

Geometry of the Gass–Saaty Parametric Cost LP Algorithm*

Victor Klee¹ and Peter Kleinschmidt²

¹ Department of Mathematics, University of Washington, Seattle, WA 98195, USA

² Institut für Mathematik und Informatik, Universität Passau, Passau,
Federal Republic of Germany

Abstract. In previous discussions of the Gass–Saaty algorithm, the possibility of cycling is avoided by making strong nondegeneracy assumptions or by incorporating a lexicographic decision rule. By analyzing the geometric ideas on which the algorithm is based, it is shown here that even without any “lexicography,” cycling is impossible unless the two objective functions are related in a very special way to each other or to the constraints defining the feasible region P . In particular, the avoidance of cycling does not require any restriction on the facial structure of P or on the algebraic relationships among the linear equalities and inequalities by means of which P is defined.

Introduction

The Gass–Saaty algorithm [6], [7] is concerned with the maximization of linear objective functions on a feasible region P defined by linear inequality constraints in nonnegative real variables. In addition to two distinct objective functions f and g , it involves affine combinations of f and g —functions of the form $f_\mu = (1 - \mu)f + \mu g$. The computation starts from a basic feasible solution (bfs) s_0 that maximizes f on P , and then uses the pivots of the simplex method to produce a sequence s_0, s_1, \dots, s_k of bfs and a subdivision of the unit interval

$$0 = \sigma_{-1} \leq \sigma_0 \leq \dots \leq \sigma_k = 1 \quad (1)$$

* Work on this paper by the first author was supported in part directly by the National Science Foundation, and in part by the Mathematical Sciences Research Institute and the Institute for Mathematics and Its Applications with funds provided by the National Science Foundation. Work by the second author was supported in part by the Alexander von Humboldt Stiftung.

such that the following conditions are satisfied:

- (a) for $1 \leq i \leq k$, s_i is obtained from s_{i-1} by means of a single pivot;
- (b) for $0 \leq i < k$, s_i provides a P -maximum of f_μ for all $\mu \in [\sigma_{i-1}, \sigma_i]$;
- (c) s_k either provides a P -maximum of g or makes it easy to recognize that $\sup\{g(p) : p \in P\} = \infty$.

The first part of (c) applies when P is bounded, and the inequalities in (1) are strict when P and its representation are “nondegenerate.”

The Gass-Saaty algorithm is useful when both f and g provide plausible ways of measuring the “value” of a point of P . If, as often happens, f and g cannot be maximized simultaneously, it may be desirable to inspect each of the bfs s_0, s_1, \dots, s_k and then choose among them by means of supplementary criteria. Also, when s_0 is known from a previously solved problem that required maximization of f , the parametric procedure is useful for maximizing an objective function g that results from a slight perturbation of f . Finally, when the sole aim is to maximize g , every choice of f and s_0 (actually, as we shall see, *almost* every choice) provides a pivot rule that determines the progression from one bfs to another until a g -maximizing bfs is reached (if one exists).

In studying the algorithm from the last-mentioned viewpoint. Murty [10] shows that its worst-case behavior is exponential, while Borgwardt [2]-[4] and Haimovich [9] show that under reasonable assumptions on the distribution of input data, the average-case behavior is polynomially bounded. The sharp average-case results of Adler *et al.* [1] are also based on the parametric procedure of Gass and Saaty [6], [7]. All of these authors make strong nondegeneracy assumptions to avoid the possibility of cycling (also, in some cases, for deeper reasons) and thus assure that the algorithm does indeed progress from the given f -maximizing bfs to a g -maximizing bfs. As shown by Murty [11], cycling can also be avoided by the use of a suitable lexicographic decision rule.

In [11] Murty describes an example of Gana [5] which shows that unless some precautions (such as a lexicographic rule) are taken, cycling can occur in the “parametric right-hand-side simplex algorithm.” This is dual to the “parametric cost simplex algorithm” (which we have called the Gass-Saaty algorithm), and Murty remarks that cycling can occur in the latter as well. When Gana’s cycling example is converted into an example of cycling in the Gass-Saaty algorithm, it is seen that each of the two objective functions f and g is a negative multiple of the other. Our geometric analysis shows that this is in a sense typical. Cycling in the Gass-Saaty algorithm cannot arise from special geometric properties of the feasible region or from special algebraic relationships among the constraints defining the region. It can arise only from degeneracy in the relationship of the two objective functions to each other or to the constraints.

In the absence of degeneracy, the geometry of the Gass-Saaty algorithm is very simple, and it has already been discussed in some of the papers mentioned above. The degenerate case is also not very complicated, but it does require a clear geometric understanding of the basic pivot process. In order to explain the geometry of the Gass-Saaty algorithm as clearly as possible, the first two sections

below provide a purely geometric formulation with no computational details. Computational aspects are treated in the third section, and the fourth section discusses the example of Gana [5]. Some of the proofs have been left to the reader.

Our section headings are as follows: 1. Pointed polyhedra and their support systems; 2. Line crossing a system of polyhedral cones; 3. Computational aspects; 4. An example of cycling.

1. Pointed Polyhedra and Their Support Systems

The Gass-Saaty algorithm is of course an algebraic procedure, but the aspect that we want to discuss is best understood in geometric terms. To prepare for the geometric analysis in Section 2, the present section collects the relevant properties of pointed polyhedra and their support systems. Since the properties are all well known, they are stated here without proof. Some detailed arguments can be found in the books of Grünbaum [8], Murty [10], and Schrijver [12].

Prefixes indicate dimension, and real d -space is denoted by \mathbb{R}^d . A *polyhedron* is a subset P of \mathbb{R}^d that is the intersection of a finite family of closed halfspaces. A set is said to be *polyhedral* if it is a polyhedron. A face of a polyhedron P is P itself, the empty set \emptyset , or P 's intersection with a supporting hyperplane. The 0-, 1-, and $(d-1)$ -dimensional faces of a d -polyhedron P are respectively its *vertices* (or *extreme points*), *edges*, and *facets*. The number of faces is finite, and P 's faces (aside from P and perhaps \emptyset) are precisely the intersections of facets.

A polyhedron is *pointed* if it has at least one vertex or, equivalently, is nonempty and contains no line. The unbounded edges of a pointed polyhedron P are its *extreme rays*, and it is known that P is the convex hull of the union of its extreme rays and extreme points. A set that is a nonempty union of rays issuing from the origin is called a cone. As is well known, a set is a pointed polyhedral cone if and only if it contains a ray but no line and is the intersection of a finite family of closed halfspaces whose bounding hyperplanes pass through the origin. A *simple* cone is one that consists of all nonnegative linear combinations of a linearly independent set; equivalently, it is a k -cone that has exactly k extreme rays. It is known that each polyhedral k -cone C is the union of all the (simple) cones of the form $\text{con}(R_1 \cup \cdots \cup R_k)$ (equivalently, $R_1 + \cdots + R_k$, the $+$ indicating vector addition), where the R_i 's are extreme rays of C .

The usual inner product in \mathbb{R}^d is denoted by $\langle \cdot, \cdot \rangle$. In terms of this inner product, points of \mathbb{R}^d may be regarded as linear functionals on \mathbb{R}^d , and vice versa.

The family \mathcal{C} of Theorem 1.1 is sometimes called the *support system* of P . The conclusion of Theorem 1.1 is denoted by SH1 because it is one of the standing hypothesis in Section 2.

Theorem 1.1. *Suppose that V is the vertex-set of a pointed d -polyhedron P in \mathbb{R}^d . For each $v \in V$, let*

$$C_v = \{w \in \mathbb{R}^d : \langle v, w \rangle = \max_{p \in P} \langle p, w \rangle\}.$$

Then with $\mathcal{C} = \{C_v : v \in V\}$, it is true that

SH1: \mathcal{C} is a nonempty finite family of pointed polyhedral d -cones in \mathbb{R}^d such that the union $\bigcup \mathcal{C}$ is convex and the intersection of any two members of \mathcal{C} is a face of each.

Note that C_v consists, in effect, of all the linear functionals on \mathbb{R}^d whose P -maximum is attained at the vertex v . (It is interesting also that the vector sum $v + C_v$ consists of all points $p \in \mathbb{R}^d$ such that v is a point of P nearest to p .) It is known that whenever a linear function on \mathbb{R}^d is bounded above on a pointed polyhedron P , it attains its P -maximum at a vertex of P . Hence $\bigcup \mathcal{C}$ is just the set of all linear functions that are bounded above on P . This set is obviously convex, it is closed, and when P is bounded it is all of \mathbb{R}^d .

In Theorem 1.1 P is merely a geometric object sitting in \mathbb{R}^d . However, for algorithmic purposes we must consider an algebraic representation of P .

Theorem 1.2. Suppose that z_1, \dots, z_h are points of $\mathbb{R}^d \setminus \{0\}$, none of which is a positive multiple of another, and that $\alpha_1, \dots, \alpha_h$ are scalars. For each i let

$$J_i = \{u \in \mathbb{R}^d : \langle u, z_i \rangle \leq \alpha_i\} \quad \text{and} \quad H_i = \{u \in \mathbb{R}^d : \langle u, z_i \rangle = \alpha_i\},$$

a closed half-space and its bounding hyperplane. With $P = \bigcap_1^h J_i$, assume that P is pointed and d -dimensional, and let V and \mathcal{C} be as in Theorem 1.1. For each $v \in V$, let

$$Z_v = \{z_i : v \in H_i\},$$

and set

$$Z = \bigcup_{v \in V} Z_v = \{z_i : H_i \text{ intersects } V\},$$

$$\mathcal{R} = \{[0, \infty[z : z \in Z\}.$$

Then

SH2: \mathcal{R} is a finite set of rays issuing from 0 such that $\bigcup \mathcal{R} \subset \bigcup \mathcal{C}$ and $\bigcup \mathcal{R}$ includes each extreme ray of each member of \mathcal{C} .

In effect, Z_v consists of all the z_i 's for which the associated constraint is "tight" or "binding" at v . By definition,

$$[0, \infty[z = \{\lambda z : 0 \leq \lambda < \infty\},$$

so \mathcal{R} is just the set of rays that issue from the origin and pass through points of Z . Note that we permit an unlimited (finite) number of constraints of each of the following sorts:

nonbinding—half-spaces J_i for which $z_i \in Z$;

binding but redundant—half-spaces J_i such that $z_i \in Z$ but the ray $[0, \infty[z_i$ is not an extreme ray of any member of \mathcal{C} ;

nonredundant associated with the same $v \in V$ —half-spaces J_i such that $z_i \in Z$ and the ray $[0, \infty[z_i$ is an extreme ray of C_v .

The usual nondegeneracy assumptions require that no binding constraint is redundant and that each $v \in V$ is associated with precisely d binding constraints—in other words, each member of \mathcal{C} is a simple d -cone and \mathcal{R} consists precisely of the extreme rays of the members of \mathcal{C} . In contrast, aside from requiring that the polyhedron P is pointed and full-dimensional, we do not restrict its combinatorial structure or the algebraic structure of its representation. It is certainly appropriate to assume pointedness, because whenever an LP problem is formulated in standard or canonical form, its feasible region is (if nonempty) pointed by virtue of being a polyhedral subset of a nonnegative orthant. However, the assumption of full dimensionality is temporary, designed to make the geometric picture easier to follow. As is shown at the end of Section 2 and in Section 3, it is not needed for our analysis of the Gass–Saaty algorithm.

In addition to the members of \mathcal{C} , two other collections of cones are essential to our analysis. For $h = d$ and for $h = d - 1$, let $(\mathcal{C}, \mathcal{R})_h$ denote the collection of all simple h -cones S such that the h extreme rays of S all belong to \mathcal{R} and are all contained in the same member of \mathcal{C} . Thus the members of $(\mathcal{C}, \mathcal{R})_{d-1}$ are just the facets of the members of $(\mathcal{C}, \mathcal{R})_d$. Each member C of \mathcal{C} is the union of the members of $(\mathcal{C}, \mathcal{R})_d$ that are contained in C , and the boundary ∂C is the union of the members of $(\mathcal{C}, \mathcal{R})_{d-1}$ that are contained in ∂C . Also, each member of $(\mathcal{C}, \mathcal{R})_d$ is contained in a unique member of \mathcal{C} , by the second part of SH1. Under the usual nondegeneracy assumptions, \mathcal{C} is equal to $(\mathcal{C}, \mathcal{R})_d$.

As is explained in the Introduction and in Section 3, the Gass–Saaty algorithm starts from a bfs s_0 that represents an f -maximizing vertex of P , and then attempts by suitable pivots to find a bfs that maximizes a different objective function g or to decide that g is not bounded above on P . Under the usual self-duality of \mathbb{R}^d with respect to the inner product, f and g may be regarded as points of \mathbb{R}^d , and they determine an oriented line L (oriented, say, from f toward g). There is no point to the algorithm when g is a positive multiple of f , and the example in Section 4 shows that cycling may occur when g is a negative multiple of f . For the analysis in Section 2, we make the following nondegeneracy assumption:

SH3: *L is an oriented line in $\mathbb{R}^d \setminus \{0\}$ that intersects $\bigcup \mathcal{C}$ but does not intersect the relative boundary of any member of $(\mathcal{C}, \mathcal{R})_{d-1}$.*

(The *relative boundary* and *relative interior* of a convex subset K of \mathbb{R}^d are respectively the boundary and the interior of K relative to the smallest flat (affine subspace) containing it. In particular, the relative boundary of a member $R_1 + \cdots + R_{d-1}$ of $(\mathcal{C}, \mathcal{R})_{d-1}$ is the union of the $d - 1$ simple $(d - 2)$ -cones formed by adding up all but one of the R_i 's.)

Finally, the initial f -maximizing bfs s_0 is brought into the picture by means of the following assumption:

SH4: *The point f of L belongs to the member S_0 of $(\mathcal{C}, \mathcal{R})_d$.*

For the situation described in Theorems 1.1 and 1.2, condition SH3 is unrealistic because the vertex-set V , the collection \mathcal{C} of polyhedral cones, and the collection

\mathcal{R} of rays are usually not immediately available from the presentation of the feasible region $P = \bigcap_1^h J_i$. However, SH3 is implied by the condition that no point of L can be expressed as a nonnegative combination of $d - 2$ of the points z_i , and this in turn is implied by the following easily tested condition:

SH3': For each linearly independent set Y consisting of $d - 2$ of the points z_i , the linear hull of $Y \cup L$ is the entire space \mathbb{R}^d .

2. Line Crossing a System of Polyhedral Cones

Three conditions from Section 1—SH1, SH2, and SH3—are assumed as standing hypotheses in the present section. Let $<$ denote the antireflexive linear ordering of the line L that agrees with L 's orientation, so that $f < g$ when L is oriented from f toward g . For each $u \in L$, let $L_{<}(u)$ (resp. $L_{>}(u)$) denote the open ray consisting of all points $q \in L$ such that $q < u$ (resp. $q > u$).

Now consider a sequence of the form

$$\dots, T_{-1}, S_{-1}, T_0, S_0, T_1, S_1, \dots, \quad (2)$$

where only S_0 is required to be present and the sequence may begin or end with either an S or a T . (For example, $S_{-2}, T_{-1}, S_{-1}, T_0, S_0, T_1$ and S_0, T_1, S_1, T_2, S_2 are both sequences of the indicated sort.) A sequence of the form (2) is called a $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain if the following four conditions are satisfied whenever the mentioned sets exist:

- (a) $S_i \in (\mathcal{C}, \mathcal{R})_d$;
- (b) T_i and T_{i+1} are facets of S_i ;
- (c) T_i includes a unique point q_i of L ;
- (d) $\dots < q_{i-1} < q_0 < q_1 < \dots$.

Condition (d) merely says that the points q_i march along L according to the given orientation. In particular, there is no repetition in the sequence (d) and hence (by (c)) none in the sequence of T_i 's. From (a)–(d) and SH1–SH3 it follows that when T_i and T_{i+1} are both present, L intersects the interior of the simple d -cone S_i in an open segment $]q_i, q_{i+1}[$ whose ends belong to the relative interiors of the simple $(d - 1)$ -cones T_i and T_{i+1} . It follows also that $T_i = S_{i-1} \cap S_i$ when all three of these sets are present. In particular, there is no repetition in the sequence of S_i 's.

When the system $(\mathcal{C}, \mathcal{R}, L, S_0)$ arises from a parametric LP problem under the usual strong nondegeneracy assumptions, each S_i in (2) is a member of \mathcal{C} . However, under the present hypotheses there may be many successive S_i 's in (2) that are properly contained in the same member of \mathcal{C} . (This occurs when many successive bfs correspond to the same vertex of the feasible region P .) But even under the present weak assumptions, the following is true:

Each sequence of pivots in the Gass–Saaty algorithm is associated with a $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain, and there is no repetition in such a chain. This is the crucial

point in explaining why degeneracies in the structure or representation of the feasible region cannot be responsible for cycling in the algorithm.

With SH3 as the only nondegeneracy condition, Theorem 2.1 below contains the geometric essence of the Gass-Saaty algorithm. Computational aspects are discussed in Section 3. To correlate the theorem with the steps of the algorithm, think of S_0 as the simple d -cone corresponding to the nonbasic variables associated with the initial f -maximizing bfs, and think of $[q_0, q_1]$ as a segment in L that contains f . The usual algorithm progresses along the line L in only one direction—from f toward g —and halts when it reaches g . However, the underlying geometry leads naturally to a subdivision of the entire line L , and that is described below.

A $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain (2) is said to be *full* if the following conditions are satisfied by the chain and the associated sequence (d) of q_i 's:

- (α) the chain starts with T_{-k} for $k \geq 0$, and the ray $L_{<}(q_k)$ misses $\bigcup \mathcal{C}$; or
- (α') the chain starts with S_{-k} for $k \geq 0$, and $L_{<}(q_{-k+1}) \subset S_{-k}$;
- (ω) the chain ends with T_l for $l \geq 0$, and the ray $L_{>}(q_l)$ misses $\bigcup \mathcal{C}$; or
- (ω') the chain ends with S_l for $l \geq 0$, and $L_{>}(q_{l-1}) \subset S_l$.

For a quadruple $(\mathcal{C}, \mathcal{R}, L, S_0)$ as in Section 1, condition (ω) arises when each point of $L_{>}(q_l)$ represents an objective function that is unbounded above on P . Condition (ω') arises when each point of $L_{>}(q_{l-1})$ represents an objective function whose P -maximum is attained at the vertex of P corresponding to the bfs whose set of nonbasic variables is associated with the simple cone S_l . Similar comments apply to the conditions (α) and (α').

Theorem 2.1. *When the standing hypotheses SH1-SH4 are all satisfied, each $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain that is not full can be extended to a full $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain by appending suitable S 's and T 's at one or both ends of the chain.*

Proof. We consider first the case in which the chain consists of S_0 alone. Since the point f of L belongs to S_0 by SH4, the line L certainly intersects S_0 . By the nondegeneracy condition SH3, L does not intersect the relative boundary of any facet of S_0 , and from this it follows that L intersects the interior of S_0 . Since S_0 contains no line, the intersection of L with the interior of S_0 must be one of the following: an open ray $L_{<}(q_1)$, where $f \leq q_1$; an open segment $]q_0, q_1[$, where $q_0 \leq f \leq q_1$; an open ray $L_{>}(q_0)$, where $q_0 < f$. Each of the q_i 's belongs to a unique facet T_i of S_0 , and that makes it possible to begin the extension process. (Section 3 discusses the algorithmic aspect of choosing T_i .)

Henceforth we discuss the extension process only for the right end of the chain. The situation for the left end is similar. We consider two cases, according to whether the (current stage of the) chain under discussion ends with T_{l-1}, S_l or with S_l, T_l . In the first case, the ray $L_{>}(q_{l-1})$ is either contained in the simple d -cone S_l —in which case the extension process at the right end of the chain is terminated—or it emerges from S_l at a point q_{l+1} of a simple $(d-1)$ -cone $F \in (\mathcal{C}, \mathcal{R})_{d-1}$ that is a facet of S_l , and in this case the chain is extended by appending $T_l = F$ at the right end.

In the second case, T_i and q_i are already present. Then the ray $L_{>}(q_i)$ either misses $\bigcup \mathcal{C}$ —in which case the extension process at the right end of the chain is terminated—or it enters a simple d -cone $G \in (\mathcal{C}, \mathcal{R})_d$ of which T_i is a facet, and in this case the chain is extended by appending $S_{i+1} = G$ at the right end. \square

Until now we have assumed that the pointed polyhedron $P \subset \mathbb{R}^d$ is d -dimensional because that leads to a simpler geometric picture and hence to easier understanding. Now it is time to review the arguments used earlier, in order to explain why the assumption of full dimensionality can be discarded. Suppose, then, that P is a pointed k -polyhedron in \mathbb{R}^d , with $k < d$. Let A denote the affine hull of P , so that A is a k -flat (k -dimensional affine subspace) of \mathbb{R}^d . Let N denote the set of all linear functionals on \mathbb{R}^d that are constant on A , so that N is a $(d - k)$ -dimensional linear subspace of \mathbb{R}^d . Just as before, let V denote the vertex-set of P , and for each $v \in V$ define the set C_v as before. Then C_v is still a polyhedral cone, but it is no longer pointed because of course it contains the subspace N . The cone C_v has no extreme rays of its own, but can be expressed in various ways as the direct sum of the subspace N and a pointed polyhedral cone which of course does have extreme rays. However, these extreme rays have no intrinsic importance. In the use of SH2, the requirement that

\mathcal{R} includes each extreme point of each member of \mathcal{C}

was significant only because it guaranteed that

for each $v \in V$, the cone C_v is the union of the members of the system $(\mathcal{C}, \mathcal{R})_d$ that are contained in C_v .

The latter condition is satisfied even when P is not required to be full-dimensional. And the use of SH1 did not depend on the pointedness of the members of the collection \mathcal{C} of polyhedral cones, but only on the finiteness of the collection and the fact that the intersection of any two members is a face of each. Those conditions are still satisfied when P is of dimension less than d .

3. Computational Aspects

The aim of this section is to relate the computational steps of the Gass-Saaty algorithm to the geometric description contained in the preceding sections. Since the algorithm starts from a basic feasible solution, we assume without loss of generality that the feasible region is given by

$$[A | I_m]x = b, \quad x \geq 0, \quad (3)$$

where A is an $m \times (n - m)$ matrix, I_m is the $m \times m$ identity matrix, and the m -vector b is nonnegative.

Relative to (3), let x_1, \dots, x_{n-m} denote the variables associated with the columns of A and let y_1, \dots, y_m denote the variables associated with the columns of I_m . Thus initially the x_r 's are nonbasic variables and the y_r 's are basic variables. Let X denote the set of "x-variables" and Y the set of "y-variables." We shall

construct a system of cones in which each cone of maximum dimension corresponds to a basic feasible solution of (3). This system will not necessarily satisfy all the conditions stated in hypotheses SH1-SH4, but it will fulfill all the conditions that were actually used in the proof of Theorem 2.1. (In this connection, see the last paragraph of Section 2.) Hence Theorem 2.1 will apply to the constructed system of cones.

Let \mathcal{B} denote the set of all feasible bases for (3), and for each $B \in \mathcal{B}$ let $X_B \cup Y_B$ denote the set of nonbasic variables, where $X_B \subset X$, $Y_B \subset Y$, and $\text{card}(X_B \cup Y_B) = n - m$. With e_i denoting the i th member of the standard basis for \mathbb{R}^{n-m} and a_k denoting the k th row of the matrix A , the cone $S(B)$ associated with the feasible basis B is defined as follows:

$$S(B) = \text{pos}(\{-e_j : x_j \in X_B\} \cup \{a_k : y_k \in Y_B\}).$$

In other words, $S(B)$ is the set of all nonnegative linear combinations of the indicated vectors $-e_j$ and a_k . We claim that the $n - m$ generators $-e_j$ and a_k of $S(B)$ are linearly independent, whence $S(B)$ is a simple $(n - m)$ -cone in the sense of Section 1.

In proving linear independence, we assume without loss of generality that

$$X_B = \{x_1, \dots, x_r\} \quad \text{and} \quad Y_B = \{y_1, \dots, y_{n-m-r}\}.$$

Then the vectors $-e_1, \dots, -e_r, a_1, \dots, a_{n-m-r}$ can be arranged as follows in an $(n - m) \times (n - m)$ matrix:

$$\begin{array}{l} a_1 \rightarrow \\ \vdots \\ a_{n-m-r} \rightarrow \\ -e_1 \rightarrow \\ \vdots \\ -e_r \rightarrow \end{array} \left[\begin{array}{ccc|ccc} x_1 & \cdots & x_r & x_{r+1} & \cdots & x_{n-m} \\ * & \cdots & * & & & D \\ \vdots & & \vdots & & & \\ * & \cdots & * & & & \\ \hline -1 & & & & & \\ & \ddots & & & & 0 \\ & & 0 & & & \\ & & & -1 & & \end{array} \right].$$

To establish the desired independence, it suffices to note that the submatrix D (whose columns correspond to the basic variables x_{r+1}, \dots, x_{n-m}) has full rank, and this in turn follows from the fact that the $m \times m$ submatrix of $[A | I_m]$ corresponding to the basic columns for B has the structure

$$\left[\begin{array}{ccc|ccc} x_{r+1} & \cdots & x_{n-m} & y_{n-m-r+1} & \cdots & y_m \\ & & & & & \\ \hline & & D & & & 0 \\ * & \cdots & * & 1 & & 0 \\ \vdots & & \vdots & & \ddots & \\ * & \cdots & * & & 0 & \ddots & 1 \end{array} \right]$$

and is invertible.

From the system \mathcal{B} of feasible bases for the system (3), we obtain a corresponding system of simple $(n - m)$ -cones in \mathbb{R}^{n-m} . To compare this system directly with

the system $(\mathcal{C}, \mathcal{R})_d$ of Section 1, we pass to the inequality version of (3), which is

$$Ax \leq b, \quad x \geq 0 \quad (\text{with } x \in \mathbb{R}^{n-m}). \quad (4)$$

In connection with (4), the y -variables are interpreted as slack variables, and this leads to the usual bijective correspondence between the sets of vertices of the feasible regions of (3) and (4).

Let P denote the feasible region of (4), and let the vectors

$$-e_1, \dots, -e_r, a_1, \dots, a_{n-m-r} \in \mathbb{R}^{n-m}$$

be associated as before with a feasible basis for (3). Then the associated vertex of P is the intersection of hyperplanes

$$v = \bigcap_{j=1}^r H_{-e_j} \cap \bigcap_{k=1}^{n-m-r} H_{a_k},$$

where

$$H_{-e_j} = \{x \in \mathbb{R}^{n-m} : x_j = 0\} \quad \text{and} \quad H_{a_k} = \{x \in \mathbb{R}^{n-m} : \langle a_k, x \rangle = 0\}.$$

Thus the cone $S(B)$ is spanned by the normals of the $n-m$ hyperplanes in \mathbb{R}^{n-m} whose intersection is v . Setting $d := n-m$, this is exactly the situation described in Section 1, where C_v is the union of all cones of the form $S(B)$ for which B is a feasible basis associated with the vertex v . This shows that the system $(\mathcal{C}, \mathcal{R})_d$ of Section 1 corresponds precisely to the set of all $S(B)$ where B belongs to the collection \mathcal{B} of all feasible bases of (4) (or (3)). Each extreme ray of the cone $S(B)$ is associated with a unique nonbasic variable in B and is a member of the set Z of Theorem 1.2.

For expository reasons, P was assumed in Sections 1 and 2 (except for the last paragraph of Section 2) to be of full dimension d . For purposes of the present discussion, d is equal to $n-m$. Note, however, that the system $(\mathcal{C}, \mathcal{R})_d$ of simple d -cones is obtained (as described in the preceding paragraph) whenever P is described as in (3), even if P is of dimension less than d . If the feasible region P is nonempty and is described initially in the standard form

$$A'x = b', \quad x \geq 0,$$

where A' is an $m \times n$ matrix, then whenever A' is of rank m the usual conversion to canonical form yields a system such as (3). This leads to a system of the form (4) and hence to the desired system of cones. (As was explained at the end of Section 2, pointedness of the cones C_v is not essential.)

Now let $S_i := S(B_i)$ for some basis $B_i \in \mathcal{B}$, and let Z_i consist of the vectors that span the extreme rays of S_i . (Thus each member of Z_i is a $-e_j$ or an a_k , and corresponds to a nonbasic variable for B_i .)

Recalling the pivot procedure that is customary for edge-following primal LP algorithms, we see that the usual sort of feasible pivot step for (3) or (4) amounts to a transition from some S_i to some S_{i+1} , where this is accomplished by removing an element of Z_i (the corresponding variable enters the basis) and replacing it by another element (corresponding to the variable leaving the basis) to obtain Z_{i+1} . The sequence of pivots chosen by the Gass-Saaty method can be interpreted in terms of a line intersecting the sequence of the S_i 's.

In the version of the parametric cost LP considered by Gass and Saaty [6], [7], two linear functionals c and c' on \mathbb{R}^d are given, along with a basic feasible solution that maximizes c on the feasible region P defined by (4). The aim is to find basic feasible solutions that maximize the various functionals of the form

$$c + \lambda c' \quad \text{with } \lambda \geq 0, \tag{5}$$

and this is equivalent (setting $\mu = \lambda / (1 + \lambda)$) to maximizing the convex combinations

$$f_\mu = (1 - \mu)c + \mu c' \quad (0 \leq \mu < 1). \tag{6}$$

The linear functionals in (6) all belong to the segment $[c, c']$ in the space of linear functionals on \mathbb{R}^d , and of course this segment is contained in the line L determined by c and c' . (With respect to the usual inner product, points of \mathbb{R}^d are regarded as linear functionals on \mathbb{R}^d , and vice-versa.)

The following nondegeneracy condition corresponds to condition SH3' of Section 1:

Each set consisting of c, c' , and $d - 2$ linearly independent columns of the matrix $[A^T | I_{n-m}]$ is linearly independent.

This condition restricts only the position of the line L with respect to the feasible region P ; it does not in any way restrict P itself. Once the condition has been checked, it will hold throughout the computation. There is no need to recheck it for the updated tableaus that result from changes in the set of basic variables.

The Gass-Saaty algorithm starts from a basic feasible solution s_0 (hence from a vertex of P) that maximizes the objective function $c_0 := c$. The bfs s_0 corresponds to a simple d -cone—call it S_0 —that includes the point c_0 of L , and (as was noted in Section 2) the intersection of the line L with the interior of S_0 is an open segment $]q_0, q_1[$, a ray $L_{<}(q_1)$, or a ray $L_{>}(q_0)$. The algorithm then successively pivots from S_i to S_{i+1} by finding a point q_{i+1} which belongs to the boundary of S_i and also to the ray $L_{>}(q_i)$ in L that issues from q_i and passes through the other objective function c' . The cone S_{i+1} is associated with another bfs that is obtained by replacing just one member of Z_i by an element of Z_{i+1} , and this corresponds to the geometric fact that S_{i+1} intersects S_i in a $(d - 1)$ -cone T_{i+1} that contains q_{i+1} . The algorithm stops at the l th stage, where l is such that one of the following holds:

- (i) c' is contained in $[q_l, q_{l+1}]$ and the bfs s_l maximizes the objective function c' .

(ii) No S_i exists, which means that the linear functional c' is not bounded above on P .

(iii) No q_{i+1} exists, which means that s_i maximizes c' and $L_{>}(q_i) \subset S_i$.

The cases (ii) and (iii) correspond respectively to the cases (ω) and (ω') of Section 2. The sequence

$$S_0, T_1, S_1, \dots, S_{i-1}, T_i \quad \text{or} \quad S_0, T_1, S_1, \dots, T_i, S_i$$

is a $(\mathcal{C}, \mathcal{R}, L, S_0)$ -chain in the sense of Section 2. The algorithm usually does not produce what we have called a full chain, because it is concerned only with the segment $[f, g] \subset L$ and not with the entire line L .

We showed in Section 2 that there is no repetition in the sequence of cones S_0, S_1, \dots , and thus it follows (under the assumption SH3') that the Gass-Saaty algorithm does not cycle. And it follows from the remarks preceding Theorem 2.1 that the vertex associated with the cone S_i maximizes all the linear functionals in the segment $[q_i, q_{i+1}]$.

In computational practice [6], [7], the tableau defining the feasible region P is augmented by rows that express $-c$ and $-c'$ in terms of the nonbasic variables. And rather than moving along the segment in L given by (6), it is customary to parametrize the procedure along the ray given by (5). This is of course equivalent to the procedure described above.

Let us denote by β_k and α_k the entries in the k th column of the rows that represent the respective functionals $-c$ and $-c'$. At the pivot step from S_i to S_{i+1} , the value

$$\begin{aligned} \lambda_i &= -\frac{\beta_r}{\alpha_r} \\ &= \min \left\{ -\frac{\beta_k}{\alpha_k} : \alpha_k < 0 \text{ and column } k \text{ is nonbasic} \right\} \end{aligned}$$

is computed, the r th column of the tableau is chosen as the pivot column, and the pivot element in the column is chosen as usual [6], [7]. The procedure terminates when a c' -maximum is attained (or it is realized that no finite c' -maximum exists), and this is checked in the usual way.

At each pivot step, the value λ_i determines the critical functional $q_i = c + \lambda_i c'$. The vertex determined by S_i maximizes all $q = c + \lambda c'$ for which $\lambda_i \leq \lambda \leq \lambda_{i+1}$. This sequence of q_i 's corresponds to the one described in Section 2.

For the Gass-Saaty procedure, the amount of computation per pivot is comparable to that for the ordinary simplex algorithm, because the only extra work consists of updating the additional row. Computational experience (with a series of randomly generated problems) seems to indicate that the number of pivots for the Gass-Saaty rule is slightly higher than for the usual gradient rule when both start with a basic feasible solution that maximizes c and the sole aim is to find a bfs that maximizes c' . However, the extra information supplied by the Gass-Saaty algorithm is useful in the circumstances for which the algorithm was developed.

4. An Example of Cycling

The Gass-Saaty algorithm deals with “parametric cost” LP problems, and these are dual to “parametric right-hand-side” LP problems. Murty [11] describes an example of Gana [5] that produces a cycle of eight pivots for the “parametric right-hand-side simplex algorithm,” and remarks that cycling can also occur in the “parametric cost simplex algorithm” (i.e., the Gass-Saaty algorithm). We shall dualize Gana’s example, and demonstrate that the reason for cycling is precisely the violation of our assumption SH3. This illustrates the conclusion of our general analysis—that cycling cannot arise from any special algebraic relationship that involves the constraints alone, but only from “degeneracy” in the relationship of the two objective functions to each other or to the constraints.

The example in [11] can be stated as follows:

Minimize $0x$ subject to

$$x_1 + 2x_2 \geq \lambda - 1,$$

$$x_2 + 2x_3 \geq \lambda - 1,$$

$$2x_1 + x_3 \geq \lambda - 1,$$

$$\text{where } x_1, x_2, x_3, \lambda \geq 0.$$

The dual problem is:

Maximize $-(y_1 + y_2 + y_3) + \lambda(y_1 + y_2 + y_3)$ subject to

$$y_1 + 2y_3 \leq 0,$$

$$2y_1 + y_2 \leq 0,$$

$$2y_2 + y_3 \leq 0,$$

$$\text{where } y_1, y_2, y_3, \lambda \geq 0.$$

After the introduction of slack variables, the dual problem assumes the form (3), with $c = (-1, -1, -1)$ and $c' = (1, 1, 1) = -c$. Thus condition SH3 is violated, even independently of the constraints. The Gass-Saaty algorithm produces a cycle consisting of eight pivots.

References

- [1] I. Adler, R. Karp, and R. Shamir, A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivots, *J. Complexity* 3 (1987), 372-387.
- [2] K.-H. Borgwardt, Untersuchungen zur Asymptotik der mittleren Schrittzahl von Simplexverfahren in der linearen Optimierung, *Oper. Res. Verf.* 28 (1978), 333-345.
- [3] K.-H. Borgwardt, Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method, *Math. Oper. Res.* 7 (1982), 441-462.
- [4] K.H. Borgwardt, The average number of pivot steps required by the simplex-method is polynomial, *Z. Oper. Res. Ser. A-B* 26 (1983), A157-A177.

- [5] A. Gana, Studies in the complementarity problem, Ph.D. Dissertation, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor (1982).
- [6] S. Gass, *Linear Programming*, McGraw-Hill, New York, 1975.
- [7] S. Gass and T. Saaty, The computational algorithm for the parametric objective function, *Naval Res. Logist. Quart.* **2** (1955), 39-45.
- [8] B. Grünbaum, *Convex Polytopes*, Wiley-Interscience, London, 1967.
- [9] M. Haimovich, The simplex algorithm is very good! On the expected number of pivot steps and related properties of random linear programs, Unpublished manuscript, Columbia University, New York (1983).
- [10] K. Murty, Computational complexity of parametric linear programming, *Math. Programming* **19** (1980), 213-219.
- [11] K. Murty, *Linear Programming*, Wiley, New York, 1983.
- [12] A. Schrijver, *Theory of Linear and Integer Programming*, Wiley-Interscience, New York, 1986.

Received December 4, 1986, and in revised form October 19, 1987.