A Superlinearly Convergent $O(\sqrt{n}L)$-Iteration Algorithm for Linear Programming

Y. Ye, [*] *R. A. Tapia* [†] *and Y. Zhang* [‡]

July 1991

Abstract

In this note we consider a large step modification of the Mizuno-Todd-Ye $O(\sqrt{n}L)$ predictor-corrector interior-point algorithm for linear programming. We demonstrate that the modified algorithm maintains its $O(\sqrt{n}L)$-iteration complexity, while exhibiting superlinear convergence for general problems and quadratic convergence for nondegenerate problems. To our knowledge, this is the first construction of a superlinearly convergent algorithm with $O(\sqrt{n}L)$-iteration complexity.

Key words: Linear programming, primal and dual, superlinear and quadratic convergence, polynomiality

Abbreviated title: A superlinearly convergent $O(\sqrt{n}L)$ algorithm for LP

---

[*] *Department of Management Sciences, The University of Iowa, Iowa City, Iowa, 52242. This author was supported in part by NSF Grant DDM-8922636, University of Iowa and in part by NSF Coop. Agr. No. CCR-8809615 as a visiting member of the Center for Research in Parallel Computation, Rice University, Houston, Texas, 77251-1892.*

[†] *Department of Mathematical Sciences and Center for Research in Parallel Computation, Rice University, Houston, Texas, 77251-1892. This author was supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.*

[‡] *Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD, 21228. This author was supported in part by NSF Coop. Agr. No. CCR-8809615 as a visiting member of the Center for Research in Parallel Computation, Rice University, Houston, Texas, 77251-1892.*

1. Introduction

Consider the primal linear program (LP):

$$\begin{aligned} &\min \;\; c^T x \\ &s.t. \quad Ax = b, \;\; x \geq 0, \end{aligned}$$

and its dual (LD):

$$\begin{aligned} &\max \;\; b^T y \\ &s.t. \quad A^T y + s = c, \;\; s \geq 0, \end{aligned}$$

where $A \in \mathbf{R}^{m \times n}$ has full rank, $c \in \mathbf{R}^n$, and $b \in \mathbf{R}^m$. Feasible points x^* and (y^*, s^*) are optimal for (LP) and (LD), respectively, if and only if

$$x_j^* s_j^* = 0 \quad \text{for} \quad j = 1, 2, ..., n.$$

We assume that optimal solutions and strictly feasible points exist for both problem (LP) and problem (LD).

Recently, superlinear convergence results have been obtained for LP interior-point algorithms by several authors, e.g. Kojima et al. [2], and Zhang et al. [6],[7],[8]. These authors have shown how to select the parameters in a primal-dual LP interior-point algorithm so as to achieve superlinear convergence under the assumption that the iterates converge, or quadratic convergence under the assumption of nondegeneracy. Clearly, it is desirable for an algorithm to possess the global property of polynomial complexity and the local property of superlinear convergence. Interestingly, the complexity of these algorithms is at best $O(nL)$-iteration, while the best known complexity bound is $O(\sqrt{n}L)$-iteration. Here, L is interpreted as the precision of the final solution. Hence, at this juncture it seems reasonable to conjecture that the price one must pay for superlinear convergence is at least an increase in the complexity bound by a factor of $\sqrt{n}$. The purpose of this technical note is to show that this is not the case. We accomplish this objective by demonstrating that a large step strategy first proposed by Ye [5] can be adapted to the Mizuno-Todd-Ye predictor-corrector interior-point algorithm in a manner that maintains its $O(\sqrt{n}L)$ iteration complexity and induces superlinear convergence for general problems and quadratic convergence for nondegenerate problems.

2. The Predictor-Corrector Algorithm

In this section, we briefly describe the Mizuno-Todd-Ye predictor-corrector LP algorithm [4] and adapt it with a large step strategy. We employ the notation $X = \mathrm{diag}(x)$, $S = \mathrm{diag}(s)$, etc. and we let Ω denote the collection of all strictly feasible points (x, s), i.e., $(x, s) > 0$, x is feasible for (LP), and s is feasible for (LD). Consider the neighborhood

$$\mathcal{N}(\alpha) = \{(x, s) \in \Omega : \|Xs/\mu - e\| \leq \alpha\},$$

where $\|.\|$ represents the l_2 norm, $\mu = x^T s/n$, e is the vector of all ones, and α is a constant between 0 and 1.

To begin with choose $0 < \beta < 1/2$ (a typical choice would be $1/4$). All search directions d_x, d_s, and d_y will be defined as solutions of the following system of linear equations (Kojima et al. [3])

$$\begin{aligned} X d_s + S d_x &= \gamma \mu e - X s \\ A d_x &= 0 \\ A^T d_y + d_s &= 0. \end{aligned} \tag{1}$$

A typical iteration of the algorithm proceeds as follows. Given $(x^k, s^k) \in \mathcal{N}(\beta)$, we solve the system (1) with $(x, s) = (x^k, s^k)$ and $\gamma = 0$. For some step length $\theta \geq 0$ let $x(\theta) = x^k + \theta d_x$, $s(\theta) = s^k + \theta d_s$, and $\mu(\theta) = x(\theta)^T s(\theta)/n$. This is the predictor step. Our specific choice for θ will be stated after we consider the following lemma

Lemma 1. If for some positive $\theta^k \leq 1$ we have

$$\|X(\theta)s(\theta)/\mu(\theta) - e\| \leq \alpha < 1 \quad \text{for all} \quad 0 \leq \theta \leq \theta^k, \tag{2}$$

then $(x(\theta^k), s(\theta^k)) \in \mathcal{N}(\alpha)$.

The proof of Lemma 1 follows directly from a continuity argument. Lemma 1 basically says that the feasibility (positivity) of $(x(\theta^k), s(\theta^k))$ is guaranteed as long as (2) is satisfied. Thus, we can choose the *largest* step length $\theta^k \leq 1$ such that (2) is satisfied for $\alpha = 2\beta$, and let

$$\hat{x}^k = x(\theta^k) \quad \text{and} \quad \hat{s}^k = s(\theta^k).$$

This choice of step length was first suggested by Ye [5].

Now we solve the system (1) with $(x, s) = (\hat{x}^k, \hat{s}^k) \in \mathcal{N}(2\beta)$, $\mu = (\hat{x}^k)^T\hat{s}^k/n$, and $\gamma = 1$. Let $x^{k+1} = \hat{x}^k + d_x$ and $s^{k+1} = \hat{s}^k + d_s$. It has been proved that $(x^{k+1}, s^{k+1}) \in \mathcal{N}(\beta)$ (Lemma 3 [4]). This is the corrector step.

Observe that the algorithm generates a sequence of points satisfying

$$(\hat{x}^k)^T\hat{s}^k = (1-\theta^k)(x^k)^T s^k \tag{3.1}$$

and

$$(x^{k+1})^T s^{k+1} = (\hat{x}^k)^T\hat{s}^k. \tag{3.2}$$

Mizuno et al. (Lemmas 1, 2, and 4 [4]) showed that for the predictor step

$$\theta^k \geq \min\left\{\frac{1}{2}, (\frac{\mu^k}{8\|D_x d_s\|})^{0.5}\right\}$$

and

$$\|D_x d_s\|/\mu^k \leq \sqrt{2}n/4.$$

Thus, this inequality together with (3) implies that the iteration complexity of the algorithm is $O(\sqrt{n}L)$. Note that, unfortunately, the algorithm requires that the linear system (1) be solved twice at each iteration.

The above lower bound is not sufficient to demonstrate superlinear convergence since it is at most $1/2$. Thus, we derive another bound for θ^k.

Lemma 2. If θ^k is the largest θ^k satisfying the conditions of Lemma 1 with $\alpha = 2\beta$, then

$$\theta^k \geq \frac{2}{\sqrt{1+4\|D_x d_s/\mu^k\|/\beta}+1}.$$

Proof. In the predictor step

$$\begin{aligned}
\|X(\theta)s(\theta) - \mu(\theta)e\| &= \|(1-\theta)(X^k s^k - \mu^k e) + \theta^2 D_x d_s\| \\
&\leq \|(1-\theta)(X^k s^k - \mu^k e)\| + \theta^2\|D_x d_s\| \\
&= (1-\theta)\|(X^k s^k - \mu^k e)\| + \theta^2\|D_x d_s\| \\
&\leq (1-\theta)\beta\mu^k + \theta^2\|D_x d_s\|
\end{aligned}$$

for $0 \le \theta \le 1$. Thus,

$$(1-\theta)\beta\mu^k + \theta^2\|D_x d_s\| \le 2\beta(1-\theta)\mu^k = 2\beta\mu(\theta) \tag{4}$$

guarantees the satisfaction of inequality (2) for $\alpha = 2\beta$. The largest θ satisfying (4) is the positive root θ^+ of the quadratic equation

$$(1-\theta)\beta\mu^k + \theta^2\|D_x d_s\| = 2\beta(1-\theta)\mu^k,$$

which is

$$\theta^+ = \frac{2}{\sqrt{1 + 4\|D_x d_s/\mu^k\|/\beta} + 1}.$$

Note for all $0 \le \theta < \theta^+$ strict inequality holds in (4), and also in (2) with $\alpha = 2\beta$. This concludes the proof of Lemma 2. ∎

3. Superlinear Convergence

From (3), we see that if $(1-\theta^k) \to 0$ then the duality gap $(x^k)^T s^k$ converges to zero Q-superlinearly. Moreover, if $(1-\theta^k) = O((x^k)^T s^k)$ then the duality gap converges to zero Q-quadratically. Hopefully, we can accomplish these objectives using the bound given in Lemma 2.

At the kth predictor step, for convenience, let $\delta^k = D_x d_s/\mu^k$. If θ^k is the largest θ^k satisfying the conditions of Lemma 1 with $\alpha = 2\beta$, then

$$\begin{aligned} 1-\theta^k &\le 1 - \frac{2}{\sqrt{1+4\|\delta^k\|/\beta}+1} \\ &= \frac{\sqrt{1+4\|\delta^k\|/\beta}-1}{\sqrt{1+4\|\delta^k\|/\beta}+1} \\ &= \frac{4\|\delta^k\|/\beta}{(\sqrt{1+4\|\delta^k\|/\beta}+1)^2} \\ &\le \|\delta^k\|/\beta. \end{aligned} \tag{5}$$

We are now in a position to state our main result.

Algorithm (Large-step predictor-corrector)

By the large-step predictor-corrector algorithm we mean the Mizuno-Todd-Ye algorithm defined in Section 2 adapted with the step length choice as the *largest* θ^k satisfying the conditions of Lemma 1 with $\alpha = 2\beta$.

The choice of θ^k in our algorithm will involve finding the roots of a quartic polynomial. Following the proof of our theorem we will point out that the choice for θ^k need not be this involved and it suffices to choose θ^k as the lower bound given in Lemma 2.

Theorem.

(i) The Algorithm has iteration complexity $O(\sqrt{n}L)$.

Let $\{(x^k, s^k), (\hat{x}^k, \hat{s}^k)\}$ be generated by the Algorithm. Assume that

$$\lim(x^k, s^k) = \lim(\hat{x}^k, \hat{s}^k) = (x^*, s^*).$$

Then

(ii) $1 - \theta^k \to 0$.

(iii) $(x^k)^T s^k \to 0$ Q-superlinearly.

(iv) $X^k s^k \to 0$, component-wise, Q-superlinearly.

Assume that (x^*, s^*) is nondegenerate. Then

(v) $1 - \theta^k = O((x^k)^T s^k)$.

(vi) $(x^k)^T s^k \to 0$ Q-quadratically.

Proof. The proof of (i), i.e., the $O(\sqrt{n}L)$-iteration complexity of the algorithm is as before.

Since the iteration sequence $\{(x^k, s^k)\}$ converges to (x^*, s^*), then from a theorem of Güler and Ye [1] (x^*, s^*) must be a strict complementarity solution of (LP) and (LD). This is guaranteed because all (x^k, s^k) and $(\hat{x}^k, \hat{s}^k)$ lie in $\mathcal{N}(2\beta)$. Without loss of generality, assume that

$$x_j^k \geq \xi \quad \text{and} \quad s_j^k \to 0 \tag{6}$$

for some fixed positive number ξ. Now, at the kth predictor step, we have from (1)

$$x_j^k (d_s)_j + s_j^k (d_x)_j = -x_j^k s_j^k$$

or

$$\frac{(d_s)_j}{s_j^k} + \frac{(d_x)_j}{x_j^k} = -1.$$

From (6), we have

$$\lim_{k\to\infty} \frac{(d_x)_j}{x_j^k} = \frac{1}{\theta^k} \lim_{k\to\infty} \frac{\hat{x}_j^k - x_j^k}{x_j^k} = 0.$$

Hence,

$$\lim_{k\to\infty} \frac{(d_s)_j}{s_j^k} = -1.$$

Thus,

$$\lim_{k\to\infty} \frac{(d_x)_j (d_s)_j}{x_j^k s_j^k} = 0,$$

so that

$$\lim_{k\to\infty} \delta_j^k = \lim_{k\to\infty} \frac{(d_x)_j (d_s)_j}{\mu^k} = 0,$$

since $x_j^k s_j^k = O(\mu^k)$. The above limit holds for all $1 \le j \le n$. Hence,

$$\lim_{k\to\infty} \|\delta^k\| = 0,$$

which together with (5) implies

$$\lim_{k\to\infty} (1 - \theta^k) = 0.$$

This establishes (ii). Conclusion (iii) follows from (3) and (ii), and in turn (iv) follows from (iii) exactly as in the proof of Corollary 3.2 of Zhang et al. [7]. If we assume that (x^*, s^*) is nondegenerate, then from Lemma 3.2 of Zhang et al. [7],

$$(d_x)_j / x_j^k = O((x^k)^T s^k)$$

for j such that $x_j^* > 0$ and

$$(d_s)_j / s_j^k = O((x^k)^T s^k)$$

for j such that $x_j^* = 0$. Thus, for all j

$$\frac{|(d_x)_j(d_s)_j|}{x_j^k s_j^k} = O((x^k)^T s^k)$$

and

$$|\delta_j^k| = \frac{|(d_x)_j(d_s)_j|}{\mu^k} = O((x^k)^T s^k).$$

This estimate together with (5) gives (v). From (3) we see (vi) follows from (v). This proves the theorem. ∎

4. Further Remarks

The assumption that the iteration sequence converges is used in all other superlinear convergence results (Kojima et al. [2], and Zhang et al. [6],[7],[8]). Although this property is "observed" in practice for LP problems whose solution set is bounded, it is a challenging open question as to what conditions imply the convergence of the iteration sequence. Under the assumption of nondegeneracy the LP solution pair is unique. This coupled with the facts that the level sets of the duality gap function are bounded and the duality gap is decreasing monotonically to zero implies that the iteration sequence converges. This is implicitly used in Theorem 7.1 of Zhang and Tapia [6]. Hence conclusions (v) and (vi) of our theorem do not require the convergence assumption made in the theorem.

From Lemma 2 and the proof of the Theorem, it follows that one can simply choose

$$\theta^k = \frac{2}{\sqrt{1 + 4\|\delta^k\|/\beta} + 1}$$

and not search for the largest θ^k satisfying the conditions of Lemma 1 with $\alpha = 2\beta$. The global and local behavior of the algorithm will remain the same.

It has been observed in practice that the $O(\sqrt{n}L)$ algorithm is generally slower than some of the $O(nL)$ algorithms (or other non-polynomial algorithms). One conclusion that follows from this note is that the practical efficiency of these $O(nL)$ algorithms is not entirely due to their local behavior.

References

[1] O. Güler and Y. Ye, "Convergence behavior of some interior-point algorithms," Working Paper 91-4, The College of Business Administration, The University of Iowa (Iowa City, IA.), 1991.

[2] M. Kojima, Y. Kurita and S. Mizuno, "Large step interior-point algorithms for linear complementarity problems," Research Report B-243, Department of Information Sciences, Tokyo Institute of Technology (Tokyo, Japan), 1991.

[3] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," *Mathematical Programming* **44** (1989) 1-26.

[4] S. Mizuno, M. J. Todd, and Y. Ye, "On adaptive-step primal-dual interior-point algorithms for linear programming," Technical Report No. 944, School of ORIE, Cornell University (Ithaca, NY.), 1990, to appear in *Mathematics of Operations Research.*

[5] Y. Ye, "Further developments on interior algorithms for convex quadratic programming," unpublished manuscript, Department of Engineering-Economic Systems, Stanford University (Stanford, CA.), 1987.

[6] Y. Zhang and R. A. Tapia, "A quadratically convergent polynomial primal-dual interior-point algorithm for linear programming," TR90-40, Department of Mathematical Sciences, Rice University (Houston, TX.), 1990.

[7] Y. Zhang, R. A. Tapia and J. E. Dennis, "On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms," TR90-6, Department of Mathematical Sciences, Rice University (Houston, TX.), 1990.

[8] Y. Zhang, R. A. Tapia and F. Potra, "On the superlinear convergence of interior-point algorithms for a general class of problems," TR90-7, Department of Mathematical Sciences, Rice University (Houston, TX.) 1990.