Reducing Concave Programs with Some Linear Constraints

Arthur M. Geoffrion

SIAM Journal on Applied Mathematics, Vol. 15, No. 3 (May, 1967), 653-664.

Stable URL:
http://links.jstor.org/sici?sici=0036-1399%28196705%2915%3A3%3C653%3ARCPWSL%3E2.0.CO%3B2-U

SIAM Journal on Applied Mathematics is currently published by Society for Industrial and Applied Mathematics.

SIAM J. APPL. MATH.
Vol. 15, No. 3, May, 1967
Printed in U.S.A.

REDUCING CONCAVE PROGRAMS WITH SOME LINEAR CONSTRAINTS*

ARTHUR M. GEOFFRION†

Abstract. The problem of maximizing a concave function over a general convex set subject to linear inequality constraints is reduced to a finite sequence of subproblems involving linear *equality* constraints. This reduction can be expected to be computationally useful when there are but a few constraints, or when at most a few constraints are binding at the optimal solution of the original problem, or when prior (though possibly fallible) information is available concerning which constraints are likely to be binding. For quadratic programs the procedure specializes to an improved version of the Theil-van de Panne method. Computational considerations and experience are discussed, and a graphical example is given. The theory and viewpoint developed herein provide the foundation for related reduction procedures that may prove computationally useful even for large problems in the absence of a priori information.

1. Introduction. Consider the problem:

$$(\mathrm{P}) \quad \text{Maximize}_{x \in X} f(x) \quad \text{subject to} \quad a_i x + b_i \geqq 0, \qquad i \in M,$$

where f is strictly concave on the convex set $X \subseteq E^n$ and M is a finite set of linear constraint indices.[1] Assume that (P) admits an optimal solution x^*. It will be shown that (P) can be reduced to a finite sequence of subproblems of the form:

$$(\mathrm{P}_S) \quad \text{Maximize}_{x \in X} f(x) \quad \text{subject to} \quad a_i x + b_i = 0, \qquad i \in S,$$

where $S \subseteq M$. When a subproblem (P_S) arises that is infeasible, one considers instead the following subproblem:

$$(\mathrm{L}_S^j) \quad \text{Maximize}_{x \in X} a_j x + b_j \quad \text{subject to} \quad a_i x + b_i = 0, \qquad i \in S - j,$$

where ^{j} is a certain index in S, a feasible x is available, and the (linear) maximand is bounded above by 0. We assume the attainment of the constrained suprema of the subproblems that actually arise in the course of the reduction. To execute the reduction numerically, one requires means for achieving these suprema and finding the associated generalized Lagrange multipliers, which are known to exist (see §3).

It is easy to show by means of the Kuhn-Tucker-Karlin saddle point

* Received by the editors August 12, 1965, and in final revised form November 18, 1966.

† Western Management Science Institute, University of California, Los Angeles, California 90024. This work was supported partially by the Office of Naval Research under Task NR 047-041, Contract Nonr 233(75), and by the Western Management Science Institute under a grant from the Ford Foundation.

[1] Any nonlinear constraints must be incorporated into the definition of X.

characterization of (P) and (P_S) (see §3) that there exists a collection $\mathcal{O}$ of *optimal* subsets of M with the property that $x^S = x^*$ and $u^S \geqq 0$, $S \in \mathcal{O}$. Here x^S is the optimal solution of (P_S) and u^S is an associated vector of generalized Lagrange multipliers. For example, the set

$$B = \{i \in M: a_i x^* + b_i = 0\}$$

of binding constraints at x^* is in $\mathcal{O}$, and every set in $\mathcal{O}$ is a subset of B. We shall give rules for generating a sequence $\langle S^0, S^1, \cdots\rangle$, based on the solutions of the corresponding subproblems, so that this sequence reaches $\mathcal{O}$ in a finite number of steps, thereby leading to the optimal solution of (P). S^0 can be any subset of M such that (P_{S^0}) is feasible, and each trial set will differ from one of its predecessors by exactly one index.

Reducing (P) to a finite sequence of subproblems can be advantageous, obviously, when the subproblems can be solved more readily than (P) and when not too many need be solved. The difference between (P) and (P_S) is that (P_S) involves only linear *equality* constraints extracted from M, and is therefore a "less constrained" problem than (P). When X is "simple"—as when it comprises all of E^n, or a linear manifold, or the nonnegative orthant—(P_S) may be amenable to analytical solution by calculus or by simple but effective search procedures [7] that are frustrated by ordinary linear inequality constraints. When X is not "simple", it is at least possible to eliminate some variables by substituting out the equations. And whether or not X is simple, if M is populous but B is sparse, then a sparse S^0 will lead to subproblems having many fewer constraints than (P).

As will be developed in the final subsection, the present version of the reduction procedure is ultra-conservative. The result is that although termination occurs within a finite number of subproblems, the number of subproblems depends crucially on the choice of the initial subproblem (P_{S^0})—approximately exponentially on the distance (according to the symmetric difference metric) from S^0 to $\mathcal{O}$, in fact. For computational effectiveness, therefore, S^0 must be chosen to differ from some set in $\mathcal{O}$ by at most a half-dozen indices or so. When (P) has only a handful of constraints (or more, if it is known that a relatively small proportion of them are binding at x^*), then $S^0 = \varnothing$ probably satisfies this condition. Otherwise, prior (although possibly fallible) information must be available regarding which of the constraints are likely to be in B. Fortunately such information is frequently available when (P) is familiar or amenable to insight on the part of the analyst, or when it has been solved for slightly different values of the coefficients as in sensitivity analysis applications.

The theory and viewpoint developed herein leads to a number of closely related but less conservative reduction procedures. There is reason to

believe that some of these will better tolerate a poor choice of the initial subproblem.

In the next section we present the procedure in detail. A proof of convergence is given in §3. Section 4 hosts a graphical example, and §5 applies the procedure to quadratic programming and demonstrates a relation to an algorithm of Theil and van de Panne. The concluding section includes a discussion of the assumptions and how they can be weakened, some computational considerations, and a report on preliminary computational experience.

2. The reduction procedure. Fig. 1 gives details for the reduction of (P) to a finite sequence of subproblems of the form (P_S) or ($L_S{}^j$).

A formal definition of (x^S, u^S) at Step 1a is:

$$f(x) + \sum_S u_i{}^S(a_i x + b_i)$$
$$\leqq f(x^S) + \sum_S u_i{}^S(a_i x^S + b_i) \leqq f(x^S) + \sum_S u_i(a_i x^S + b_i)$$

for all $x \in X$ and u_i, $i \in S$. A similar definition holds for $(\hat{x}^S, z^S)$ at Step 1b. Note that when X is polyhedral, $(L_S{}^j)$ is a linear program and z^S is then the optimal dual vector.

At any given time, $\mathcal{S}$ comprises the current *generation* of trial sets. The zeroth generation is comprised of S^0 alone, the first generation of all sets of the form $S = S^0 \pm t$ for some $t \in T_{S^0}$, and so on. The symbol $S \pm t$ denotes $S \cup t$ when $t \notin S$ and $S - t$ otherwise. Thus if S is a trial set in the kth generation ($k \geqq 1$), then $S = S' \pm t$ for some trial set S' in the $(k-1)$st generation and t in $T_{S'}$. The set S' is called the immediate *lineal predecessor* of S, and either $S \subset S'$ or $S \supset S'$. Clearly S^0 is a lineal predecessor of every trial set at every generation. See the example in §4.

A few remarks that may be helpful in achieving efficient computational implementation are in order.

Remark 1. When S^0 is chosen so that $S^0 \subseteq B$ (e.g., $S^0 = \varnothing$), then termination will be hastened by redefining T_S at Step 1a as $\{i \in M - S: a_i x^S + b_i < 0\}$ and by ignoring any trial sets that arise for which (P_S) is infeasible. Note that this completely eliminates the need for u^S and Step 1b. The justification for this change is evident from Lemma 1 below. Evidently $S^0 = \varnothing$ is a propitious choice in complete ignorance of B. Similarly, if S^0 can be chosen so that $S^0 \supseteq B$ (remember that (P_{S^0}) must be feasible), it is advantageous to redefine T_S at Step 1a as $\{i \in S: u_i{}^S < 0\}$. In this case, Steps 4 and 1b can be eliminated and Step 3 can return control directly to Step 1a.

Remark 2. Let X be a linear manifold represented by the linearly inde-

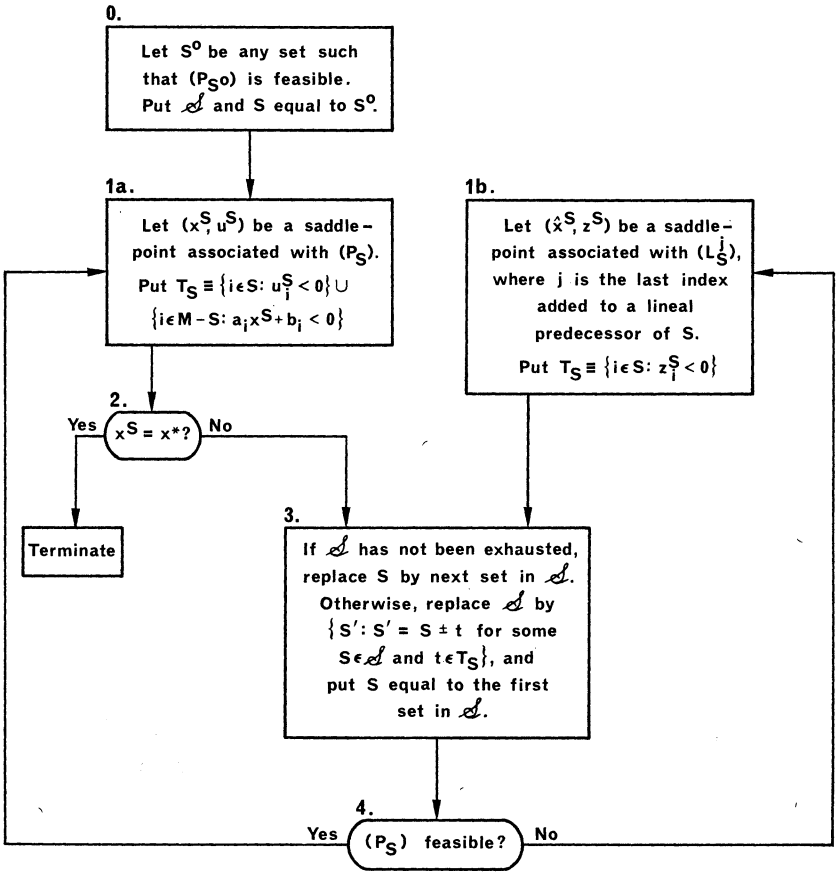

FIG. 1. *The reduction procedure*[2]

pendent constraints $a_i x + b_i = 0$, $i \in N$, where $N \cap M = \varnothing$ ($N = \varnothing$ is permissible). Then if we take S^0 to be "independent" in the sense that the collection $\{a_i : i \in N \cup S^0\}$ is linearly independent, it can be shown that the procedure has the following properties:

(a) at Step 1a, $N \cup S$ is always independent;

(b) at Step 1b, z^S is the unique solution to the equations

$$a_j + \sum_{N \cup S - j} z_i a_i = 0.$$

This result simplifies the implementation of Step 1.

Remark 3. At Step 2, $\{i \in M - S: a_i x^S + b_i < 0\} = \varnothing$ is necessary and $T_S = \varnothing$ is sufficient for $x^S = x^*$. Under various natural additional hy-

[2] See the text of §2 for clarification and some suggestions for efficient computational implementation.

potheses, $T_S = \varnothing$ is both necessary and sufficient. When f is differentiable in some neighborhood of x^S and X is determined by differentiable concave inequality ($\geqq 0$) constraints, $x^S = x^*$ if and only if there exist multipliers u^S that satisfy the *linear* (in u^S) first order differential conditions for (P) at x^S.

Remark 4. When a new generation S of trial sets is being defined at Step 3, one may, of course, eliminate any sets that appeared in any previous generation, although it may be inefficient in terms of storage and comparison time to attempt to avoid *all* duplications.

Remark 5. Step 4 can often be essentially by-passed. For example, Step 3 can return directly to Step 1a if the method used there for finding (x^S, u^S) automatically detects infeasibility; in this event, control can just be transferred to Step 1b. On the other hand, if the method employed at Step 1a cannot handle infeasibility effectively, then Step 3 must return to Step 1b. Unless, of course, S happens to be a subset of its immediate lineal predecessor S' and $(P_{S'})$ was feasible, in which case Step 3 can return to Step 1a. If (P_S) is indeed feasible, this will be signalled by the fact that $a_j x + b_j$ can be driven to 0 or above in $(L_S{}^j)$; and control can be transferred to Step 1a. When the method employed at Step 1a requires a feasible point to get started, the natural thing to do when $x^{S'}$ is not feasible in (P_S) is to obtain a feasible point (if one exists) via Step 1b and then go to Step 1a. In this regard it may be worth noting that a feasible solution to $(L_S{}^j)$ at Step 1b is always available from the solution to the subproblem associated with S'.

Remark 6. It is often possible to take advantage, when organizing the computations at Step 1, of the fact that each subproblem differs from one of its lineal predecessors by exactly one constraint. Solving a sequence of very similar problems is usually much easier than the same number of dissimilar problems.

3. Proof of convergence. The saddle point (x^S, u^S) required at Step 1a exists because it has been assumed that (P_S) has an optimal solution whenever it is feasible, thanks to the saddle point characterization for such programs. A similar assertion holds regarding $(L_S{}^j)$, which is always feasible when encountered at Step 1b. For completeness we quote without proof the version of the saddle point theorem for concave programs that best suits our needs (see [5, Theorem 7.1.1. ff.]).

THEOREM (Kuhn-Tucker-Karlin). *Let $F(x)$ be concave on the convex set X. Then x^0 is an optimal solution of the problem*:

Maximize$_{x \in X}$ $F(x)$ *subject to*

$$(1) \qquad \begin{aligned} a_i x + b_i &\geqq 0, \qquad i = 1, \cdots, m_1, \\ a_i x + b_i &= 0, \qquad i = m_1 + 1, \cdots, m, \end{aligned}$$

if and only if there exists an m-vector u^0 (*"generalized Lagrange multipliers"*) *such that*

$$F(x) + \sum_{i=1}^{m} u_i^0(a_i x + b_i) \leqq F(x^0) + \sum_{i=1}^{m} u_i^0(a_i x^0 + b_i)$$

$$\leqq F(x^0) + \sum_{i=1}^{m} u_i(a_i x^0 + b_i)$$

for all $x \in X$ *and* u *such that* $u_i \geqq 0$, $i = 1, \cdots, m_1$.

The pair (x^0, u^0) is a saddle point of the Lagrangian associated with (1).

Evidently the reduction procedure is well defined under the assumptions stated in the Introduction. That termination occurs in a finite number of steps is a consequence of the following theorem, which in turn follows from the two lemmas below. We formally define the collection $\mathcal{O}$ of optimal subsets of M as the subsets R satisfying $\{i \in M: u_i^* > 0\} \subseteq R \subseteq \{i \in M: a_i x^* + b_i = 0\}$ for some saddle point (x^*, u^*) associated with (P).

THEOREM. *Let* S *be any trial set that could arise in the course of executing the procedure of Fig.* 1. *If* $x^S \neq x^*$, *then* $S \pm t$ *is one unit of distance closer to* $\mathcal{O}$ *than* S *is for at least one element* t *of* T_S.

We employ the symmetric difference metric: e.g., $d(S, B) = \mu(S - B) + \mu(B - S)$ is the distance from S to B, where μ denotes the number of indices in a finite set. Let $d^0 \equiv \min\{d(S^0, R): R \in \mathcal{O}\}$ denote the distance from S^0 to $\mathcal{O}$ (i.e., to the nearest set in $\mathcal{O}$). The theorem implies that termination will occur in exactly d^0 generations of trials with the discovery of the optimal set nearest S^0. Evidently the phenomenon of *degeneracy*, which in the present context can be identified with the existence of more than one optimal set, can only accelerate termination.

LEMMA 1. *Let* (P_S) *be feasible,* (x^S, u^S) *an associated saddle point, and* R *an arbitrary set in* $\mathcal{O}$ (*e.g., the one nearest to* S). *If* $x^S \neq x^*$, *then either* $a_i x^S + b_i < 0$ *for some* $i \in R - S$ *or* $u_i^S < 0$ *for some* $i \in S - R$, *or possibly both.*

LEMMA 2. *Let* (P_S) *be feasible,* $a_j x^S + b_j < 0$, $(P_{S \cup j})$ *infeasible, and* R *an arbitrary set in* $\mathcal{O}$. *Then* $S - R \neq \varnothing$ *and* $z_i^S < 0$ *for some* i *therein, where the* z_i^S *are generalized Lagrange multipliers associated with the problem:*

Maximize$_{x \in X}$ $a_j x + b_j$ *subject to* $a_i x + b_i = 0$, $i \in S$. *Let* $z_{i_1}^S < 0$. *If* $(P_{S \cup j - i_1})$ *is not feasible, then* $\{S \cup j - i_1\} - R \neq \varnothing$ *and* $z_i^{S - i_1} < 0$ *for some* i *therein. Let* $z_{i_2}^{S - i_1} < 0$. *If* $(P_{S \cup j - i_1 - i_2})$ *is not feasible, then* $\{S \cup j - i_1 - i_2\} - R \neq \varnothing$ *and* $z_i^{S - i_1 - i_2} < 0$ *for some* i *therein. And so on.*

Proof of Lemma 1. Suppose to the contrary that (i) $a_i x^S + b_i \geqq 0$ for $i \in R - S$, and (ii) $u_i^S \geqq 0$ for $i \in S - R$. Then, defining

$u_i^S = 0$, $i \in R - S$, we see that (x^S, u^S) is also a saddle point associated with the problem:

Maximize$_{x \in X} f(x)$ subject to

$$(2) \qquad \begin{aligned} a_i x + b_i &= 0, \qquad i \in S \cap R, \\ a_i x + b_i &\geqq 0, \qquad i \in (S - R) \cup (R - S). \end{aligned}$$

Let (x^*, u^*) be a saddle point associated with (P) and R. It follows that (x^*, u') is a saddle point for (2), where u' is defined by extraction from u^* in the obvious way. Since both x^S and x^* are optimal in (2), and (2) must have a unique optimal solution in view of the strict concavity of f, we have the contradiction $x^S = x^*$.

Proof of Lemma 2. When (P_S) is feasible, $a_j x^S + b_j < 0$, and $(P_{S \cup j})$ infeasible, the problem:

$$\text{Maximize}_{x \in X}\, a_j x + b_j \quad \text{subject to} \quad a_i x + b_i = 0, \qquad i \in S,$$

is feasible (x^S satisfies the constraints) and its objective function is bounded above by 0. Let $(\hat{x}^S, z^S)$ be a saddle point associated with this problem. Now,

$$\begin{aligned} 0 \leqq a_j x^* + b_j &= -a_j(\hat{x}^S - x^*) + (a_j \hat{x}^S + b_j) \\ &< -a_j(\hat{x}^S - x^*) \\ &= -(a_j \hat{x}^S + b_j) + (a_j x^* + b_j) \\ &\leqq -[(a_j x^* + b_j) + \sum_S z_i^S (a_i x^* + b_i)] + (a_j x^* + b_j) \\ &= -\sum_S z_i^S (a_i x^* + b_i) = -\sum_{S-R} z_i^S (a_i x^* + b_i). \end{aligned}$$

We have used the left-hand inequality of the saddle point definition for $(\hat{x}^S, z^S)$ at x^*. Thus we have $0 < -\sum_{S-R} z_i^S (a_i x^* + b_i)$. We conclude $S - R \neq \varnothing$ and $z_i^S < 0$ for some i therein. This proves the first part of the lemma. The remaining parts are proved by repetition of this argument.

4. An example. The example of Fig. 2 is designed to illustrate the operation of the procedure of Fig. 1. For convenience, we present a graphical, rather than numerical, example in two dimensions ($n = 2$). We take $f(x)$ to be the Euclidean distance from x to a fixed point x_0, and $X = E^2$. Hence (P_S) yields as its solution x^S the orthogonal projection of x_0 onto the manifold $\{x: a_i x + b_i = 0, i \in S\}$, and the multipliers u_i^S are the representation of $-\nabla f(x^S)$ in terms of the gradients a_i, $i \in S$. Loci of the four constraints are drawn and labeled with constraint indices; their gradients are also drawn in at selected points. The feasible region is hatched and x^* is marked by a heavy dot. For convenience of discussion we call

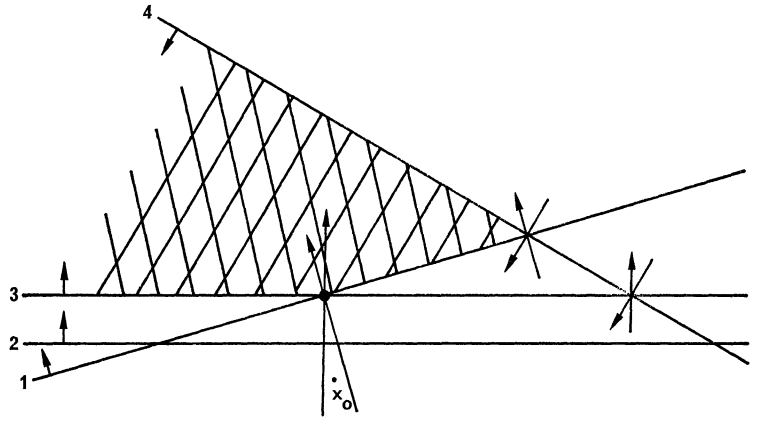

FIG. 2. *A two-dimensional graphical example of the reduction procedure*

$\{i \in S\colon z_i^S < 0\}$ or $\{i \in S\colon u_i^S < 0\}$ the set of *optimality alarms*, and $\{i \in M - S\colon a_i x^S + b_i < 0\}$ the set of *feasibility alarms*.

Let the initial trial set be arbitrarily chosen as, say, $S^0 = \{3, 4\}$. S^0 is easily seen to yield a feasibility alarm for $i = 1$ and optimality alarms for $i = 3$ and 4. Hence the first generation trial sets are $\{3, 4, 1\}$, $\{4\}$, and $\{3\}$. Now $\{4\}$ yields only one alarm, an optimality alarm for $i = 4$; $\{3\}$ also yields one alarm, a feasibility alarm for $i = 1$; and $\{3, 4, 1\}$ leads to an infeasible subproblem. Solving the dual constraints $z_3 a_3 + z_4 a_4 + a_1 = 0$ of $(L^1_{\{3,4,1\}})$ for their unique solution, one obtains optimality alarms for both $i = 3$ and 4. Hence the second generation trial sets are $\varnothing$; $\{3, 1\}$; $\{4, 1\}$ and $\{3, 1\}$. We find that $\varnothing$ yields feasibility alarms for $i = 1$, 2, and 3; that $\{4, 1\}$ yields optimality alarms for $i = 1$ and 4; and that $\{1, 3\}$ yields no alarms at all. Hence $\{1, 3\}$ is optimal and the computations terminate.

A diagrammatic summary of the trials is given in Fig. 3. Of course, $\mathcal{O} = \{1, 3\}$ by inspection of Fig. 2, so that an optimal set has indeed been found. Note that $d(\{3, 4\}, \{1, 3\}) = 2$, and that an optimal set was found in two generations.

5. Application to quadratic programming. The strictly concave quadratic programming problem can be written as follows:

$$\text{Maximize}_x \tfrac{1}{2}x^t C x + c^t x \quad \text{subject to}$$

$$\text{(QP)} \qquad a_i x + b_i \geqq 0, \qquad i = 1, \cdots, m_1,$$

$$a_i x + b_i = 0, \qquad i = m_1 + 1, \cdots, m,$$

where C is negative definite. A convenient choice for (P) is to put $X = \{x \in E^n\colon a_i x + b_i = 0, i = m_1 + 1, \cdots, m\}$, $f(x) = \frac{1}{2}x^t C x + c^t x$, and $M = \{1, \cdots, m_1\}$. If (QP) is feasible, then the assumptions stated in the

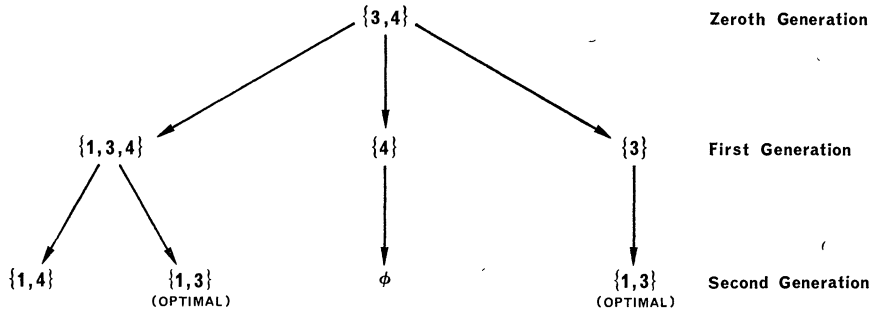

FIG. 3. *Summary of the sequence of trial sets for the graphical example*

Introduction are satisfied, for the constrained suprema of (P) and of the feasible (P_S) are achieved due to the well-known fact that a concave quadratic polynomial bounded above on a convex polyhedron achieves its constrained supremum. A saddle point (x^S, u^S) associated with any feasible (P_S) can be found by solving the equivalent linear system (the Lagrange multiplier equations)

$$Cx + c + \sum_S u_i a_i + \sum_{i=m_1+1}^{m} v_i a_i = 0,$$

$$a_i x + b_i = 0, \qquad i \in S \cup \{m_1 + 1, \cdots, m\};$$

and ($L_S{}^j$) is a linear program. If some of the inequalities are simple non-negativity constraints, these can either (a) be left in M, with resulting simplifications in the Lagrange multiplier equations, or (b) be included in X, at the expense of introducing a nonlinear "complementary slackness" requirement in the otherwise linear system that is now equivalent to finding a saddle point associated with (P_S).

Remarks 2, 3, 5 and 6 of §2 are particularly applicable here, as are partitioning and bordering methods for maintaining the relevant matrix inverses from subproblem to subproblem.

If $S^0 = \varnothing$ and Remark 1 of §2 is applied, the algorithm of Theil and van de Panne [6] is recovered and is seen to be valid even in the absence of their antidegeneracy assumption.[3] Our prerogative to begin with $S^0 \neq \varnothing$ can make the difference between the computational practicality and impracticality of this approach when prior information is available.

6. Discussion.

6.1. Discussion of the assumptions and extensions. We required f to be *strictly* concave in order to have the computational and theoretical advan-

[3] Boot [1] has rederived their algorithm using the Kuhn-Tucker conditions, but he also resorts to an antidegeneracy assumption.

tages of uniqueness in the optimal solutions of (P) and (P_S) and (2). It can be shown that a nonstrictly concave f can be accommodated, when it fails to provide the necessary uniqueness, by modifying Step 2 to determine whether *some* optimal solution of (P_S) also solves (P), and Step 1a to include $\{i \in S: u_i^S = 0\}$ in T_S.

Concerning the attainment of the constrained suprema, we note that if X is compact and f is continuous on its boundary (by concavity, f is automatically continuous on the interior of X), there is no question about attainment in (P), (P_S) or (L_S^j) when they are feasible. It can also be shown that these suprema are attained if X is closed and f achieves its supremum over X, for then, by strict concavity, attention can be restricted to a compact subset of X. Certain types of nonattained suprema can, however, be accommodated by an appropriate modification of Step 1a; for example, the situation in which (P_S) has an unbounded optimal value. In this case x^S would not exist, and it can be shown that one should put T_S equal to the indices of the constraints that are violated by any sequence $\langle x^r \rangle$ feasible in (P_S) for which $\langle f(x^r) \rangle \to \infty$. Discretion can be employed in choosing the sequence $\langle x^r \rangle$, of course, in an effort to make T_S small.

Regarding the availability of means for solving the subproblems, one must resort to the extensive literature on optimization and mathematical programming (see, for example, Graves and Wolfe [3], Hadley [4], and Wilde [7]). The differentiability conditions of Remark 3 in §2 are required by many of the optimization methods that might be used to solve (P_S). Such methods often produce u^S as an automatic by-product of the discovery of x^S; but when this is not the case, u^S can be found as the solution to the first order differential conditions for an optimum of (P_S), which comprise a *linear* system once x^S is known. A similar remark holds for (L_S^j).

6.2. Numerical stability. It is a computational fact of life that ordinarily one can obtain only arbitrarily close approximations to (x^S, u^S) or z^S in finite time. What are the implications for the present approach? Fortunately, the convergence of the procedure is quite unlikely to be disrupted by moderate random errors in x^S, u^S, or z^S. This is due to the extremely conservative and "locally exhaustive" nature of the scheme used to direct the sequence of trial sets. Note also that only the *signs* of u_i^S, $a_i x^S + b_i$, and z_i^S are used. It follows that there is a strictly positive tolerance on the accuracy required of x^S, u^S and z^S for every "computational" T_S to include the true T_S, and, therefore, for convergence to be assured (although it could be prolonged by a proliferation of unnecessary trial sets). The effect due to inaccuracies of x^S on Step 2 cannot be discussed until its mechanism is specified, although one would expect that Step 2 would employ a mechanism that could recognize a good approximate solution to (P) as such.

TABLE 1

Summary of computational results for three quadratic test problems (estimates are based on the assumption that no redundant trial sets are discarded)

d^0	Estimated Average Number of Subproblems to be Solved before Termination		
	Problem 1 (20 × 9)	Problem 2 (10 × 15)	Problem 3 (50 × 25)
1	2.3	2.1	2.0
2	4.8	5.6	4.0
3	19	20	15.3
4	87	75	79
5	452	301	448
6	2,824	1,442	3,718
7	15,300	9,500	43,100
⋮	⋮	⋮	⋮

6.3. Computational experience. Preliminary computational experience has been acquired with three quadratic test problems. Test problems 1 and 3 were adapted from problems of practical origin that were kindly made available to the author by Leola Cutler. They were 20 × 9 (20 variables and 9 constraints) and 50 × 25, respectively. Test problem 2, 10 × 15, was methodically generated from a random number table. Based on this experience, estimates of the average number of subproblems that must be solved before termination were derived for each problem for $d^0 \leqq 7$.[4] They are presented in Table 1. Each average is over all possible initial sets with a given value of d^0, and it should be noted that the estimates are based on the assumption that no test is made on a newly generated trial set to see whether it has been tried before and can therefore be discarded. To what extent this assumption inflates the estimates over the true averages when redundant trial sets are discarded is not clear, although obviously the number of subproblems could not exceed 2^m (e.g., 512 in the first problem) in this case. Even when redundant trial sets are excluded, however, it is doubtful that the average number of required subproblems would increase with d^0 at less than an exponential rate over the range of interest.

Computing time on the IBM 7094 was well below one second per subproblem for all three problems.

If this experience is any guide, and we suspect that it may be, the present version of the reduction procedure can be recommended only for applications in which a set of truly restrictive constraints can be identified a priori with fewer than about six errors.

The reason why a poor choice of the initial subproblem leads to such an

[4] Recall that d^0 is the distance from the initial set S^0 to the nearest optimal set.

inefficient reduction seems clear enough: Step 3 is extremely conservative in the way it determines the order of trial subproblems. The motive for a conservative strategy is due to the fact that the phrase "at least one" in the theorem of §3 cannot, in general, be strengthened to "all" or even "most". In the author's experience, however, typically $\frac{3}{4}$ or more of the elements in T_S are in the symmetric difference set $\{S - R\} \cup \{R - S\}$, where R is the nearest optimal set! This empirical observation suggests that a less conservative version of Step 3 might lead to an efficient reduction even for the poorest choices for S^0. One of the least conservative versions of Step 3 imaginable would be simply: $S \leftarrow S \pm t$ for some randomly chosen element t of T_S. Storage would be negligible, and although eventual termination can be assured only with probability 1, the expected number of subproblems to be solved before termination may well be quite reasonable even for poor choices of S^0. This strategy will be investigated more fully in another paper [2].

Additional evidence for the desirability, when a "good" choice of S^0 is improbable, of less conservative versions of Step 3 stems from the following observation. The primal and dual simplex methods for linear programming can be very naturally described as specializations of the procedure of Fig. 1 (modified as indicated in §6.1 so as to apply to linear programs, of course) that utilize certain very simple versions of Step 3. This observation also suggests that algorithms for concave programming based on the reduction procedure of this paper can be viewed as generalizations of the simplex method.

Acknowledgment. The author is indebted to Joseph Naruishi for his computational assistance, which was rendered under a grant from the Division of Research of the Graduate School of Business Administration, University of California at Los Angeles, and to the Western Data Processing Center for making its facilities available.

REFERENCES

[1] J. C. G. Boot, *Quadratic Programming*, North Holland, Amsterdam, 1965.

[2] A. Geoffrion, *A Markovian procedure for strictly concave programming with some linear constraints*, Proceedings of the Fourth International Conference on Operational Research, Cambridge, Massachusetts, 1966, to appear.

[3] R. L. Graves and P. Wolfe, *Recent Advances in Mathematical Programming*, McGraw-Hill, New York, 1963.

[4] G. Hadley, *Nonlinear and Dynamic Programming*, Addison-Wesley, Reading, Massachusetts, 1964.

[5] S. Karlin, *Mathematical Methods and Theory in Games, Programming, and Economics*, vol. I, Addison-Wesley, Reading, Massachusetts, 1959.

[6] H. Theil and C. van de Panne, *Quadratic programming as an extension of classical quadratic maximization*, Management Sci., 7 (1960), pp. 1–20.

[7] D. J. Wilde, *Optimum Seeking Methods*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.