SECANT APPROXIMATION METHODS FOR CONVEX OPTIMIZATION

by

C. Y. Kao and R. R. Meyer

Computer Sciences Technical Report #352

April 1979

SECANT APPROXIMATION METHODS FOR CONVEX OPTIMIZATION[1)]

C. Y. Kao and R. R. Meyer

Abstract

The methods discussed are based on local piecewise-linear secant approximations to continuous convex objective functions. Such approximations are easily constructed and require only function evaluations rather than derivatives. Several related iterative procedures are considered for the minimization of separable objectives over bounded closed convex sets. Computationally, the piecewise-linear approximation of the objective is helpful in the case that the original problem has only linear constraints, since the subproblems in this case will be linear programs. At each iteration, upper and lower bounds on the optimal value are derived from the piecewise-linear approximations. Convergence to the optimal value of the given problem is established under mild hypotheses. The method has been successfully tested on a variety of problems, including a water supply problem with more than 900 variables and 600 constraints.

---

1) Research supported by National Science Foundation Grant MCS74-20584 A02.

1. Introduction

Several classes of "secant" optimization methods will be considered for convex nonlinear programs of the form

$$\begin{aligned} &\min_{x} \quad f(x) \\ &\text{s.t.} \quad x \in C \cap [\ell,u], \end{aligned} \tag{1.1}$$

where $x = (x_1,\ldots,x_n)^T \in \mathbb{R}^n$, C is a closed convex set, $[\ell,u]$ denotes the hyper-rectangle corresponding to the constraints $\ell \leq x \leq u$, and f is a continuous convex function on $[\ell,u]$. (To avoid trivial cases we assume $\ell < u$.) The term "secant" is used because the basic idea of these methods is to approximate f at each iteration by piecewise-linear functions that, from a geometric viewpoint, are determined by secants in the graph of f. It should be emphasized that because of the nature of these approximations, these methods require only function values of f, and not first or second derivatives. For notational convenience we let $S \equiv C \cap [\ell,u]$, and to avoid trivial cases we assume that S is non-empty. Note that the assumptions made with respect to (1.1) imply that it has an optimal solution. In the case that (1.1) has only linear constraints (i.e., C is polyhedral), the piecewise-linear approximation of f is computationally useful because the corresponding approximating problem is easily reduced to a linear program.

In Sections 2 and 3 we will consider the case in which f is separable, i.e., $f(x) = \sum_{i=1}^{n} f_i(x_i)$, and relate our approach to other techniques for separable programming and to other local approximation methods. Section 4 indicates how the techniques may be extended to differentiable non-separable f. Computational experience in the linearly constrained case and directions for further research are described in the concluding sections.

2. Global Approximation Methods for Separable Programming

There are many optimization problems that may be stated in the format (1.1) with f separable, examples being data fitting [Bachem and Korte (1977)], electrical networks [Rockafellar (1976)], and water supply applications [Collins, et al (1978)]. Many mathematical programs under uncertainty (stochastic programs) may also be converted to convex separable programs ([Dantzig (1963)], [Symonds (1967)], [Wets (1966)]). [Hadley (1964)] and [Wagner (1967)] have discussed the technique of using transformations to convert a problem which is not separable in its original form to a separable program. Some problems arising in personnel assignment and logistics ([Gross (1956)] and [Saaty (1970)]) turn out to be separable.

Although general nonlinear programming algorithms ([Avriel (1976)]) may be applied to the separable convex case, these algorithms do not take maximum advantage of separability. We will first discuss the traditional algorithms for separable programming and then consider in Section 3 some new iterative separable programming algorithms that have several advantages over existing techniques.

Piecewise-linear approximation for separable programs was apparently first proposed in [Charnes and Lemke (1954)]. The idea of this approach is simply to use piecewise-linear functions $\tilde{f}_i$ to approximate the f_i over the intervals $[\ell_i, u_i]$. The problem $\min_x \sum_{i=1}^{n} \tilde{f}_i(x_i)$, s.t. $x \in S$ can then be converted to an equivalent linear program (LP) if S is polyhedral. (Although the algorithms to be described below are theoretically valid for the case of non-polyhedral S, the corresponding nonlinear subproblems would generally not be any easier to solve than the original problems in such cases, so the algorithms are practical only for the case of linear constraints. Although piecewise-linear approximation can also be applied to nonlinear constraints as in [Thakur (1978)] to yield linear subproblems, the convergence properties of such an approach have so far only been considered under restrictive assumptions in the separable case.)

Two significant questions associated with piecewise-linear approximation are:

(1) How many "grid points" are to be used to determine the $\tilde{f}_i$?

(2) How "good" is the resulting approximation?

In essence, with a global approximation approach we are faced with the trade-off between the accuracy of the approximation and the size of the problem solved. [Thakur (1978)] gives bounds on the maximum deviation between the piecewise-linear approximation and a Lipschitz continuous function, and uses them to establish a bound on the optimal objective value of the original problem. [Geoffrion (1977)] and [Meyer (1977), 1979)] also give results relating approximation error to optimal value error in more general cases.

If the error associated with a given approximation is too large, the traditional approach is to use a finer grid for the approximation of the f_i. This procedure greatly increases the size of the constraint matrix of the equivalent LP, and problems of storage and efficiency may occur. Similar problems may arise when the "interpolatory" method of [Beale (1968)] is applied to the grids at each iteration. (The interpolatory approach also leads to difficulties in obtaining good error bounds, an issue to be considered in Section 3.) For these reasons we will consider in the next section an iterative separable programming method based on local approximations in which the number of grid points for each variable is never more than three.

3. Local Approximation Methods for Separable Programming

Rather than using global piecewise-linear approximations to the objective functions, it turns out to be sufficient to utilize a sequence of appropriately chosen "local" piecewise-linear approximations over appropriately chosen "neighborhoods". In order to make these notions precise, we will introduce notation that permits compact descriptions of the type of approximating problems that we will consider.

Specifically, we will employ approximations that correspond to ordered triples $(\hat{\ell},\hat{m},\hat{u})$ that satisfy

$$(3.1) \qquad \ell \leq \hat{\ell} \leq \hat{m} \leq \hat{u} \leq u,$$

$$(3.2) \qquad \hat{\ell}_i < \hat{m}_i \quad \text{if} \quad \ell_i < \hat{m}_i, \quad \text{and}$$

$$(3.3) \qquad \hat{m}_i < \hat{u}_i \quad \text{if} \quad \hat{m}_i < u_i.$$

Such a triple will be said to be <u>admissible</u>, and the corresponding approximating problem $P(\hat{\ell},\hat{m},\hat{u})$ is defined as

$$\min_x \quad \sum \hat{f}_i(x_i)$$

$$\text{s.t.} \quad x \in S \cap [\hat{\ell},\hat{u}],$$

where

$$\hat{f}_i(x_i) = \begin{cases} f_i(\hat{m}_i) + \left[\dfrac{f_i(\hat{m}_i)-f_i(\hat{\ell}_i)}{\hat{m}_i - \hat{\ell}_i}\right](x_i-\hat{m}_i) & \text{if} \quad \hat{\ell}_i \leq x_i < \hat{m}_i \\ f_i(\hat{m}_i) & \text{if} \quad x_i = \hat{m}_i \\ f_i(\hat{m}_i) + \left[\dfrac{f_i(\hat{u}_i)-f_i(\hat{m}_i)}{\hat{u}_i - \hat{m}_i}\right](x_i-\hat{m}_i) & \text{if} \quad \hat{m}_i < x_i \leq \hat{u}_i \end{cases}$$

for $x_i \in [\hat{\ell}_i,\hat{u}_i]$.

Note that except for the degenerate cases in which $\ell_i = \hat{\ell}_i = \hat{m}_i$ or $u_i = \hat{u}_i = \hat{m}_i$, $\hat{f}_i$ is the two-segment, piecewise-linear, convex approximation to f_i determined by the values of f_i at the three points $\hat{\ell}_i$, $\hat{m}_i$, and $\hat{u}_i$ (see Figure 1). In the degenerate cases (see Figure 2), $\hat{f}_i$ will be an affine function determined by two values of f_i. (From the definition of an admissible triple and the fact that $\ell_i < u_i$, it is impossible to have $\hat{\ell}_i = \hat{m}_i = \hat{u}_i$.) For x_i not in the interval $[\hat{\ell}_i, \hat{u}_i]$, $\hat{f}_i(x_i)$ is defined by extending the segments to the left and to the right in the non-degenerate case, and extending the affine function in the degenerate case. Observe that convexity implies that $f_i(x_i) \leq \hat{f}_i(x_i)$ for any x_i that is part of a feasible solution of $P(\hat{\ell},\hat{m},\hat{u})$ and $f_i(x_i) \geq \hat{f}_i(x_i)$ for $x_i \notin [\hat{\ell}_i, \hat{u}_i]$. Letting $\hat{f}(x) \equiv \sum_{i=1}^{n} \hat{f}_i(x_i)$ and $\hat{S}$ be the feasible set of $P(\hat{\ell},\hat{m},\hat{u})$ we thus have $f(x) \leq \hat{f}(x)$ for all $x \in \hat{S}$. Since $\hat{f}(\hat{m}) = f(\hat{m})$, this property implies that if $\tilde{x} \in \hat{S}$ and $\hat{f}(\tilde{x}) < \hat{f}(\hat{m})$, then $f(\tilde{x}) < f(\hat{m})$. The latter inequality guarantees strict monotonicity of the iterates, since in the algorithms to be developed $\hat{m}$ will play the role of the most recently generated feasible solution of (1.1). From this viewpoint $\hat{f}$ is a "local" piecewise-linear approximation in the neighborhood $[\hat{\ell},\hat{u}]$ of the most recent iterate, and $\hat{S}$ is non-empty since it contains $\hat{m}$. Before developing additional properties of these approximations, we will contrast them with related local approximation methods.

A heuristic two-segment approach was proposed in the early separable programming paper, [Dantzig, et al (1958)]. This method reduces the size of the interval for each variable by a factor of at least ½ at each iteration, but need not converge to an optimal solution except in the case that S is the product of intervals. (See [Meyer (1980)]

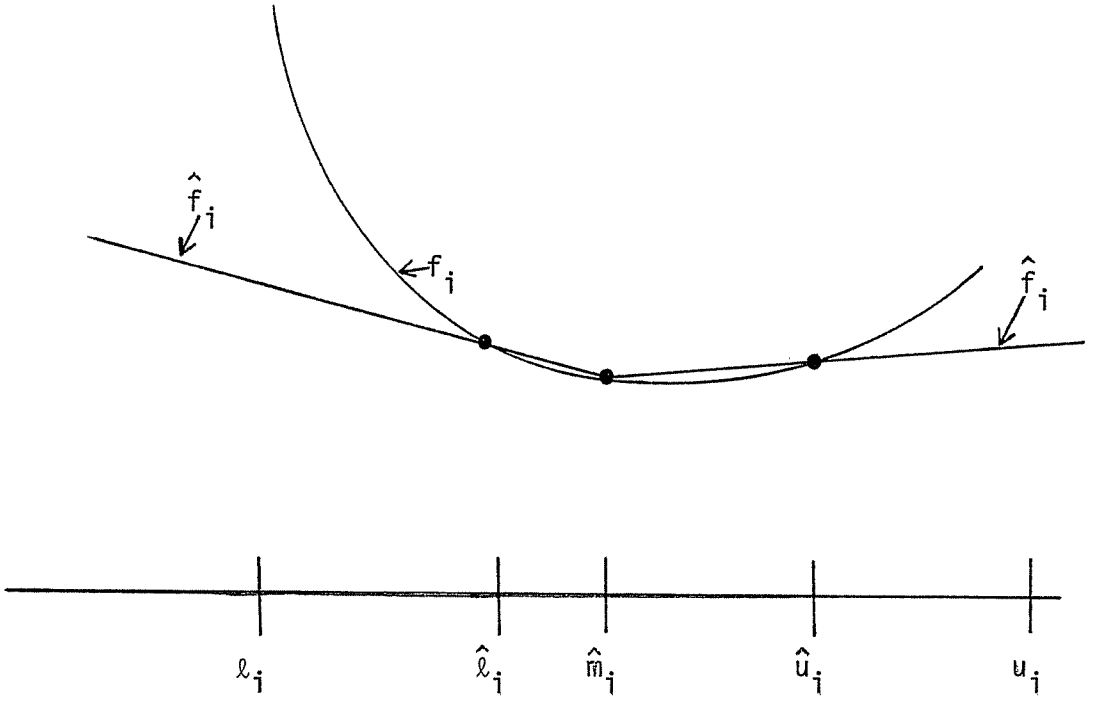

Figure 1. f_i and $\hat{f}_i$ in the non-degenerate case

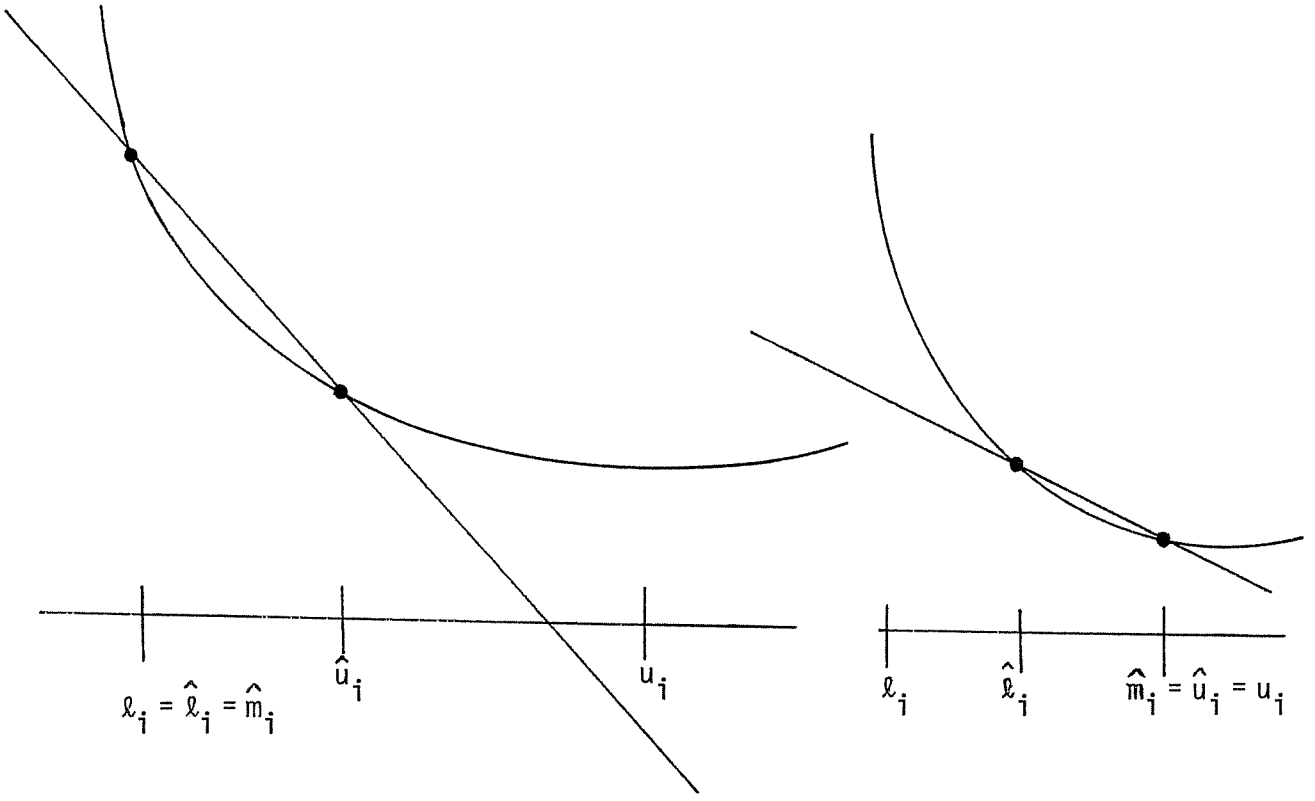

Figure 2. The two degenerate cases

for details of the method and an example in which this heuristic does not converge to the optimal value.) In essence, the convergence problem associated with this method is a consequence of the fact that the intervals are forced to shrink too quickly.

[Dantzig (1963)] describes a local approximation algorithm related to "generalized programming". This technique uses line searches in descent directions generated via the local approximations, but no convergence proof is given, and the efficacy of this algorithm appears to be unknown.

The two-segment methods developed in [Meyer (1977)], [Meyer and Smith (1978)], and [Meyer (1979), (1980)] were shown to be finitely convergent to optimal solutions of certain fine-grid approximating problems, and the algorithms to be developed in this section may be thought of as extensions that ensure convergence to the optimal value of the given problem (1.1).

Related local approximation algorithms include the boxstep method [Marsten, Hogan, and Blankenship (1975)], the MAP method [Griffith and Stewart (1961)], and the method of hypercubes [Fletcher (1972)]. The boxstep method also employs additional bounds on the variables, but keeps the corresponding "box" fixed in size, and optimizes the original function (rather than an approximation) over the resulting restriction of the feasible set. The MAP method uses variable "box" sizes that are selected in an adaptive manner, but uses first-order linear approximations to the objective and the constraints. The method of hypercubes employs quadratic, quasi-Newton approximations of the objective in a hypercube centered about the current iterate.

In the algorithms and convergence proofs to follow, error bounds play a crucial role. Since $\hat{f}_i$ is a piecewise-linear approximation determined by $(\hat{\ell}_i,\hat{m}_i,\hat{u}_i)$, note that the error bound defined by $\hat{E}_i \equiv \max_{\hat{\ell}_i \le x_i \le \hat{u}_i} [\hat{f}_i(x_i)-f_i(x_i)]$ satisfies $\hat{E}_i \le \max_{[\hat{\ell}_i,\hat{u}_i]} \hat{f}_i(x_i) - \min_{[\hat{\ell}_i,\hat{u}_i]} f_i(x_i)$ $\le \max_{[\hat{\ell}_i,\hat{u}_i]} f_i(x_i) - \min_{[\hat{\ell}_i,\hat{u}_i]} f_i(x_i)$. Similar results of course hold for each segment of $\hat{f}_i$ if $\hat{f}_i$ has two segments, and the bounds for the segments may be combined to obtain a sharper bound for $\hat{E}_i$. Note, however, that even the crude bound above implies that $\hat{E}_i$ must tend to zero as the size of $[\hat{\ell}_i,\hat{u}_i]$ tends to 0. Moreover, since $E(\hat{\ell},\hat{m},\hat{u}) \equiv \max_{[\hat{\ell},\hat{u}]} [\hat{f}(x)-f(x)] = \sum_{i=1}^{n} \hat{E}_i$, we have the following result; where $E(\ell^k,m^k,u^k)$ is defined to be the maximum over $[\ell^k,u^k]$ of the difference of the approximation generated by (ℓ^k,m^k,u^k) and the function f.

Lemma 3.1: If $\{(\ell^k,m^k,u^k)\}$ is a sequence of admissible triples such that $\ell^k \to \bar{m}$, $m^k \to \bar{m}$, and $u^k \to \bar{m}$, then $E(\ell^k,m^k,u^k) \to 0$.

In the convergence results to be developed, error estimates $\bar{e}_i$ are used to provide bounds on the optimal value z^{**} of (1.1).

Theorem 3.1: If $\bar{x}$ is an optimal solution of an approximating problem $P(\bar{\ell},\bar{x},\bar{u})$ where $(\bar{\ell},\bar{x},\bar{u})$ is an admissible triple, then the following lower and upper bounds hold for the optimal value z^{**} of (1.1).

$$f(\bar{x}) - \sum_{i=1}^{n} \bar{e}_i \le z^{**} \le f(\bar{x}),$$

where $\bar{e}_i \ge \max_{\bar{\ell}_i \le x_i \le \bar{u}_i} (\bar{f}_i(x_i)-f_i(x_i))$.

Proof: Because of the convexity of C and the admissibility of $(\bar{\ell},\bar{x},\bar{u})$, it is easily verified that for any $x\in S$, $\lambda x + (1-\lambda)\bar{x}$ is feasible for $P(\bar{\ell},\bar{x},\bar{u})$ for all sufficiently small positive λ. From this property it is easily seen that $\bar{x}$ is also optimal for the problem $\min_{x\in S} \bar{f}(x)$. Thus, if x^{**} is optimal for (1.1) we have

$$f(\bar{x}) \leq \bar{f}(\bar{x}) \leq \bar{f}(x^{**}) = f(x^{**}) + [\bar{f}(x^{**})-f(x^{**})] \leq z^{**} + \sum_{i=1}^{n} \bar{e}_i,$$ from which the lower bound follows. ▲

These error bound results lead to the following necessary and sufficient optimality conditions for the separable case.

Theorem 3.2: Let $\{(\bar{\ell}^k,\bar{x},\bar{u}^k)\}$ be a sequence of admissible triples such that $\bar{\ell}^k \to \bar{x}$, $\bar{u}^k \to \bar{x}$. A point $\bar{x}\in S$ is an optimal solution of (1.1) if and only if it is an optimal solution of each problem in the family of problems $P(\bar{\ell}^k,\bar{x},\bar{u}^k)$, $(k=1,2,\ldots)$.

Proof: ($\Leftarrow$) By the previous theorem we have $f(\bar{x}) - E(\bar{\ell}^k,\bar{x},\bar{u}^k) \leq z^{**} \leq f(\bar{x})$. However, by Lemma 3.1, $E(\bar{\ell}^k,\bar{x},\bar{u}^k) \to 0$ as $k \to +\infty$, so that $f(\bar{x}) = z^{**}$.

($\Rightarrow$) Since the objective function of $P(\bar{\ell}^k,\bar{x},\bar{u}^k)$ is at least $f(x)$ on $[\bar{\ell}^k,\bar{u}^k]$, but coincides with $f(x)$ at $\bar{x}$, it follows that $\bar{x}$ also solves each problem $P(\bar{\ell}^k,\bar{x},\bar{u}^k)$. ▲

We will refer to the preceding optimality conditions as secant optimality conditions at $\bar{x}$. These optimality conditions are the key

to the algorithm, which uses a corresponding search procedure at each iteration in order to obtain a feasible solution with an improved objective function value if the current iterate is non-optimal. That is, if $\bar{x}$ is the current iterate, a sequence of approximating problems satisfying the conditions of Theorem 3.2 is considered. If $\bar{x}$ is optimal, its optimality is established; otherwise a new iterate is obtained by selecting the "best" solution to the sequence of approximating problems. We will first state the most basic and straightforward algorithm of this sort, prove its convergence, and then indicate a number of refinements that are computationally important. Although an arbitrary procedure may be used to generate a starting feasible solution for the algorithm, note that one possibility that is in keeping with the spirit of the algorithm is to solve the two-segment approximating problem $P(\ell,\frac{1}{2}(\ell+u),u)$. (However, if information is available that leads to estimates for optimal values of some variables, these may be used instead of the corresponding midpoint values $\frac{1}{2}(\ell_j+u_j)$.) The parameter $\alpha\in(0,1)$ to be employed in the algorithm will be called the contraction factor for obvious reasons, and the search procedure that uses powers of α will be termed a contraction search.

Algorithm 1:

(a) Let $\alpha\in(0,1)$ and let $\bar{x}$ be the feasible solution available at the start of the current iteration.

(b) Denoting by $\bar{x}(\lambda)$ an optimal solution of the problem $P(\lambda\ell+(1-\lambda)\bar{x}, \bar{x}, \lambda u + (1-\lambda)\bar{x}) \equiv P_1(\lambda)$ for $\lambda\in(0,1)$ (with the understanding that $\bar{x}(\lambda)$ is taken to be $\bar{x}$ if $\bar{x}$ is optimal for $P_1(\lambda)$), determine $\bar{x}(\lambda)$ for $\lambda = \alpha, \alpha^2, \ldots$.

(c) If $\bar{x}(\alpha^k) = \bar{x}$ for $k = 1,2,\ldots,$ then $\bar{x}$ is an optimal solution of the given problem (1.1) by the secant optimality conditions.

(d) Otherwise let $\bar{\alpha}$ be a power of α such that $\bar{x}(\bar{\alpha})$ is the "best" point generated by the search procedure in the sense that if $\omega_1(\lambda)$ denotes the optimal value of $P_1(\lambda)$, then $\omega_1(\bar{\alpha}) \leq \omega_1(\alpha^k)$ for $k = 1,\ldots$. The point $\bar{x}(\bar{\alpha})$ is used as the starting feasible solution for the next iteration.

Theorem 3.3: If Algorithm 1 generates a sequence of points $\{x^j\}$, then $f(x^0) > f(x^1) > \ldots$, and $f(x^j) \to z^{**}$, the optimal value of (1.1). Each accumulation point of $\{x^j\}$ is an optimal solution of (1.1).

Proof: See Appendix. ▲

The basic idea of the proof is a principle frequently used in convergence analysis. If x^* is an accumulation point of $\{x^j\}$ and x^* does not solve (1.1), then the search procedure of the algorithm, applied at x^*, would produce a feasible solution whose objective function was strictly less than $f(x^*)$. Because of continuity properties of the approximating problems, this would imply that the search procedure would produce similar solutions near x^*, contradicting the fact that $f(x^*) < f(x^j)$ for all j. (In the case that

S is polyhedral, some rather strong continuity properties of the approximating problems are developed; but in the case that S is a general closed convex set, some weaker (but nonetheless adequate) continuity properties are established.

Algorithm 1, while easy to state, has some obvious computational disadvantages. One of these is the placing of the bounds at fractions of the distance to the left and right endpoints. This is generally not desirable in those cases in which, e.g., $\bar{x}_i = \ell_i$ for some i, so that the trial values of the upper bound for x_i would be located far to the right of $\bar{x}_i$ until the contraction factor α^k became quite small. A more reasonable alternative is to use multiples of some <u>fixed step</u> δ_i in placing the lower and upper bounds to the left and right of $\bar{x}_i$. This is the approach used in Algorithm 2 below, in which the problem $P_1(\lambda)$ is replaced by $P_2(\lambda)$, defined as follows: $P_2(\lambda) \equiv P(\bar{\ell}(\lambda), \bar{x}, \bar{u}(\lambda))$, where $\bar{\ell}_i(\lambda) \equiv \max\{\ell_i, \bar{x}_i - \lambda\delta_i\}$, $\bar{u}_i(\lambda) \equiv \min\{u_i, \bar{x}_i + \lambda\delta_i\}$, and δ is a fixed positive vector. The <u>optimal value</u> of $P_2(\lambda)$ is denoted by $\omega_2(\lambda)$. A second drawback of Algorithm 1 is the consideration of an infinite collection of approximating problems in step (b) of the Algorithm. From a computational viewpoint this may be circumvented by specifying an <u>improvement tolerance</u> $\mu > 0$ such that $\bar{x}(\alpha^r)$ is accepted as the next iterate if r is the first integer such that the optimal value of $P_2(\alpha^r)$ is less than $f(\bar{x}) - \mu$. Since μ may be chosen smaller than the smallest positive number representable on a given computer, use of such an improvement tolerance

is computationally equivalent to letting r be the first integer such that $\bar{x}$ is non-optimal for $P_2(\alpha^r)$. In practice, $\bar{x}$ is generally not optimal for $P_2(\alpha)$, so the next iterate is usually obtained by solving $P_2(\alpha)$, using $\bar{x}$ as a starting feasible solution. (The necessity of considering an infinite number of problems in part (b) of the algorithm if $\bar{x}$ is optimal for (1.1) may be averted by specifying an optimality tolerance ε^* such that the algorithm terminates if the condition $f(\bar{x}) \leq z^{**} + \varepsilon^*$ is known to hold. The use of error estimates in achieving finite convergence to an ε^*-optimal solution will be discussed below in connection with Algorithm 3.)

Algorithm 2 (fixed step; improvement tolerance)

(a) Let $\alpha \in (0,1)$ and let $\delta > 0$ and $\mu > 0$ be given. Let $\bar{x}$ be the feasible solution available at the start of the current iteration.

(b) Denoting by $\bar{x}(\lambda)$ an optimal solution of $P_2(\lambda)$ for $\lambda \in (0,1)$ (with the understanding that $\bar{x}(\lambda)$ is taken to be $\bar{x}$ if $\bar{x}$ is optimal for $P_2(\lambda)$), $\bar{x}$ is optimal for (1.1) if $\bar{x}(\alpha^k) = \bar{x}$ for $k = 1,2,\ldots$.

(c) Otherwise, let $\hat{\alpha}$ be the first power of α such that $\omega_2(\hat{\alpha}) \leq f(\bar{x}) - \mu$, provided that such an $\hat{\alpha}$ exists. If such an $\hat{\alpha}$ does not exist, choose $\hat{\alpha}$ as a power of α such that $\omega_2(\hat{\alpha}) \leq \omega_2(\alpha^k)$ for $k = 1,2,\ldots$. The point $\bar{x}(\hat{\alpha})$ is used as the starting feasible solution for the next iteration. ▲

The convergence properties of sequences generated by Algorithm 2 coincide with those described in Theorem 3.3. See the Appendix for details.

After a problem $P_2(\lambda)$ is solved, it is possible to derive with little additional computation an error bound $\varepsilon(\bar{x}(\lambda))$ for the corresponding solution, i.e., $\varepsilon(\bar{x}(\lambda))$ may be constructed so that $f(\bar{x}(\lambda)) \leq z^{**} + \varepsilon(\bar{x}(\lambda))$. Note that since $f(\bar{x}(\lambda)) \leq f(\bar{x})$, we may assume that $\varepsilon(\bar{x}(\lambda)) \leq \varepsilon(\bar{x})$. (Keep in mind that this notation does not reflect the fact that the error bound is computed using the data of $P_2(\lambda)$, so that if $\bar{x}(\lambda) = \bar{x}$, we may expect $\varepsilon(\bar{x}(\lambda)) < \varepsilon(\bar{x})$.) The basic property of convex functions used in the computation of the error bound is the lower bound for f_i provided by the extension of a segment of a secant approximation beyond the interval in which it dominates f_i. With such linear lower bounds, the error bound $\bar{e}_i$ of Theorem 3.1 is easily obtained by computing the maximum difference between two linear functions on $[\bar{\ell}_i, \bar{u}_i]$. (By performing a post-optimality analysis as described in [Meyer (1980)], bounds $\bar{\ell}$ and $\bar{u}$ may be constructed for each iterate $\bar{x}$ so that $\bar{x}$ will be optimal for $P(\bar{\ell}, \bar{x}, \bar{u})$, where $(\bar{\ell}, \bar{x}, \bar{u})$ is an admissible triple.)

Algorithm 3 (finite convergence to an ε^*-optimal solution) (See Fig. 3)

(a) Let $\alpha \in (0,1)$, let $\mu > 0$, and $\varepsilon^* > 0$ be given constants, and let $\delta > 0$ be a given vector of $\mathbb{R}^n$. Let $\bar{x}$ be the feasible solution available at the start of the current iteration.

(b) Let $\bar{x}(\lambda)$ be as in Algorithm 2, and let $\varepsilon(\bar{x}(\lambda))$ be the corresponding error bound. If $\bar{x}(\alpha^k) = \bar{x}$ for $k = 1,2,\ldots,s$, where α^s is the first power of α such that $\varepsilon(\bar{x}(\alpha^s)) \leq \varepsilon^*$, then the algorithm terminates with $\bar{x}$ as an ε^*-optimal solution.

(c) Otherwise, let $\hat{\alpha}$ be the first power of α such that $\omega_2(\hat{\alpha}) \leq f(\bar{x}) - \mu$, provided that such an $\hat{\alpha}$ exists. If such an $\hat{\alpha}$ does not exist, choose $\hat{\alpha}$ as a power of α such that $\omega_2(\hat{\alpha}) \leq \omega_2(\alpha^k)$ for $k = 1,2,\ldots$ (as noted below, only a finite number of $\omega_2(\alpha^k)$ need to be calculated to obtain $\hat{\alpha}$). If $\varepsilon(\bar{x}(\hat{\alpha})) \leq \varepsilon^*$, then the algorithm terminates with $\bar{x}(\hat{\alpha})$ as an <u>ε^*-optimal solution</u>. Otherwise, $\bar{x}(\hat{\alpha})$ is used as the starting feasible solution for the next iteration. ▲

To see that only a finite number of values of k need be considered to establish $\hat{\alpha}$ in the second part of step (c), note that failure of the Algorithm 3 to terminate at step (b) implies the existence of a power $\tilde{\alpha}$ of α such that $\omega_2(\tilde{\alpha}) < f(\bar{x})$. The availability of lower bounds on f then allows the computation of a $\bar{t}$ with the property that $\omega_2(\tilde{\alpha}) \leq \omega_2(\alpha^k)$ can be guaranteed for $k \geq \bar{t}$, so that at most $\omega_2(\alpha),\ldots, \omega_2(\alpha^{(\bar{t}-1)})$ need be computed.

It is easily seen that Algorithm 3 is finitely convergent to an ε^*-optimal solution under the assumption that $f(x^j) \to z^{**}$ implies $\varepsilon(x^j) \to 0$.

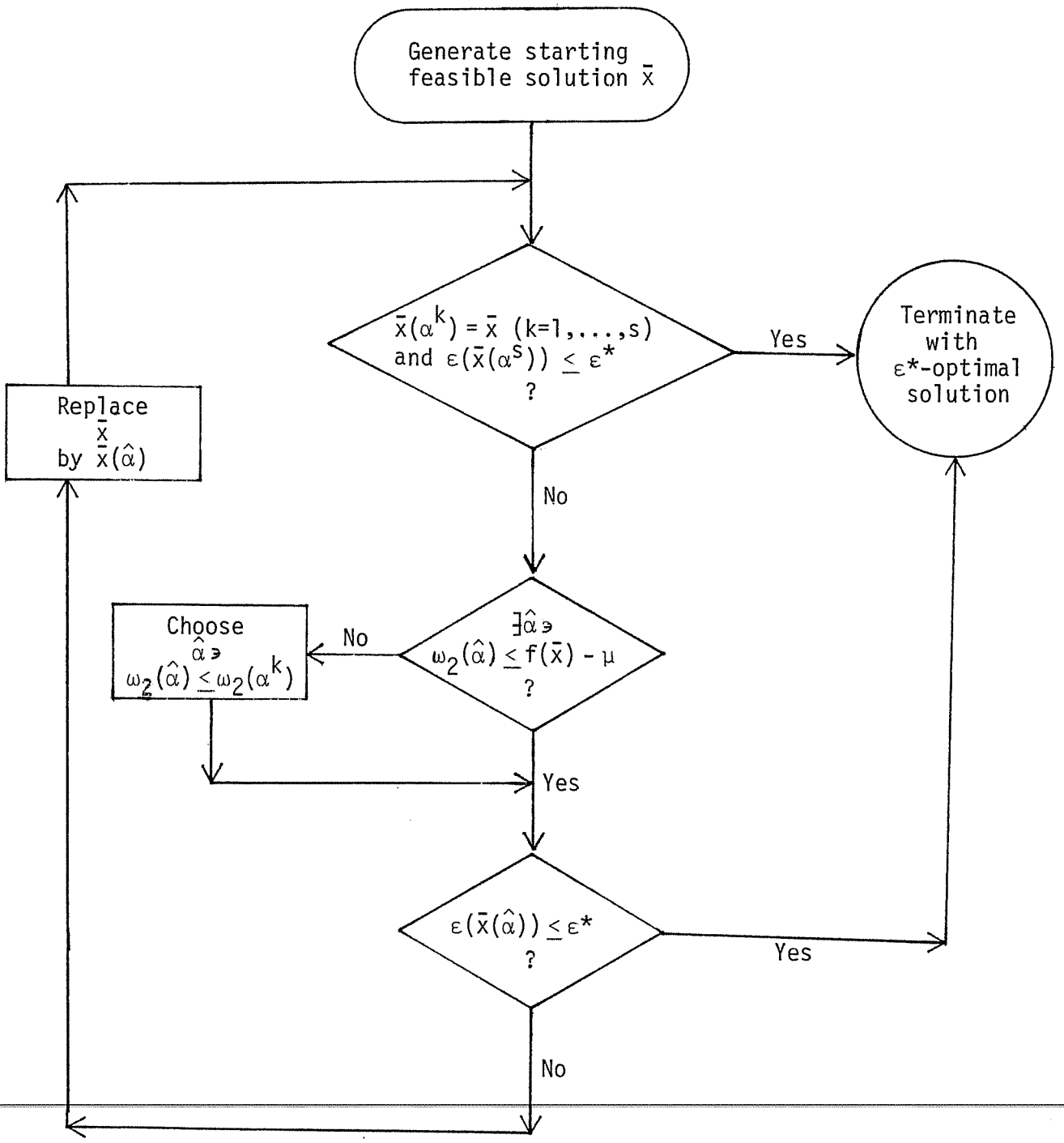

Figure 3. Flowchart for Algorithm 3

4. Extensions to the Non-Separable Case

The approach of the preceding section may be extended in a relatively straightforward manner to the case in which the objective function f of (1.1) is not assumed to be separable. Suppose that f is finite and convex on all of $\mathbb{R}^n$ and that the set of vectors $D \equiv \{d^1,\ldots,d^q\}$ has the property that the convex cone generated by D is $\mathbb{R}^n$, i.e., any vector in $\mathbb{R}^n$ may be expressed as a non-negative linear combination of the vectors of D. (If $q = n + 1$, then the vectors of D are the vertices of a simplex in n-space, but it is not necessary to construct D in this way. In fact, the two-segment method of the preceding section corresponds to setting $D = \{e^1,-e^1,\ldots,e^n,-e^n\}$, where e^i is the i^{th} unit vector.) An analog of the objective function of $P_2(\lambda)$ is obtained by using the following convex, piecewise-linear approximation of f, where D is analogous to δ:

$$\tilde{f}(\lambda,D,\bar{x},x) \equiv \min_{\theta} \; f(\bar{x})\theta_0 + \sum_{j=1}^{q} f(\bar{x}+\lambda d^j)\theta_j$$

$$\text{s.t. } \bar{x}\theta_0 + \sum_{j=1}^{q} (\bar{x}+\lambda d^j)\theta_j = x \tag{4.1}$$

$$\sum_{j=0}^{q} \theta_j = 1, \; \theta_j \geq 0 \; (j=0,\ldots,q)$$

Using (4.1), the problem (1.1) is approximated by

$$\min_{x,\theta} \; f(\bar{x})\theta_0 + \sum_{j=1}^{q} f(\bar{x}+\lambda d^j)\theta_j$$

$$\text{(4.2)} \qquad \text{s.t. } x\in S,\; x = \bar{x}\theta_0 + \sum_{j=1}^{q} (\bar{x}+\lambda d^j)\theta_j,$$

$$\sum_{j=0}^{q} \theta_j = 1,\; \theta_j \geq 0 \;(j=0,\ldots,q),$$

so that (4.2) is analogous to $P_2(\lambda)$. (Note that the problem (4.2) is an LP is S is polyhedral.) Moreover, since the constraints of (4.2) include those of (4.1), the objective function of (4.2) dominates $f(x)$ for all x feasible for (4.2). By substituting (4.2) for $P_2(\lambda)$ in Algorithm 2, a local piecewise-linear approximation method is obtained for the non-separable case. It should be pointed out, however, that the error analysis essential to establish the validity of this method in the non-separable case requires more than just continuity of f. In particular, examples are easily constructed to show that the analog of Algorithm 2 for the non-separable case may actually stall at a non-optimal point in step (b). However, it may be shown that if f is assumed to be differentiable, the error bound behavior and convergence analysis in the non-separable case are analogous to the separable case.

5. Computational Results

The secant approximation methods described in Section 3 have been applied to a variety of test problems with uniformly excellent results. Here we will cite only cases of particular interest. The first set of test problems cited arose from a statistical application and contains non-differentiable functions. The second example is a very large problem involving more than 900 variables and 600 constraints.

[Test Problem 1].

This class of problems is obtained from the calculation of the "Large deviation probability" of the Wilcoxson and the Yates test in statistics [Teng (1978)]. Such problems can be reduced to convex separable programs of the form:

$$\min \quad C + 0.5 \sum_{i=1}^{n} x_i \cdot \log x_i$$

$$\text{s.t.} \quad Ax = b$$

$$0 \leq x \leq e,$$

where $e = (1,\ldots,1)^T$.

Because of the $x_i \cdot \log x_i$ terms, the objective function is non-differentiable at any point having any $x_i = 0$. Even though an optimal solution x^{**} will satisfy $x^{**} > 0$, the non-differentiability property leads to difficulties in methods requiring derivatives, since the iterates may have some 0 components. Moreover, these problems were supplied to us by Teng after the use of a commercial separable programming package (FMPS), employing a fixed grid approach, produced results that were unsatisfactory. (The solutions that were generated by the package had some $x_i = 0$, a value that was unacceptable to the problem formulator given the interpretation he attached to the variables; the optimal values of the fixed grid approximating problems were also clearly not close to the values that were anticipated and were shown by this study to be as much as 35% too large when a grid size of 0.1 was used.)

We have tested three problems of this class, and obtained the following results:

Case	Iterations	Contractions	C	A	b	Upper bound	Lower bound	Final x	Error bound
1	3	11	1.0486538	A^1	b^1	.26394222	.26394222	x^1	$.199060 \times 10^{-15}$
2	8	2	1.1924368	A^2	b^2	.167257560	.167257558	x^2	$.218307 \times 10^{-8}$
3	16	0	1.6619357	A^3	b^3	.149409878	.149409876	x^3	$.243360 \times 10^{-8}$

where

$$b^1 = \begin{pmatrix} 0.6285 \\ 0.45 \\ 0.2 \\ 0.35 \\ 1.0 \end{pmatrix} \quad A^1 = \begin{pmatrix} 0.225 & 0.5 & 0.825 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 \\ 1.0 & 1.0 & 1.0 & 0 & 0 & 0 \end{pmatrix}$$

$$x^1 = (.12801 \quad .23888 \quad .63310 \quad .77199 \quad .16112 \quad .066895)^T,$$

$$b^2 = \begin{pmatrix} 0.6475 \\ 1.0 \\ 0.5 \\ 0.35 \\ 0.15 \\ 1.0 \end{pmatrix} \quad A^2 = \begin{pmatrix} 0.25 & 0.625 & 0.8375 & 0.9625 & 0 & 0 & 0 & 0 \\ 1.0 & 0 & 0 & 0 & 1.0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 & 0 & 0 & 0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$x^2 = (0.23242 \quad .33226 \quad .29809 \quad .13722 \quad .76758 \quad .16774 \quad .051906 \quad .012777)^T,$$

$$b^3 = \begin{pmatrix} .646875 \\ 0.7 \\ 0.35 \\ 0.35 \\ 0.25 \\ 0.2 \\ 0.15 \\ 1.0 \end{pmatrix} \quad A^3 = \begin{pmatrix} 0.175 & 0.4375 & 0.6125 & 0.7625 & 0.875 & 0.9625 & 0 \\ 1.0 & & & & & & 1 \\ & 1.0 & & & & & 1 \\ & & 1.0 & & & & 1 \\ & & & 1.0 & & & 1 \\ & & & & 1.0 & & 0 \\ & & & & & 1.0 & 1 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 0 \end{pmatrix}$$

$$x^3 = (0.13440 \quad .15211 \quad .21945 \quad .19170 \quad .16893 \quad .13341 \quad .56560 \quad .19789 \quad .13055 \quad .058303 \quad .031071 \quad .016585)^T.$$

[Test Problem 2].

This is a model of the Dallas Water supply network with data provided to us by Jeff Kennington of SMU. A detailed description of the formulation of this problem as an optimization problem is given in [Collins, et al (1978)]. This problem has 666 constraints, 906 variables, and an objective function involving 18 linear functions, 16 integrals $(\int_0^x -\sqrt{B(A-t^2)}dt \equiv -\frac{\sqrt{B}}{2}(x\cdot\sqrt{A-x^2} + A\sin^{-1}\frac{x}{\sqrt{A}}))$ and 872 terms of the form $C_i|x_i|^{2.85}$. In 19 iterations a feasible solution with objective value -206156 was obtained along with a lower bound on the optimal value of -206162. This lower bound was obtained via an estimation technique that used only the three final available values of each f_i, and could be improved through the use of first and second derivative information on the f_i. (Our objective values differ from those in the Collins report because of certain scale factors that were introduced into nonlinear objective terms.)

A rough comparison of these results may be made with the numerical studies reported in [Collins, et al (1978)]. In that paper results for four different approaches were presented for a smaller version of the problem containing 452 constraints and 530 arcs. The four methods compared were piecewise-linear approximations (with a fixed grid of points chosen by an error minimization approach), the Frank-Wolfe method, the convex simplex method, and Newton's method. The authors of that paper concluded that their piecewise-linear approximation method was superior to the Frank-Wolfe and convex simplex methods because of

its significantly smaller computer time, and was superior to the commercial (Systems Control, Inc.) version of Newton's method because of the frequent failure of the latter to converge when started outside of a small (±5%) neighborhood of the optimum. We believe that our piecewise-linear approach has several advantages relative to that of Collins, et al. In our method, the number of segments is limited to at most two, so the size of each approximating problem is smaller than that of Collins, who employed 8 segments. Rather than using an a priori selection of grid points, most of which will be of little ultimate value because of their distance from the optimal solution, we employ an adaptive strategy for grid point selection that adds only grid points in a neighborhood of the most recent estimate of optimal solution. Finally, when a relative error bound significantly better than 1% is required, the number of points required by an a priori approach will be prohibitive, whereas this goal poses no difficulties for the iterative piecewise-linear method.

Iteration	Upper Bound	Lower Bound	Error Bound	Objective Decrease
1	-165263			
2	-175262			99999
3	-185379			10117
4	-193969			8590
5	-201536			7567
6	-203809			2273
7	-205177	-273963	68786	1368
8	-205780	-220676	14896	603
9	-206052	-211212	5160	272
10	-206130	-207352	1222	78
11	-206147	-206728	481	17
12	-206153	-206313	160	6
13	-206155	-206242	87	2
14	-206155	-206206	51	0.7
15	-206156	-206205	49	0.2
16	-206156	-206167	11	0.2
17	-206156	-206167	11	0.2
18	-206156	-206164	8	0.1
19	-206156	-206162	6	0.02

6. Directions for Further Research

There are numerous issues related to the preceding discussion that are currently under investigation. From a numerical viewpoint, a key question is how best to conduct the "contraction search" at each iteration. For example, additional computational experience is needed to determine what contraction factor α should be used, and how accurate a search should be performed for an approximate minimum of the function $\omega_2(\lambda)$. Another possibility for a further refinement in the search procedure is to use the outcome of the contraction search as a descent direction along which a line search could be performed to determine the next iterate. (One disadvantage of this refinement would be the difficulties that this would cause with respect to the current error estimation procedure, which is applicable only at solutions of problems in which the objective function of (1.1) has been replaced by an approximation. However, the error estimates are really useful only in the terminal iterations, so these line searches could be employed in the initial iterations.) Interpolation or extrapolation could be used to determine "initial" bounds at each iteration, i.e., δ_i could be allowed to be iteration-dependent, with its value chosen so that $\bar{x}_i + \alpha\delta_i$ or $\bar{x}_i - \alpha\delta_i$ would coincide with an estimate for the optimal value of x_i. For example, single-variable quadratic interpolation (which is superlinearly convergent in the unconstrained case) could be applied term-by-term to a suitable Lagrangian.

In order to reduce the computation time per iteration, specialized variants of the simplex method could be used to take advantage of the two-segment structure in the case that S is polyhedral. Since the two variables associated with the segments of each piecewise-linear approximation correspond to columns (in the simplex tableau) that are nearly identical, it is certainly not necessary to deal with these columns as if they were completely distinct, and existing network and separable programming techniques such as those described in [Beale (1968)] and [Müller-Merbach (1970)] take advantage of this property.

Alternative procedures for computing lower bounds should also be considered. In particular, it might be possible to obtain tighter bounds by generating feasible solutions to a suitable dual of (1.1). The relationship between the "primal" error analysis approach of [Meyer (1979)] and the error bounds obtainable via duality theory should be investigated in order to assess the potential for computational improvements in this area.

Bounds may also be derived on the optimal values of the variables by generalizing the approach of [Thakur (1978)]. Additional computational experience is needed in order to determine if these bounds are tight enough to be useful.

The algorithms may also be generalized in a straightforward manner to handle nonlinear constraints under suitable convexity assumptions (see, e.g., [Dantzig (1963)], [Müller-Merbach (1970)]), but the convergence properties and computational efficiency of this generalization of the two-segment approach and its non-separable extension are yet to be determined.

Finally, it appears that the secant optimality conditions (Theorem 3.2) and their generalization to the non-separable case may be further extended to local optima in the non-convex case by taking into account the values of the function f at the optimal solutions of the piecewise-linear approximations.

Appendix

Continuity properties of the objective and the feasible sets of the approximating problems are essential to the convergence proofs of the algorithm. We will first establish continuity of the objective approximations, and then consider the continuity properties of the feasible sets.

Lemma A.1: Let $\{(\ell^j,m^j,u^j)\}$ be a sequence of admissible triples with corresponding piecewise-linear approximations f^j, and let $\ell^j \to \ell^*$, $m^j \to m^*$, and $u^j \to u^*$, where (ℓ^*,m^*,u^*) form an admissible triple whose corresponding objective function is denoted by f^*. If $\{y^j\}$ is a sequence such that $y^j \in [\ell^j,u^j]$ for all j and $y^j \to y^*$, then $f^j(y^j) \to f^*(y^*)$.

Proof: The proof consists of establishing the result $f_i^j(y_i^j) \to f_i^*(y_i^*)$ for each i by considering several cases.

Case 1 $(\ell_i^* \leq y_i^* < m_i^*)$

Clearly, $y_i^j < m_i^j$ for all sufficiently large j, and it is easily seen from the continuity of f_i that $f_i^j(y_i^j) \to f_i^*(y_i^*)$ in this case.

Case 2 $(m_i^* < y_i^* \leq u_i^*)$

Analogous to Case 1.

Case 3 $(\ell_i^* \leq y_i^* = m_i^* < u_i^*)$

Let J be an infinite subsequence of the integers such that $f_i^j(y_i^j) \overset{J}{\to} q^*$. If there is a subsequence J' of J such $y_i^j = m_i^j$,

then clearly $q^* = f_i^*(y_i^*)$. If there is a subsequence J' of J such that $m_i^j < y_i^j$ the same result follows from Case 2. If $\ell_i^* < m_i^*$ and there exists a subsequence such that $y_i^j < m_i^j$, the result follows from Case 1. Finally, if $\ell_i^* = m_i^*$ and $y_i^j < m_i^j$ for all sufficiently large $j \in J$, then the result follows from $\min \{f_i(\ell_i^j), f_i(m_i^j)\} \leq f_i^j(y_i^j) \leq \max \{f_i(\ell_i^j), f_i(m_i^j)\}$.

Case 4 $(\ell_i^* < m_i^* = y_i^* \leq u_i^*)$

Analogous to Case 3. ▲

Our next results deal with the continuity properties of the feasible sets of the approximating problems. Given a sequence of non-empty sets S^j, we say that $\lim_{j\to\infty} S^j = S^*$ if (a) $y^j \in S^j$ and $y^j \to y^*$ implies $y^* \in S^*$, and (b) $y^* \in S^*$ implies that there exists a sequence $\{y^j\}$ with $y^j \in S^j$ such that $y^j \to y^*$. Property (a) is called upper semi-continuity (u.s.c.) since it implies that S^* is at least as big as the "limit" of the S^j, and property (b) is called lower semi-continuity (l.s.c.) since it implies the opposite limit property.

Lemma A.2: If $\{(\ell^j, m^j, u^j)\}$ and (ℓ^*, m^*, u^*) satisfy the assumptions of the preceding lemma, if S is polyhedral, and if $m^j \in S$ for all j, then $\lim_{j\to\infty} S^j = S^*$, where $S^j \equiv S \cap [\ell^j, u^j]$ and $S^* \equiv S \cap [\ell^*, u^*]$.

Proof: In this case the sets S^j may be defined by a system of linear inequalities. It is known that the feasible sets of linear systems

are continuous under right-hand side (RHS) perturbations that preserve feasibility (see [Dantzig, et al (1967)]). For the sake of completeness, however, we will present a compact proof of this results.

Upper semi-continuity follows trivially from the definitions, so we need only verify l.s.c. Let $x^* \in S^*$ and consider the problem

$$\begin{aligned} &\min_{y} \quad \|y - x^*\|_\infty \\ &\text{s.t. } y \in S,\ \ell^j \leq y \leq u^j \end{aligned} \tag{A.1}$$

Denoting the optimal value by $\mu(\ell^j, u^j)$, note that l.s.c. may be established by showing that $\mu(\ell^j, u^j) \to 0$. Since $\mu(\ell^*, u^*) = 0$, it is sufficient to demonstrate an appropriate continuity property of μ. To establish this, first note that the problem (A.1) is equivalent to

$$\begin{aligned} &\min_{y,z} \quad z \\ &\text{s.t.} \quad y \in S,\ \ell^j \leq y \leq u^j \\ &\qquad -ze \leq y - x^* \leq ze, \end{aligned} \tag{A.2}$$

where e is a vector of 1's. Since (A.2) represents a family of LP's which are defined parametrically in terms of RHS perturbations, it is easily seen that the corresponding optimal value function μ is continuous with respect to bounds on y that ensure feasibility of the LP. ▲

If we drop the polyhedrality hypothesis and assume merely that S is a closed convex set, then the previous proof does not apply, because the optimal value function μ need not be continuous. However, by taking into account special properties of the sequences (ℓ^j,u^j) associated with Algorithm 2, an analogous continuity property may be established. (A somewhat weaker continuity property holds with respect to Algorithm 1.)

Theorem A.1: Let (ℓ^*,m^*,u^*) be an admissible triple such that $m^* \in S$, and for $i = 1,\ldots,n$, $\ell_i^* = \max \{\ell_i, m_i^* - \alpha^*\delta_i\}$, and $u_i^* = \min \{u_i, m_i^* + \alpha^*\delta_i\}$ where $\alpha^*\delta_i > 0$. If $\{m^j\} \subset S$ is such that $m^j \to m^*$, and for $i = 1,\ldots,n$ and all j, $\ell_i^j \equiv \max \{\ell_i, m_i^j - \alpha^*\delta_i\}$ and $u_i^j \equiv \min \{u_i, m_i^j + \alpha^*\delta_i\}$, then $\lim S^j = S^*$, where $S^j \equiv S \cap [\ell^j,u^j]$ and $S^* \equiv S \cap [\ell^*,u^*]$.

Proof: Upper semi-continuity follows trivially from the definitions, so we need only verify l.s.c. If $x^* \in S^*$, then $\lambda x^* + (1-\lambda)m^j \in S$ for all j and all $\lambda \in [0,1]$ because of the convexity of S. To see how "close" x^* is to $[\ell^j,u^j]$, consider the problem

$$\begin{aligned} &\min_{y,\lambda} \ \|y - x^*\|_\infty \\ &\text{s.t. } y = \lambda x^* + (1-\lambda)m^j \\ &\qquad \ell^j \leq y \leq u^j,\ 0 \leq \lambda \leq 1. \end{aligned} \tag{A.3}$$

Denoting the optimal value by μ^j, note that lower semi-continuity of the sequence $\{S^j\}$ may be established by showing that $\mu^j \to 0$. The problem (A.3) may be re-written as

$$\min_{\theta} \ \|\theta(m^j - x^*)\|_\infty$$

(A.4)

$$\text{s.t. } \ell^j - x^* \leq \theta(m^j - x^*) \leq u^j - x^*,$$

$$0 \leq \theta \leq 1$$

where $\theta = 1 - \lambda$, since $y - x^* = \theta(m^j - x^*)$. Note that $\theta = 1$ is always a feasible solution, so that (A.4) has an optimal solution for all j. Let J be chosen such that $\mu^* \equiv \limsup \mu^j = \lim_{j \in J} \mu^j$.

Without loss of generality we may assume that there is an r such that for $j \in J$, the optimal value θ^j of (A.4) is determined by the constraint

$$\ell_r^j - x_r^* \leq \theta(m_r^j - x_r^*), \tag{A.5}$$

so that

$$\theta^j = (\ell_r^j - x_r^*)/(m_r^j - x_r^*). \tag{A.6}$$

(If none of the constraints $\ell^j - x^* \leq \theta(m^j - x^*) \leq u^j - x^*$ have the property that they bound θ from below, then $\theta^j = 0$ and the desired result is proved. If a constraint involving u^j bounds θ from below, the argument is similar to that in the ℓ^j case.) If $\ell_r^j < x_r^* < m_r^j$, the preceding expression for θ^j is negative; if $x_r^* = m_r^j$, the expression is undefined (and it is clear that θ is not bounded from below by the corresponding constraint); and if $m_r^j < x_r^*$, then the constraint (A.5) provides an upper bound on θ rather than a lower bound. Thus, when θ^j is determined by (A.6), it is the case

that $x_r^* \leq \ell_r^j \leq m_r^j$, and thus $\ell_r^* \leq x_r^*$ implies $\ell_r^j \overset{J}{\to} x_r^*$. Clearly, if $x_r^* < m_r^*$, then $\theta^j \overset{J}{\to} 0$, so suppose $x_r^* = m_r^*$. Since $\ell_r^* = x_r^*$, admissibility implies $\ell_r^* = x_r^* = m_r^* = \ell_r$. However, since $\ell_r^j = \max \{\ell_r, m_r^j - \alpha^* \delta_r\}$, where α^* and δ_r are fixed positive constants, $\ell_r^j = \ell_r$ for sufficiently large $j \in J$, and (A.6) implies $\theta^j = 0$ for such j. ▲

(In the case of Algorithm 1, it may be deduced from (A.6) that $\lim_{j \to \infty} \theta^j = 1 - \alpha^*$ in the worst case, but this property turns out to be adequate for the convergence proof of Algorithm 1.)

Corollary A.1: If the hypotheses of Theorem A.1 hold, and ω^*, the optimal value of $P(\ell^*, m^*, u^*)$, satisfies $\omega^* < f(m^*)$, then the optimal value of $P(\ell^j, x^j, u^j)$ is also strictly less than $f(m^*)$ for j sufficiently large.

Proof: By Theorem A.1, if y^* is an optimal solution of $P(\ell^*, m^*, u^*)$, then there exists a sequence $\{y^j\}$ with $y^j \in S^j$ such that $y^j \to y^*$. By Lemma A.1, $f^j(y^j) \to f^*(y^*) = \omega^*$. Since the optimal value of $P(\ell^j, x^j, u^j)$ is not greater than $f^j(y^j)$, the conclusion follows. ▲

Convergence Proof for Algorithm 2:

If $\{x^j\}$ is a sequence generated by Algorithm 2, then $\{f(x^j)\}$ is monotone decreasing, so let $z^* = \lim_{j \to \infty} f(x^j)$, and assume $z^{**} < z^*$. Let x^* be an accumulation point of $\{x^j\}$ so that $f(x^*) = z^*$ and x^* is not optimal for (1.1). Let α^* be a power of α such that x^* is not optimal for $P(\ell^*, x^*, u^*)$, where $\ell_i^* \equiv \max \{\ell_i, x_i^* - \alpha^* \delta_i\}$,

and $u_i^* \equiv \min \{u_i, x_i^* + \alpha^* \delta_i\}$. By Corollary A.1 with $m^* = x^*$, the optimal value of the problems $P(\ell^j, x^j, u^j)$ is strictly less than $f(x^*)$ for x^j sufficiently close to x^*, so that, for such x^j, $f(x^j) < f(x^*)$, a contradiction. ▲

Convergence Proof for Algorithm 1:

The proof is analogous to that of Algorithm 2, except that an appropriate modification of the proof of Corollary A.1 is needed in the non-polyhedral case, as noted in the remarks following Theorem A.1.

References

1. Avriel, M., Nonlinear Programming Analysis and Methods, Prentice-Hall, Inc., Englewood Cliff, N.J., 1976.

2. Bachem, Achim and Korte, Bernhard, "Quadratic Programming over Transportation Polytopes", Report No. 7767-OR, Institut für Okonometrie und Operations Research, Bonn, 1977.

3. Beale, E. M. L., Mathematical Programming in Practice, Sir Isaac Pitman & Sons Ltd., London, 1968.

4. Charnes, A. and Lemke, C. E., Minimization of Nonlinear Separable Convex Functionals, Naval Res. Logistics Quart., 1, pp. 301-312, 1954.

5. Collins, M., Cooper, L., Helgason, R., Kennington, J., and LeBlanc, L., "Solving the Pipe Network Analysis Problem Using Optimization Techniques", Man. Sci., 24, pp. 747-760, 1978.

6. Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, N.J., 1963.

7. Dantzig, G. B., Folkman, J., and Shapiro, N., "On the Continuity of the Minimum Set of a Continuous Function, JMAA, 17, 519-548, 1967.

8. Dantzig, G. B., Johnson, S., White, W., A Linear Programming Approach to the Chemical Equilibrium Problem, Man. Sci., 5, 38-43, 1958.

9. Fletcher, R., "An Algorithm for Solving Linearly Constrained Optimization Problems", Mathematical Programming, 2, 133-165, 1972.

10. Geoffrion, A. M., Objective Function Approximations in Mathematical Programming, Mathematical Programming, 13, 23-27, 1977.

11. Griffith, R. E. and Stewart, R. A., "A Nonlinear Programming Technique for the Optimization of Continuous Processing Systems", Man. Sci., 7, 379-392, 1961.

12. Gross, O., Class of Discrete Type Minimization Problem, RM-1644, RAND Corp., Santa Monica, California, 1956.

13. Hadley, G., Nonlinear and Dynamic Programming, Addison-Wesley Publishing Co., Inc., 1964.

14. Marsten, R. E., Hogan, W. W., Blankenship, J. W., "The Boxstep Method for Large-Scale Optimization", Operations Research, 23, pp. 389-405, 1975.

15. Meyer, R. R., "A Class of Nonlinear Integer Programs Solvable by a Single Linear Program", SICOP, 15, pp. 935-946, 1977.

16. Meyer, R. R., Two-Segment Separable Programming, Management Science, 25, pp. 385-395, 1979.

17. Meyer, R. R., "Computational Aspects of Two-Segment Separable Programming", in preparation, 1980.

18. Meyer, R. R. and Smith, M. L., "Algorithms for a Class of 'Convex' Nonlinear Integer Programs", in Computers and Mathematical Programming", ed. by W. W. White, NBS Special Publication 502, 1978.

19. Miller, C. E., "The Simplex Method for Local Separable Programming", in Recent Advances in Mathematical Programming", ed. by R. L. Graves and P. Wolfe, McGraw-Hill, New York, 1963.

20. Müller-Merbach, H., "Die Methode der 'direkten Koeffizientanpassung' (μ-Form) des Separable Programming", Unternehmensforschung, Band 14, Heft 3, pp. 197-214, 1970.

21. Rockafellar, R. T., Optimization in Networks, Lecture Notes, University of Washington, 1976.

22. Saaty, T. L., Optimization in Integers and Related Extremal Problems, McGraw-Hill, New York, 1970.

23. Symonds, G. H., "Chance-Constrained Equivalents of Some Stochastic Program", Operation Research, 15, pp. 495-512, 1967.

24. Teng, James Z, "Exact Distribution of the Kruskal-Wallis H Test and the Asymptotic Efficiency of the Wilcoxon Test with Ties", Ph.D. Thesis, University of Wisconsin-Madison, 1978.

25. Thakur, L. S., "Error Analysis for Convex Separable Programs: The Piecewise Linear Approximation and the Bounds on the Optimal Objective Value", SIAM J. App. Math., pp. 704-714, 1978.

26. Wagner, H. M., Principles of Operational Research-with Applications to Managerial Decisions, Prentice-Hall, Inc., 1969.

27. Wets, R., "Programming Under Uncertainty: The Equivalent Convex Program", SIAM J. Appl. Math., 14, pp. 89-105, 1966.