_j^{k-1}}{x_j^k}+\cdots+\frac{\beta x_j^0\delta_j^0}{x_j^1}\right]\\&=\beta^{-1}\left[\frac{1}{x_j^{k+1}}\left(\frac{x_j^{k+1}}{x_j^k}-1\right)+\frac{1}{x_j^k}\left(\frac{x_j^k}{x_j^{k-1}}-1\right)+\cdots+\frac{1}{x_j^1}\left(\frac{x_j^1}{x_j^0}-1\right)\right]\\&=\beta^{-1}\left[\frac{1}{x_j^0}-\frac{1}{x_j^{k+1}}\right]\to-\infty.\end{aligned}$$

This together with (2.15) then implies that

$$[A^{\mathrm{T}}\bar{p}^*-\nabla q(x^*)]_j\leq\frac{\beta\gamma 2^{3L}}{1-\beta}. \tag{2.16}$$

The inequalities (2.11) and (2.16) ensure that the conditions (2.10) of Lemma 2 for x^* and $\bar{p}^*$ with $\varepsilon=\beta\gamma 2^{3L}/(1-\beta)$. The conclusion of Theorem 3 is therefore valid due to Lemma 2. □

3. Conclusions and final remarks

Dikin [2, 3], Barnes [1], Vanderbei and Lagarias [8], and Vanderbei, Meketon and Freeman [9] showed convergence of various affine-scaling algorithms for LP under certain nondegeneracy assumptions. Tseng and Luo [5] and Tsuchiya [6] provided convergence proofs for some affine-scaling methods for LP without using non-degeneracy assumptions. This paper is concerned with Ye and Tse's primal afine-scaling algorithm for CQP. We show that the convergence of the convergence of the algorithm does not depend on nondegeneracy assumptions and that the dual iterative solutions have an ergodically convergent property. However, our proof does not imply that the algorithm has polynomial complexity. Notice that Monteiro, Adler and Resende [7] have proven polynomial complexity for a primal–dual affine-scaling method for CQP, but the proof requires the initial point close to the central path. An interesting open question then is whether there is a polynomial-time affine-scaling method for LP or CQP that allows a natural start point. Another direction of possible future research is to investigate the convergence properties of affine-scaling methods in more general (e.g. convex) or more special (e.g. network) settings than CQP.

Acknowledgements

The author is grateful to the associate editor and the referees for their valuable comments in revising this paper.

References

[1] E.R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," *Mathematical Programming* 36 (1989) 174–182.
[2] I.I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Mathematics Doklady* 8 (1967) 674–675.
[3] I.I. Dikin, "On the speed of an iterative process," *Upravlyaemye Sistemi* 12 (1974) 54–60.
[4] R.C. Monteiro, I. Adler and M.G.C. Resende, "A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension," *Mathematics of Operations Research* 15 (1990) 191–214.
[5] P. Tseng and Z.-Q. Luo, "On the convergence of the affine-scaling algorithm," Preprints CICS-P-169, M.I.T. (Cambridge, MA, 1989).
[6] T. Tsuchiya, "Global convergence of the affine scaling methods for degenerate linear programming problems," Research Memorandum No. 373, The Institute of Statistical Mathematics (Tokyo, Japan, 1990).
[7] T. Tsuchiya, "Global convergence property of the affine scaling methods for primal degenerate linear programming problems," Research Memorandum No. 367, The Institute of Statistical Mathematics (Tokyo, Japan, 1989).
[8] R.J. Vanderbei and J.C. Lagarias, "Dikin's convergence result for the affine-scaling algorithm," in: J. Lagarias and M. Todd, *Contemporary Mathematics No.* 114 (AMS, Providence, RI, 1990) pp. 109–119.
[9] R.J. Vanderbei, M.S. Meketon and B.A. Freeman, "A modification of Karmarkar's linear programming algorithm," *Arithmica* 1 (1986) 395–407.
[10] Y. Ye, "An extension of Karmarkar's projective algorithm and the trust region method for quadratic programming," in: N. Megiddo, ed., *Progress in Mathematical Programming – Interior Point and Related Methods* (Springer, Berlin, 1989).
[11] Y. Ye and E. Tse, "An extension of Karmarkar's projective algorithm for convex quadratic programming," *Mathematical Programming* 44 (1989) 157–180.