© DISCRETE MATHEMATICS 4 (1973) 367–377. North-Holland Publishing Company

THE SIMPLEX ALGORITHM WITH THE PIVOT RULE OF MAXIMIZING CRITERION IMPROVEMENT

R.G. JEROSLOW

Carnegie–Mellon University, Pittsburgh, USA

Received 20 March 1972*

Abstract. We extend a result of Klee and Minty by showing that the Simplex Algorithm with the pivot rule of maximizing criterion improvement is not a good algorithm in the sense of Edmonds. The method of proof extends to other similar pivot rules.

0. Introduction

In their landmark contribution to the theory of maximization over polyhedra [6], V. Klee and G.J. Minty show that it is possible for the Simplex Algorithm of Dantzig [1] to require in the order of $n^{[d/2]}$ pivots before optimization occurs in a linear program with $(n-d)$ linear equality constraints in n non-negative variables, if the most commonly used pivot rule is employed (see [1]).[1]

Their result naturally leads to the question: can one do better (i.e., lessen the number of pivots) if other pivot rules are used? In their concluding paragraphs, Klee and Minty write: " ... our methods could probably be used exhibit the same bad behavior for many other pivot rules. Indeed, we do not believe there exists a pivot rule that turns the simplex method into a 'good algorithm' in the sense of Edmonds, though the rule calling at each stage for the greatest possible improvement ... of the objective function would seem to merit further study ...".

Here we wish to write a postscript to [6] which confirms the opin-

* Original version received 7 July 1971.

[1] As Klee and Minty note in [6], Gale (How to solve linear inequalities, Am. Math. Monthly 76 (1969) 589–599) has regarded the determination of the computational complexity of linear programming as a task which "has stood as a challenge to workers in the field for twenty years now and remains, in my opinion, the principal open question in the theory of linear computation."

ions of Klee and Minty; using their techniques, we shall derive the identical result for the rule they mention as meriting further study. In fact, it will be evident, after our discussion, that, for any fixed integer k, the following rule can be made to advance up the polyhedra with the same order of slowness as the usual pivot rule: as next pivot one chooses the first element in an optimal k-sequence, where the latter nomenclature denotes a sequence of k pivots which increases the criterion value at least as much as any other sequence of k pivots.

No doubt the reason why Klee and Minty suggest an investigation of the rule which calls for the greatest increase in the criterion function, is because their constructions in [6] do not exhibit programs which behave badly under this rule. In fact, if this rule is used on the hypercube constructed in their proof that $H(d+1, n+2) \geq 2H(d, n) + 1$, the optimum is reached in one pivot step, although $2^d - 1$ pivot steps are required if the usual rule in [1] is used.[2] Similarly, their second and major construction in [6] yields polyhedra in which the optimum is reached in k steps (uniformly in P), if this new rule is used. The main devices that force the Simplex Algorithm with the usual rule to pursue an excessively long tour of vertices are given in [6]: essentially, the usual rule is sensitive not only to the polyhedron described by the linear program, but also to the *representation* (in terms of inequalities) *for the facets* of the polyhedron, and by adjusting the representation, one can "fool" the usual pivot rule. However, the rule of maximizing criterion improvement, which we now proceed to examine, is independent of representation and is an intrinsic of the polyhedron (as imbedded in Euclidean space).

We follow the very cautious and conservative approach of [6] by admitting that we do not know the significance of our results for practical linear programming computation. After all, experience with the Simplex Algorithm is very good, so the polytopes we construct below do not occur in the applications (to date); but why not?

[2] The reader should consult [6] for any unexplained notation or terminology. A polytope is of *class* (d, n) if it is d-dimensional and has precisely n facets. It is *simple* if each of its vertices is incident to precisely d facets, and it is d-dimensional. Simple polytopes correspond to nondegenerate linear programming problems. $H(d, n)$ is the maximum number of pivots which can be encountered in a linear program deriving from a simple (d, n)-polytope, where *any* criterion-increasing pivot may be chosen.

1. The main construction

Let a (d, n) simple polytope P be given (our notation is from [6]). We shall assume that P is *reversible of length t*, by which we mean that the following conditions hold for P: There exists two vertices p, p^* of P and a linear functional ϕ such that:

(i) when the Simplex Algorithm with the pivot rule of maximizing criterion improvement starts at p with ϕ as criterion function, it defines a unique path $p = p_0, p_1, \ldots, p_t = p^*$ of adjacent vertices ending at p^* such that

$$\phi(p_0) < \phi(p_1) < \ldots < \phi(p_t);$$

(ii) when the Simplex Algorithm with the same pivot rule starts at p^* with $-\phi$ as criterion function, we obtain a unique path $p^* = q_0, q_1, \ldots, q_t = p$ ending at p of the same length with

$$-\phi(q_0) < -\phi(q_1) < \ldots < -\phi(q_t).$$

(A polytope P may have several lengths.)

For this given polytope P, we shall construct a polytope $V \subseteq \mathbf{R}^2$, and then, following a perturbation of $V \times P$ to a polytope $Q \subseteq \mathbf{R}^{d+2}$ which is combinatorially equivalent to $V \times P$, show that Q is a reversible polytope of type $(d+2, n+4k+3)$ and of length at least $2kt+4k$, where k is the number of facets in V.[3] Since we shall be able to obtain such a V for any given k, this will prove that

$$(1) \qquad M(d+2, n+4k+3) \geq 2kM(d, n) + 4k,$$

where $M(d, n)$ is the maximum of the lengths of reversible polytopes of type (d, m) with $m \leq n$. Assuming (1), we can prove the following result (which is our main result) exactly as Klee and Minty use their inequality $H(d+2, n+k+1) \geq kH(d, n) + k - 1$ to obtain their main results in [6].

Theorem 1. $\lim_n \inf M(d, n)/n^{[d/2]} \geq 1/6^{[d/2]^2}$, *whenever* $d \geq 2$,

[3] Since P is simple and V is also, $V \times P$, and hence its combinatorial equivalent Q, is simple.

and hence there is a constant $\gamma_d > 0$ *such that*

$$M(d, n) \geq \gamma_d \, n^{[d/2]}, \; d \geq 2$$

Proof (assuming (1)). The proof is by induction on d. It is evident that

$$M(2, n) \geq n - 4$$

and an easy geometrical construction (which we leave to the reader) establishes

$$M(3, n) \geq \tfrac{1}{4} n ,$$

so that the "ground cases" $d = 2$ and $d = 3$ pose no difficulty.

We establish the result for $d + 2$ by using the result for d and the following inequalities:

$$\begin{aligned}
\liminf_{n \to +\infty} \frac{M(d+2, n)}{n^{[(d+2)/2]}} &= \liminf_{n \to +\infty} \frac{M(d+2, 5n+3)}{(5n+3)^{[d/2]+1}} \\
&\geq \liminf_{n \to +\infty} \frac{2n \, M(d, n) + 4n}{(6n)(6n)^{[d/2]}} \\
&\geq \tfrac{1}{3} \, \frac{1}{6^{[d/2]}} \liminf_{n \to +\infty} \frac{M(d, n)}{n^{[d/2]}} \\
&\geq \frac{1}{6^{[d/2]+1}} \cdot \frac{1}{6^{[d/2]^2}} \\
&\geq \frac{1}{6^{([d/2]+1)^2}} = \frac{1}{6^{[((d/2)/2]^2}}
\end{aligned}$$

In the remainder of this section, we shall be working toward obtaining (1).

As in [6], the main difficulty to overcome is to insure that the deformed polytope Q is combinatorially equivalent to $V \times P$. It is generally *false* that a small perturbation of a polyhedron does not change its combinatorial type; visualize, for instance, a cube in three dimensions, in which two diagonally opposite points on the top face are "pushed down," so that the top face becomes two faces. The cross-product construction is useful precisely because it allows us to obtain many vertices with only few facets in the polytope (so that long vertex paths can exist), since the cross-product of a (d, n) and a (c, m) polytope is a $(d+c, m+n)$ polytope, so that the facets add, while the vertices multiply. If the deformed polytope Q has many more faces than $V \times P$, we cannot obtain Theorem 1.

In [6], Klee and Minty required only a very small deformation of the cross product. In small deformations, the extreme points of a polytope are unchanged. Using this fact, and the following lemma they developed, they were able to show that the combinatorial type of the polytope they constructed was that of the cross product.

Lemma 2 (see [6]). *Let X and Y be polytopes having the same number m of vertices, the vertices of X being $x_1, ..., x_m$ and those of Y being $y_1, ..., y_m$. Suppose that for each index set $I \subseteq \{1, 2, ..., m\}$, whenever the convex hull of $\{x_i | i \in I\}$ is a facet of X, then the convex hull of $\{y_i | i \in I\}$ is a facet of Y. Then X and Y are combinatorially equivalent.*

In the construction below, we shall make use of the observation that, even if one does deform the cross product polytope quite substantially, as long as one does it in the manner of Klee and Minty, involving certain considerations of parallelism (we shall be explicit below), then the vertices of the perturbed polytope do correspond to those of the cross product, so that Lemma 2 can be applied exactly in the way that Klee and Minty apply it to obtain combinatorial equivalence of the perturbed polytope with the cross product. We shall need to employ substantial, rather than small, deformations, to obtain the polytope Q; and the ideas of large deformations and the construction of a V that depends on P are really the only new devices we bring to the subject matter.

With P given and as described, let us now begin the construction of the polytope $V \subseteq \mathbf{R}^2$. We will simultaneously define two sequences of points $v_0, v_1, ..., v_{4k}, v_{4k+1}, v_{4k+2}$ and $w_0, w_1, ..., w_{4k}, w_{4k+1}, w_{4k+2}$; the former points will be the vertices of V, all points w_i will be interior to V, and the condition will be met that

$$v_i v_{i+1} \text{ is parallel to } w_i w_{i+1} \quad i = 0, ..., 4k+1 .$$

The reader may wish to follow our construction with paper and pencil, since we shall refer to geometrical aspects of it.

We chose as v_1 the point (0, 1). Letting α denote the linear functional of $x \in \mathbf{R}^2$ which gives the first co-ordinate, we chose w_1 on the line $y = 1$ so that $\alpha(w_1) = \alpha(v_1) + 1$ (and hence $w_1 = (1, 1)$). We shall assume, without loss of generality, that $\phi(P) = [0, 1]$, so that $\phi(p_1) = 0$ and $\phi(p_t) = 1$. In the following, let $\lambda > 0$ be the minimum of the positive numbers $\phi(p_{i+1}) - \phi(p_i)$ and $\phi(q_i) - \phi(q_{i+1})$ for $i = 1, ..., t-1$.

To every point $p \in P$, we are going to assign the point

$$p^1 = ((1 - \phi(p))v_1 + \phi(p)w_1, p)$$

in a deformation Q of the Cartesian product $V \times P$. We wish to arrange it so, that, if p_0^1 $(=(p_0)^1)$ is the initial solution of the linear program over the polytope Q with α as criterion function, the Simplex Algorithm, under the pivot rule of maximizing criterion improvement, will proceed up through the points $p_0^1, p_1^1, ..., p_t^1$. To do so, we wish to make no points of Q adjacent to p_0^1 "less attractive" under this pivot rule than (v_2, p_0) and (v_0, p_0). It will be possible to arrange things so that the points adjacent to p_0^1 are (v_2, p_0) and (v_0, p_0) (where we are about to choose both v_0 and v_2), and the points p^1, where p is adjacent to p_0 in P. To make all these latter points "more attractive" than the two former points mentioned, it will suffice to take v_2 and v_0 so that we have both $\alpha(v_2) - \alpha(v_1) < \lambda$ and $\alpha(v_0) - \alpha(v_1) < \lambda$. We further restrict v_2 so that the line $v_1 v_2$ has positive slope and restrict v_0 so that the line $v_1 v_0$ has negative slope.

Let us assume that these restrictions have been met and we have chosen v_0 and v_2. Then we choose the points w_i so that $v_i w_i$ is parallel to $v_1 w_1$ and so that $\alpha(w_i) = \alpha(v_i) + 1$ for $i = 0, 2$. For each $p \in P$ and $i = 0, 2$, we associate the point

$$p^i = (\phi(p)w_i + (1 - \phi(p))v_i, p)$$

of Q. This same relation shall also be used in the further when the other points $v_3, ..., v_{4k+2}$ and $w_3, ..., w_{4k+2}$ have been defined, to define points

p^i for $p \in P$ and $i = 3, \ldots, 4k+2$; Q is then defined as the convex hull of the points $p^i \in R^{d+2}$.

Taking for granted that the points p^i are (as we shall show later) vertices of Q, by our choices of v_0, v_2, w_0, w_2, it is clear that the Simplex Algorithm will proceed through the vertices $p_0^1, \ldots, p_t^1$ as desired. However, when p_t^1 is reached, no further improvement can be obtained by moving through the P component (i.e., the last d co-ordinates) of Q, so a change in the V component (i.e., the first two co-ordinates) is needed for improvement. Thus the points then considered by the algorithm are p_t^2 and p_t^0; we want it to proceed to p_t^2, and to do so we need only that

$$\alpha(v_0) - \alpha(v_1) < \alpha(v_2) - \alpha(v_1) ,$$

which we can certainly assume, without any loss of generality.

Once at the point p_t^2, all points p^2 represent no increase in the criterion value, as must be, since $v_1 w_1 w_2 v_2$ is a parallelogram. It therefore again pays only to move through the V component. Now let us choose v_3 so that $v_2 v_3$ has positive slope, but slope less than that of $v_1 v_2$; this will insure that v_3 is an extreme point of V, once $v_3 v_4$ is chosen also to have positive slope, but slope less than that of $v_2 v_3$. Fixing some slope meeting these requirements for the line $v_2 v_3$, we have freedom as to where exactly we shall place v_3 on that line. Extend through the point w_2 a line L parallel to the line which is to be $v_2 v_3$, and draw any point w_3 on L strictly to the right of w_2. Then v_3 is chosen on $v_2 v_3$ so as to insure that $\alpha(v_3) - \alpha(w_3) = 1$. Note that $v_2 w_2 w_3 v_3$ is a trapezoid (usually not isosceles).

Returning to the behavior of the algorithm, we see that, when it is pivoting at p_t^2, the adjacent point p_t^3 will be chosen, since it is the only adjacent point with criterion improvement. We now want the algorithm to proceed to take the long route $p_t^3 = q_0^3, q_1^3, q_2^3, \ldots, q_t^3 = p_0^3$. To do so, we have to make all alternatives at p_t^3 worse than the adjacent point q_1^3. This is easily done by choosing v_4 so that the slope $v_3 v_4$, while positive, is less than that of $v_2 v_3$, choosing w_4 so that $v_3 w_3 w_4 v_4$ is a parallelogram, and insuring that $\alpha(v_4) - \alpha(v_3) < \lambda$.

When the point $p_0^3 = q_t^3$ is reached by the algorithm, the only adjacent point offering improvement is p_0^4. Once at p_0^4, improvement can again be obtained only by moving through the V component of Q. We choose a line M of positive slope less than that of v_3v_4 passing through v_4, and put v_5 on this line any place to the right of v_4. Then w_5 is choosen so that $w_4v_4v_5w_5$ is a trapezoid with parallel sides v_4v_5 and w_4w_5, and so that $\alpha(w_5)-\alpha(v_5) = 1$.

Now the pattern repeats, as it does every four vertices; v_5 is treated like v_1; v_6 and w_6 are defined so that $v_5w_5w_6v_6$ is a parallelogram and $\alpha(v_6)-\alpha(v_5) < \lambda$; w_7 and v_7 are defined so that $v_6w_6w_7v_7$ is a trazoid and $\alpha(v_7)-\alpha(w_7) = 1$; v_8 and w_8 are defined so that $v_7w_7w_8v_8$ is a parallelogram and $\alpha(v_8)-\alpha(v_7) < \lambda$; etc. The pattern ends with the construction of the points v_{4k} and w_{4k} (where k could have been chosen arbitrarily). Then the point v_{4k+1} is chosen so that v_{4k+1} lies on $y = 1$ and has $\alpha(v_{4k+1}) > \alpha(v_{4k})$; w_{4k+1} is chosen so that $\alpha(v_{4k+1})-\alpha(w_{4k+1}) = 1$. Finally, the point v_{4k+2} is chosen so that v_0v_{4k+2} is parallel to the x-axis and $v_{4k+2}v_{4k+1}$ has (say) the same slope as v_1v_0, but with negative sign, and w_{4k+2} is chosen so that $\alpha(v_{4k+2})-\alpha(w_{4k+2}) = 1$.

Now, provided that we have indeed insured that Q is combinatorially equivalent to $V \times P$ and that the adjacent vertices are as we described them, it is clear that in traversing every four vertices of Q the algorithm takes $2t + 4$ pivots, so that at least $2kt + 4k$ are required in all. Furthermore, an investigation of the behavior of the algorithm for criterion function when started at initial solution p_0^{4k+1} (assuming the facts on vertices and adjacency are correct) will reveal the same number of pivots so long as we have chosen v_{4k+1} so that $\alpha(v_{4k+1})-\alpha(v_{4k}) \geq 1$, say (which can always be done), so that Q is a reversible polytope of length $2kt + 4k$, justifying our inequality $M(d+2, n+4k+3) \geq 2kM(d, n)+4k$ and thereby Theorem 1.

What remains is to prove that p^i for $i = 0, ..., 4k+2$, p an extreme point of P, are precisely the vertices of their convex hull, and then use this fact combined with Lemma 2 to give the combinatorial equivalence. The adjacency relations will automatically be satisfied because the correspondence between vertices of Q and $V \times P$ which we now assert is that p^i correspond to (v_i, p), and hence the adjacency relations in Q can be easily read off from those in $V \times P$, which is readily seen to satisfy our assertions in this regard.

If our claim regarding the vertices of Q is false, then there is an extreme point p of P and an i for which a convex combination of the following form holds:

$$p^i = \sum_{j \neq i} \lambda_p^j p^j + \sum_{j=0}^{4k+2} \{\lambda_q^j q^j \mid q \text{ an extreme point of } P, q \neq p\},$$

$$1 = \sum_{j \neq i} \lambda_p^j + \sum_{j=0}^{4k+2} \{\lambda_q^j \mid q \text{ an extreme point of } P, q \neq p\},$$

where $\lambda_q^j \geq 0$. Since $p \in P$ is an extreme point of P, the fact that the last d components of points q^j are $q \in P$, shows that $\lambda_q^j = 0$ for all $q \neq p$ and all j. Thus, we now have that (taking first two co-ordinates)

$$(1-\phi(p))\,v_i + \phi(p) w_i = \sum_{j \neq i} \lambda_p^j (1-\phi(p))\,v_j + \phi(p)\,w_j.$$

To refute this conclusion, thus establishing the claim, we need only show that for each $p \in P$ the points $(1-\phi(p))v_j + \phi(p)w_j$, which we shall call ${}^jp$, are precisely the vertices of their convex hull in $\mathbf{R}^2$.

We proceed to this latter issue as follows. Since $v_1 v_2$ is parallel to $w_1 w_2$, and the proportion of $v_1\,{}^1p$ to $v_1 w_1$ is the same as the proportion of $v_2\,{}^2p$ to $v_2 w_2$ (and is, namely, $\phi(p)$), we see that ${}^1p\ {}^2p$ is parallel to $v_1 v_2$. Similarly, ${}^2p\ {}^3p$ is parallel to $v_2 v_3$, ${}^3p\ {}^4p$ parallel to $v_3 v_4$, etc. Thus the slopes of all lines ${}^ip\ {}^{i+1}p$ are equal to the corresponding slopes of lines $v_i v_{i+1}$. For the very same reason that our choices of slopes in the lines $v_i v_{i+1}$ made the points v_i the extreme points of their convex hull, the points ${}^jp$ will also be the extreme points of their convex hull. We conclude that the points p^j are indeed the extreme points of Q.

With the correspondence of p^i to (v_i, p) between the extreme points of Q and $V \times P$, respectively, we show that the hypotheses of the Lemma 1 are satisfied precisely as Klee and Minty do in [6] for the polytope they construct in their proof that $H(d+2, n+k+1) \geq kH(d, n) + k - 1$. Essentially, the very same functionals defining the faces of Q and $V \times P$, respectively, can be employed. This proves Theorem 1.

2. Concluding comments

In our construction, the pivot rule which calls for examining the next (say) two possible pivots, and choosing as the best pivot the one which begins the best sequence of two pivots, would have proceeded to the optimum in Q in only 3 pivots, independent of P. But this pivot rule can also be made to slow up, by the following devices. Instead of just one point v_0 to the right and down from P a miniscule distance, two extreme points are to be put to the right and down from P a negligible distance. Thus one makes

$$\alpha(v_2)-\alpha(v_1) < \lambda, \ \alpha(v_3)-\alpha(v_2) < \lambda\,,$$

but $\alpha(v_4)-\alpha(v_3) > 2$; and then

$$\alpha(v_5)-\alpha(v_4) < \lambda, \ \alpha(v_6)-\alpha(v_5) < \lambda\,.$$

Hence, when started at p_0^1, the best possible gain in two pivots is to go p_1^1 and p_2^1 so the first pivot would be p_1^1. By an analysis similar to the one in Section 1, it can easily be shown that the beginning of the path chosen by the pivot rule under discussion will be

$$p_0^1, p_1^1, p_2^1, \ldots, p_t^1, p_t^2, p_t^3, p_t^4 = q_0^4, q_1^4, \ldots, q_t^4 = p_0^4, p_0^5, p_0^6$$

and that this pattern repeats every 6 vertices of V, which is to be chosen to have $6k+4$ vertices.

The construction would be similar for a pivot rule which proceeds by examining the sequences of all possible next r pivots for fixed r; by insuring that the next r extreme points of V give bad improvement compared to moving through P, one forces the algorithm first up and then down P, the cycle repeating every $2(r+1)$ points. Thus one obtains inequalities similar to Theorem 1 in which the constant may be less than $\frac{1}{6}$, but the order of magnitude of the number of pivots, namely $n^{[d/2]}$, does not change.

It seems that any algorithm which proceeds in a purely local manner across a polytope will suffer from the same deficiencies as the pivot rule we have examined. But what of algorithms which simultaneously explore different local regions of the polytope, seeking to combine the local knowledge into a global estimate of the shape of the polytope? This question, while interesting, appears to be purely academic, because the Simplex Algorithm works well in practice.

Acknowledgments

We wish to thank Professor J. Goldman for stimulating us to undertake the work reported here and Professor V. Klee for his interest in our work and results.

The referees provided decidedly helpful and detailed criticisms which we used in preparing the final draft of this paper.

References

[1] G.B. Dantzig, Maximization of a linear function of variables subject to linear inequalities, in: Activity analysis of production and allocation, T.C. Koopmans, ed., Cowles Commission Monograph 13 (Wiley, New York, 1950).

[2] G.B. Danzig, Linear programming and extensions (Princeton Univ. Press, Princeton, N.J., 1963).

[3] B. Grünbaum, Convex polytopes (Wiley, New York, 1967).

[4] V. Klee, A class of linear programs requiring a large number of iterations, Numer. Math. 7 (1965) 313–321.

[5] V. Klee, Convex polytopes and linear programming, in: Proc. IBM sci. computing symp. on combinatorial problems (IBM Data Processing Division, White Plains, N.Y., 1966) 123–158.

[6] V. Klee and G.J. Minty, How good is the Simplex Algorithm? , in: Inequalities III, O. Shisha, ed. (Academic Press, New York, 1971).