Mathematical Programming 36 (1986) 174-182
North-Holland

A VARIATION ON KARMARKAR'S ALGORITHM FOR SOLVING LINEAR PROGRAMMING PROBLEMS

Earl R. BARNES
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

Received 4 June 1985
Revised manuscript received 3 June 1986

The algorithm described here is a variation on Karmarkar's algorithm for linear programming. It has several advantages over Karmarkar's original algorithm. In the first place, it applies to the standard form of a linear programming problem and produces a monotone decreasing sequence of values of the objective function. The minimum value of the objective function does not have to be known in advance. Secondly, in the absence of degeneracy, the algorithm converges to an optimal basic feasible solution with the nonbasic variables converging monotonically to zero. This makes it possible to identify an optimal basis before the algorithm converges.

Key words: Linear programming, nondegeneracy, dual variables.

1. The algorithm

Consider the linear programming problem

$$\begin{aligned} &\text{minimize} && c^{\mathrm{T}}x \\ &\text{subject to} && Ax=b, \\ & && x\geq 0, \end{aligned} \tag{1.1}$$

where c and x are n-dimensional column vectors, b is an m-dimensional column vector, and A is an $m\times n$ matrix of rank m. The jth column of A will be denoted by a_j. We make the assumption that (1.1) has no degenerate basic feasible solutions and that its dual

$$\begin{aligned} &\text{maximize} && b^{\mathrm{T}}\lambda \\ &\text{subject to} && A^{\mathrm{T}}\lambda\leq c \end{aligned} \tag{1.2}$$

has no degenerate basic solutions. This means that b cannot be written as a positive combination of fewer than m columns of A and that at most m of the equations

$$c_i-a_i^{\mathrm{T}}\lambda=0,\quad i=1,\dots,n,$$

can be satisfied simultaneously. Clearly, (1.1) and (1.2) remain nondegenerate under small perturbations in b and c. In fact, the following is true. If (1.1) is nondegenerate, there exists a number $\varepsilon_1>0$ such that any feasible solution of (1.1) has at least m

components $> \varepsilon_1$. If (1.2) is nondegenerate, there exists a number $\varepsilon_2 > 0$ such that at most m of the inequalities

$$|c_i - a_i^{\mathrm{T}}\lambda| < \varepsilon_2, \quad i = 1, \ldots, n,$$

can be satisfied simultaneously. Clearly, numbers ε_1 and ε_2 can be found such that these conditions hold when b and c are perturbed slightly.

Let $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ be a feasible solution of (1.1) satisfying $y_i > 0$, $i = 1, \ldots, n$. If $0 < R < 1$, the ellipsoid

$$\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \leq R^2 \tag{1.3}$$

lies in the interior of the positive orthant in E_n. To see this observe that if $x_j \leq 0$ for some j, then

$$\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \geq \frac{(x_j - y_j)^2}{y_j^2} \geq 1 > R^2.$$

This implies that we can obtain a feasible solution of (1.1) satisfying $c^{\mathrm{T}}x < c^{\mathrm{T}}y$ by solving the following problem:

$$\begin{aligned} &\text{minimize} \quad c^{\mathrm{T}}x \\ &\text{subject to} \quad Ax = b, \\ &\qquad\qquad \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i^2} \leq R^2. \end{aligned} \tag{1.4}$$

Note that the constraint $x \geq 0$ has been replaced by (1.3), which is easier to handle.

To solve (1.4) let $\lambda = (\lambda_1, \ldots, \lambda_m)^{\mathrm{T}}$ be a vector of Lagrange multipliers corresponding to the constraints $Ax = b$. Let $D = \mathrm{diag}(y_1, \ldots, y_n)$. Since $y > 0$, D is nonsingular. For any x satisfying (1.4) we have

$$\begin{aligned} c^{\mathrm{T}}y - c^{\mathrm{T}}x &= \{c - A^{\mathrm{T}}\lambda\}^{\mathrm{T}}(y - x) = [D(c - A^{\mathrm{T}}\lambda)]^{\mathrm{T}} D^{-1}(x - y) \\ &\leq \|D(c - A^{\mathrm{T}}\lambda)\| \, \|D^{-1}(x - y)\| \leq \|D(c - A^{\mathrm{T}}\lambda)\| R, \end{aligned} \tag{1.5}$$

the inequalities being obtained from Schwartz's inequality and (1.4). Equality holds throughout (1.5) if

$$D(c - A^{\mathrm{T}}\lambda) = \gamma D^{-1}(x - y) \tag{1.6}$$

for some constant γ, and if

$$\|D^{-1}(x - y)\| = R.$$

These conditions imply that

$$\gamma = \frac{\|D(c - A^{\mathrm{T}}\lambda)\|}{R}.$$

Substituting this in (1.6) gives

$$x = y - R\frac{D^2(c - A^{\mathrm{T}}\lambda)}{\|D(c - A^{\mathrm{T}}\lambda)\|}. \tag{1.7}$$

The condition $Ax = Ay = b$ implies that $AD^2(c - A^{\mathrm{T}}\lambda) = 0$, or

$$\lambda = (AD^2A^{\mathrm{T}})^{-1}AD^2c. \tag{1.8}$$

Writing (1.5) as

$$c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}y - R\|D(c - A^{\mathrm{T}}\lambda)\| \tag{1.9}$$

we see that the minimum is given by the right-hand side of (1.9) and is attained when x and λ are given by (1.7) and (1.8), respectively. This suggests an algorithm for iteratively finding the solution of (1.1). The algorithm can be stated formally as follows.

Algorithm. Let $x^0 > 0$ satisfying $Ax^0 = b$ be given. In general, if x^k is known, define

$$D_k = \mathrm{diag}(x_1^k, \ldots, x_n^k)$$

and compute $x^{k+1} > 0$ by the formula

$$x^{k+1} = x^k - R\frac{D_k^2(c - A^{\mathrm{T}}\lambda_k)}{\|D_k(c - A^{\mathrm{T}}\lambda_k)\|} \tag{1.10}$$

where

$$\lambda_k = (AD_k^2A^{\mathrm{T}})^{-1}AD_k^2c.$$

Theorem 1.1. *If* (1.1) *has a bounded solution, the sequence* $\{x^k\}$ *defined by* (1.10) *converges to a solution of* (1.1) *that is an extreme point of the constraint set defined by* $Ax = b$, $x \geqslant 0$.

Proof. We have seen that

$$c^{\mathrm{T}}x^{k+1} = c^{\mathrm{T}}x^k - R\|D_k(c - A^{\mathrm{T}}\lambda_k)\|$$

for each k. Since the numbers $c^{\mathrm{T}}x^k$ are decreasing and bounded below, the sequence $\{c^{\mathrm{T}}x^k\}$ converges. This implies that

$$\lim_{k\to\infty}\|D_k(c - A^{\mathrm{T}}\lambda_k)\| = 0.$$

Let $\varepsilon > 0$ be given satisfying $\varepsilon < \varepsilon_1$ and $\varepsilon < \varepsilon_2$—see our discussion of non-degeneracy of (1.1) and (1.2). Choose k so large that

$$\|D_r(c - A^{\mathrm{T}}\lambda_r)\| < \varepsilon^2$$

for $r \geqslant k$. There then exist n-m values of i for which

$$|c_i - a_i^{\mathrm{T}}\lambda_r| > \varepsilon_2 > \varepsilon. \tag{1.11}$$

Assume $r \geq k$ is fixed and, without loss of generality, assume that (1.11) holds for $i = m+1, \ldots, n$.

For each i we have

$$x_i^r |c_i - a_i^{\mathrm{T}} \lambda_r| \leq \left\{ \sum_{j=1}^{n} (x_j^r)^2 (c_j - a_j^{\mathrm{T}} \lambda_r)^2 \right\}^{1/2} = \|D_r(c - A^{\mathrm{T}} \lambda_r)\| < \varepsilon^2. \tag{1.12}$$

It therefore follows from (1.11) that

$$x_i^r < \varepsilon \quad \text{for } i = m+1, \ldots, n. \tag{1.13}$$

Since $Ax^r = b$ and (1.1) is not degenerate, x^r has at least m components $> \varepsilon_1 > \varepsilon$. It therefore follows from (1.13) that

$$x_i^r > \varepsilon_1, \quad i = 1, \ldots, m. \tag{1.14}$$

It now follows from (1.12) and the condition $\varepsilon_1 > \varepsilon$ that

$$|c_i - a_i^{\mathrm{T}} \lambda_r| < \varepsilon, \quad i = 1, \ldots, m.$$

This condition implies that the vectors $a_1, \ldots, a_m$ are linearly independent if ε is sufficiently small, for it is easy to see that if these vectors were dependent, the dual problem (2), with c slightly perturbed, would have a degenerate basic solution.

Let B denote the feasible basis $(a_1, \ldots, a_m)$ and let $x_B^r = (x_1^r, \ldots, x_m^r)^{\mathrm{T}}$. We then have

$$x_B^r - B^{-1} b = - \sum_{i=m+1}^{n} x_i^r B^{-1} a_i \tag{1.15}$$

and it follows from (1.13) that x_B^r is very close to the basic feasible solution $B^{-1}b$ for ε sufficiently small or, equivalently, for k sufficiently large and $r \geq k$. In this way we can associate each of the feasible solutions x^r, $r \geq k$, with a basic feasible solution of (1.1). If z and y are extreme points of the set defined by $Ax = b$, $x \geq 0$, we clearly have $\|z - y\| > \sqrt{2}\,\varepsilon_1$ by our nondegeneracy assumption. Thus for ε sufficiently small x^r is close to a unique extreme point of the constraint set for (1.1).

From (1.3) we see that

$$\frac{(x_i^{r+1} - x_i^r)^2}{(x_i^r)^2} \leq R^2,$$

which means that $x_i^{r+1} \geq (1-R) x_i^r$ for each i and each r. This means that if ε is sufficiently small and $x_i^r > \varepsilon_1$, we cannot have $x_i^{r+1} < \varepsilon$. Thus we see from conditions (1.13) and (1.14) that the basic feasible solution of (1.1) associated with x^r will also be associated with x^{r+1} for r sufficiently large. Thus conditions (1.13) and (1.14) hold for all r sufficiently large and for some basis B, which we will continue to assume to be $B = (a_1, \ldots, a_m)$. We will call the variables satisfying (1.13) nonbasic variables and those satisfying (1.14) basic variables. Since ε in (1.13) can be chosen arbitrarily small, we have

$$\lim_{r \to \infty} x_i^r = 0 \tag{1.16}$$

for each nonbasic variable x_i^r.

Next we will show that the Lagrange multipliers λ_r converge to the simplex multiplier $(B^{\mathrm{T}})^{-1}c_B$ as $r\to\infty$, where c_B is the m-dimensional vector $c_B=(c_1,\ldots,c_m)^{\mathrm{T}}$ corresponding to the basic variables.

Let $\hat{D}_r=\operatorname{diag}(x_1^r,\ldots,x_m^r)$. Then

$$AD_r^2A^{\mathrm{T}}=\sum_{i=1}^{n}(x_i^r)^2a_ia_i^{\mathrm{T}}=B\hat{D}_r^2B^{\mathrm{T}}+\sum_{i=m+1}^{n}(x_i^r)^2a_ia_i^{\mathrm{T}}=B\hat{D}_r^2B^{\mathrm{T}}+M_1$$

where $\|M_1\|=\mathrm{O}(\varepsilon^2)$. This implies that

$$\begin{aligned}(AD_r^2A^{\mathrm{T}})^{-1}&=(I+(B\hat{D}_r^2B^{\mathrm{T}})^{-1}M_1)^{-1}(B\hat{D}_r^2B^{\mathrm{T}})^{-1}\\&=\{I-(B\hat{D}_r^2B^{\mathrm{T}})^{-1}M_1+((B\hat{D}_r^2B^{\mathrm{T}})^{-1}M_1)^2-\cdots\}(B\hat{D}_r^2B^{\mathrm{T}})^{-1}\\&=(B\hat{D}_r^2B^{\mathrm{T}})^{-1}+M_2\end{aligned}$$

where $\|M_2\|=\mathrm{O}(\varepsilon^2)$.

Let N denote the $m\times(n-m)$ matrix

$$N=[(x_{m+1}^r)^2a_{m+1},\ldots,(x_n^r)^2a_n]$$

and let $c_N=(c_{m+1},\ldots,c_n)^{\mathrm{T}}$. Then

$$AD_r^2c=B\hat{D}_r^2c_B+Nc_N$$

and

$$\lambda_r=(AD_r^2A^{\mathrm{T}})^{-1}AD_r^2c=(B\hat{D}_r^2B^{\mathrm{T}})^{-1}B\hat{D}_r^2c_B+e_r=(B^{\mathrm{T}})^{-1}c_B+e_r$$

where $\|e_r\|=\mathrm{O}(\varepsilon^2)$. Since we can let $\varepsilon\to0$ as $r\to\infty$, we have

$$\lim_{r\to\infty}\lambda_r=(B^{\mathrm{T}})^{-1}c_B \tag{1.17}$$

as we claimed.

Now consider the ith component of equation (1.8). We have

$$x_i^{r+1}=x_i^r-R\frac{(x_i^r)^2(c_i-a_i^{\mathrm{T}}\lambda_r)}{\|D_r(c-A^{\mathrm{T}}\lambda_r)\|}.$$

This shows that

$$x_i^{r+1}>x_i^r\quad\text{if } c_i-a_i^{\mathrm{T}}\lambda_r<0 \tag{1.18}$$

and

$$x_i^{r+1}<x_i^r\quad\text{if } c_i-a_i^{\mathrm{T}}\lambda_r>0.$$

(Compare this with the condition for nonbasic variables to increase in the simplex algorithm.)

If x_i^r is a nonbasic variable, it is impossible for $c_i-a_i^{\mathrm{T}}\lambda_r$ to change signs as r increases for sufficiency large values of r. This follows from our nondegeneracy assumption together with the fact that $c_i-a_i^{\mathrm{T}}\lambda_r$ cannot change signs without coming close to zero because of (1.17). It follows from (1.16) and (1.18) that

$$c_i-a_i^{\mathrm{T}}\lambda_r>0 \tag{1.19}$$

for each nonbasic variable if r is sufficiently large. This means that eventually each nonbasic variable decreases monotonically to zero.

Let $x_B = B^{-1}b$. It follows from (1.15) and (1.16) that $x_B^r \to x_B$ as $r \to \infty$. To complete the proof of the theorem we will show that x_B is an optimal basic feasible solution of (1.1). We have

$$c_B^{\mathrm{T}} x_B = c_B^{\mathrm{T}} B^{-1} b = b^{\mathrm{T}} \lambda$$

where $\lambda = (B^{\mathrm{T}})^{-1} c_B$ is the limit (1.15). Thus the vectors $x = (x_B, 0, \ldots, 0)^{\mathrm{T}}$ and λ make the primal and dual objectives in (1.1) and (1.2) equal. Also, complementary slackness

$$x^{\mathrm{T}}(c - A^{\mathrm{T}}\lambda) = 0$$

holds. Finally, we have

$$c_i - a_i^{\mathrm{T}}\lambda = \lim_{r\to\infty}(c_i - a_i^{\mathrm{T}}\lambda_r) \geq \varepsilon_2 > 0$$

for x_i as nonbasic variable. It follows that

$$A^{\mathrm{T}}\lambda \leq c;$$

hence λ is feasible for the dual. Thus x and λ satisfy the necessary and sufficient conditions for x to be a basic feasible solution of (1.1). This completes the proof of the theorem.

2. The rate of convergence

Theorem 2.1. *Let x^* denote the solution of* (1.1). *The sequence $\{x^k\}$ generated by* (1.10) *satisfies*

$$c^{\mathrm{T}}x^{k+1} - c^{\mathrm{T}}x^* \leq \left(1 - \frac{R}{\sqrt{n - m + \varepsilon_k}}\right)(c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^*) \tag{2.1}$$

where $\{\varepsilon_k\}$ is a sequence of positive numbers converging to 0 as $k \to \infty$.

Proof. From our nondegeneracy assumptions we know that $x^* = \lim_{k\to\infty} x^k$ has $n - m$ components equal to 0. For simplicity, assume that $x^* = (x_1^*, \ldots, x_m^*, 0, \ldots, 0)$. Then

$$\begin{aligned} c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^* &= [D_k(c - A^{\mathrm{T}}\lambda_k)]^{\mathrm{T}} D_k^{-1}(x^k - x^*) \leq \|D_k(c - A^{\mathrm{T}}\lambda_k)\|(n - m + \varepsilon_k)^{1/2} \\ &= \frac{1}{R}(c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^{k+1})(n - m + \varepsilon_k)^{1/2} \end{aligned} \tag{2.2}$$

where

$$\varepsilon_k = \sum_{i=1}^{m} \left(\frac{x_i^k - x_i^*}{x_i^k}\right)^2 \to 0$$

as $k \to \infty$. This inequality can be written as

$$c^{\mathrm{T}}x^{k+1} - c^{\mathrm{T}}x^* - (c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^*) \leq -\frac{R}{\sqrt{n-m+\varepsilon_k}}(c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^*),$$

which is equivalent to (2.1).

3. A practical version of the method

In applying our method to (1.1), we first add a new column a_{n+1} to A and a corresponding variable x_{n+1} to x. We define

$$a_{n+1} = b - \sum_{i=1}^{n} a_i.$$

The constraints

$$Ax = \sum_{i=1}^{n+1} a_i x_i = b, \quad x \geq 0 \tag{3.1}$$

are then satisfied by the vector $x = (1, \ldots, 1)^{\mathrm{T}} \in E_{n+1}$. We add a large positive component c_{n+1}, corresponding to x_{n+1}, to c.

Now consider the problem

$$\text{minimize} \quad \sum_{i=1}^{n+1} c_i x_i \tag{3.2}$$

subject to (3.1). We can apply our method to this problem starting at $x^0 = (1, \ldots, 1)^{\mathrm{T}}$. If c_{n+1} is sufficiently large, the solution of (3.2) will have $x_{n+1} = 0$. The variables $(x_1, \ldots, x_n)$ will then form a solution of (1.1).

It is clear from our proof of the theorem in Section 1 that the amount by which the objective $c^{\mathrm{T}}x$ decreases at each step of the algorithm increases if R is increased. This suggests that at each step of the algorithm we should increase R as much as possible subject to the condition that all variables remain nonnegative. From (1.8) we see that $x_i^{k+1} = 0$ if $c_i - a_i^{\mathrm{T}}\lambda_k > 0$ and

$$R = \frac{\|D_k(c - A^{\mathrm{T}}\lambda_k)\|}{x_i^k(c_i - a_i^{\mathrm{T}}\lambda_k)}.$$

In order to insure that $x^{k+1} > 0$ we choose R at the kth step by the formula

$$R = \min_{c_i - a_i^{\mathrm{T}}\lambda_k > 0} \frac{\|D_k(c - A^{\mathrm{T}}\lambda_k)\|}{x_i^k(c_i - a_i^{\mathrm{T}}\lambda_k)} - \alpha, \tag{3.3}$$

where α is a small positive number. In our experiments we have used $\alpha = 1/10$.

We have seen that for k sufficiently large the nonbasic variables, those for which $c_i - a_i^{\mathrm{T}}\lambda_k > 0$, are decreasing monotonically to zero. The choice (3.3) of R drives these variables to zero very fast.

Let k be chosen such that each nonbasic variable x_i^r is less than some small positive number ε for $r \geq k$. From the derivation of (1.15) we see that

$$\|\lambda_r - (B^T)^{-1} c_B\| = O(\varepsilon^2)$$

where B is an optimal basis. Comparing this with (1.13) we see that λ^r is a much more accurate estimate of the optimal dual variable than x^r is of the optimal primal variable. This suggests a stopping criterion for the algorithm. We iterate until m of the reduced costs $c_i - a_i^T \lambda_r$ are small in absolute value and the remaining n-m are positive. To be specific, assume that for some r the numbers $\varepsilon_1 = c_1 - a_1^T \lambda_r, \ldots, \varepsilon_m = c_m - a_{m_{\lambda_r}}^T$ are small in absolute value and the numbers $c_{m+1} - a_{m+1}^T \lambda_r, \ldots, c_n - a_n^T \lambda_r$ are positive. Solve the equations $a_i^T \lambda = c_i$, $i = 1, \ldots, m$ for $\lambda = (B^T)^{-1} c_B$ where $B = (a_1, \ldots, a_m)$ and $c_B = (c_1, \ldots, c_m)$. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)^T$. Then $\|\lambda_r - \lambda\| = O(\|\varepsilon\|)$ and therefore $c_i - a_i^T \lambda > 0$, $i = m+1, \ldots, n$, if r is large enough. This means that $A^T \lambda \leq c$. If r is large enough, the numbers $c_i - a_i^T \lambda_r$, $i = m+1, \ldots, n$, remain positive as r increases. It then follows from (1.13) that $x_B^r \to x_B = B^{-1} b$ as $r \to \infty$. Since $x_B^r > 0$ we have $B^{-1} b > 0$. Also

$$c_B^T x_B = (B^T \lambda)^T B^{-1} b = b^T \lambda.$$

It follows that B is an optimal feasible basis.

The important point to observe here is that we are not suggesting iterating the algorithm until the sequence $\{x^r\}$ converges. We are suggesting that some λ_r, presumably for a reasonably small value of r, be used to determine a basis B. Then x_B and λ are determined from the equations $Bx_B = b$, $B^T \lambda = c_B$. If $x_B > 0$ and $A^T \lambda \leq c$, B is an optimal feasible basis.

4. Acknowledgments

It has been called to my attention that the algorithm discussed in this paper was discovered independently by Vanderbei, Meketon and Freedman [4], and also by Cavalier and Soyster [1]. Our work remains of interest since it presents a different point of view and a more complete convergence analysis than that given in [4] and [1].

A similar algorithm has also been described in [2]. These authors show that Karmarkar's algorithm can be viewed as a special case of a barrier-function method for solving nonlinear programming problems. If the barrier parameter is set to zero in computing their Lagrange multipliers, the algorithm described in this paper is obtained.

References

[1] T.M. Cavalier and A.L. Soyster, "Some computational experience and a modification of the Karmarkar algorithm," The Pennsylvania State University, ISME Working Paper 85-105, 1985.

[2] P.E. Gill, W. Murray, M.A. Saunders, J.A. Tomlin and M.H. Wright, "On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method," Manuscript, Stanford University, 1985.

[3] N. Karmarkar, "A new polynomial-time algorithm for linear programming," Proceedings of the 16th Annual ACM Symposium on Theory of Computing, 1984, pp. 302-311.

[4] R.J. Vanderbei, M.S. Meketon and B.A. Freedman, "A modification of Karmarkar's linear programming algorithm," Manuscript, AT & T Bell Laboratories, Holmdel, New Jersey, June 1985.