COMPUTATIONAL ASPECTS OF TWO-SEGMENT SEPARABLE PROGRAMMING

by

R. R. Meyer

Computer Sciences Technical Report #382

March 1980

Computational Aspects of Two-Segment Separable Programming

By

R. R. Meyer

Computer Sciences Department
University of Wisconsin
Madison, WI, USA, 53706

Abstract

Recursive separable programming algorithms based on local, two-segment approximations are described for the solution of separable convex programs. Details are also given for the computation of lower bounds on the optimal value by both a primal and a dual approach, and these approaches are compared. Computational comparisons of the methods are provided for a variety of test problems, including a water supply application (with more than 600 constraints and more than 900 variables) and an econometric modelling problem (with more than 2000 variables).

Key words: Separable Programming, Network Optimization, Piecewise-linear Approximation, Error Bounds

Abbreviated title: R.R. Meyer/Two-Segment Separable Programming

Research supported by National Science Foundation Grants MCS74-20584 A02 and MCS-7901066.

1. Introduction

In a previous paper [10] convergence theorems were developed for recursive separable programming algorithms for problems of the form

$$\text{(1.1)}\qquad \min_{x} \sum_{i=1}^{n} f_i(x_i) \quad \text{s.t. } x \in C \cap [\ell, u],$$

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 each f_i is a continuous convex function on the interval $[\ell_i,u_i]$. Here we will consider computational aspects and experience with these and newly developed algorithms for the special case in which $C = \{x|Ax=b\}$, where A is an $m\times n$ matrix and $b \in \mathbb{R}^m$. (These methods also apply with obvious modifications to the case in which C is given by any finite set of linear equations and inequalities, but for notational simplicity we will consider only the case of C given by linear equations.) In this case each iteration requires only the solution of a linear programming problem.

It is notationally convenient to denote the feasible set of (1.1) by S. To avoid trivial cases we will assume that S is non-empty (if it was empty, this fact would be established on the first iteration) and that the bounds satisfy $\ell < u$. Under the assumptions made, (1.1) has an optimal value, denoted by z^{**}. Finally, we denote $\sum_{i=1}^{n} f_i(x_i)$ by $f(x)$.

Problems of this form arise in numerous applications, including econometric data fitting [1], electrical networks [17], water supply systems [5], logistics [18], and statistics [19]. Computational experience with problems arising from several of these areas is described in Section 8.

At each iteration, in addition to the generation of a feasible solution and a corresponding upper bound for the optimal value as described in Sections 2-4, a lower bound on the optimal value may also be computed by using only function value information. If each f_i is differentiable and if equations of the form $f_i'(x) = \alpha$ may be easily solved, lower bounds based on Lagrangian relaxation may also be computed. It is shown that the "primal" lower bounds described in Section 5 are not tighter than the "dual-based" lower bounds from the Lagrangian relaxation of Section 6. The algorithm may be terminated when the difference between the upper and lower bounds is less than a prescribed tolerance. This termination criterion guarantees that the feasible solution giving rise to the upper bound differs (in an objective function sense) from the optimal solution by less than the tolerance.

2. The Linear Programming Subproblems

The iterative method to be described for the given problem requires the construction at each iteration of a piecewise-linear approximation of at most two segments for each of the f_i. Except for the initial iteration, in which a feasible starting solution is not assumed, the piecewise-linear approximations $\tilde{f}_i$ are determined by function values at points $(\tilde{\ell},\tilde{m},\tilde{u})$, where $\tilde{m}$ is the optimal solution of the preceding iteration, and $\tilde{\ell}$ and $\tilde{u}$ are "temporary" lower and upper bounds with the admissibility properties:

(2.1) $\ell \leq \tilde{\ell} \leq \tilde{m} \leq \tilde{u} \leq u$,

(2.2) $\tilde{m}_i > \tilde{\ell}_i$ if $\tilde{m}_i > \ell_i$, and

(2.3) $\tilde{m}_i < \tilde{u}_i$ if $\tilde{m}_i < u_i$.

From a computational viewpoint, these admissibility properties may be thought of as allowing a decrease in x_i (thought of as starting the iteration with a value $\tilde{m}_i$) if x_i is not at the true lower bound ℓ_i, and allowing an increase if x_i is not at the true upper bound. (These properties are also used in convergence proofs for the method.) The precise manner in which $\tilde{\ell}$ and $\tilde{u}$ are generated at each iteration will be described below. (Note that when $\ell_i = \tilde{\ell}_i = \tilde{m}_i$ or $\tilde{m}_i = \tilde{u}_i = u_i$, f_i is approximated by the secant determined by two function values rather than by a two-segment approximation determined by three function values.)

In the case in which $\tilde{f}_i$ is comprised of two segments, the function corresponding to $[\tilde{\ell}_i,\tilde{m}_i]$ will be denoted as $\tilde{f}_i^L$ and the function corresponding to $[\tilde{m}_i,\tilde{u}_i]$ will be denoted as $\tilde{f}_i^U$. The key property of $\tilde{f}_i$ that is used to guarantee monotonicity of the algorithm is that

$f_i(x_i) \leq \tilde{f}_i(x_i)$ for $x_i \in [\tilde{\ell}_i, \tilde{u}_i]$. To exploit this property the constraints $x \in [\tilde{\ell}, \tilde{u}]$ are imposed in the subproblem. Thus, the triple $(\tilde{\ell}, \tilde{m}, \tilde{u})$ not only determines the approximations $\tilde{f}_i$, but determines additional constraints as well. For this reason, the corresponding subproblem denoted by $P(\tilde{\ell}, \tilde{m}, \tilde{u})$ may be written as

$$P(\tilde{\ell},\tilde{m},\tilde{u}) \equiv \begin{cases} \min\limits_{x} \sum\limits_{i=1}^{n} \tilde{f}_i(x_i) \\ \\ \text{s.t. } x \in S \cap [\tilde{\ell},\tilde{u}]. \end{cases}$$

Since each $\tilde{f}_i$ is piecewise-linear and convex, $P(\tilde{\ell},\tilde{m},\tilde{u})$ my be converted to and solved as a linear program (LP) by standard techniques (for details, see [13]. However, instead of directly decomposing each variable x_i corresponding to a two-segment approximation into the sum of two variables x_i^+ and x_i^-, it should be noted that with the proper handling of the pricing out operation, this LP also can be taken to be of size $m \times n$ except that an additional cost coefficient must be stored for a variable corresponding to a two-segment approximation. $P(\tilde{\ell},\tilde{m},\tilde{u})$ will have an optimal solution since it is feasible (the optimal basic feasible solution $\tilde{m}$ from the previous iteration may be used as a starting basic feasible solution) and bounded.

For the <u>initial</u> subproblem, in which a feasible starting solution $\tilde{m}$ is <u>not</u> assumed, the problem $P(\ell,(\ell+u)/2,u)$ may be solved. This is the problem corresponding to the two-segment approximations generated by the endpoints and midpoint of each segment $[\ell_i,u_i]$. (Note that the feasible set for this problem is simply the <u>original</u> feasible set, so the feasibility of the initial subproblem is a consequence of the feasibility of the original problem.) If estimates or "target" values (to be described below) are available for any of the x_i, they may be used in the first iteration instead of the midpoint values $(\ell_i+u_i)/2$, which may be thought of as "default" values. In later iterations $\tilde{m}$ will be feasible, but the values used for $\tilde{\ell}$ and $\tilde{u}$ need not be.

3. An Overview of the Algorithms

In essence, the algorithms to be considered for (1.1) consist of the solution of a sequence of problems of the form $P(\tilde{\ell},\tilde{m},\tilde{u})$, where $\tilde{m}$ is an optimal solution of the preceding iteration, and $\tilde{\ell}$ and $\tilde{u}$ are appropriately chosen temporary bounds with the admissibility properties (2.1)-(2.3). If x^k denotes the optimal solution of the k-th iteration, it may be shown (see [12]) that the iterates have the property that $f(x^k) \geq f(x^{k+1})$, and that convergence of the sequence $\{f(x^k)\}$ to the optimal value of (1.1) is guaranteed when the values of $\tilde{\ell}$ and $\tilde{u}$ are chosen by a procedure called contraction search. In this section the contraction search procedure will be described under the assumption that initial estimates for $\tilde{\ell}$ and $\tilde{u}$ are given. Procedures for generating these initial estimates are outlined below and described in detail in the following sections.

The idea underlying contraction search is that a feasible solution $\tilde{m}$ is an optimal solution for the original problem (1.1) if and only if it is optimal for a family of problems of the form $P(\tilde{\ell}^j,\tilde{m},\tilde{u}^j)$, $(j=1,2,\ldots)$, where the initial triple $(\tilde{\ell}^1,\tilde{m},\tilde{u}^1)$ is admissible and the others are defined by $\tilde{\ell}^j \equiv \tilde{m} - \beta(\tilde{m}-\tilde{\ell}^{j-1})$, $\tilde{u}^j \equiv \tilde{m} + \beta(\tilde{u}^{j-1}-\tilde{m})$ for $j \geq 2$, where $\beta < 1$ is a given constant (typically $\beta = \frac{1}{2}$). Thus, unless $\tilde{m}$ is optimal for the original problem, as the hypercube $[\tilde{\ell}^j,\tilde{u}^j]$ contracts toward $\tilde{m}$, a problem of the form $P(\tilde{\ell},\tilde{m},\tilde{u})$ will be generated with the property that $\tilde{m}$ is not optimal for $P(\tilde{\ell},\tilde{m},\tilde{u})$. In that case, if x^* is an optimal solution of $P(\tilde{\ell},\tilde{m},\tilde{u})$, it follows that $f(x^*) \leq \tilde{f}(x^*) < \tilde{f}(\tilde{m}) = f(\tilde{m})$, so that x^* is a feasible solution with a better objective value than $\tilde{m}$. Both from a theoretical and a computational viewpoint certain additional properties are required for the initial bounds $\tilde{\ell}^1$ and $\tilde{u}^1$. Theoretical properties sufficient to guarantee convergence were described in [10]. Computationally, several rather different strategies have proved successful and will be described below. In essence, these strategies endeavor to construct intitial bounds that are neither so "far" from $\tilde{m}$ that $\tilde{m}$ is optimal for $P(\tilde{\ell}^1,\tilde{m},\tilde{u}^1)$ (with the result that a contraction is needed) nor so "close" to $\tilde{m}$ that only a small improvement in the objective value is possible.

4. Adaptive Strategies

An adaptive strategy (referred to as AS1 below) that has been found effective for the determination of the initial bounds $\tilde{\ell}^1$ and $\tilde{u}^1$ for the k-th major iteration uses the final set of bounds $\hat{\ell}$ and $\hat{u}$ of the $(k-1)$st iteration and the optimal solution $\tilde{m}$ of the corresponding problem $P(\hat{\ell},\hat{m},\hat{u})$. We first consider the case of a variable x_i that is not artificially bounded, i.e., $\tilde{m}_i = \hat{\ell}_i$ implies $\hat{\ell}_i = \ell_i$ and $\tilde{m}_i = \hat{u}_i$ implies $\hat{u}_i = u_i$, so that only an original bound may be active at the optimal value of x_i. In this case the initial bounds for the next iteration are defined by $\tilde{\ell}_i^1 \equiv \max\{\ell_i, \tilde{m}_i - \beta(\hat{m}_i - \hat{\ell}_i) - \gamma|\tilde{m}_i - \hat{m}_i|\}$ and $\tilde{u}_i^1 \equiv \min\{u_i, \tilde{m}_i + \beta(\hat{u}_i - \hat{m}_i) + \gamma|\tilde{m}_i - \hat{m}_i|\}$, where β and γ are constants with $\beta < 1$ and $\beta + \gamma > 1$ ($\beta = 0.5$ and $\gamma = 1$ were used). These formulas reduce the half-range $(\hat{u}_i - \hat{\ell}_i)/2$ for x_i by a factor of β if the optimal value of x_i coincides with the median point $\hat{m}_i$, and increase the half-range by nearly a factor of $\beta + \gamma$ if the optimal value is nearly at a bound, with linear interpolation between these extremes being used for intermediate values. If it is the case that $\ell_i < \hat{\ell}_i = \tilde{m}_i$ or $\tilde{m}_i = \hat{u}_i < u_i$, provision is made for a somewhat larger increase in the half-ranges. Details for these cases are given in [Meyer (1980)].

An alternative approach that has also proved successful involves the direct use of error bounds. As will be shown in Sections 5-7, it is possible at iteration $(k-1)$ to derive a lower bound $\underline{z}_{k-1}$ on the optimal value of (1.1), so that $f(x^{k-1}) - z^{**} \leq f(x^{k-1}) - \underline{z}_{k-1} \equiv B_{k-1}$. The error bound from iteration $(k-1)$ may be employed to derive the initial estimates for iteration k by using the expressions

$\tilde{\ell}_i^1 = \max\ \{\ell_i, \tilde{m}_i - \theta\cdot(B_{k-1}/f(\tilde{m}))^{\frac{1}{2}}(u_i-\ell_i)/2\}$, where θ is a given positive constant (0.4 was used) and $\tilde{u}_i^1 = \min\ \{u_i, \tilde{m}_i + \theta\cdot(B_{k-1}/f(\tilde{m}))^{\frac{1}{2}}(u_i-\ell_i)/2\}$. Thus, as the error bound tends to 0, the width of the initial half-ranges also tends to 0. It is easily shown that this choice of $\tilde{\ell}^1$ and $\tilde{u}^1$ also guarantees convergence of $\{f(x^k)\}$ to z^{**}.

The two adaptive strategies may be combined by beginning with the first and then switching over to the error-bound strategy once the error bound is sufficiently small (say, $B_{k-1}/f(\tilde{m}) \leq 0.01$). This combined strategy will be referred to as AS2 below. (For stability purposes the ratio $(\tilde{u}_i^1-\tilde{\ell}_i^1)/(\hat{u}_i-\hat{\ell}_i)$ is also bounded from above and below in this strategy.)

5. Error Bounds from $P(\tilde{\ell},\tilde{m},\tilde{u})$

Given an optimal solution x^* of $P(\tilde{\ell},\tilde{m},\tilde{u})$, a lower bound on the optimal value z^{**} of (1.1) may be obtained. We will first derive a lower bound under the assumption that x^* is not artificially bounded, and then show how the other case ($x_i^*=\tilde{\ell}_i>\ell_i$ or $x_i^*=\tilde{u}_i<u_i$ for some i) may be transformed to the first through the use of the optimal values of the dual variables and the modification of the bounds $\tilde{\ell}$ and/or $\tilde{u}$. Related ideas may be found in [7] and [20].

It was shown in [12] that, when x^* is not artificially bounded by $\tilde{\ell}$ or $\tilde{u}$, a lower bound on z^{**} is given by $\tilde{f}(x^*) - \sum_{i=1}^{n} \tilde{e}_i$, where $\tilde{e}_i \geq \max_{\tilde{\ell}_i\leq x_i\leq\tilde{u}_i} (\tilde{f}_i(x_i)-f_i(x_i))$. For those functions f_i for which the equations $f_i'(x_i) = c_i^+$ and $f_i'(x_i) = c_i^-$ may be solved, $\tilde{e}_i$ may be taken to be $\tilde{E}_i \equiv \max_{\tilde{\ell}_i\leq x_i\leq\tilde{u}_i} (\tilde{f}_i(x_i)-f_i(x_i))$.

In cases in which the maximum error is not easily computed, $\tilde{e}_i$ may be obtained by using the fact that a lower bound for f_i on $[\tilde{\ell}_i,\tilde{m}_i]$ is given by $\tilde{f}_i^U$ on that interval, and a lower bound for f_i on $[\tilde{m}_i,\tilde{u}_i]$ is given by $\tilde{f}_i^L$ on the latter interval (see Figure 1). This yields an upper bound

$$\tilde{e}_i = \max \{ \max_{[\tilde{\ell}_i,\tilde{m}_i]} (\tilde{f}_i^L(x_i)-\tilde{f}_i^U(x_i)), \max_{[\tilde{m}_i,\tilde{u}_i]} (\tilde{f}_i^U(x_i)-\tilde{f}_i^L(x_i))\}$$

which, although seemingly rather crude, has proved rather good in practice. (In the case in which $\tilde{f}_i$ consists of only a single segment, any value of f_i outside of the segment can be used to generate a similar upper bound on the approximation error. If first derivatives

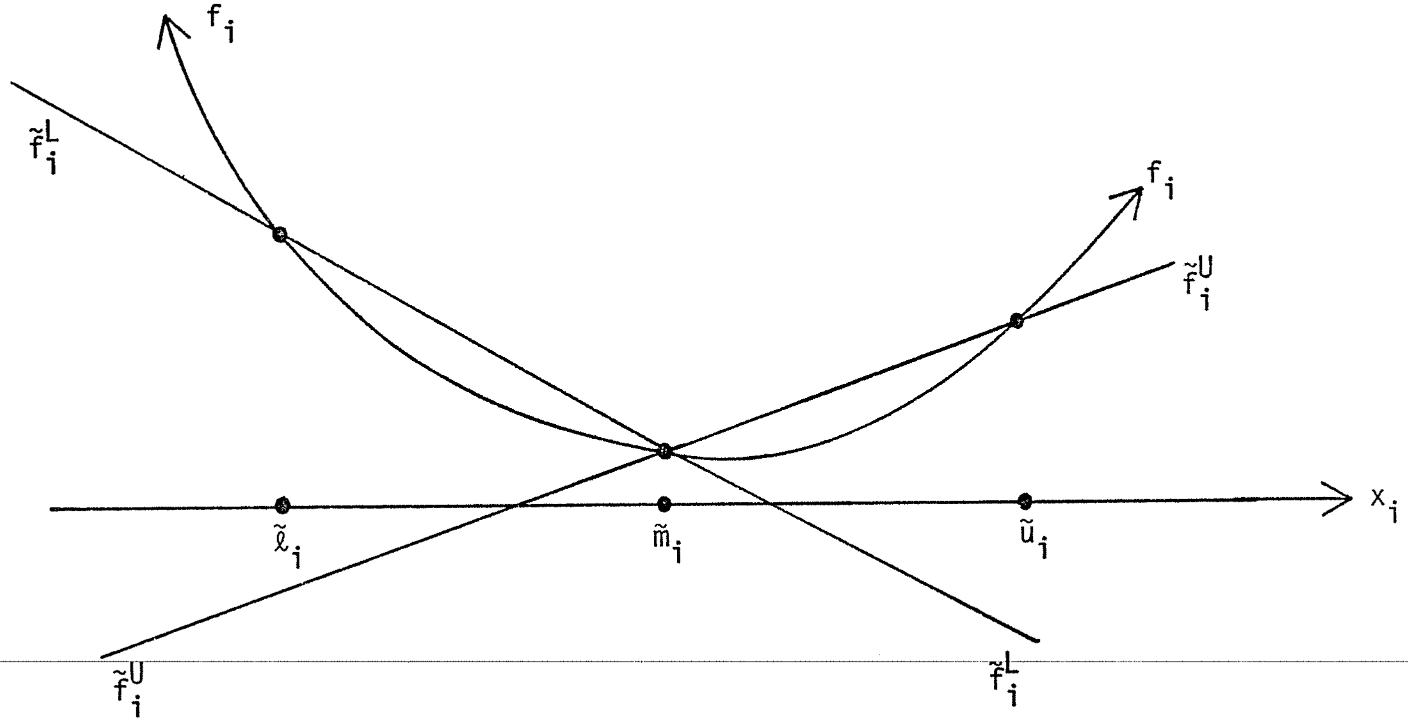

Figure 1. Using the two-segment approximations to obtain under-estimates for f_i

of f_i are available or if bounds on the second derivative of f_i are known, then more accurate under-estimates of f_i may be used to obtain smaller values of $\tilde{e}_i$ (see [20], but in such cases $\tilde{E}_i$ itself may usually be computed.) ▲

We will now complete the discussion of the error analysis by considering the case in which $x_i^* = \tilde{\ell}_i > \ell_i$ or $x_i^* = \tilde{u}_i < u_i$ for some i (x_i is said to be <u>artificially bounded</u>). In this instance, for each i such that x_i is artificially bounded, the approximation $\tilde{f}_i$ will be replaced by a new two-segment approximation f_i^E (the superscript E is used to suggest that the new approximation <u>extends</u> outside of $[\tilde{\ell}_i,\tilde{u}_i]$).

<u>Theorem 5.1</u>: If a variable x_i is at its upper bound $\tilde{u}_i < u_i$ in an optimal solution x^* of $P(\tilde{\ell},\tilde{m},\tilde{u})$ and if $\rho_{i,+}^*$ is an optimal value of the dual variable for the corresponding bound constraint, then x^* is also an optimal solution of the problem obtained from $P(\tilde{\ell},\tilde{m},\tilde{u})$ by (1) replacing $\tilde{u}_i$ by u_i and (2) replacing $\tilde{f}_i$ by the function f_i^E defined on $[\tilde{\ell}_i,\tilde{u}_i]$ by $\tilde{f}_i^U$ and on $[\tilde{u}_i,u_i]$ by $\tilde{f}_i^U(x) - \rho_{i,+}^*(x_i-\tilde{u}_i)$.

<u>Proof</u>: This result may be established by considering the optimality conditions for an appropriate LP. For details refer to [13]. ▲

Geometrically, this implies the optimality of x^* with respect to the two-segment approximations f_i^E in which $\tilde{f}_i^U(x_i)$ (with slope c_i^+) becomes the left segment of f_i^E and $\tilde{f}_i^U(x_i) - \rho_{i,+}^*(x_i-\tilde{u}_i)$ becomes the right segment (see Figure 2). An analogous argument holds for the case in which a variable x_i is at a lower bound $\tilde{\ell}_i$ satisfying $\ell_i < \tilde{\ell}_i < \tilde{m}_i$ (see Figure 3). In this case the new two-segment approximation f_i^E has

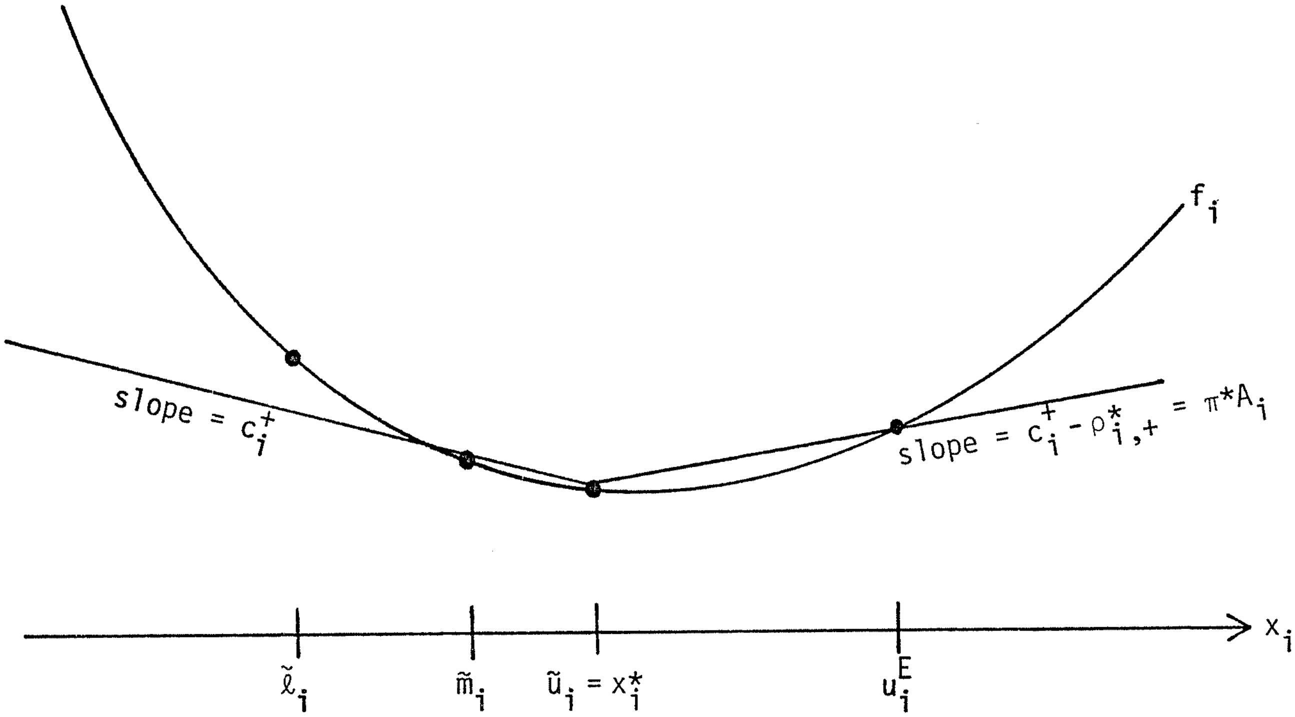

Figure 2. Construction of the new two-segment approximation f_i^E in the case $x_i^* = \tilde{u}_i$

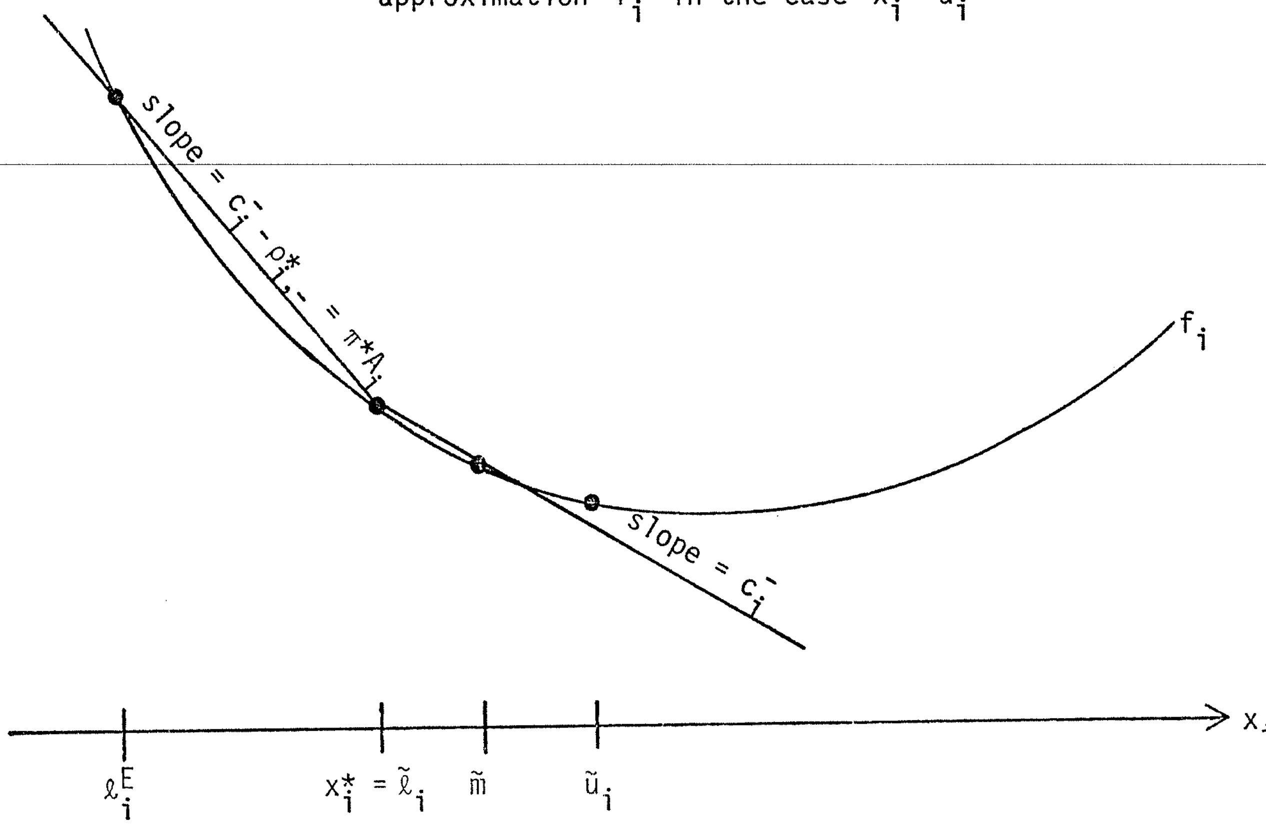

Figure 3. Construction of the new two-segment approximation f_i^E in the case $x_i^* = \tilde{\ell}_i$

$\tilde{f}_i^L(x_i)$ (with slope c_i^-) as its right segment, and $\tilde{f}_i^L(x_i) - \rho_{i,-}^*(x_i-\tilde{\ell}_i)$ as its left segment. If this construction is carried out for each variable that is artificially bounded in the optimal solution, then x^* is still optimal for the resulting set of two-segment approximations, but x^* is no longer artificially bounded with respect to the bounds associated with the new approximations, so the previous error analysis applies.

There is an interesting relationship between the estimation of the maximum of the approximation error $\tilde{f}_i - f_i$ and the solution of a Lagrangian relaxation of the original problem. This relationship will be described in the next section.

6. Lagrangian Relaxation

An alternative approach for the construction of error bounds and initial bounds for the contraction search is based upon Lagrangian relaxation. For every m-vector π we may define a Lagrangian relaxation $P_\pi(x)$ by multiplying the equality constraints of (1.1) by π and subtracting this product from the objective:

$$P_\pi(x) \equiv \begin{cases} \min\limits_x \; f(x) - \pi(Ax-b) \\ \text{s.t. } \ell \leq x \leq u \end{cases}$$

We denote by $L(x,\pi)$ the objective function $f(x) - \pi(Ax-b)$ and by $\omega(\pi)$ the optimal value of $P_\pi(x)$, and observe that, for any π, $\omega(\pi) \leq f(x^{**}) - \pi(Ax^{**}-b) = z^{**}$, so that $\omega(\pi)$ is a lower bound for the optimal value z^{**}. From duality theory (see, for example, [17] it is easily shown that there exists a π^{**} such that $\omega(\pi^{**}) = z^{**}$, and therefore "good" choices of π will provide tight lower bounds on z^{**}. We will refer to these as "dual" lower bounds to distinguish them from the "primal" lower bounds of the preceding section.

From a computational viewpoint, the solution of the approximating problem $P(\tilde{\ell},\tilde{m},\tilde{u})$ as an LP provides a set of optimal values π^* for the dual variables corresponding to the constraints $Ax = b$. Moreover, the separability of f implies that an optimal solution of $P_\pi(x)$ may be obtained by separately solving n one-dimensional optimization problems, since $L(x,\pi)$ is also separable and the constraints of $P_\pi(x)$ are simply bounds on the individual variables. In order to give a geometric interpretation to these problems, we define $s_i^* = \pi^* A^i$, where A^i is the ith column of A, and let $P_{\pi^*}(x_i)$ denote the problem

$$\min_{\ell_i \leq x_i \leq u_i} f_i(x_i) - (s_i^* x_i + k_i^*),$$

where k_i^* is a constant chosen so that $\tilde{f}_i(x_i^*) = s_i^* x_i^* + k_i^*$. (Observe that the value of the constant term in $P_{\pi^*}(x_i)$ has no effect on the set of optimal solutions of this problem.) For notational convenience we denote an optimal solution of $P_{\pi^*}(x_i)$ as $x_i^*(\pi^*)$ and the optimal value of that problem as $\omega_i(\pi^*)$. Note that $-\omega_i(\pi^*) = \max_{\ell_i \leq x_i \leq u_i} [(s_i^* x_i + k_i^*) - f_i(x_i)]$, and that $-\omega_i(\pi^*) \geq 0$ since the functions $s_i^* x_i + k_i^*$ and $\tilde{f}_i(x_i)$ agree at x_i^*. Thus, the value $-\omega_i(\pi^*)$ may be interpreted as the maximum amount (which must be non-negative) by which $s_i^* x_i + k_i^*$ over-estimates $f_i(x_i)$ on the interval $[\ell_i, u_i]$, and $x_i^*(\pi^*)$ is the point at which this maximum error occurs. The following two lemmas establish that the lower bound $\omega(\pi^*)$ may be written as $\tilde{f}(x^*) - \sum_{i=1}^{n} (-\omega_i(\pi^*))$, so that it is simply $\tilde{f}(x^*)$ minus the error bounds associated with the approximation of $f_i(x_i)$ by $s_i^* x_i + k_i^*$. We will then show that because of the optimality conditions corresponding to the problem $P(\tilde{\ell}, \tilde{m}, \tilde{u})$, these error bounds are not greater than those associated with $\tilde{f}$. Moreover, from the standpoint of reducing the error bound term, the use of $x_i^*(\pi^*)$ in constructing the approximation for the next iteration serves to reduce to 0 the approximation error at the point at which it was largest.

Lemma 6.1: $\pi^*(Ax-b) = \sum_{i=1}^{n} (s_i^* x_i + k_i^*) - \tilde{f}(x^*)$

Proof: By definition, $s_i^* = \pi^* A^i$, so that $\sum_{i=1}^{n} s_i^* x_i = s^* x = \pi^* Ax$, and thus the coefficients of x agree in the equation to be verified. Since x^* is feasible for (1.1), we have $Ax^* - b = 0$. This implies $\pi^*(Ax^*-b) = 0$, but, by definition, $0 = \sum_{i=1}^{n} (s_i^* x_i^* + k_i^* - \tilde{f}_i(x_i^*)) = \sum_{i=1}^{n} (s_i^* x_i^* + k_i^*) - \tilde{f}(x^*)$. ▲

Lemma 6.2: $\omega(\pi^*) = \sum_{i=1}^{n} \omega_i(\pi^*) + \tilde{f}(x^*)$

Proof: Clearly the vector $x^*(\pi^*)$ is an optimal solution of $P_{\pi^*}(x)$, so the relation is established by applying the preceding lemma to $\omega(\pi^*)$ as follows:

$$\omega(\pi^*) = f(x^*(\pi^*)) - \pi^*(Ax^*(\pi^*)-b) =$$

$$f(x^*(\pi^*)) - \sum_{i=1}^{n} (s_i^* x_i^*(\pi^*) + k_i^*) + \tilde{f}(x^*)$$

The result now follows from the definition of $\omega_i(\pi^*)$. ▲

The next theorem shows that the approximation error $-\omega_i(\pi^*)$ associated with $s_i^* x_i + k_i^*$ is not greater than that associated with $\tilde{f}_i$ on $[\tilde{\ell}_i, \tilde{u}_i]$.

Theorem 6.1: If x^* is an optimal solution of $P(\tilde{\ell}, \tilde{m}, \tilde{u})$ and π^* is a set of optimal values of the dual variables for the constraints $Ax = b$ in $P(\tilde{\ell}, \tilde{m}, \tilde{u})$, then $s_i^* x_i + k_i^* \leq \tilde{f}_i(x_i)$ for $x_i \in [\tilde{\ell}_i, \tilde{u}_i]$.

Proof: From the optimality of π^* it follows that x^* is also an optimal solution of the problem

$$\min_x \ \tilde{f}(x) - \pi^*(Ax-b)$$

$$\text{s.t. } \tilde{\ell} \le x \le \tilde{u}.$$

However, this implies that, for $i = 1,\ldots,n$, x_i^* is an optimal solution of

$$\min_{\tilde{\ell}_i \le x_i \le \tilde{u}_i} \tilde{f}_i(x_i) - (s_i^* x_i + k_i^*).$$

Since $\tilde{f}_i(x_i^*) - (s_i^* x_i^* + k_i^*) = 0$ by construction, it follows that $\tilde{f}_i(x_i) - (s_i^* x_i + k_i) \ge 0$ for all $x_i \in [\tilde{\ell}_i, \tilde{u}_i]$ from which the result follows. ▲

Corollary 6.1: If $x_i^* \in (\tilde{\ell}_i, \tilde{m}_i)$, then $s_i^* x_i + k_i^* = \tilde{f}_i^L(x_i)$. If ~~$x_i^* \in (\tilde{m}_i, \tilde{\ell}_i)$, then $s_i^* x_i + k_i^* = \tilde{f}_i^U(x_i)$.~~

Proof: If $x_i^* \in (\tilde{\ell}_i, \tilde{m}_i)$, then the linear function $s_i^* x_i + k_i^*$ dominates $\tilde{f}_i^L(x_i)$ on $[\tilde{\ell}_i, \tilde{m}_i]$, but it also coincides with $\tilde{f}_i^L(x_i)$ at a point in the interior of that interval, so the two functions must agree. The analogous result holds in the $\tilde{f}_i^U$ case. ▲

From a geometric viewpoint, the possible relationships between $s_i^* x_i + k_i^*$ and $\tilde{f}_i$ correspond (in the differentiable case) to the eight cases illustrated in Figures 4-11.

Note that in cases 3a-3d in which $x_i^* = \tilde{\ell}_i$ or $x_i = \tilde{u}_i$, the function $s_i^* x_i + k_i^*$ coincides with the extension f_i^E for $x_i < \tilde{\ell}_i$

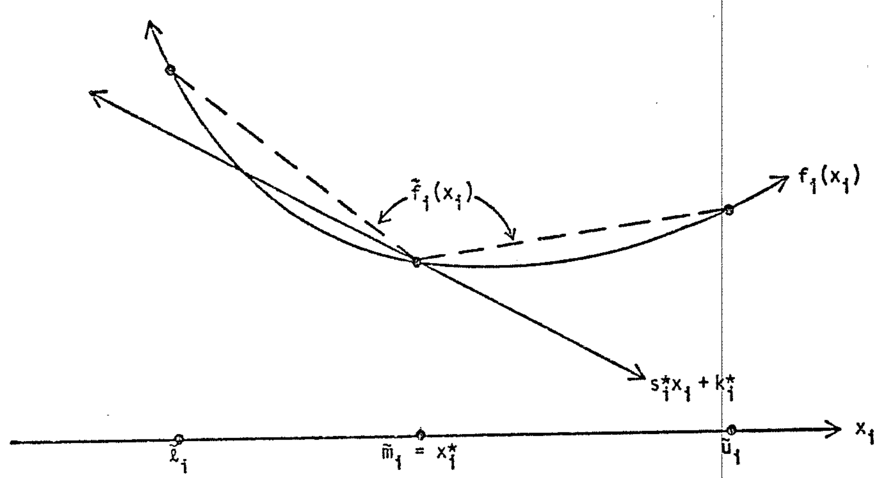

Figure 4. Case 1a: $x_i^* = \tilde{m}_i \epsilon (\tilde{\ell}_i, \tilde{u}_i)$, $s_i^* < f_i'(x_i^*)$

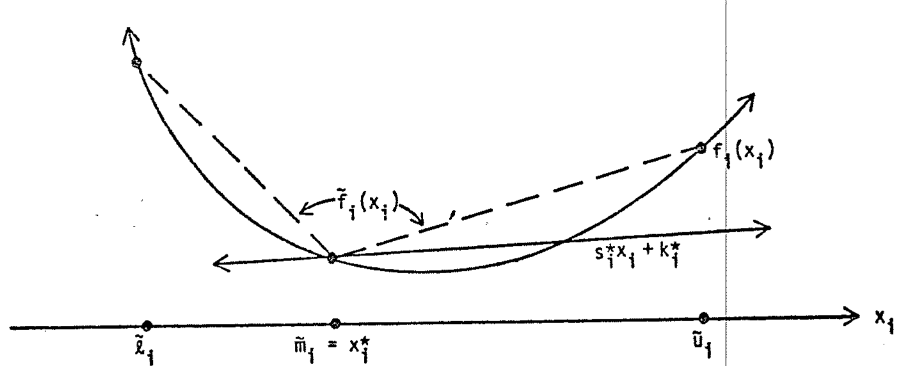

Figure 5. Case 1b: $x_i^* = \tilde{m}_i \epsilon (\tilde{\ell}_i \tilde{u}_i)$, $s_i^* > f_i'(x_i^*)$

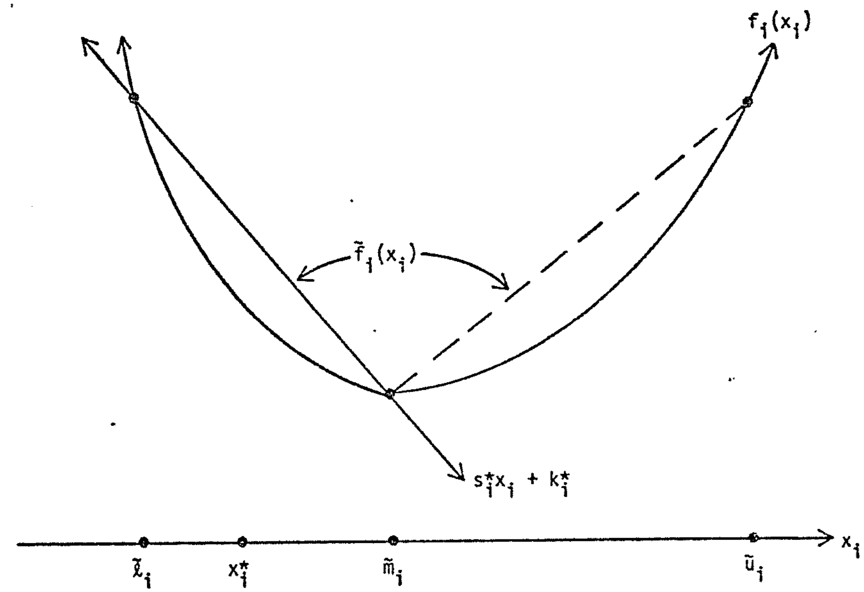

Figure 6. Case 2a: $\tilde{\ell}_i < x_i^* < \tilde{m}_i$

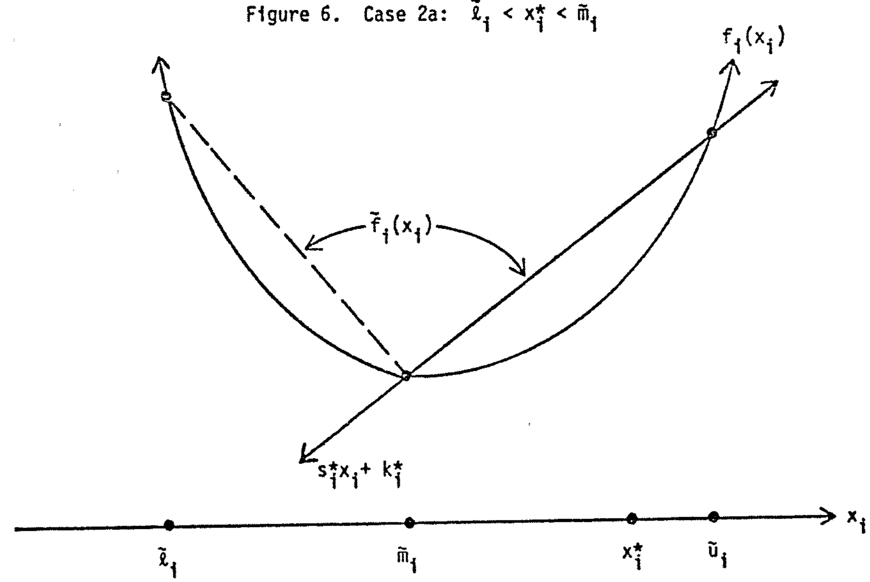

Figure 7. Case 2b: $\tilde{m}_i < x_i^* < \tilde{u}_i$

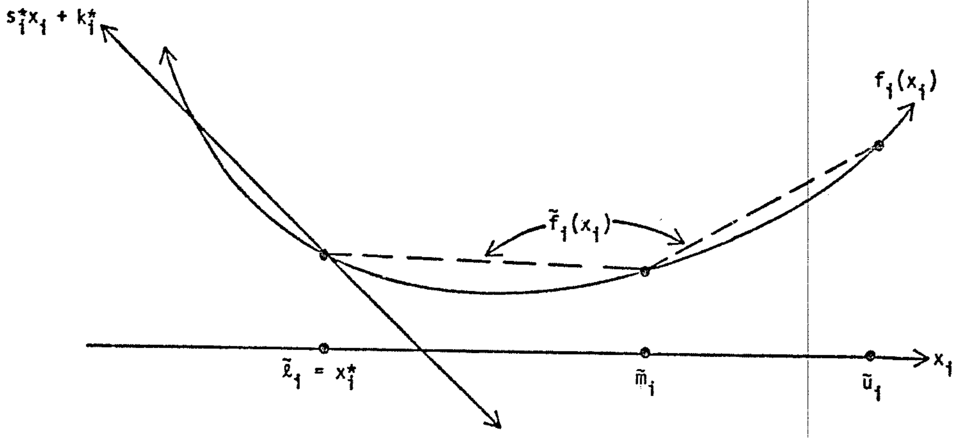

Figure 8. Case 3a: $x_i^* = \tilde{\ell}_i$, $s_i^* < f_i'(x_i^*)$

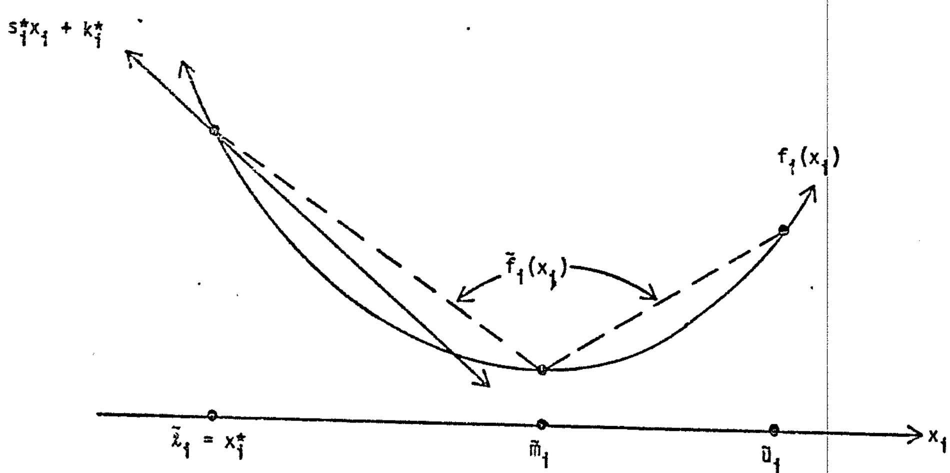

Figure 9. Case 3b: $x_i^* = \tilde{\ell}_i$, $s_i^* > f_i'(x_i^*)$

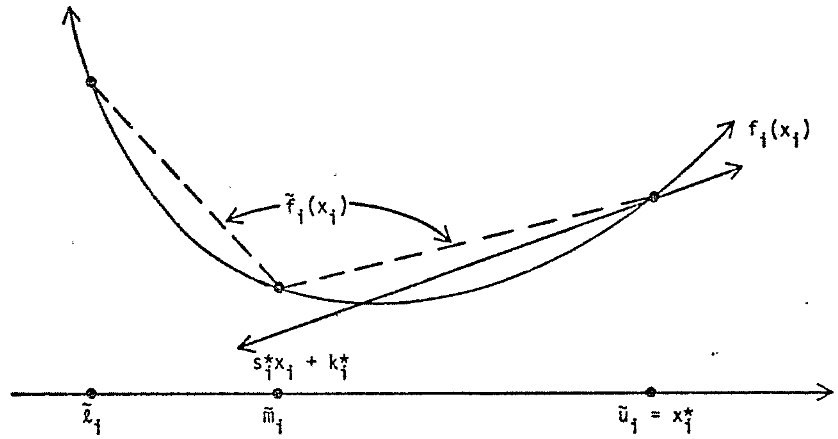

Figure 10. Case 3c: $x_i^* = \tilde{u}_i$, $s_i^* < f_i'(x_i^*)$

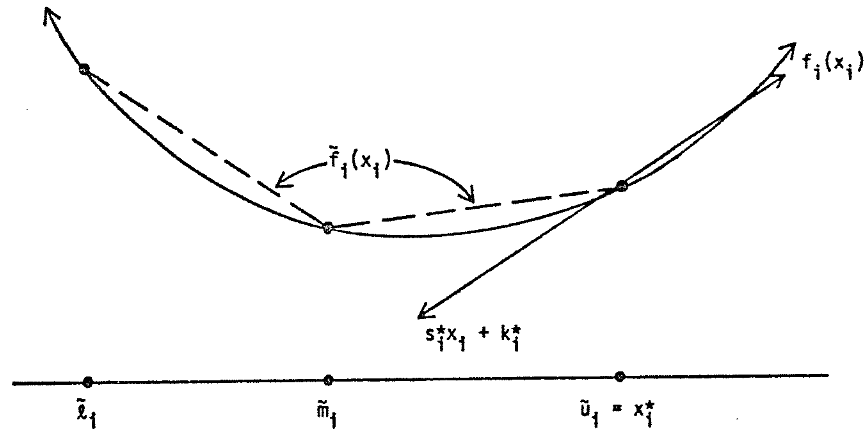

Figure 11. Case 3d: $x_i^* = \tilde{u}_i$, $s_i^* > f_i'(x_i^*)$

or $x_i > \tilde{u}_i$ respectively, so that the dominance property carries over to these cases also. (In the case of non-differentiable f_i, convexity implies the existence of finite left and right (directional) derivatives for $x_i \in (\ell_i, u_i)$, and obvious extensions of the differentiable cases may be made.) It should also be observed that in cases 3a or 3d, if the active bound coincides with an original bound $(\ell_i$ or $u_i)$, then the error term $\omega_i(\pi^*)$ will be 0. (The error term will also be 0 if $x_i^* = \tilde{m}_i \in (\tilde{\ell}_i, \tilde{u}_i)$ and $s_i^* = f_i'(x_i^*)$, but this case would be exceptional.) In many practical problems the f_i are differentiable, and $P_{\pi^*}(x_i)$ may be solved by letting $x_i^*(\pi^*)$ be a point satisfying $f_i'(x_i) = s_i^*$ if such a point exists within $[\ell_i, u_i]$, and otherwise letting $x_i^*(\pi^*) = \ell_i$ or u_i according to the value that maximizes the error function.

In addition to yielding a lower bound on z^{**}, the solutions $x_i^*(\pi^*)$ may be used to provide the initial bounds for the next iteration. If $x_i^*(\pi^*) < x_i^*$, then $x_i^*(\pi^*)$ is used as the initial lower bound, and if $x_i^*(\pi^*) > x_i^*$, then $x_i^*(\pi^*)$ is used as the initial upper bound, with the remaining bound being located symmetrically with respect to x_i^*. In this way, information from both the primal and dual problems may be taken into account in constructing the next approximating problem. (From another point of view, an approximation constructed in this manner combines information from both feasible (i.e., x^*) and infeasible (i.e., $x^*(\pi^*)$) solutions, since $x^*(\pi^*)$ is infeasible unless it is optimal for (1.1).) In practice, tolerances based on the error bound are imposed in order to safeguard against having $x_i^*(\pi^*)$ too close or too far from x_i^*.

7. Line Searches

One key feature of the algorithm as presented is that it does not require a line search in order to determine a better feasible solution. Because of the dominance property of the approximations, the solution x^* of $P(\tilde{\ell},\tilde{m},\tilde{u})$ is guaranteed to satisfy $f(x^*) < f(\tilde{m})$ provided that $x^* \neq \tilde{m}$. However, in certain instances it has been found helpful to use a line search in order to obtain further improvements in the objective value. Note that $f(x^*) - f(\tilde{m}) < 0$ implies that the direction $(x^*-\tilde{m})$ is a descent direction when f is differentiable and convex. Thus, a line search may be carried out in the direction $(x^*-\tilde{m})$ to determine the best value of f on $L \equiv \{x | x \in S, x=\tilde{m}+\lambda(x^*-\tilde{m}), \lambda>0\}$. Since $A(x^*-\tilde{m}) = 0$, the constraints $Ax = b$ need not be taken into account in determining L, and only the bounds ℓ, u will serve to limit λ. The line search has the nice property of compensating for overly restrictive temporary bounds. Moreover, if the final three values of λ used in the line search are λ', λ'', and λ''' with $\lambda' < \lambda'' < \lambda'''$ and $f(\tilde{m}+\lambda''(x^*-\tilde{m})) < f(\tilde{m}+\lambda'(x^*-\tilde{m}))$ and $f(\tilde{m}+\lambda''(x^*-\tilde{m})) < f(\tilde{m}+\lambda'''(x^*-\tilde{m}))$, the values of $\tilde{m} + \lambda'(x^*-\tilde{m})$ and $\tilde{m} + \lambda'''(x^*-\tilde{m})$ may be used in the obvious manner in setting up the initial bounds for the next iteration. (In the case of 0 components of $(x^*-\tilde{m})$, the procedures previously described would be used to set up bounds.) In some preliminary studies made with the numerical test problems to be described in the next section, the additional function value improvement associated with the use of the line search was too small to justify the additional computational effort. However, the line search option still holds some promise with regard to strategies that do not fully solve each subproblem (see section 9), and line searches have proved effective in the extension of this piecewise-linear approximation approach to the case of non -separable objectives as described in [10].

8. A Comparison of Numerical Results

In this section we present results to show the effectiveness of the strategies on a variety of test problems. Test problem set 1 is a collection of problems arising from an application in statistics [10, 19]. The second test problem set contains some of the econometric modelling problems described in [1]. The third is a water supply system application [5,6] and the fourth is a set of optimal control problems given in [3].

The following notation is used in the tables below:

I = total number of (major) iterations

n = number of variables

m = number of linear constraints (excluding bounds)

AS1 = objective value of the feasible solution found with adaptive strategy 1

AS2 = objective value of the feasible solution found with adaptive strategy 2

LR = objective value of the feasible solution using Lagrangian relaxation for construction of the initial bounds at each iteration

PLB2 - lower bound on the optimal value computed using the primal approach and adaptive strategy 2

LLB - Lagrangian lower bound

[Test Problem Set 1]

The objective functions for these test problems contain terms of the form $x_i \log x_i$ and are thus non-differentiable for $x_i = 0 = \ell_i$. A complete statement of the problems is given in [10].

Case	(m,n)	I	AS1	AS2	PLB2
1	(5,6)	11	.263949	.26394222	.26394222
2	(6,8)	10	.167263	.167257560	.167257558
3	(8,12)	16	.149422	.149409878	.149409876

Table 1. Summary of results for test problem set 1

[Test Problem Set 2]

These problems are quadratic transportation problems of the type described in [1]. The problem format is

$$\min_{x} \sum_{i,j} (x_{i,j}-t_{i,j})^2$$

$$\text{s.t.} \sum_{i=1}^{k} x_{i,j} = d_j \qquad (j=1,\ldots,k)$$

$$\sum_{j=1}^{k} x_{i,j} = s_i \qquad (i=1,\ldots,k)$$

$$x_{i,j} \geq 0,$$

where the $t_{i,j}$ are constants giving "target" flows on the arcs, d_j is the demand at node j, and s_i is the supply available at node i.

The initial problem in this group was constructed to have a known optimal solution given in [13]; the others are based on real econometric data used by Bachem and Korte. Except for the first problem, the adaptive strategies proved inferior to the Lagrangian relaxation approach, so only results for the latter are reported. The superiority of the Lagrangian method also held for the remaining test problems.

Case	(m,n)	I	LR	LLB
1	(20,100)	13	6600.0001	6599.9998
2	(24,144)	13	7302243.46	7302243.17
3	(24,144)	12	2714896.45	2714895.60
4	(109,2202)	13	5634201.64	5634200.88
5	(109,2238)	13	12022985.19	12022984.64

Table 2. Summary of results for test problem set 2

[Test Problem Set 3]

The linear constraints for these hydraulic equilibrium problems are network constraints, so a network optimization code was used to solve the subproblems. The largest problem in this set has 906 variables (n=906) and 666 constraints (m=666), 18 linear objective terms, and 888 nonlinear objective terms (mostly of the form $c_i|x_i|^{2.85}$). Further details are given in [5] and [6].

Case	(m,n)	I	LR	LLB
1	(30,46)	15	-32392.730	-32392.731
2	(150,196)	17	-48199.858	-48199.864
3	(666,906)	23	-206175.21	-206175.67

Table 3. Summary of results for test problem set 3

[Test Problem Set 4]

This set of optimal control problems for a reservoir is described in [3]. Water release from the reservoir is to be scheduled so as to come as close as possible to certain target figures. There are periodic inputs to the reservoir and bounds on the total volume of water in the reservoir. Two types of objective functions are considered and the corresponding results are presented in Table 4. The case numbers without primes correspond to quadratic objective functions, while the primed numbers correspond to problems with the same constraints but exponential objective functions.

Case	(m,n)	I	LR	LLB
1	(13,23)	11	-1975.6491	-1975.6492
1′	(13,23)	12	12.6412	12.6411
2	(53,103)	11	-8731.0258	-8731.0264
2′	(53,103)	12	56.5603	56.5594
3	(105,207)	10	-17393.553	-17393.558
3′	(105,207)	13	124.758	124.757
4	(366,729)	11	-60750.488	-60750.491
4′	(366,729)	17	476.266	476.265

Table 4. Summary of results for test problem set 4

9. Directions for Further Research

Although excellent computational experience has been obtained with the algorithms in their present form, there are a number of ideas under study that may further improve efficiency. One obvious strategy is to terminate the solution of the subproblems $P(\tilde{\ell},\tilde{m},\tilde{u})$ prior to optimality, particularly in the initial major iterations in which these problems require numerous pivots. In this case the termination criterion for the early subproblems could be a fixed number of pivots or a tolerance on the reduced cost of candidates to enter the basis or a combination of these two strategies. Such a tolerance would avoid pivots that would have only a marginal effect on the objective function value in favor of pivots (in the next major iteration) with a more significant effect. In later major iterations the termination criterion could be the achievement of a certain percentage reduction of the error bound. Note that the limiting case of this strategy would be the use of only one pivot per iteration (except for the first iteration, in which this approach would be postponed until feasibility had been attained). While it would probably be inefficient in this limiting case to calculate new objective function values for all variables after each pivot, the algorithm could be further modified by calculating a new value for a variable only if the variable is driven to a bound $\tilde{\ell}_i \neq \ell_i$ or $\tilde{u}_i \neq u_i$ as a result of the pivot, in which case the new evaluation of f_i would allow the variable to continue its change past the temporary bound $\tilde{\ell}_i$ or $\tilde{u}_i$ (provided that the variable corresponding to the new segment of the cost function priced out with the proper sign). This approach is equivalent to what might be termed an "implicit grid" strategy in which the function values at the implicit grid points are calculated only as needed when a variable reaches the limits of its initial range $[\tilde{\ell}_i, \tilde{u}_i]$. (With regard to the calculation of function values, it should also be noted that, in the case that a subproblem is not to be solved to optimality, the only objective coefficients required to apply the primal simplex method are those of the basic variables and those non-basics to which the pricing out operation is applied. In other words, it is possible to avoid many of the additional function evaluations that would otherwise be required to compute cost coefficients for non-basics.)

Along similar lines, note that if, as a result of a pivot, a variable x_i^+ has been driven to 0, the next variable to be priced out should be x_i^-, since there is a good possibility that a further decrease in x_i may lead to additional improvement of the objective function. (Analogous observations apply if x_i^- is driven to 0.) Moreover, the pricing out operation as applied to x_i^+ yields the "reduced cost" of x_i^- as well except that the sign must be changed and the term $c_i^+ - c_i^-$ must be added. (This observation also establishes that when a variable $x_i^+(x_i^-)$ is basic, its complement $x_i^-(x_i^+)$ will price out in such a way that it will not be a candidate to enter the basis, so that those non-basics need not be priced out.) Thus, algorithms with special provisions for taking these pricing strategies into account (such as the GNET code of [4] and the algorithm described in [16]) should be considerably more efficient for this problem class than those that do not take advantage of the problem structure.

In a strategy in which the subproblems are not solved to optimality, the use of a line search could be expected to be of greater value than it is in the current strategies. Note that convexity guarantees that any point yielding an improvement in the approximating function furnishes a descent direction.

For special problem classes such as those of test problem set 2, the ability of the algorithm to utilize both feasible and infeasible points in the construction of the objective approximations could be further exploited in the initial iteration. For example, in those quadratic transportation problems an initial feasible solution may be easily generated that is relatively "close" to the target flows $t_{i,j}$. This feasible solution could be used as the initial $\tilde{m}$, and the $t_{i,j}$ themselves could serve as initial values for the $\tilde{\ell}_i$ or $\tilde{u}_i$ (depending on whether they were above or below the values $\tilde{m}_i$).

The idea of using local piecewise-linear approximation may, of course, be easily extended to non-separable objective functions as described in [10] and to nonlinear constraints as well. However, the convergence properties of such algorithms in the nonlinear constraint case are still under study. Note also that many of the ideas dealing with primal and dual error bounds also carry over to the non-separable case. Further details and computational experience will be given in [14].

Acknowledgements

The computational results in this paper were obtained by C. Y. Kao and P. Kamesam. The RNET minimum cost network flow subroutines (see [8]) used to solve the linear network subproblems were provided by M. D. Grigoriadis.

References

1. A. Bachem and B. Korte, "Quadratic programming over transportation polytopes", Report No. 7767-OR, Institute für Okonometrie and Operations Research (Bonn, 1977).

2. E. M. L. Beale, Mathematical programming in practice (Sir Isaac Pitman & Sons Ltd., London, 1968).

3. D. P. Bertsekas, "On the Goldstein-Levitin-Poljak gradient projection method", IEEE Transactions on Automatic Control 21 (1976) 174-184.

4. G. H. Bradley, G. G. Brown, and G. W. Graves, "Design and implementation of large scale primal transshipment algorithms", Management Science 24 (1977) 1-34.

5. M. Collins, L. Cooper, R. Helgason, J. Kennington, and L. LeBlanc, "Solving the pipe network analysis problem using optimization techniques", Management Science 24 (1978), 747-760.

6. R. Dembo and J. Klincewicz, "An approximate second order algorithm for network flow problems with convex separable costs", Series #21, School of Business, Yale University (New Haven, Conn., 1978).

7. A. M. Geoffrion, "Objective function approximations for mathematical programming", Mathematical Programming 13 (1977) 23-27.

8. M. D. Grigoriadis and T. Hsu, "RNET: The Rutgers minimum cost network flow subroutines", SIGMAP Bulletin (1979) 17-18.

9. C. Y. Kao, private communication, 1979.

10. C. Y. Kao and R. R. Meyer, "Secant approximation methods for convex optimization", Mathematical Programming Study 14 (1981) 143-162.

11. R. R. Meyer, "A class of nonlinear integer programs solvable by a single linear program", SIAM Journal on Control and Optimization 15 (1977) 935-946.

12. R. R. Meyer, Two-segment separable programming, Management Science 25 (1979) 385-395.

13. R. R. Meyer, "Computational aspects of two-segment separable programming", University of Wisconsin-Madison Computer Sciences Department Technical Report #382 (Madison, WI, 1980).

14. R. R. Meyer, "Piecewise-linear approximation methods for optimization problems", University of Wisconsin-Madison Computer Sciences Department Technical Report, in preparation, (Madison, WI, 1982).

15. R. R. Meyer and M. L. Smith, "Algorithms for a class of 'convex' nonlinear integer programs". in: W. W. White ed., Computers and mathematical programming (National Bureau of Standards Special Publication 502, Washington, 1978).

16. H. Müller-Merbach, "Die methode der 'direkten Koeffizientanpassung' (μ-Form) des separable programming", Unternehmensforschung, Band 14, Heft 3 (1970) 197-214.

17. R. T. Rockafellar, Optimization in networks, Lecture notes, Department of Mathematics, University of Washington (Seattle, WA, 1976).

18. T. L. Saaty, Optimization in integers and related extremal problems (McGraw-Hill, New York, 1970).

19. J. Z. Teng, "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, WI, 1978).

20. L. S. Thakur, "Error analysis for convex separable programs: The piecewise linear approximation and the bounds on the optimal objective value", SIAM Journal on Applied Mathematics 34 (1978) 704-714.