ON SCOLNIK'S PROPOSED POLYNOMIAL-TIME LINEAR PROGRAMMING ALGORITHM

David M Gay

TR 73-190

December 1973

Computer Science Department
Cornell University
Ithaca, New York 14850

ON SCOLNIK'S PROPOSED POLYNOMIAL-TIME
LINEAR PROGRAMMING ALGORITHM*

David M Gay

Department of Computer Science

Cornell University

Abstract:

At a recent symposium, Hugo Scolnik expressed some ideas leading to an algorithm which he thought might solve the linear programming problem in polynomial time. We examine the algorithm and find that it often fails to solve the linear programming problem, even in the special cases considered by Scolnik. We conclude that the algorithm probably cannot be modified to work properly.

* The author is a National Science Foundation Graduate Fellow. This work was also partially supported by National Science Foundation Grant GJ40903.

ON SCOLNIK'S PROPOSED POLYNOMIAL-TIME LINEAR PROGRAMMING ALGORITHM

David M Gay

Department of Computer Science
Cornell University
Ithaca, N.Y. 14850

1. Introduction

At the Eighth International Symposium on Mathematical Programming (August 1973 at Stanford), Hugo Scolnik suggested a line of reasoning leading to an algorithm which he thought might solve the linear programming problem in polynomial time.* The algorithm which Scolnik contemplated would construct an optimal basis one column at a time, in such a way that once a column is selected to be in this basis, it need never be thrown out later; after t basic columns had been selected, the algorithm would first use two rejection criteria (a) and (b) to reduce the number of candidates for the $(t+1)^{\underline{st}}$ column and would then use a selection criterion (c) to choose one of the remaining candidates.

In §2 we state the linear programming problem precisely and review some properties of its solutions which will be useful in the sequel; we also review some facts about pseudoinverses and introduce some terminology and notation. In §3 we summarize Scolnik's suggestions, and in §4 we examine them in more detail.

---

* My knowledge of Scolnik's ideas on this subject comes from a handout, which he distributed at the symposium, and from an addendum that he wrote thereto, a copy of which was kindly sent to me by Gordon Bradley.

We conclude in §5 that Scolnik's suggestions have apparently shed little new light on the linear programming problem.

2. Preliminaries

2.1 The Problem

The linear programming problem may be stated in the following form: given an $m \times n$ matrix A with $m < n$ and vectors $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, let

(1a) $$S = \{x \in \mathbb{R}^n \mid x \geq 0 \text{ (componentwise) and } Ax = b\}$$

and find $x^* \in S$ such that

(1b) $$z(x^*) \equiv c^T x^* = z^* \equiv \inf\{c^T x \mid x \in S\} .$$

Of course, it may happen that $z^* = +\infty$ (i.e., S is empty) or $z^* = -\infty$; but if A has full rank (which we henceforth assume) and z^* is finite, then it is possible to reorder the columns of A (and correspondingly the components of c and x) in such a way that $A = (B,R)$, where B is a nonsingular $m \times m$ matrix, R is $m \times (n - m)$, $x^* = (x_B^*, x_R^*)$ (partitioned like A) with $x_B^* = B^{-1}b$ and $x_R^* = 0$ is a solution to (1), so that, in particular

(2a) $$B^{-1}b \geq 0,$$

and if $c^T = (c_B^T, c_R^T)$ is partitioned like A, then

(2b) $$c_B^T B^{-1} R - c_R^T \leq 0.$$

Conditions (2) are sufficient to guarantee that $x^* = \binom{B^{-1}b}{0}$

solves (1); B is called an "optimal basis". The usual simplex algorithms for solving (1) proceed by changing a trial basis (i.e., subset of m linearly independent columns of A) one column at a time until conditions (2) hold. Scolnik wants to avoid exchanging one column for another by building up an optimal basis from scratch, as briefly described in the introduction.

2.2 Pseudoinverses

Below, we shall frequently refer to the pseudoinverse M^+ of a (real) matrix M. M^+ has the dimensions of M^T (the transpose of M) and is the unique such real matrix satisfying:

$$(3) \quad MM^+M = M; \quad M^+MM^+ = M^+; \quad (MM^+)^T = MM^+; \quad (M^+M)^T = M^+M.$$

(The matrices M with which we shall deal will always have full column rank; in this case $MM^+M = M \Rightarrow M^+M = I$, i.e., M^+ is a left inverse of M.) Important for us are the facts that $(MM^+)(MM^+) = MM^+$, so that MM^+ is a projection matrix, and $(MM^+)^T(I - MM^+) = 0$, so that, in fact, MM^+ projects orthogonally (with respect to the standard inner product) onto the column space of M, and $(I - MM^+)$ projects orthogonally onto the orthogonal complement of the column space of M. In general, $y^* = M^+b$ is the least-squares solution of smallest 2-norm $(||y||_2 = (y^Ty)^{\frac{1}{2}})$† to the possibly inconsistent system $My = b$. (When M has full column rank, M^+b is the only least-

---

† It is possible to compute a pseudoinverse $\underline{M}^+$ of M with respect to arbitrary inner product norms $||y||_P = (y^TPy)^{\frac{1}{2}}$ on the domain and $||w||_Q = (w^TQw)^{\frac{1}{2}}$ on the range of M, where P and

squares solution to My=b.) If, for example, $M = v \in \mathbb{R}^m - \{0\}$, i.e. $M = v$ is a non-zero m×1 matrix, then $M^+ = (v)^+ = v^T/||v||_2^2$; cf formula (7) below (in the proof of theorem 2). The algorithm which Scolnik suggests requires us to compute M^+b for certain M; for an efficient and stable way to do this, see [1]. For more on pseudoinverses generally, see [2].

2.3 Terminology and Notation

We shall say that an $m \times m$ matrix B is a basis for (1) iff the columns $b_1,\ldots,b_m$ of B are columns of A and B is nonsingular; in this case we also say that $\{b_1,\ldots,b_m\}$ is a basis for (1). (It will usually be clear that we are talking about (1), so we will often omit the phrase "for (1)".) If B is a basis for (1), then there exists an $n \times n$ permutation matrix Π_n such that $A\Pi_n = (B,R)$; to B corresponds the basic solution $x = \Pi_n\binom{B^{-1}b}{0}$. We say that B is feasible iff (2a) holds and that B is optimal iff both (2a) and (2b) hold. If B is a basis for (1) and Π_m is an $m \times m$ permutation matrix, then $B\Pi_m$ is a basis which gives the same basic solution and enjoys the same properties as B; when we say, e.g., "$\{b_1, \ldots, b_m\}$ is the only optimal basis," we

---

Q are symmetric and positive definite. In this case $y^* = \underline{M}^+b$ is the vector of least $||\cdot||_P$ norm which minimizes $||My - b||_Q$. $\underline{M}^+$ is unique and is characterized by:

$$M\underline{M}^+M = M;\quad \underline{M}^+M\underline{M}^+ = \underline{M}^+;\quad (M\underline{M}^+)^TQ = QM\underline{M}^+;\quad (\underline{M}^+M)^TP = P\underline{M}^+M.$$

There exist symmetric, positive definite matrices $P^{-\frac{1}{2}}$ and $Q^{\frac{1}{2}}$ such that $(P^{-\frac{1}{2}})(P^{-\frac{1}{2}}) \equiv (P^{-\frac{1}{2}})^2 = P^{-1}$ and $(Q^{\frac{1}{2}})^2 = Q$, and we have $\underline{M}^+ = P^{-\frac{1}{2}}(Q^{\frac{1}{2}}MP^{-\frac{1}{2}})^+Q^{\frac{1}{2}}$, where the pseudoinverse on the right is computed with respect to the 2-norm.

mean that any optimal basis for (1) has the form $B\Pi_m$.

Let a_i denote the i^{th} column of A, so that $A = (a_1, a_2, \ldots, a_n)$. Following Scolnik, we let $A_i = (a_1, \ldots, a_i)$ denote the submatrix consisting of the first i columns of A, and we let $A_i^j = (a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_i)$ denote the matrix obtained by removing the j^{th} column from A_i. Let $\langle.,.\rangle$ denote the standard inner product: $\langle x,y\rangle = y^T x$. Let I denote the identity matrix of dimension appropriate to the context; let $P_i = I - A_i A_i^+$ denote the orthogonal projection onto the orthogonal complement of the column space of A_i, and, similarly, let $P_j^i = I - A_j^i (A_j^i)^+$.

3. Scolnik's Observations and the Algorithm They Seem to Suggest

3.1 Some Theorems and the Suggested Algorithm

Scolnik made the following observations; we defer proofs and detailed discussions until §4. (The designations "theorem 1" and "theorem 2" are here reversed with respect to Scolnik's symposium handout. The theorems and algorithm are also stated in slightly different form.)

Theorem 1: Assume that $A_j^+ b \geq 0$ for some j with $2 \leq j \leq m$. If

$$||P_j^k b||_2 = \min\{||P_j^i b||_2 : 1 \leq i \leq j\},$$

then

$$(A_j^k)^+ b \geq 0.$$

Hence, if A_m is a feasible basis, i.e., $A_m^+ b = A_m^{-1} b \geq 0$, then it is possible to reorder the first m columns of A so that $A_j^+ b \geq 0$ for $1 \leq j \leq m$.

<u>Theorem 2</u>: Assume that A_j has full column rank. Then

$$A_j^+ b \geq 0 \quad \text{iff} \quad \langle P_j^i a_i, b\rangle \geq 0 \quad \text{for} \quad 1 \leq i \leq j.$$

In particular (since P_j^i is symmetric and $P_m^m = P_{m-1}$), if A_m is a feasible basis, then $\langle a_j, P_{m-1}b\rangle \geq 0$ for $j = m$. Theorem 3 gives a condition under which this holds for other values of j.

<u>Theorem 3</u>: If A_m is a feasible basis $(A_m^{-1}b \geq 0)$ and $a_1, \ldots, a_{m-1}$ span a facet of the cone $K = \{Ax \mid x \geq 0\}$ generated by the columns of A, then

$$\langle a_j, P_{m-1}b\rangle \geq 0 \quad \text{for } 1 \leq j \leq n.$$

That $a_1, \ldots, a_{m-1}$ span a facet of K means that there exists a unit vector u orthogonal to $a_1, \ldots, a_{m-1}$ such that $\langle a_j, u\rangle \geq 0$ for $m \leq j \leq n$, and if $\langle a_j, u\rangle = 0$ for some $j \geq m$, then $A_{m-1}^+ a_j \geq 0$.

Henceforth we assume that there exists an optimal basis B, $m-1$ of whose columns span a facet of $K = \{Ax \mid x \geq 0\}$. This is, of course, a special case, but Scolnik thought that an algorithm like the one described below might become part of a unified algorithm, for which he was working on a theorem (and which he did not further describe). This assumption allows us to use theorem 3; Scolnik suggested that theorem 3 might in this case generalize to:

<u>Theorem 4</u>: For $1 \leq k \leq m-1$, if $a_1, \ldots, a_k$ span a facet of the

subcone $\{A_{k+1}x \mid x \geq 0, x \in \mathbb{R}^{k+1}\}$ generated by $a_1,\ldots,a_{k+1}$, then

$$\langle a_j, P_k b\rangle \geq 0 \quad \text{for } 1 \leq j \leq n.$$

It now appears that the columns of A might be so orderable that A_m is an optimal basis,

(4a) $$A_j^+ b \geq 0 \quad \text{for } 1 \leq j \leq m, \text{ and}$$

(4b) $$\langle a_j, P_k b\rangle \geq 0 \quad \text{for } 1 \leq k \leq m-1 \text{ and } 1 \leq j \leq n.$$

Thus we have two criteria (4a) and (4b) for reducing the number of candidates for a_{t+1}, once $a_1,\ldots,a_t$ are known. Neither (4a) nor (4b) takes into account the objective function $z(x) = \langle x,c\rangle$, so it seems reasonable to introduce a selection criterion based on the the objective function to choose one of the remaining candidates. Scolnik suggested that perhaps a_{t+1} might render a best least-squares solution to the linear programming problem in the sense that

(4c) $$(c_1,\ldots,c_{t+1})A_{t+1}^+ b \leq (c_1,\ldots,c_t,c_j)(A_t,a_j)^+ b \quad \text{for } t+2 \leq j \leq n.$$

In case equality held for some j in (4c), it might be necessary to look ahead, i.e., to repeat subsequent computations with each candidate in turn as a_{t+1} until some one of the candidates yielded a smaller value of $(c_1,\ldots,c_{t+k})A_{t+k}^+ b$ than the others; of course, this might jeopardize the polynomial running time. Thus we arrive at

<u>The Algorithm</u>

0. $t \leftarrow 0$ (iteration counter).

1. <u>Comment</u>: At this point the first t columns of an

optimal basis have been selected, and the columns of A have been reordered so that these first t basic columns are $a_1,\ldots,a_t$, i.e., the columns of A_t.

(a) Determine $R_1 = \{a_j | t < j \leq n,\ P_t a_j \neq 0$, and $(A_t,a_j)^+ b \geq 0$.

<u>Comment</u>: Because $P_t a_j \neq 0$, (A_t,a_j) has full rank for each $a_j \in R_1$.

(b) Determine $R_2 = \{a_j \in R_1 | \langle a_i, P_{t,j} b\rangle \geq 0$ for $t < i \leq n$ but $i \neq j$, where $P_{t,j} = I - (A_t,a_j)(A_t,a_j)^+\}$. If R_2 is empty, then terminate unsuccessfully.

(c) Find the element $a_j{}^*$ of R_2 such that $(c_1,\ldots,c_t,c_j{}^*)(A_t,a_j{}^*)^+ b$ is minimized (where c_i = the $i^{\underline{th}}$ component of c); if there are two or more candidates for $a_j{}^*$, then look ahead.

Interchange a_{t+1} and $a_j{}^*$ (as well as c_{t+1} and $c_j{}^*$).

$t \leftarrow t+1$.

If $t = m$ then terminate successfully; otherwise go to 1.

Notice that $R_2 = R_1$ in the last iteration $(t = m-1)$.

Recall (2b); Scolnik proved:

<u>Theorem 5</u>: Assume that the algorithm has successfully run. If

(5a) $$\langle a_j, P_{m-1} b\rangle > 0 \text{ and}$$

(5b) $$(c_1,\ldots,c_{m-1},c_j)(A_{m-1},a_j)^{-1} b \geq (c_1,\ldots,c_m)A_m^{-1} b$$

for some $j > m$, then $(c_1,\ldots,c_m)A^{-1}a_j - c_j \leq 0$.

Corollary: If the algorithm has successfully terminated and (5) holds for all j with $m \leq j \leq n$, then A_m is an optimal basis.

(In his symposium handout, Scolnik specified that, in the final iteration, a_j^* minimize $(c_1,\ldots,c_{m-1},c_j)(A_{m-1},a_j)^+b$. As theorem 5 and (2b) show, this choice of a_m renders A_m an optimal basis if (5a) holds for all $j > m$.)

3.2 Example

Scolnik offered the following example of how this algorithm works (on a problem constructed by Gass):

$$A = \begin{pmatrix} 1 & 3 & -1 & 0 & 2 & 0 \\ 0 & -2 & 4 & 1 & 0 & 0 \\ 0 & -4 & 3 & 0 & 8 & 1 \end{pmatrix},\ b = \begin{pmatrix} 7 \\ 12 \\ 10 \end{pmatrix},\ c = (0,\ 1,\ -3,\ 0,\ 2,\ 0)^T.$$

The optimal basis is $\{a_2, a_3, a_6\}$, which gives $x^* = (0,\ 4,\ 5,\ 0,\ 0,\ 11)^T$ and $z(x^*) = -11$.

Iteration 1: $(t = 0)$

$(a_1)^+b = 7$; $(a_2)^+b \doteq -1.48$; $(a_3)^+b \doteq 2.73$; $(a_4)^+b = 12$; $(a_5)^+b \doteq 1.38$; $(a_6)^+b = 10$; thus $R_1 = \{a_1, a_3, a_4, a_5, a_6\}$. $\langle a_i, P_{0,3}b\rangle > 0$ for $i \neq 3$ and $\langle a_2, P_{0,j}b\rangle < 0$ for $j = 1, 4,5,6$; thus $R_2 = \{a_3\}$ and a_3 is selected as the first basic column. To avoid confusion, we will not interchange a_1 and a_3.

Iteration 2: $(t = 1)$

We find $(a_3,a_j)^+b > 0$ for $j = 1,2,4,5,6$, so $R_1 = \{a_1, a_2,$

$a_4, a_5, a_6\}$. $\langle a_i, P_{1,j}b\rangle < 0$ for $(i,j) = (2,1), (4,2), (6,5)$, and we find $R_2 = \{a_4, a_6\}$. Since $(c_3,c_6)(a_3,a_6)^+b \doteq -7.23 < -6.9 = (c_3,c_4)(a_3,a_4)^+b$, a_6 is selected as the second basic column.

Iteration 3: $(t := 2)$

We find:

j	$((a_3,a_6,a_j)^{-1}b)^T$			$(c_3,c_6,c_j)(a_3,a_6,a_j)^{-1}b$
1	3	1	10	-9
2	5	11	4	-11
4	-7	31	40	21
5	3	-39	5	1

Thus $R_1 = R_2 = \{a_1, a_2\}$ and a_2 is selected as the third basic column.

4. Proofs and Discussion

Theorems 1, 2, 3, and 5 are valid. We prove them in §4.1, but the proofs are not relevant to the discussion which follows (§4.2). The proofs of theorems 3 and 5 are essentially those which Scolnik gave.

4.1 Proofs

Proof of Theorem 1: Let $\underline{b} = A_jA_j^+b$: then $b = \underline{b} + \bar{b}$, where $\bar{b} = b - \underline{b} = P_jb$. The following proves the geometrically obvious facts that

(6) $$||P_j^k\underline{b}||_2 = \min\{||P_j^i\underline{b}||_2 : 1 \le i \le j\}$$

and if $(A_j^k)^+b = (A_j^k)^+\underline{b}$ had a negative component, then there would be a point $\underline{b}^*$ on the line segment between $\underline{b}$ and $A_j^k(A_j^k)^+\underline{b}$ such that $\underline{b}^*$ were in the column space of $A_j^{k'}$ for some $k' \neq k$, $k' \leq j$, and hence $||P_j^{k'}\underline{b}||_2 < ||P_j^k\underline{b}||_2$.

Since $\bar{b}$ lies in the orthogonal complement of the column space of A_j and, hence, of A_j^i $(1 \leq i \leq j)$, $P_j^i\bar{b} = \bar{b}$ and $\langle\bar{b},\underline{b}\rangle = \langle\bar{b},A_j^i(A_j^i)^+b\rangle = 0$, so $P_j^ib = \bar{b} + P_j^i\underline{b}$ and $\langle\bar{b},P_j^i\underline{b}\rangle = \langle\bar{b},\underline{b} - A_j^i(A_j^i)^+\underline{b}\rangle = 0 - 0 = 0$; hence $||P_j^ib||_2^2 = ||\bar{b}||_2^2 + ||P_j^i\underline{b}||_2^2$, which implies (6). Because the columns of A_j^i are linearly independent and $A_j^i\ (A_j^i)^+\bar{b} = \bar{b} - P_j^i\bar{b} = 0$, $(A_j^i)^+\bar{b} = 0$ and so $(A_j^i)^+b = (A_j^i)^+\underline{b}$; similarly $A_j^+b = A_j^+\underline{b}$. Let $(\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_j)^T = (A_j^k)^+\underline{b}$ and $\alpha_k=0$, so that $A_j^k(A_j^k)^+\underline{b} = \sum_{i=1}^{j} \alpha_i a_i$. Now if $(A_j^k)^+b = (A_j^k)^+\underline{b}$ had a negative component $\alpha_{k'} < 0$ for some $k' \leq j$, $k' \neq k$, then we would have $P_j^k\ \underline{b} \neq 0$, since $P_j^k\ \underline{b} = 0$ would imply $A_j^+\ b = A_j^+\ \underline{b} = (\alpha_1,\ldots,\alpha_j) \not\geq 0$. Setting $\beta_i = (A_j^+b)_i = i^{\underline{th}}$ component of $A_j^+b \geq 0$, we have $\underline{b} = A_j(A_j^+b) = \sum_{i=1}^{j} \beta_i a_i$. Let $\lambda = \dfrac{\beta_{k'}}{\beta_{k'}-\alpha_{k'}} \in [0,1)$: then $\lambda\alpha_{k'} + (1-\lambda)\beta_{k'} = 0$, so $\underline{b}^* \equiv \lambda A_j^k(A_j^k)^+\underline{b}$ $+ (1-\lambda)\underline{b}$ lies in the column space of $A_j^{k'}$. But $A_j^{k'}(A_j^{k'})^+\underline{b}$ is the closest point in the column space of $A_j^{k'}$ to $\underline{b}$, so $||P_j^{k'}\ \underline{b}||_2 = ||\underline{b} - A_j^{k'}(A_j^{k'})^+\underline{b}||_2 \leq ||\underline{b} - \underline{b}^*||_2 = \lambda||\underline{b} - A_j^k(A_j^k)^+\underline{b}||_2 = \lambda||P_j^k\underline{b}||_2 < ||P_j^k\underline{b}||_2$, which contradicts (6). ∎

<u>Proof of Theorem 2</u>: Since A_j has full rank, a_j is not in the column space of A_{j-1}, so $P_{j-1}a_j \neq 0$ and (since $P_j = P_jP_j = P_j^T$ is idempotent and symmetric) $a_j^TP_{j-1}a_j = ||P_{j-1}a_j||_2^2 > 0$.

Thus

$$(7)^{\dagger} \qquad A_j^+ = \begin{pmatrix} A_{j-1}^+\left(I - \dfrac{a_j a_j^T P_{j-1}}{\|P_{j-1}a_j\|_2^2}\right) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_j^T P_{j-1}/\|P_{j-1}a_j\|_2^2 \end{pmatrix},$$

as can be verified by substitution into (3). If we apply this reasoning to $(A_j\Pi)^+ = \Pi^T A_j^+$, where Π is a permutation matrix, we find that

(8) the $i^{\underline{th}}$ row of A_j^+ is just $a_i^T P_j^i/\|P_j^i a_i\|_2^2$,

from which the theorem follows. ∎

Proof of Theorem 3: Let u be as in the remark which immediately follows theorem 3: then u spans the (one dimensional) orthogonal complement of the column space of A_{m-1}, so $P_{m-1}b = \langle b,u\rangle u$ and $\langle a_j, P_{m-1}b\rangle = \langle a_j,u\rangle\langle b,u\rangle$; since $b = A_m A_m^{-1} b$ is a nonnegative linear combination of $a_1,\ldots,a_m$, $\langle b,u\rangle \geq 0$ and the theorem follows. ∎

Proof of Theorem 5: Since $\langle a_j, P_{m-1}b\rangle = \langle P_{m-1}a_j, b\rangle > 0$, $P_{m-1}a_j \neq 0$ and so (A_{m-1}, a_j), like A_m, has full rank. Thus (7) implies that hypothesis (5b) is equivalent to

$$\left\langle A_{m-1}^+ b - \frac{\langle a_m, P_{m-1}b\rangle}{\|P_{m-1}a_m\|_2^2} A_{m-1}^+ a_m, \begin{pmatrix} c_1 \\ \vdots \\ c_{m-1}\end{pmatrix}\right\rangle + \frac{\langle a_m, P_{m-1}b\rangle}{\|P_{m-1}a_m\|_2^2} c_m$$

$$\leq \left\langle A_{m-1}^+ b - \frac{\langle a_j, P_{m-1}b\rangle}{\|P_{m-1}a_j\|_2^2} A_{m-1}^+ a_j, \begin{pmatrix} c_1 \\ \vdots \\ c_{m-1}\end{pmatrix}\right\rangle + \frac{\langle a_j, P_{m-1}b\rangle c_j}{\|P_{m-1}a_j\|_2^2}, \text{ so}$$

---

†Scolnik calls (7), or something similar to it, "Greville's formula.

$$(9)\qquad \frac{\langle a_m, P_{m-1}b\rangle}{\|P_{m-1}a_m\|_2^2}\left(c_m - \langle A_{m-1}^{+}a_m, \begin{pmatrix} c_1\\ \vdots \\ c_{m-1}\end{pmatrix}\rangle\right)$$

$$+\ \frac{\langle a_j, P_{m-1}b\rangle}{\|P_{m-1}a_j\|_2^2}\left(\langle A_{m-1}^{+}a_j, \begin{pmatrix} c_1\\ \vdots \\ c_{m-1}\end{pmatrix}\rangle - c_j\right) \le 0.$$

Because A_m is a feasible basis, theorem 2 (or step (b) of the penultimate iteration of the algorithm) implies $\langle a_m, P_{m-1}b\rangle \geq 0$, and, since $P_{m-1}b \neq 0$, we must have $\langle a_m, P_{m-1}b\rangle > 0$. Also, if u is a unit vector orthogonal to $a_1, \ldots, a_{m-1}$, then $P_{m-1}y = \langle y,u\rangle u$, so

$$\begin{aligned}\langle a_j, P_{m-1}b\rangle\langle a_m, P_{m-1}a_j\rangle &= (\langle b,u\rangle\langle a_j,u\rangle)(\langle a_m,u\rangle\langle a_j,u\rangle)\\ &= (\langle b,u\rangle\langle a_m,u\rangle)(\langle a_j,u\rangle)^2\\ &= \langle a_m, P_{m-1}b\rangle\|P_{m-1}a_j\|_2^2\ ;\end{aligned}$$

thus $$\frac{\langle a_m, P_{m-1}a_j\rangle}{\langle a_m, P_{m-1}b\rangle} = \frac{\|P_{m-1}a_j\|_2^2}{\langle a_j, P_{m-1}b\rangle} > 0.$$

Multiplying both sides of (9) by this positive number, we obtain:

$$\frac{\langle a_m, P_{m-1}a_j\rangle}{\|P_{m-1}a_m\|_2^2}\left(c_m - \langle A_{m-1}^{+}a_m, \begin{pmatrix} c_1\\ \vdots \\ c_{m-1}\end{pmatrix}\rangle\right) + \langle A_{m-1}^{+}a_j, \begin{pmatrix} c_1\\ \vdots \\ c_{m-1}\end{pmatrix}\rangle - c_j$$

$$= (c_1, \ldots, c_m)\begin{pmatrix} A_{m-1}^{+}\left(I - \dfrac{a_m a_m^T P_{m-1}}{\|P_{m-1}a_m\|_2^2}\right)\\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_m^T P_{m-1}/\|P_{m-1}a_m\|_2^2\end{pmatrix} a_j - c_j \le 0,$$

which, by formula (7) with $j = m$, is the desired conclusion. ∎

4.2 Discussion

Aside from possibly motivating theorem 3, theorem 2 appears to have little bearing on Scolnik's other observations and suggestions.

Following theorem 3 we assumed that the columns of A could be so ordered that A_m would be an optimal basis and $a_1,\ldots,a_{m-1}$ would span a facet of the cone K generated by all the columns of A. Following theorem 1 we noted, in essence, that we could relabel the columns of A_m, say $\underline{a}_j = a_{p(j)}$ for some permutation p, so that $(\underline{a}_1,\ldots,\underline{a}_j)^+b \geq 0$ for $1 \leq j \leq m$. In motivating the algorithm we tacitly assumed that p could be the identity permutation, i.e., $\underline{a}_j = a_j$ $(j = 1,\ldots,m)$. However, this may not be possible; for example, let

$$A = \begin{pmatrix} 1 & -1 & -2 \\ 0 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix};$$

then $\{a_1,a_2\}$ is the only optimal basis, a_1 spans a facet of k but a_2 does not, and $(a_1)^+b = (-1) \not\geq 0$. This example is illustrated in figure 1. We thus see that steps (a) and (b) may conspire to prevent the algorithm from choosing an optimal basis. In the present example, for instance, it selects the non-optimal basis $\{a_3,a_1\}$. It is always possible to choose $\underline{a}_j = a_j$ in the examples given below, which demonstrate other shortcomings in the proposed algorithm.

Theorem 4 is false. For example, let

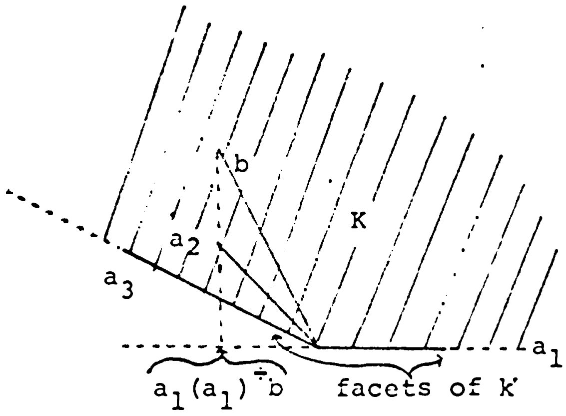

Figure 1

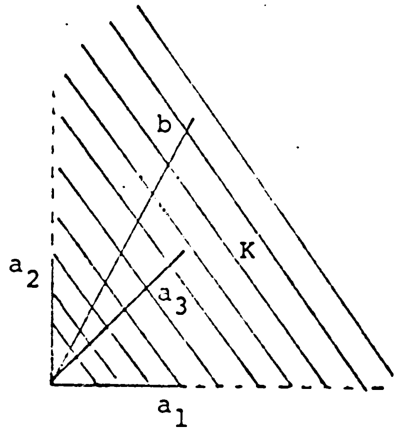

Figure 2

$$A = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \quad \text{and } c = (0,0,0,1)^T:$$

then $\{a_1,a_2,a_3\}$ is an optimal basis such that a_1 and a_2 span a facet of K (take $u = (0,0,1)^T$ in the remark following theorem 3), a_1 spans the facet $\{a_1x \mid x \geq 0,\ x \in \mathbb{R}^1\}$ of the subcone $\{A_2x \mid x \geq 0,\ x \in \mathbb{R}^2\}$, but $P_1b = (0,2,1)^T$ has $\langle a_4,P_1b\rangle = \langle(-1,-1,1)^T,(0,2,1)^T\rangle = -1 < 0$. Moreover, it is not possible to relabel a_1,a_2,a_3,a_4 so that theorem 4 holds for this example: $\{a_1,a_2,a_3\}$ is the only optimal basis for this problem; neither $\{a_1,a_3\}$ nor $\{a_2,a_3\}$ spans a facet of K; and if we interchange a_1 and a_2, then we still have $\langle a_4,P_1b\rangle = -1 < 0$.

Theorem 4 was the basis for step (b) of the proposed algorithm; if the conclusion of theorem 4 does not hold for the particular problem at hand, then step (b) will prevent the choice of a vector in an optimal basis and will thereby cause the algorithm either to unsuccessfully terminate or to find a nonoptimal feasible basis.

Even when theorem 4 holds for the problem at hand, step (c) of the algorithm may cause selection of the wrong column. For instance, let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad c = (1,1,0)^T:$$

during the first iteration, step (c) causes the algorithm to choose a_1 rather than a_2 as the first basic column; in the second iteration, only a_2 yields a feasible basis (see figure 2), so the algorithm selects the basis $\{a_1,a_2\}$, which gives $x = (1,2,0)^T$ and $z(x) = 3$; but $\{a_2,a_3\}$ is the optimal basis with $x^* = (0,1,1)^T$ and $z(x^*) = 1$. For another instance, consider the example of §3.2 with c_4 changed from 0 to $-\frac{1}{2}$: during iteration 2, step (c) causes column 4 to be chosen as the second basis vector, but there is no way to complete columns 3 and 4 to a feasible basis, so the algorithm terminates unsuccessfully in the third iteration, despite the fact that columns 3,6, and 2 still comprise an optimal basis.

As some of the above examples have shown, when the algorithm has successfully terminated, it may have found a non-optimal feasible basis. The corollary to theorem 5 gives a computationally checkable sufficient condition for optimality of the selected basis (but, depending on the circumstances, it might be slightly faster to check condition (2b) directly). If the algorithm has run successfully and (5b) holds for all $j > m$ such that (5a) holds but $\langle a_k, P_{m-1}b\rangle = 0$ for some $k > m$, then theorem 5 does not assure optimality of the selected basis; for example, let

$$A = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}, \quad c = (-2,0,1,-5)^T:$$

the algorithm selects the feasible basis A_3, which gives $x = (2,2,1,0)^T$ and $z(x) = -3$; but $\langle a_4, P_2 b\rangle = \langle(2,1,0)^T, (0,0,1)^T\rangle = 0$, and it happens that $\{a_2,a_3,a_4\}$ is the optimal basis, giving $x^* = (0,1,1,1)^T$ and $z(x^*) = -4$. On the other hand, if the algorithm has successfully run and (5a) holds for all $j > m$ but there is a $k > m$ for which (5b) fails, i.e., $(c_1,\ldots,c_{m-1},c_k)(A_{m-1},a_k)^{-1}b < (c_1,\ldots,c_m)A_m^{-1}$, then we may consider two cases: if $A_m^{-1}b > 0$ (strict inequality in all components), then the basic solution corresponding to A_m is not optimal; but if degeneracy occurs, i.e., at least one component of $A_m^{-1}b$ equals zero, then it is possible that the basic solution corresponding to A_m is optimal; for instance, let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}, \quad b = a_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad c = (-1,0,-2)^T;$$

then $S = \{(0,1,0)^T\}$ (see (1a)) and the algorithm selects the basis A_2, which, of course, gives $x = x^* = (0,1,0)^T$; but $(c_1,c_3)(A_1,a_3)^{-1}b = -1 < 0 = (c_1,c_2)A_2^{-1}b$.

Note that the only hypotheses used in the proof of theorem 5 are that A_m is a feasible basis and (5) holds. Successful termination of the algorithm guarantees that A_m is a feasible basis; aside from this, theorem 5 does not depend on the internal workings of the algorithm.

One interpretation of theorem 5 is that if the algorithm has successfully terminated, then it has found the solution of a certain subproblem, namely the one in which we require $x_j = 0$ for all j such that (5) does not hold. If we further restrict the problem so that $x_j = 0$ if $j > m$ and (A_{m-1}, a_j) is not a feasible basis, then the algorithm has found an optimal basis because it has essentially inspected all feasible bases. Specifically, assume that the columns of A have been so ordered that $\{a_1, \ldots, a_{m-1}, a_j\}$ is a feasible basis for $m \leq j \leq m+k$ (this is the only hypothesis we use in the following argument, so we may permute $a_1, \ldots, a_{m-1}$ at will): then any basic feasible solution x corresponding to a feasible basis B chosen from $a_1, \ldots, a_{m+k}$ corresponds, in fact, to a basis of the form $\{a_1, \ldots, a_{m-1}, a_j\}$. To see this, assume without loss of generality (WLOG) that $B = \{a_j | r \leq j < m+r\}$ and that $b = \sum_{j=r}^{m-1+r} \alpha_j a_j$, where $\alpha_j \geq 0$, so that the basic solution x corresponding to B has $x_j = \begin{cases} \alpha_j & \text{if } r \leq j < r+m \\ 0 & \text{otherwise} \end{cases}$. If $r \geq 2$ and $\alpha_s = 0$ for some s with $m \leq s < m+r$, say $s = m-1+r$, then because $a_1, \ldots, a_m$ are linearly independent, there is a maximum $j = j^* < r$ such that Rank $\{a_j, \ldots, a_{s-1}\} = m$, so, assuming WLOG that $j^* = r-1$, $a_{r-1}, \ldots, a_{(r-1)+m-1}$ are linearly independent and hence $\{a_{r-1}, \ldots, a_{(r-1)+m-1}\}$ is a feasible basis, to which corresponds the same basic feasible solution x as to B. Therefore we may assume WLOG that $\alpha_i > 0$ for $m \leq j < m+r$. Now let $a_o = b$. Since $P_{m-1}b \neq 0$, $(a_o, \ldots, a_{m-1})$ is non-singular; let u_j^T be the $j^{\underline{th}}$ row of $(a_o, \ldots, a_{m-1})^{-1} = \begin{pmatrix} u_o^T \\ \vdots \\ u_{m-1}^T \end{pmatrix}$, so that $\langle a_i, u_i \rangle = 1 > 0$ while $\langle a_i, u_j \rangle = 0$ for $i \neq j$ $(i, j < m)$. For each j with

$m \le j < m+r$ we have by assumption: $b = \sum_{i=1}^{m-1} \alpha_i^{(j)} a_i + \alpha_j^{(j)} a_j$,

where $\alpha_i^{(j)} \ge 0$ and (since $P_{m-1}b \ne 0$) $\alpha_j^{(j)} > 0$; thus $\begin{pmatrix} u_o^T \\ \vdots \\ u_{m-1}^T \end{pmatrix} a_j =$

$(a_o,\ldots,a_{m-1})^{-1} a_j = \frac{1}{\alpha_j^{(j)}} \begin{pmatrix} 1 \\ -\alpha_1^{(j)} \\ \vdots \\ -\alpha_{m-1}^{(j)} \end{pmatrix}$, so $\langle a_j, u_o \rangle > 0$ and

$\langle a_j, u_i \rangle \le 0$ for $1 \le i < m$ and $m \le j < m+r$. Now recall the equation $b = a_o = \sum_{j=r}^{m-1+r} \alpha_j a_j$: since $\alpha_j > 0$, we must have $\langle a_j, u_i \rangle = 0$ for $m \le j < m+r$ and $1 \le i < r$; but this also holds for $r \le j < m$, and so, since $a_r,\ldots,a_{m-1+r}$ are linearly independent and hence form a basis for R^m, we must have $r = 1$ (because otherwise u_1 would be a non-zero vector orthogonal to every element of this assumed basis).

Theorem 4 is valid for $m = 2$, since in this case it reduces to theorem 3. Thus it might appear that Scolnik's proposed algorithm might somehow be modified so that it would solve special cases of (1) with $m = 2$.* The special cases are those in which

---

* In the addendum to his symposium handout, Scolnik asserts that: if $m = 2$ and there is an i such that $c_i < 0$ and $\langle a_i, b \rangle \ge 0$ (which condition can be realized by adding a multiple of an appropriate row of A to c), then the algorithm finds an optimal basis. His proof uses (among others) the assumptions that after the algorithm has run: $c_1(a_1)^+ b \le c_i(a_i)^+ b$ for all i such that $\langle a_i, b \rangle \ge 0$ (which, because of step (b) of the algorithm, will often be invalid) and $c_1(a_1)^+ b \ge (c_1,c_2)A_2^{-1} b$. For example, let $A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 2 & 1 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 6 \\ 6 \end{pmatrix}$, $c = (-1,2,-3,2)^T$: then $c_i < 0$ and $\langle a_i, b \rangle > 0$ for $i = 1, 3$, yet the algorithm finds the basis A_2, which gives $x = (3,3,0,0)^T$ and $z(x) = 3$, instead of the optimal

(1) has an optimal basis, one column of which spans a facet of $K = \{Ax | x \geq 0\}$. If we somehow know a priori that this is the case, then we can permute the columns of A so that, say, a_1 and a_n span the facets of K (see figure 3); we then know that there is an optimal basis of the form $\{a_1, a_j\}$ or $\{a_n, a_j\}$, so we need inspect at most $2n-1$ candidate bases. If step (c) of the algorithm could be modified so that it told us with certainty which of a_1 or a_n would appear in an optimal basis, then we might save ourselves a small amount of work; as it stands, however, the present algorithm makes no contribution in even this special case.

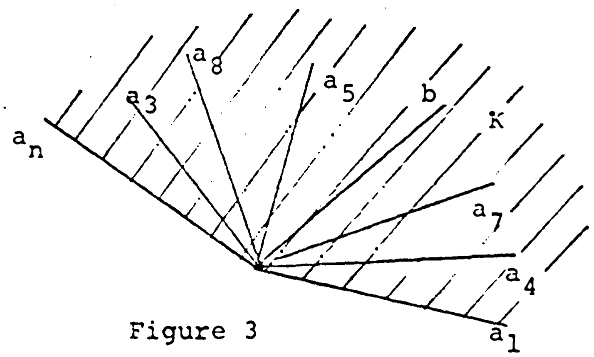

Figure 3

5. Conclusion

As mentioned in the introduction, Scolnik envisioned an algorithm which would construct an optimal basis for the linear programming problem (1) one column at a time; after selecting the first t columns of this basis, the algorithm would apply two

---

basis $\{a_3, a_4\}$, which gives $x^* = (0,0,3,3)^T$ and $z(x^*) = -3$; in this example $c_3(a_3)^+b = -10.8 < c_1(a_1)^+b = -6 < 3 = (c_1, c_2)A_2^{-1}b = z(x)$.

rejection criteria (a) and (b) to reduce the number of candidates for the $(t+1)^{st}$ column and would then use a selection criterion (c) to choose one of the remaining candidates. Unfortunately, the proposed algorithm has some shortcomings. Criteria (a) and (b) together assume the existence of a sometimes impossible ordering of the columns of an optimal basis. Worse, criteria (b) and (c) are based on conjectures (theorem 4 and inequality (4c)) which we have here shown to be false; even when criterion (b) does not cause failure by preventing the right column from being chosen, criterion (c) may select the wrong column and thus force the algorithm to terminate unsuccessfully or to find a non-optimal basis. I see no obvious way to modify criteria (b) or (c) so that they perform properly.

All three criteria involve pseudoinverses computed with respect to the 2-norm. If the algorithm were correct, we might reasonably expect it to perform just the same if the pseudoinverses were computed with respect to an arbitrary inner product norm $||y||_Q = (y^TQy)^{\frac{1}{2}}$ (where Q is positive definite) on R^m; this it does not.* If an algorithm for the linear programming problem is to use pseudoinverses, perhaps they should be computed with respect to an inner product norm which somehow depends on the given data (A, b, and c).

---

* If Q is a diagonal matrix, then computing the pseudoinverses with respect to $||\cdot||_Q$ is equivalent to replacing A and b by $Q^{\frac{1}{2}}A$ and $Q^{\frac{1}{2}}b$, i.e., performing row scaling, and then computing the pseudoinverses with respect to the 2-norm. W.W. White mentioned in a letter to me that Gordon Bradley of the Naval Postgraduate School had noted that Scolnik's algorithm is sensitive to row scaling.

Except possibly for Scolnik's observation that if A_m is a feasible basis, then its columns may be so ordered that $A_j^+ b \geq 0$ for $1 \leq j \leq m$, I see among his observations nothing which might lead to a better algorithm for solving the linear programming problem.

6. References

[1] Businger, P., and Golub, G.H., contribution I/8 (pp. 111-118) of: Wilkinson, J.H., and Reinsch, C., Handbook for Automatic Computation, vol. 2: Linear Algebra, Bauer, F.L., et al, eds., Springer Verlag (1971).
Also printed as:
-- "Linear least squares solutions by Householder transformations", Numerische Mathematic 7, pp. 269-276 (1965).

[2] Rao, C.R., and Mitra, Sujit K., Generalized Inverse of Matrices & Its Applications, Wiley (1971).