SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

TECHNICAL REPORT NO. 879

December 1989

ANTICIPATED BEHAVIOR OF KARMARKAR'S ALGORITHM

by

Michael J. Todd

Research of the author was supported in part by NSF grant DMS-8904406 and ONR contract N00014-87-K0212.

Abstract

We show that, under a plausible assumption on the distribution of the projected cost vector in an iteration of Karmarkar's linear programming algorithm, the decrease in the potential function will be $\Omega(n/\log n)$ with probability approaching one as n, the number of variables, tends to infinity. Thus the "anticipated" number of steps to obtain precision 2^{-q} is $O(q\ \ell n\ n)$. This contrasts with a worst-case bound on $O(qn)$.

1. Introduction

The aim of this paper is to attempt to explain the excellent behavior of interior point algorithms for linear programming. Several computational studies have demonstrated the potential of various variants of Karmarkar's algorithm [10]: see [1,11,12,13,16]. While the theoretical worst-case bounds of such algorithms are $O(Ln)$ iterations for projective methods such as Karmarkar's and $O(L\sqrt{n})$ iterations for various path-following (e.g., Renegar [17]) and potential-reduction (e.g., Ye [21]) methods, in practice the number of iterations required seems to grow very slowly with n. (Here n is the number of inequalities (variables in the standard form) and L the length of the (integer) data. Alternatively, L can be thought of as the number of significant figures required in the final solution.) In particular, the results of Lustig, Shanno and Gregory [12] indicate that the number of iterations required by their code grows linearly with $\ell n\ n$.

Here we consider the "anticipated" behavior of Karmarkar's original algorithm with optimal step size. By this we mean the following. At any iteration, we make a plausible assumption on the distribution of the projected cost vector, and then prove that Karmarkar's potential function can be decreased by $\Omega(n/\ell n\ n)$ with probability approaching one as $n \to \infty$ under this distribution. The anticipated number of iterations is then defined to be the number of iterations required if this high-probability event actually occurs every iteration (or at least once in every fixed number of iterations). Note that the probabilistic assumptions we make at each iteration may not be consistent with each other nor with a fixed probability distribution on the original problem; hence our results differ from those on the expected number of iterations. From well-known properties of the potential function, it follows that the anticipated number of iterations is $O(L\ \ell n\ n)$, or $O(q\ \ell n\ n)$ to obtain precision 2^{-q}, in contrast with a worst-case bound of $O(Ln)$ or $O(qn)$ respectively. In addition, we show that a step of $(1-\epsilon)$ times the distance to the boundary also achieves a similar reduction of the potential function with high probability. Similar step-size rules are often used in practice.

The concept of anticipated number of steps was introduced in Gonzaga and Todd [9], and has also been studied by Mizuno, Todd and Ye [14,15]. A related analysis of the simplex method was made by Dantzig [6]. Our analysis also applies to the Todd-Burrell algorithm [20], which adjusts lower

bounds on the optimal value, and to the standard-form variants of Anstreicher [2], Gay [7], de Ghellinck and Vial [8] (given a feasible starting solution), Jensen and Steger [19], and Ye and Kojima [22]. Anstreicher [4] has recently given an example where Karmarkar's algorithm requires $O(L\ \ell n\ n)$ iterations. This nicely complements our results, although in Anstreicher's example the potential reduction per iteration is constant.

Section 2 outlines Karmarkar's algorithm. In section 3 we introduce the probabilistic model and analyze the decrease in the potential function obtained. Section 4 addresses the step size.

2. Karmarkar's Algorithm

Consider the linear programming problem in Karmarkar's canonical form:

$$(P) \qquad \begin{aligned} \min\ & c^T x \\ Ax &= 0 \\ e^T x &= n \\ x &\geq 0 \end{aligned}$$

where e denotes a vector of ones in R^n and it is assumed that A is $m \times n$ of rank m, that $Ae = 0$ and $c^T e > 0$ so that $x^0 = e$ is an initial positive feasible solution, and that the optimal value of (P) (which is clearly bounded) is zero. Karmarkar's algorithm generates a sequence $\{x^k\}$ of positive feasible solutions with $c^T x^k \to 0$. Progress is measured via the potential function

$$f(x) := n\ \ell n\ c^T x - \textstyle\sum_j \ell n\ x_j \qquad (1)$$

defined on all positive feasible solutions. Indeed, if $f(x) < f(x^0) - \gamma$ for such an x, then it is not hard to show that $c^T x \leq \exp(-\gamma/n) c^T x^0$. Hence $c^T x^k \leq 2^{-q} c^T x^0$ in $O(qn)$ steps if we can reduce

f by a constant at each iteration, and in $O(q\ \ell n\ n)$ steps if the reduction is by $\Omega(n/\ell n\ n)$ at each iteration (or at least once in every fixed number of iterations, and is not increased in the others).

The iteration from x^k to x^{k+1} proceeds as follows. Let $X := X_k = \mathrm{diag}(x^k)$, and consider the projective transformation

$$x \to \bar{x} := \frac{nX^{-1}x}{e^T X^{-1} x} \quad \text{with inverse} \quad \bar{x} \to x := \frac{nX\bar{x}}{e^T X \bar{x}}\ . \tag{2}$$

Under this transformation (P) becomes

$$(\bar{P}') \qquad \min \frac{\bar{c}^T \bar{x}}{e^T X \bar{x}}$$
$$\bar{A}\bar{x} = 0$$
$$e^T \bar{x} = n$$
$$\bar{x} \geq 0$$

where $\bar{A} := AX$ and $\bar{c} := Xc$. Note that this is a linear fractional programming problem, but since we seek a solution with objective value zero, it is equivalent to solve

$$(\bar{P}) \qquad \min \bar{c}^T \bar{x}$$
$$\bar{A}\bar{x} = 0$$
$$e^T \bar{x} = n$$
$$\bar{x} \geq 0.$$

Note that $x = x^k$ is transformed by (2) into $\bar{x} = e$. Also, if we define $\bar{f}(\bar{x}) := n\ \ell n\ \bar{c}^T \bar{x} - \sum_j \ell n\ \bar{x}_j$, it is easy to show that when x and $\bar{x}$ correspond by (2),

$$\bar{f}(\bar{x}) = f(x) + \ell n \det X.$$

Thus potential differences in the transformed and the original spaces coincide, and we therefore want to find a feasible $\bar{x}$ with potential suitably reduced as compared to e. Let

$$\bar{x}(\alpha) := e - \alpha \frac{g}{||g||}, \tag{3}$$

where g is the projected cost vector of $(\bar{P})$,

$$\bar{g} = \bar{c}_p := P_{\begin{bmatrix} \bar{A} \\ e^T \end{bmatrix}} \bar{c}. \tag{4}$$

Here P_M denotes orthogonal projection onto the null space of M. Hence $-\bar{g}$ is the projected steepest descent direction for the linear objective function of $(\bar{P})$. More relevantly, it is also proportional to the projected steepest descent direction for the potential function $\bar{f}$ at e [20].

Karmarkar shows that a fixed value of $\alpha < 1$ gives $\bar{f}(\bar{x}(\alpha)) \leq \bar{f}(e) - \frac{1}{8}$, and then by defining x^{k+1} as the inverse image of this $\bar{x}(\alpha)$ under the transformation (2) we have our desired iterate. Todd and Burrell [20] also prove this, and suggest performing a line search on $\bar{f}$ to choose α. Note that $\alpha < 1$ ensures that $\bar{x}(\alpha)$, and hence x^{k+1}, is positive.

We will follow Anstreicher [3] in parametrizing instead by a parameter λ so that $\lambda = 1$ corresponds to the maximum feasible step. Let $\bar{g}_{min}$ and $\bar{g}_{max}$ denote the minimum and maximum components of $\bar{g}$ respectively. Since $e^T\bar{g} = 0$ and $\bar{g} \neq 0$, $\bar{g}_{min} < 0 < \bar{g}_{max}$. Then we redefine $\bar{x}(\cdot)$ by

$$\bar{x}(\lambda) := e - \lambda \frac{\bar{g}}{\bar{g}_{max}}. \tag{5}$$

The basis for our analysis in the next section is the guaranteed reduction in $\bar{f}$ proved by Anstreicher [3]:

$$\bar{f}(e) - \bar{f}(\bar{x}(\lambda)) \geq \frac{\lambda||\bar{g}||^2}{|\bar{g}_{min}|\,\bar{g}_{max}} + \sum_j \ell n \left(1 - \frac{\lambda \bar{g}_j}{\bar{g}_{max}}\right). \tag{6}$$

Now Karmarkar [10] showed that

$$\ell n(1-\theta) \geq -\theta - \frac{\theta^2}{2(1-\theta)} \quad \text{for} \quad 0 \leq \theta < 1,$$

and a Taylor expansion proves that

$$\ell n(1-\theta) \geq -\theta - \frac{\theta^2}{2} \quad \text{for} \quad \theta < 0,$$

so if $\lambda \leq 1-\epsilon$ we easily obtain from (6)

$$\bar{f}(e) - \bar{f}(\bar{x}(\lambda)) \geq \frac{\lambda \|\bar{g}\|^2}{|\bar{g}_{min}|\, \bar{g}_{max}} - \frac{\lambda^2 \|\bar{g}\|^2}{2\epsilon\, \bar{g}_{max}^2}\,. \tag{7}$$

3. Probabilistic Analysis of Potential Difference

The assumption we make is that

$$\bar{g} \text{ is distributed with a rotationally symmetric distribution in the subspace } e^{\perp} := \{u \in \mathbb{R}^n : e^T u = 0\}. \tag{8}$$

(Of course, there is no justification for this assumption; indeed, the asymptotic behavior of $\bar{g}$ is known in the nondegenerate case (Asic et al. [5]). However, we hope that the results obtained give insight into the observed behavior of Karmarkar's algorithm.) Since the scale of $\bar{g}$ is immaterial in (5)-(7), we can assume that

$$\bar{g} = h - \frac{h^T e}{n} e, \quad \text{where} \qquad h \overset{d}{=} N(O,I), \tag{9}$$

i.e. the components of h are independent random variables each with a standard normal distribution. Then

$$\begin{aligned} \|\bar{g}\|^2 &= \|h\|^2 - (h^T e)^2/n, \\ \bar{g}_{min} &= h_{min} - h^T e/n, \text{ and} \\ \bar{g}_{max} &= h_{max} - h^T e/n, \end{aligned} \tag{10}$$

where h_{min} and h_{max} have the obvious meanings.

Let $f(n)$ be a sequence of random variables and $c(\cdot)$ a function. We write

$$f(n) \sim c(n)$$

to denote: for all $\epsilon > 0$,

$$P\{f(n) \in [(1-\epsilon)c(n), (1+\epsilon)c(n)]\} \to 1 \text{ as } n \to \infty.$$

Similarly,

$$f(n) \lesssim c(n)$$

denotes: for all $\epsilon > 0$,

$$P\{f(n) \leq (1+\epsilon)c(n)\} \to 1 \text{ as } n \to \infty.$$

It is then well known that

$$\|h\|^2 \stackrel{d}{=} \chi^2(n) \sim n \text{ and}$$

$$(h^T e)^2/n \stackrel{d}{=} \chi^2(1) \lesssim n^{1/2},$$

so that

$$\|\bar{g}\|^2 \sim n \tag{11}$$

and

$$h^T e/n \stackrel{d}{=} N(0, n^{-1/2})$$

so that

$$|h^T e/n| \lesssim 1, \tag{12}$$

where $\chi^2(r)$ denotes a chi-square random variable with r degrees of freedom.

Moreover, extreme-value theory (see Resnick [18]) shows that

$$h_{max} \sim (2\ \ell n\ n)^{1/2} \quad \text{and}$$

$$|h_{min}| \sim (2\ \ell n\ n)^{1/2},$$

so that (12) gives (13)

$$\bar{g}_{max} \sim (2\ \ell n\ n)^{1/2} \quad \text{and}$$

$$|\bar{g}_{min}| \sim (2\ \ell n\ n)^{1/2}.$$

Now let $\lambda = \frac{1}{2}$, $\epsilon = \frac{1}{2}$ in (7). Then

$$\frac{\lambda||\bar{g}||^2}{|\bar{g}_{min}|\ \bar{g}_{max}} - \frac{\lambda^2||\bar{g}||^2}{2\epsilon\ \bar{g}^2_{max}} \sim \frac{n}{8\ \ell n\ n}$$

from (11) and (13). We have proved:

Theorem 1. Under assumption (8), the reduction in the potential function obtained by searching on $\bar{f}$ in the direction $-\bar{g}$ given by (4) is $\Omega(n/\ell n\ n)$ with probability converging to one as $n \to \infty$.

Indeed, such a reduction is achieved if we merely take as our step size half the maximum feasible step size.

Corollary. The anticipated number of iterations required by Karmarkar's algorithm with optimal step size to achieve $c^T x^k \leq 2^{-q} c^T x^0$ is $O(q\ \ell n\ n)$.

4. Probabilistic Analysis of the Step Size

Here we show that a reduction in the potential function of $\Omega(n/\ell n\ n)$ can also be achieved by choosing

$$\lambda = 1\text{-}\epsilon \quad \text{for any fixed} \quad \epsilon > 0,$$

i.e. by moving a fixed proportion $1-\epsilon$ of the way to the boundary, with probability approaching one as $n \to \infty$. For this we need to use (6) rather than (7) because of the ϵ in the denominator in (7).

We divide the sum in (6) into two terms:

$$\sum_j \ell n \left(1 - \frac{\lambda \bar{g}_j}{\bar{g}_{max}}\right) = \sum_{j:\, \bar{g}_j > \left(\frac{\ell n\, n}{8}\right)^{1/2}} \ell n\left(1 - \frac{\lambda \bar{g}_j}{\bar{g}_{max}}\right) + \sum_{j:\, \bar{g}_j \le \left(\frac{\ell n\, n}{8}\right)^{1/2}} \ell n\left(1 - \frac{\lambda \bar{g}_j}{\bar{g}_{max}}\right)$$

$$\ge \frac{-\lambda^2}{\bar{g}_{max}^2}\left(\sum_{j:\, \bar{g}_j > \left(\frac{\ell n\, n}{8}\right)^{1/2}} \frac{\bar{g}_j^2}{2\epsilon} + \sum_{j:\, \bar{g}_j < \left(\frac{\ell n\, n}{8}\right)^{1/2}} \frac{\bar{g}_j^2}{2\max\{\epsilon,\, 1 - (\frac{\ell n\, n}{8})^{1/2}/\bar{g}_{max}\}}\right). \qquad (14)$$

Now, using (11) and (13) we find that

$$\text{the second sum in (14)} \lesssim \frac{n}{2.\frac{3}{4}} = \frac{2n}{3}. \qquad (15)$$

Let

$$S_n := \sum \{\bar{g}_j^2 : \bar{g}_j > (\ell n\, n/8)^{1/2}\}$$

and

$$T_n(k) := \sum \{\bar{g}_j^2 : \bar{g}_j > k\}$$

for each fixed k. Note that, for any k, $S_n \le T_n(k)$ for all sufficiently large n. By the strong law of large numbers

$$\frac{T_n(k)}{n} = \frac{1}{n} \sum \bar{g}_j^2 I_{\{\bar{g}_j > k\}}$$

$$\to E\{\bar{g}_1^2 I_{\{\bar{g}_1 > k\}}\} \text{ in probability.}$$

But this expectation converges to 0 as $k \to \infty$. So for any $\eta > 0$, $P\left\{\frac{S_n}{n} \ge \eta\right\} \le P\left\{\frac{T_n(k)}{n} \ge \eta\right\} \to 0$

as $n \to \infty$. If we let $\eta < \epsilon/3$, we find

$$\text{the first sum in (14)} \lesssim \frac{\eta n}{2\epsilon} = \frac{n}{6} . \tag{16}$$

Now (13)-(16) yield

$$- \sum_j \ell n \left(1 - \frac{\lambda \bar{g}_j}{\bar{g}_{max}}\right) \lesssim \frac{5\lambda^2 n}{12\ \ell n\ n} .$$

Also, (11) and (13) show that

$$\frac{\lambda \|\bar{g}\|^2}{|\bar{g}_{min}|\bar{g}_{max}} \sim \frac{\lambda n}{2\ \ell n\ n} ,$$

and these two results prove:

Theorem 2. Under assumption 8, the reduction in the potential function obtained by moving a proportion $(1-\epsilon)$ of the way to the boundary in the direction $-\bar{g}$ given by (4) for any fixed $\epsilon > 0$ is $\Omega(n/\ell n\ n)$ with probability converging to one as $n \to \infty$.

Acknowledgement. I am grateful to L. Khachiyan and G. Samorodnitsky for helpful comments related to this work. In particular, the analysis of S_n via $T_n(k)$ is due to Samorodnitsky and substantially simplifies my earlier approach.

References

1. I. Adler, N. Karmarkar, M.G.C. Resende, and G. Veiga, "An implementation of Karmarkar's algorithm for linear programming," Mathematical Programming 44 (1989).

2. K.M. Anstreicher, "A monotonic projective algorithm for fractional linear programming," Algorithmica 1 (1986) 483-498.

3. K.M. Anstreicher, "The worse-case step in Karmarkar's algorithm", Mathematics of Operations Research, 14 (1989) 294-302.

4. K.M. Anstreicher, "On the performance of Karmarkar's algorithm over a sequence of iterations," CORE Discussion Paper No. 8934, Universite Catholique de Louvain, Belguim, December 1989.

5. M.D. Asic, V.V. Kovacevic-Vujcic, and M.D. Radosavljevic-Nikolic, (1987), "Asymptotic behavior of Karmarkar's method for linear programming," manuscript, Faculty of Organizational Science, Belgrade University, Yugoslavia, to appear in Mathematical Programming.

6. G.B. Dantzig, "Expected number of steps of the simplex method for a linear program with a convexity constraint," Technical Report SOL 80-3, Department of Operations Research, Stanford University, Stanford, CA 1980.

7. D. Gay, "A variant of Karmarkar's linear programming algorithm for problems in standard form," Mathematical Programming 37 (1987) 81-90.

8. G. de Ghellinck and J.-Ph. Vial, "A polynomial Newton method for linear programming," Algorithmica 1 (1986) 425-453.

9. C. C. Gonzaga and M.J. Todd, "An $0(\sqrt{n}\ L)$ -iteration large-step primal-dual affine algorithm for linear programming," Technical Report No. 862, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1989.

10. N. Karmarkar, "A new polynomial time algorithm for linear programming," Combinatorica 4 (1984) 373-395.

11. I.J. Lustig, R.E. Marston and D.F. Shanno, "Computational experience with a primal-dual interior point method for linear programming," Technical Report SOR 89-17, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, 1989.

12. I.J. Lustig, D.F. Shanno and J.W. Gregory, "The primal-dual interior point method on a Cray supercomputer," Technical Report SOR 89-21, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, 1989.

13. K.A. McShane, C.L. Monma and D.F. Shanno, "An implementation of a primal-dual interior point method for linear programming," ORSA Journal on Computing 1 (1989), 70-83.

14. S. Mizuno, M.J. Todd and Y. Ye, "Anticipated behavior of path-following algorithms for linear programming," Technical Report No. 878, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1989.

15. S. Mizuno, M.J. Todd and Y. Ye, "Anticipated behavior of long-step algorithms for linear programming", in preparation.

16. C.L. Monma and A.J. Morton, "Computational experience with a dual affine variant of Karmarkar's method for linear programming," Operations Research Letters 6 (1987), 261-167.

17. J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," Mathematical Programming 40 (1988), 59-93.

18. S.I. Resnick, Extreme Values, Regular Variation, and Point Processes (Springer-Verlag, New York, 1987).

19. A. Steger, "An extension of Karmarkar's algorithm for bounded linear programming problems," M.S. Thesis, SUNY at Stonybrook, New York (1985).

20. M.J. Todd and B.P. Burrell, "An extension of Karmarkar's algorithm for linear programming using dual variables," Algorithmica 1 (1986) 409-424.

21. Y. Ye, "An $0(n^3L)$ potential reduction algorithm for linear programming," manuscript, Department of Management Sciences, The University of Iowa, Iowa City, Iowa (1988, revised 1989), to appear in Mathematical Programming.

22. Y. Ye and M. Kojima, "Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming," Mathematical Programming 39 (1987) 305-317.