An Optimal Basis Identification Technique for Interior-Point Linear Programming Algorithms

R.A. Tapia
Yin Zhang

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

An Optimal Basis Identification Technique
for Interior-Point Linear Programming Algorithms [*][†]

R. A. Tapia[‡]and Yin Zhang[§]

April, 1989 (Revised September, 1990)

Abstract

This work concerns a method for identifying an optimal basis for linear program-ming problems in the setting of interior point methods. To each iterate x^k generated by a primal interior point algorithm, say, we associate an indicator vector q^k with the property that if x^k converges to a nondegenerate vertex x^*, then q^k converges to the 0–1 vector $sign(x^*)$. More interestingly, we show that the convergence of q^k is quadratically faster than that of x^k in the sense that $\|q^k - q^*\| = O(\|x^k - x^*\|^2)$. This clear-cut separation and rapid convergence allow one to infer at an intermediate stage of the iterative process which variables will be zero at optimality and which will not. We also show that under suitable assumptions this method is applicable to dual as well as primal-dual algorithms and can be extended to handle certain types of degen-eracy. Numerical examples are included to corroborate the convergence properties of the indicators. The practical limitations of the indicator technique are also discussed.

1 Introduction

This paper concerns linear programs in the standard form:

$$\text{minimize} \quad c^T x,$$
$$\text{subject to} \quad Ax = b, \tag{1.1}$$
$$x \geq 0.$$

where $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n} (m < n)$ and A has full rank m. The dual linear program of (1.1) is

$$\text{maximize} \quad b^T y,$$
$$\text{subject to} \quad A^T y + z = c, \tag{1.2}$$
$$z \geq 0,$$

where $z \in \mathbf{R}^n$ is the vector of dual slack variables.

The simplex method for linear programming can be viewed as an active set method that utilizes the combinatorial structure of linear programs and has an exponential worst-case complexity. On the other hand, interior-point methods such as the ellipsoid algorithm and the Karmarkar algorithm do not rely on the combinatorial structure and possess polynomial complexity. Recent developments have demonstrated that interior-point algorithms have the real potential to be competitive in practice with the simplex method.

Theoretically, with integer data an interior-point algorithm can be terminated when the current iterate is sufficiently close to an optimal solution and then is rounded to the nearby optimal solution. However, theoretical termination criteria of this kind are difficult to define and are usually inefficient.

A promising approach for improving the efficiency of Karmarkar-type interior point algorithms seems to be the development of reliable techniques for identifying optimal basic and nonbasic variables in the early stages of an interior point iterative process. In this way either an early termination or a reduction in problem size can be obtained. In other words, the efficiency of interior point methods may be improved by utilizing the combinatorial structure of linear programs.

Suppose that an interior point method is generating a sequence $\{x^k\}$ that is converging to an optimal solution of the linear program (1.1). For simplicity, let us assume that x^* is a nondegenerate basic feasible solution. At the k-th iteration, for example, if one can partition, using some identification technique, the current iterate x^k into a set of m likely basic variables and a set of $n - m$ likely nonbasic variables with a reasonable certainty, then one may want to set the nonbasic variable candidates to zero in the constraint equations

$Ax = b$ and solve the resulting square system for the basic variable candidates. If the solution obtained in this manner is indeed a basic feasible solution and the corresponding reduced costs are all nonnegative, then the optimal solution has been obtained and the algorithm can be terminated. Otherwise, one proceeds with the interior point algorithm to the next iteration. In a procedure of this kind, the identification technique plays the central role. In order to successfully terminate the interior point algorithm as early as possible, the identification technique must be reliable, inexpensive, and most importantly, able to identify an optimal basis in an early stage of the iterative process.

In the presence of degeneracies, the situation becomes more complicated. For example, it is no longer a straightforward matter to determine an optimal basis or even to check the optimality of a basic feasible solution in the case of primal degeneracy (i.e., when there are more than $n - m$ zero variables) even after the zero and nonzero variables have been correctly identified. Nevertheless, any information telling us which variables are zero and which are nonzero at optimality is still of some value and may be used to improve the efficiency of interior point methods. For instance, once it has been determined that some variables are zero at optimality, they may be eliminated from the problem, yielding a reduction in the problem size. Also inequality constraints may be removed if their corresponding slack variables are identified as nonzero at optimality. For large scale problems, these reductions in problem size may result in savings in computational effort.

Working primarily with the Karmarkar algorithm or one of its variants the optimal basis identification problem has been considered in recent years by several authors, including Kojima [9], Ye and Todd [19], Asic *et al.* [2], Todd [13], Ye [17] and Gay [6].

In this paper, we will propose and study a new identification technique using an indicator to identify the optimal basis. This indicator will be shown to possess several elegant mathematical properties. However, its practical applicability seems to be limited because it requires nondegeneracy assumptions, and for highly sparse problems the added expense incurred in calculating the indicator may dominate any gain obtained from its use.

This paper is organized as follows. In Section 2, we define our indicator and study its properties. The applications of the indicator to primal, dual and primal-dual algorithms for identifying an optimal basis are developed in Section 3. Section 4 deals with the numerical computation of the indicator. The extension of our indicator technique to degenerate problems is studied in Section 5. Described in Section 6 are methods for randomly generating nondegenerate and primal degenerate problems for use in numerical experimentations. Numerical examples are presented in Section 7 and some concluding remarks are given in the final section.

3

2 Definition and Properties of the Indicator

A key ingredient in essentially every interior point method motivated and influenced by Karmarkar's milestone work of 1984 [8] is a matrix of the form $DA^T(AD^2A^T)^{-1}AD$, where D is a diagonal matrix that changes at every iteration. In primal algorithms (affine variants of the Karmarkar algorithm, for example, see [3], [5] and [16]), we see $D = diag(x)$. In dual algorithms (see Adler *et al.* [1] and Monma and Morton [11], for example), we see $D = [diag(z)]^{-1}$, where z is the vector of dual slack variables. And in primal-dual algorithms (see Kojima *et al.* [10], for example), we see $D^2 = diag(x)[diag(z)]^{-1}$. In this work, we will show that under suitable assumptions, optimal basis information can be obtained from the diagonal of this matrix $DA^T(AD^2A^T)^{-1}AD$, which we shall call the indicator vector or simply the indicator.

Now we give a formal definition of the indicator. For a fixed matrix $A \in R^{m \times n}(m < n)$, consider the matrix-valued function $H : R^n \to R^{n \times n}$ defined by

$$H(d) = DA^T(AD^2A^T)^+AD, \tag{2.1}$$

where $d \in R^n$, $D = diag(d)$ and the superscript "+" denotes the generalized inverse. We will be primarily interested in the function $q : R^n \to R^n$ obtained as the diagonal of $H(d)$, i.e.,

$$q(d) = diag(H(d)) \quad \text{or} \quad q_i(d) = H_{ii}(d), \quad i = 1, 2, 3, ..., n. \tag{2.2}$$

This function $q(d)$ is defined for all $d \in R^n$; however it will not be continuous at points d where the matrix AD^2A^T changes rank in every neighborhood of d. At points d where AD^2A^T has constant rank in some neighborhood of d, $q(d)$ will be infinitely smooth.

Ye and Todd [19] were probably the first to observe that the diagonal elements of such a matrix contain valuable information. In a primal-dual context, they developed an interesting criterion which was guaranteed eventually to identify the optimal basis for a nondegenerate vertex. Their criterion involved several quantities including the diagonal of a matrix of the form (2.1). However, they did not consider the diagonal as an indicator and did not study the properties of the diagonal.

As the first step towards showing that $q(d)$ has several interesting properties, we offer the following lemma.

Lemma 2.1 *If $q(d)$ is given by (2.2), then for all $d \in R^n$ we have*

$$0 \leq q(d) \leq 1.$$

Moreover, if A has no zero columns and AD has full rank, then $q_i(d) = 0$ if and only if $d_i = 0$.

Proof: Observe that both $H(d)$ and $I - H(d)$ are orthogonal projections satisfying $P^T = P$ and $P^2 = P$; they are therefore positive semi-definite and must have nonnegative diagonals. This proves the first part.

The second part follows from the formula

$$q_i(d) = d_i^2 \, a_i^T (AD^2 A^T)^{-1} a_i,$$

where a_i is the i-th column of A and the fact that the quantity $a_i^T (AD^2 A^T)^{-1} a_i > 0$ under the given assumptions. $\square$

The following two lemmas are crucial to the development of our theory.

Lemma 2.2 *For $d \in \mathbf{R}^n$ let $q(d)$ be given by (2.2) for some A of full row rank. Consider the n-dimensional vector d^* where some components of d^* may be infinite. Assume that the components of d^* can be divided into two sets: S_α and S_β, such that*

1. S_α contains m and S_β contains $n - m$ components of d^*;

2. all elements in S_α are nonzero and

$$\frac{\max\{|d_i^*| : d_i^* \in S_\beta\}}{\min\{|d_i^*| : d_i^* \in S_\alpha\}} = 0; \tag{2.3}$$

3. the m by m submatrix of A consisting of columns corresponding to the components in S_α is nonsingular.

Then, as d converges to d^*

$$\lim_{d \to d^*} q_i(d) = q_i(d^*) = \begin{cases} 1, & \text{if } d_i^* \in S_\alpha, \\ 0, & \text{if } d_i^* \in S_\beta. \end{cases} \tag{2.4}$$

Proof: Without lose of generality, we assume that

$$S_\alpha = \{d_1^*, d_2^*, ..., d_m^*\} \quad \text{and} \quad S_\beta = \{d_{m+1}^*, d_{m+2}^*, ..., d_n^*\}.$$

Let A_α and A_β be the submatrices of A consisting of the first m and the last $n - m$ columns of A, respectively. By Assumption 3, A_α is nonsingular. Similarly, let D_α and D_β be the submatrices of D consisting of the first m and the last $n-m$ columns of D, respectively, where $D = diag(d)$ and d converges to d^*. Also let $r = 1/\min\{|d_i| : d_i^* \in S_\alpha\}$. From Assumption 2

5

we know that r is well defined for d sufficiently close to d^* and all the diagonal elements of rD_α have absolute value greater than or equal to 1. Substituting $AD = [A_\alpha D_\alpha \quad A_\beta D_\beta]$ into $H(d)$, we have for d close to d^*

$$H(d) = [rA_\alpha D_\alpha \quad rA_\beta D_\beta]^T [A_\alpha(rD_\alpha)^2 A_\alpha{}^T + A_\beta(rD_\beta)^2 A_\beta{}^T]^{-1} [rA_\alpha D_\alpha \quad rA_\beta D_\beta].$$

Since Assumption 2 implies that $\lim_{d \to d^*} rD_\beta = 0$, it is evident that

$$\lim_{d \to d^*} H(d) = H(d^*) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.5}$$

which proves (2.4). □

Since d^* may have infinite components, we will define the derivatives of q at d^* by continuity.

Lemma 2.3 *Let $q(d)$ be given as in (2.2) and d^* satisfy the conditions in Lemma 2.2. Define the derivatives of $q(d)$ at the point d^* as the limits of corresponding derivative values at $d \in \mathbf{R}^n$ as d converges to d^*. Then $q(d)$ is at least twice continuously differentiable at d^*. Moreover, the Jacobian matrix of $q(d)$ vanishes at d^*; or equivalently,*

$$\nabla q_i(d^*) = 0, \quad i = 1, 2, ..., n. \tag{2.6}$$

Proof: To verify the differentiability of $q(d)$ at d^*, we first assume that d^* is finite. By Assumption 3 of Lemma 2.2, AD^2A^T is nonsingular at d^*. It is therefore nonsingular in a neighborhood of d^*. Observe that $q(d)$ is a rational function of d near d^* with a nonzero denominator. This follows from the well-known adjoint form for the inverse matrix and the fact that all elements of AD^2A^T are quadratic functions of d. Therefore, $q(d)$ is actually infinitely differentiable at finite d^*.

Now we show that $q(d)$ has continuous second-order partial derivatives even at an infinite d^*. (Since we will only make use of second derivatives in our analysis, we will not concern ourselves with derivatives of order higher than 2.) Obviously, our definition of derivatives at d^* guarantees that if a derivative exists at d^*, then it is also continuous at d^*.

Let the matrix-valued function P be defined as

$$P(d) = A^T (AD^2A^T)^{-1} A.$$

From the definitions of $H(d)$ and $q(d)$, we have that for any finite d sufficiently close to d^*,

$$H_{ij}(d) = d_i d_j P_{ij}(d) \quad \text{and} \quad q_i(d) = H_{ii}(d) = d_i^2 P_{ii}(d), \tag{2.7}$$

6

where $P_{ij}(d)$, the (i,j)-th element of $P(d)$, is given by

$$P_{ij}(d) = a_i^T (AD^2 A^T)^{-1} a_j,$$

and a_i is the i-th column of A. By the Sherman-Morrison-Woodbury formula (see [12, p. 50], for example),

$$P_{ij}(d + \epsilon e_\ell) - P_{ij}(d) = -\frac{2\epsilon d_\ell + \epsilon^2}{1 + (2\epsilon d_\ell + \epsilon^2)P_{\ell\ell}(d)} P_{i\ell}(d)P_{j\ell}(d).$$

Dividing the right-hand side by ϵ and letting ϵ go to zero, we obtain

$$\frac{\partial P_{ij}(d)}{\partial d_\ell} = -2d_\ell P_{i\ell}(d)P_{j\ell}(d). \tag{2.8}$$

It follows from (2.7) and (2.8) that

$$\frac{\partial q_i(d)}{\partial d_j} = -2d_i H_{ij}(d)P_{ij}(d) + 2\delta_{ij}d_i P_{ii}(d), \tag{2.9}$$

where δ_{ij} is the Kronecker delta.

Let us consider S_α, S_β, A_α, A_β and D_α, D_β as in Lemma 2.2. We note that Assumption 2 in Lemma 2.2 guarantees that $d_i^* \neq 0$ for all $d_i^* \in S_\alpha$. A direct calculation gives

$$\lim_{d \to d^*} P(d) = P(d^*) = \begin{bmatrix} (D_\alpha^*)^{-2} & (D_\alpha^*)^{-2} A_\alpha^{-1} A_\beta \\ A_\beta^T A_\alpha^{-T}(D_\alpha^*)^{-2} & A_\beta^T A_\alpha^{-T}(D_\alpha^*)^{-2} A_\alpha^{-1} A_\beta \end{bmatrix}, \tag{2.10}$$

where the convention $1/\infty = 0$ is used. Clearly, $P(d^*)$ is finite.

To prove (2.6), we look at the following two different cases.

Case 1: $i = j$. If $d_i^* = 0$, then the limit of (2.9) as d goes to d^* obviously exists and is zero. If $d_i^* \neq 0$, then as d goes to d^* in (2.9) and using (2.5), we have

$$\frac{\partial q_i(d^*)}{\partial d_i} = \frac{2}{d_i^*} H_{ii}(d^*)(1 - H_{ii}(d^*)) = 0.$$

Case 2: $i \neq j$. Now we only have the first term in (2.9). Assume that $|d_i^*| = \infty$; otherwise the proof would be trivial because $H_{ij}(d^*) = 0$. Now we have $d_i^* \in S_\alpha$, i.e., $0 \leq i \leq m$. If $d_j^* \neq 0$, then from the first equation in (2.7)

$$\frac{\partial q_i(d)}{\partial d_j} = -\frac{2}{d_j}(H_{ij}(d))^2$$

which has a zero limit at d^* by (2.5). If $d_j^* = 0$, then $d_j^* \in S_\beta$, i.e., $m + 1 \leq j \leq n$. Let $d \to d^*$, then from (2.10) and (2.5) we have

$$\frac{\partial q_i(d^*)}{\partial d_j} = -\frac{2H_{ij}(d^*)[A_\alpha^{-1}A_\beta]_{i,j-m}}{d_i^*} = 0.$$

So far we have proved that the Jacobian matrix of $q(d)$ exists and is zero at d^*. A direct calculation shows that for d close but not equal to d^*,

$$\frac{\partial^2 q_\ell(d)}{\partial d_i \partial d_j} = 8 H_{i\ell}(d) H_{j\ell}(d) P_{ij}(d) - 4 \delta_{j\ell} H_{i\ell}(d) P_{i\ell}(d) - 4 \delta_{i\ell} H_{j\ell}(d) P_{j\ell}(d) \qquad (2.11)$$
$$- 2 \delta_{ij} (d_\ell P_{i\ell}(d))^2 + 2 \delta_{i\ell} \delta_{j\ell} P_{\ell\ell}(d).$$

The second-order partial derivatives in (2.11) can be shown to have finite limits as $d \to d^*$. This completes the proof. $\square$

It should not be a surprise that the Jacobian matrix of $q(d)$ is zero at d^*. According to Lemma 2.2 either the maximum $(q_i(d^*) = 1)$ or the minimum $(q_i(d^*) = 0)$ is reached at d^* for every component of $q(d)$.

3   Applications of the Indicator

In this section, we show how the indicator developed in the previous section can be used in a primal, dual or primal-dual interior point algorithm to identify an optimal basis (under appropriate nondegeneracy assumptions).

Primal Algorithms

Our first result concerns primal algorithms.

Theorem 3.1 *Let $\{x^k\} \subset \mathbf{R}^n$ converge to a nondegenerate vertex (basic feasible solution) x^* of the linear program (1.1). If $q^k = q(x^k)$ is given by (2.2), then*

$$\lim_{k \to \infty} q^k = q^* = sign(x^*). \qquad (3.1)$$

Moreover,

$$\|q^k - q^*\| = O(\|x^k - x^*\|^2).$$

Proof: Let S_α in Lemma 2.2 contain the nonzero components of x^* and let S_β contain the zero components. Then obviously Assumptions 1 and 2 of Lemma 2.2 hold. The fact that x^* is a nondegenerate basic feasible solution guarantees that Assumption 3 is also satisfied. The first part of the theorem now follows from Lemma 2.2.

The second part follows directly from the fact proved in Lemma 2.3 that the Jacobian matrix of $q(x)$ is zero at x^* and $q(x)$ is twice continuously differentiable. $\square$

The theoretical advantages of using the q_i's as indicators, as opposed to the variables x_i's themselves, are twofold. First, the q_i's provide a fixed separation which is problem

independent. This is not true of the x_i's for which the separation between zero and nonzero variables can be arbitrarily small. Second, the q_i's converge quadratically faster than the x_i's. Hence an earlier termination is guaranteed.

In our numerical experiments, if high accuracy is required, then usually in less than half of the number of iterations needed for convergence, a clear 0–1 pattern will appear. A plausible explanation for this phenomenon follows. If $\{x^k\}$ converges to x^* at an R-linear rate (which is expected and observed in practice for most primal interior point algorithms on nondegenerate problems), then there exist positive constant C_x and $r < 1$, such that for k large

$$\|x^k - x^*\| \leq C_x r^k.$$

From Theorem 3.1, there exists some constant C_q, such that

$$\|q^k - q^*\| \leq C_q r^{2k}.$$

For a given small positive number $\epsilon \ll 1$, it will take approximately

$$\frac{\ln \epsilon - \ln C_x}{\ln r} \approx \frac{\ln \epsilon}{\ln r}$$

steps for $\|x^k - x^*\| \leq \epsilon$ to be satisfied, while only about half of that number of iterations is needed for the satisfaction of $\|q^k - q^*\| \leq \epsilon$.

In the context of a primal algorithm Barnes [3] used a matrix, not exactly of the form of (2.1) but quite similar, to construct estimates of the Lagrange multipliers associated with the constraints $Ax = b$. He demonstrated that these estimates converge to the true multipliers quadratically faster than the the nonbasic variables converge to zero. While this result is not directly related to our result, it does have a similar flavor. Indeed, Barnes suggested using these multiplier estimates to identify an optimal basis.

Dual Algorithms

In the dual affine algorithms developed by a number of authors as variants of Karmarkar's algorithm (see Adler *et al.* [1] and Monma and Morton [11], for example), the matrix $H(d)$ in (2.1) appears with $d_i = 1/z_i$. For these dual algorithms, we have the following result, which is analogous to its primal counterpart Theorem 3.1.

Theorem 3.2 *Let $\{z^k\} \subset \mathbf{R}^n$ converge to a dual slack vector z^* associated with a non-degenerate vertex of the dual linear program (1.2). If $q^k = q(d^k)$ is given by (2.2) with*

9

$d_i^k = 1/z_i^k$, $i = 1, 2, ..., n$, then

$$\lim_{k \to \infty} q_i^k = q_i^* = \begin{cases} 0, & \text{if } z_i^* > 0, \\ 1, & \text{if } z_i^* = 0. \end{cases}$$

Moreover,

$$\|q^k - q^*\| = O(\|z^k - z^*\|^2).$$

Proof: The proof given below was suggested by an anonymous referee. It is considerably shorter than our original proof.

Let $B \in \mathbf{R}^{(n-m) \times n}$ have full row rank and be such that its rows are orthogonal to those of A. For any positive diagonal matrix D, the rows of BD^{-1} are othogonal to those of AD. It is straightforward to verify that

$$H(d) = DA^T(AD^2A^T)^{-1}AD = I - ZB^T(BZ^2B^T)^{-1}BZ,$$

where $Z = diag(z) = D^{-1}$. Let

$$g(z) = diag(ZB^T(BZ^2B^T)^{-1}BZ).$$

It can be shown that the nondegeneracy assumption implies that z^* has $n - m$ nonzero components and the corresponding $n - m$ columns of B are linearly independent. Applying Lemmas 2.2 and 2.3 to $g(z)$, we have

$$\lim_{k \to \infty} g_i^k = g_i^* = \begin{cases} 1, & \text{if } z_i^* > 0, \\ 0, & \text{if } z_i^* = 0. \end{cases}$$

and

$$\|g^k - g^*\| = O(\|z^k - z^*\|^2).$$

Since $q_i(d) = 1 - g_i(z)$, the conclusions of the theorem follow immediately. $\square$

Primal-Dual Algorithms

Our technique is also applicable to primal-dual algorithms (see Kojima *et al.* [10], for example) where the matrix $H(d^k)$ in (2.1) appears with $d_i^k = \sqrt{x_i^k/z_i^k}$ and x^k and z^k are kept strictly positive for all k. Since both the primal and the dual programs are involved, different results can be obtained by using different combinations of assumptions (primal nondegeneracy, dual nondegeneracy, strict complementarity). Because these results and their proofs are all very similar, we choose only to present one result that is based on primal nondegeneracy.

Theorem 3.3 *Let $\{x^k\} \subset \mathbb{R}^n$ and $\{z^k\} \subset \mathbb{R}^n$ be positive and converge to an optimal solution x^* of the primal program (1.1) and an optimal dual slack vector z^* of the dual program (1.2), respectively. Assume that (i) x^* is a nondegenerate vertex of (1.1); and (ii) there are positive constants α_i and β_i such that for k large*

$$\alpha_i z_i^k \le x_i^k \le \beta_i z_i^k, \ \text{if } x_i^* = 0 \text{ and } z_i^* = 0.$$

Let $q^k = q(d^k)$ be given by (2.2) with $d_i^k = \sqrt{x_i^k / z_i^k}$, $i = 1, 2, ..., n$. Then q^k converges to a 0–1 vector with m ones and $n - m$ zeros, and

$$\lim_{k \to \infty} q_i^k = q_i^* = \begin{cases} 0, & \text{if } x_i^* = 0, \\ 1, & \text{if } x_i^* > 0. \end{cases}$$

Proof: Define

$$d_i^* = \limsup_{k \to \infty} \sqrt{x_i^k / z_i^k}.$$

Recall that x^* is a nondegenerate vertex of (1.1). There are m nonzero components and $n - m$ zero components in x^*. Let S_α in Lemma 2.2 contain those d_i^*'s corresponding to the nonzero components of x^* and S_β contain those d_i^*'s corresponding to the zero x_i^*'s. By complementarity $x_i^* z_i^* = 0$, so we have $d_i^* = \infty$ for $d_i^* \in S_\alpha$. On the other hand, $d_i^* = 0$ for $d_i^* \in S_\beta$ and $z_i^* > 0$. In addition, $d_i^* < \infty$ for $d_i^* \in S_\beta$ and $z_i^* = 0$ by Assumption (ii). Therefore, S_α and S_β satisfy Assumptions 1 and 2 of Lemma 2.2. The assumption that x^* is a vertex guarantees that Assumption 3 of Lemma 2.2 is also satisfied. It is not difficult to see from the proof of Lemma 2.2 that as long as the three assumptions in Lemma 2.2 are satisfied, we still have that $q(d_k)$ converges to the 0–1 vector with m ones and $n - m$ zeros as given in (2.4) even though we have $d_i^* = \limsup d_i^k$ instead of $d_i^* = \lim d_i^k$ for some $d_i^* \in S_\beta$. It also follows that $q_i^* = 1$ whenever $x_i^* > 0$. This completes the proof of the theorem. □

The second assumption in the above theorem basically assumes that if both x_i and z_i converge to zero, then they do so at the same rate. This assumption seems to be quite reasonable.

Basis Identification Criterion

Our optimal basis identification criterion based on the previous theorems is defined as follows: *Given a small positive number ϵ, if*

$$\{i : q_i(d^k) > 1 - \epsilon, 1 \le i \le n\} \bigcup \{i : q_i(d^k) < \epsilon, 1 \le i \le n\} = \{1, 2, ..., n\}, \qquad (3.2)$$

then test to see if an optimal basis has been identified.

11

If an orthonormal basis $U(d)$ of DA^T (i.e., of the column space of DA^D) is computed by a QR method or an SVD method, then obviously $H(d) = U(d)U(d)^T$ and

$$q_i(d) = u_i(d)^T u_i(d), \quad i = 1, 2, ..., n,$$

where $u_i(d)$ is the i-th row of $U(d)$. In this case, the cost of computing the indicator vector is $O(mn)$ for dense matrices.

If a Cholesky factorization $AD^2 A^T = L(d)L(d)^T$ is computed (as is done currently in most implementations of variants of the Karmarkar algorithm) instead of a QR factorization of DA^T, then to obtain an orthonormal basis $U(d)$ for DA^T, one needs to solve the lower triangular system $L(d)U(d)^T = AD$ which requires $O(m^2 n)$ operations for dense matrices. As we can see, this cost is one order of m higher than what is required when an orthonormal basis of DA^T is available.

4   Extension to Degenerate Problems

In general, the indicator as defined in (2.2) is not directly applicable to degenerate problems. However, we will show in this section that for primal interior point algorithms (i.e., $d = x$) a modified primal indicator can be devised to handle problems with only primal degeneracy, or more precisely, primal problems that have a unique solution. We will focus on the primal indicator for primal algorithms only, though the results obtained also apply to the dual indicator for dual algorithms.

We consider a sequence $\{x^k\}$ that converges to a degenerate basic feasible solution x^* with r nonzero components, where $r < m$. We first assume that the first r rows of AX^* are linearly independent, i.e., $A_r X^*$ has full row rank, where A_r is the matrix that consists of the first r rows of A. Then the diagonal of $XA_r^T(A_r X^2 A_r^T)^+ A_r X$, which we will call $q^{(r)}(x)$, will tend to $sign(x^*)$ and can be used as the indicator. It is fortunate that one need not know the number r in advance and the indicator $q^{(r)}(x)$ is readily available if an orthonormal basis $U(x)$ of XA^T is computed (by a QR or an SVD method, for example). One only need look through $q^{(j)}(x)$ for $j = 1, 2, ..., m$ to search for a 0–1 pattern, where

$$q_i^{(j)}(x) = q_i^{(j-1)}(x) + (U_{ij}(x))^2 \quad \text{and} \quad q_i^{(0)}(x) = 0, \quad i = 1, 2, ..., n. \qquad (4.1)$$

The cost is still $O(mn)$ for dense matrices.

The assumption that the first r rows of AX^* are linearly independent can be removed in the following way. Let $X^k A^T P^k = U^k R^k$, where U^k is an orthonormal basis of $X^k A^T$, R^k is upper triangular and P^k is a permutation matrix which forces the diagonal elements

of R^k to appear in descending order by absolute value (this can be done during the QR decomposition). Then as x^k converges to x^*, the last $m - r$ diagonal elements of R^k will tend to zero and the first r rows of $P^k A X^k$ will be linearly independent for all sufficiently large k. If we denote U_r^k as the first r columns of U^k, then

$$\lim_{k \to \infty} q^{(r)}(x^k) = \lim_{k \to \infty} diag(U_r^k U_r^{k\,T}) = sign(x^*). \qquad (4.2)$$

The indicator $q^{(r)}(x^k)$ can be computed recursively using (4.1).

We state the above discussion formally as a theorem.

Theorem 4.1 *Let $\{x^k\} \subset \mathbf{R}^n$ converge to a degenerate basic feasible solution x^* of the linear program (1.1) with r nonero components where $r < m$. Let A_r be an $r \times n$ submatrix of A such that $A_r X^*$ has rank r. If $\tilde{q}^k = \tilde{q}(x^k)$ is given by (2.2) with A replaced by A_r, then*

$$\lim_{k \to \infty} \tilde{q}^k = \tilde{q}^* = sign(x^*).$$

Moreover,

$$\|\tilde{q}^k - \tilde{q}^*\| = O(\|x^k - x^*\|^2).$$

Proof: The proof is essentially the same as that of Theorem 3.1, so we omit it. □

It is worth observing from (4.1) that once $q_i^{(j_0)}(x^*) = 1$ for some $j_0 < m$, then $q_i^{(j)}(x^*) = 1$ for all $j_0 < j \le m$. This is so because of the monotonicity of $q_i^{(j)}(x)$ with respect to j and the fact that $q(x) \le 1$.

It is unfortunate that for problems with both primal and dual degeneracy, our indicator is incapable of identifying all the zero and nonzero variables as $x^k \to x^*$ (or $z^k \to z^*$) because some components of the indicator vector may not have limits at optimality. This drawback undoubtedly limits the practical usefulness of the indicator because most real-world problems do have both primal and dual degeneracies.

5 Numerical Behavior of the Indicator

To corroborate our theory, we have performed some numerical experiments to explore the numerical behavior of the indicator. In our experiments we used randomly generated problems, fully aware that they are by no means representative of real-world problems. We stress that the numerical results have not demonstrated the effectiveness of our indicator approach, but have corroborated our theoretical convergence results.

Both nondegenerate problems and degenerate problems with only primal degeneracy are constructed. The methods of construction are described below.

To generate the cost vector c and the first $m - 1$ rows of the constraint matrix A, we use the Matlab M-file "rand" to obtain uniformly distributed random numbers in the unit interval $(0, 1)$ and then use the tangent function $tan(\pi(x - 1/2))$ to map $(0, 1)$ onto the entire real line $(-\infty, +\infty)$. For the m-th row of A, we apply the mapping $tan(\pi x/2))$ to make all the elements of that row strictly positive so that the feasible set will be bounded. The strictly positive initial feasible point x_0 is also obtained this way. Caution is taken to ensure that the generated A matrices are always of full row rank. The right-hand side vector b is set equal to Ax_0. After A, b, c and x_0 are generated, we replace A by AX_0, c by X_0c and x_0 by e — the vector of all ones, where $X_0 = diag(x_0)$. These replacements are equivalent to applying an affine transformation to the problem so that e is feasible for the transformed problem. Although there is no guarantee in theory that random problems generated in this manner should be nondegenerate, in practice the chances of getting degenerate problems seem to be extremely small.

To construct primal degenerate problems, we first solve a nondegenerate problem and obtain the solution x^*. Then we generate an ℓ by n random matrix B for $1 < \ell < n - m$ and redefine the data A, b and c by

$$A := \begin{bmatrix} A & 0 \\ B & B(x^* - e) \end{bmatrix}, \quad b := \begin{pmatrix} b \\ Bx^* \end{pmatrix}, \quad \text{and} \quad c := \begin{pmatrix} c \\ 0 \end{pmatrix}. \tag{5.1}$$

We also extend $x^0 = e$ to the $(n + 1)$-dimensional vector of all ones which is feasible for the new problem. The solution to the new problem is $(x^{*T}, 0)^T$ which is obviously primal degenerate because there are still m nonzeros in the solution but there are now $m + \ell$ constraints. Also we know that the first m rows of AX^* (with new A and X^*) are linearly independent. Obviously, the problem has a unique solution.

In our tests, we implemented the so-called standard form variant of the Karmarkar projective algorithm, a primal algorithm, developed independently by a number of authors. In this implementation we use the procedure suggested by de Ghellinck and Vial [4] and by Ye and Kojima [18] (both based on an earlier work of Todd and Burrell [14]) which uses duality to construct and update lower bounds η for the objective function so the duality gap is minimized. In our numerical experiments, we set the initial lower bound η^0 to -10^6, which happens to be adequate for our experiments. We have observed that the performance of the algorithm is not sensitive to the values of the initial lower bound.

Instead of trying to minimize the Karmarkar potential function in the search direction at each iteration by a line search, we used a simple back-tracking technique to ensure that the potential function was reduced by a fixed amount at each iteration. The initial step length was set to 0.95 times the step length that takes the iterate to the boundary of the simplex.

The program was written in Matlab and run on a SUN workstation network at Rice University with a machine epsilon of about 2.22×10^{-16}.
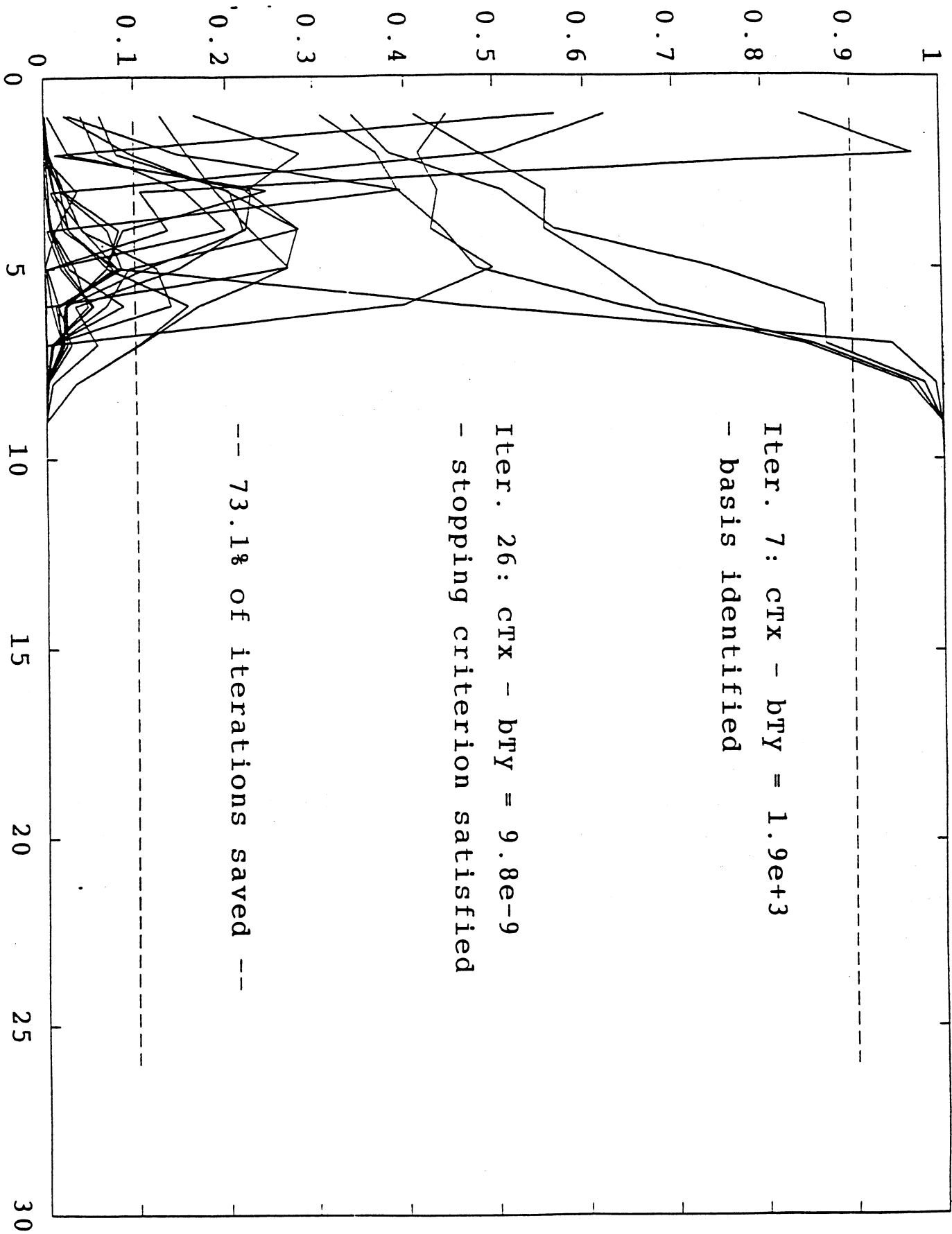
The stopping criterion used in our tests was that the duality gap be less than 10^{-8}, i.e.,

$$c^T x^k - \eta^k < 10^{-8}. \tag{5.2}$$

We tested the optimal basis identification criterion (3.2) for $\epsilon = 0.1$ on 10 randomly generated nondegenerate problems with n ranging from 20 to 200. We also tested the modified identification procedure using the recursive search as prescribed by (4.1) for 10 randomly generated primal degenerate problems. The numerical results are included in the following two tables. In the tables, the iteration numbers at which an optimal basis is identified and the stopping criterion is satisfied, respectively, are given in the fourth and sixth columns. The corresponding duality gaps are listed in the fifth and seventh columns, respectively. The rest of the tables are self-explanatory. We also include several figures which illustrate the fast convergence of the indicator. Each curve in the figures represents the history of a component of the indicator during the iterative process. As one can see from these graphs, the convergence of the indicator vectors (i.e., the separation of the two groups of indicator components converging to either zero or one) is indeed much faster than that of the iterates. From Tables 1 and 2, we see savings of about 30% to 70% in the number of iterations.

Although these numerical experiments have confirmed our convergence analysis, the computational efficiency of the indicator approach depends on how effectively the indicator can be calculated and used. For dense matrices, the cost of computing the indicator, given a Cholesky factor, is comparable with the cost of forming AD^2A^T, which is acceptable. However, recently David Gay [7] demonstrated that for sparse problems, the computation of the indicator was the dominant work in an iteration and in some cases this cost was prohibitively high. His results suggest that despite its fast convergence, the indicator may not lend itself to efficient implementations for solving sparse problems in the framework of Cholesky factorization.

History of the indicator

Iter. 7: cTx - bTy = 1.9e+3
 - basis identified

Iter. 26: cTx - bTy = 9.8e-9
 - stopping criterion satisfied

-- 73.1% of iterations saved --

number of iterations

History of the indicator

Iter. 10: c^Tx - b^Ty = 1.9e+1
- basis identified

Iter. 24: c^Tx - b^Ty = 6.3e-9
- stopping criterion satisfied

-- 58% of iterations saved --
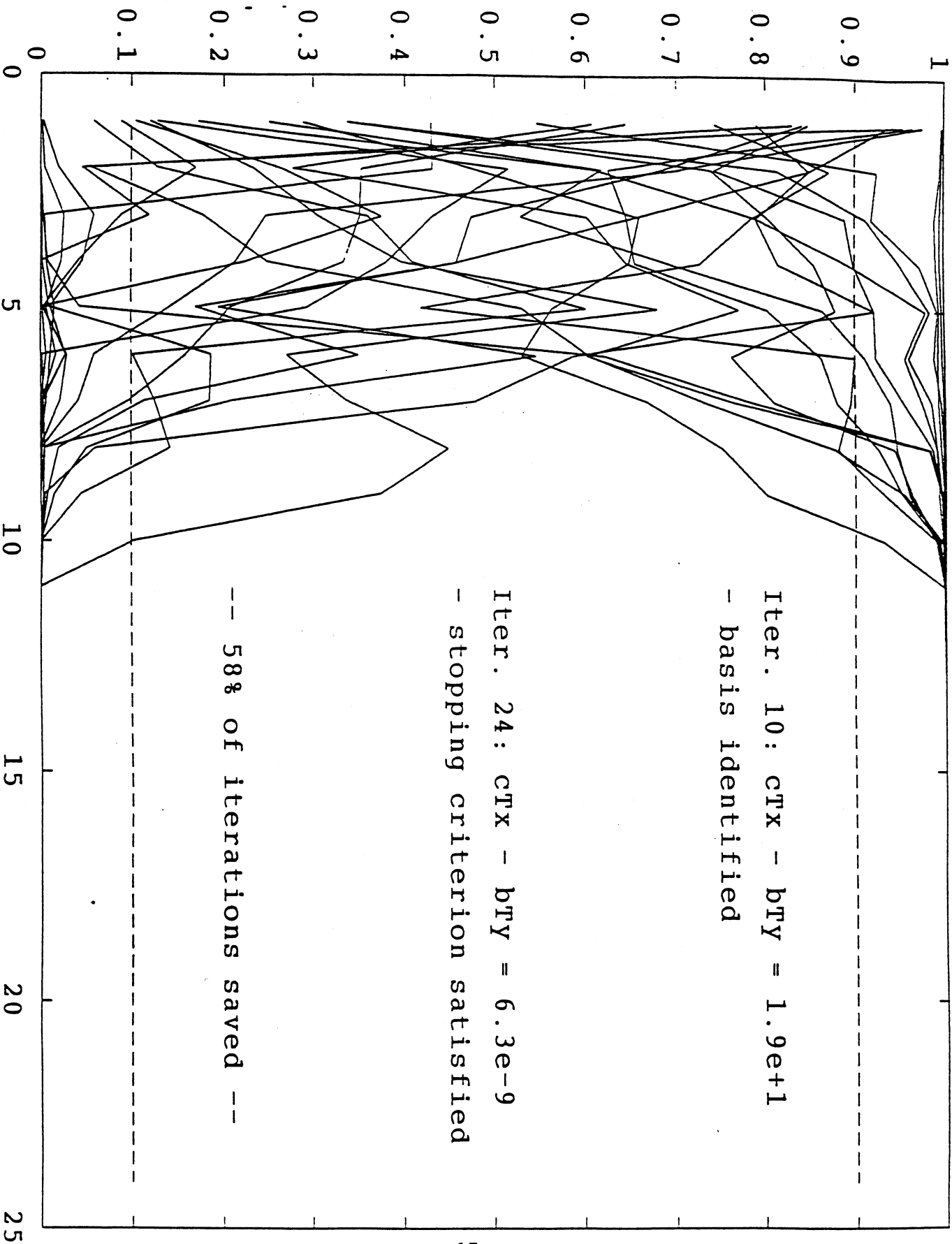
Table 1: Numerical results for nondegenerate problems

Problem number	variables n	constraints m	basis identified iter.	basis identified gap	algorithm stopped iter.	algorithm stopped gap	Iterations saved
1	20	10	8	.20E+2	24	.42E-8	67%
2	40	20	13	.23E+1	27	.94E-8	52%
3	60	30	14	.46E+2	29	.35E-8	52%
4	80	40	16	.85E+0	30	.32E-8	47%
5	100	50	13	.12E+2	29	.63E-8	55%
6	120	60	19	.66E+0	33	.26E-8	42%
7	140	70	18	.18E+0	30	.95E-8	40%
8	160	80	22	.69E-1	34	.83E-8	35%
9	180	90	19	.16E+0	32	.30E-8	40%
10	200	100	27	.22E-1	38	.55E-8	29%

Table 2: Numerical results for degenerate problems

Problem number	variables n	constraints m	basis identified iter.	basis identified gap	algorithm stopped iter.	algorithm stopped gap	Iterations saved
1	20	12	8	.69e+1	23	.71e-8	65%
2	40	24	11	.76e+1	27	.34e-8	59%
3	60	36	13	.10e+1	27	.62e-8	52%
4	80	48	15	.38e+0	28	.96e-8	46%
5	100	60	14	.18e+1	29	.52e-8	52%
6	120	72	15	.72e+1	31	.42e-8	52%
7	140	84	16	.13e+0	29	.51e-8	45%
8	160	96	16	.49e+1	32	.41e-8	50%
9	180	108	15	.72e+0	29	.69e-8	48%
10	200	120	18	.32e+1	34	.35e-8	47%

6 Concluding Remarks

In this paper, we have studied an indicator for identifying optimal bases in the setting of interior point linear programming algorithms. This indicator has the theoretical properties of being problem-independent and rapidly convergent for linear programming problems with unique solutions. It is applicable to primal, dual and primal-dual algorithms. Our randomly generated numerical examples have confirmed our theoretical analysis, showing that the use of the indicator can reduce the number of iterations by a large percentage.

From a theoretical point of view, we believe that the main result of this work is the establishment of the convergence properties of the indicator on problems without general degeneracy. However, the practical applicability of the indicator to real-world problems is severely limited by two factors. First, it is not applicable to problems with both primal and dual degeneracies. Second, the relative cost of computing the indicator can be very high for large sparse problems. Although it is still possible that the method may find application in some very special problems, at this point it seems to be mainly of theoretical interest.

Acknowledgment

References

[1] I. Adler, N. Karmarkar, M.G.C. Resende, and G. Veiga. An implementation of Karmarkar's algorithm for linear programming. *Math. Prog.*, 44:297-335, 1989.

[2] M.D. Asic, V.V. Kovacevic-Vujcic, and M.D. Radosavljevic-Nikolic. Asymptotic behaviour of Karmarkar's method for linear programming, 1987. *Math. Prog.*, 46:173-190, 1990.

[3] E. R. Barnes. A variation of Karmarkar's algorithm for solving linear programming problems. *Math. Prog.*, 36:174–182, 1986.

[4] G. de Ghellinck and J.-P. Vial. A polynomial Newton method for linear programming. *Algorithmica*, 1:425–453, 1986.

[5] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Soviet Math. Dokl.*, 8:674–675, 1967.

[6] D. M. Gay. Stopping tests that compute optimal solutions for interior-point linear programming algorithms. Numerical Manuscript 89-11, AT & T Bell Laboratories, Murray Hill, NJ, 1989.

[7] D. M. Gay. Private communication.

[8] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[9] M. Kojima. Determining basic varibles of optimal solutions in Karmarkar's new LP algorithm. *Algorithmica*, 1:499–515, 1986.

[10] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point method for linear programming, in progress in *Mathematical Programming*, editor N. Megiddo, Springer-Verlag, 1989, pp. 29-47.

[11] C.L. Monma and A.J. Morton. Computational experience with a dual affine variant of Karmarkar's method for linear programming. *OR Letters*, 6:261–267, 1987.

[12] J. M. Otega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[13] M. J. Todd. Improved bounds and containing ellipsoids in Karmarkar's linear programming algorithm. *Math. of Operations Research*, 13:650–659, 1988.

[14] M. J. Todd and B. P. Burrell. An extension of Karmarkar's algorithm for linear programming using dual variables. *Algorithmica*, 1:409–424, 1986.

[15] W. B. van Dam, J. B. G. Frenk, and J. Telgen. Randomly generated polytopes for testing mathematical programming algorithms. *Math. Prog.*, 26:172–181, 1983.

[16] R. J. Vanderbei, M. S. Meketon, and B. A. Freeman. On a modification of Karmarkar's linear programming algorithm. *Algorithmica*, 1:395–407, 1986.

[17] Yinyu Ye. Recovering optimal basis in Karmarkar's polynomial algorithm for linear programming, 1987. Manuscript, Integrated Systems, Inc., Sunnyvale, California.

[18] Yinyu Ye and M. Kojima. Recovering optimal dual solutions in Karmarkar's polynomial algorithm for linear programming. *Math. Prog.*, 39:305–317, 1987.

[19] Yinyu Ye and M. Todd. Containing and shrinking ellipsoids in the path-following algorithm, *Math. Prog.*, 47:1–9, 1990.