TRUST REGION METHODS FOR
PIECEWISE-LINEAR APPROXIMATION

by

R. R. Meyer

Computer Sciences Technical Report #626

December 1985

TRUST REGION METHODS FOR PIECEWISE-LINEAR APPROXIMATION*

R. R. Meyer
Computer Sciences Department and Mathematics Research Center
The University of Wisconsin-Madison
Madison, Wisconsin, USA 53706

Abstract

Trust region methods are analyzed for piecewise-linear approximation algorithms for linearly-constrained nonlinear optimization. Convergence to the optimal value is demonstrated for continuously differentiable convex objectives and for certain classes of nondifferentiable convex objectives. Computationally, this approach has the nice property that the approximation is generally more accurate than a linear approximation yet the subproblems to be solved at each iteration remain linear programs. The method is also well-suited to convex network optimization, since it preserves the network structure of the subproblems, thereby allowing the use of the very fast techniques available for linear network optimization. For problem classes such as traffic assignment in which the critical coupling between variables occurs in the objective function, the separability of the approximation makes possible a decomposition into independent subproblems that may be solved in parallel in a multiprocessing environment.

* This research was supported in part by NSF grant DCR8503148 and ARO contract DAAG29-80-C-0041 .

1. Introduction

We consider trust region methods for linearly-constrained nonlinear optimization problems of the form

$$\text{(NLP)} \qquad \begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & Ax \le b, \end{aligned}$$

where we assume initially that f is a continuosly differentiable convex function on the feasible set $X := \{x | Ax \le b\} \subset \Re^n$ and A is an $m \times n$ matrix. (The differentiability and convexity hypotheses will be relaxed in sections 4 and 5).

The algorithms to be considered utilize a separable convex, piecewise-linear approximation to f (note that f itself is *not* assumed to be separable). Section 2 develops this approximation and contrasts it with trust region methods based on linear and quadratic approximation The algorithm as described in section 3 involves the construction at each iteration of such an approximation to f in a suitably chosen neighborhood of the current feasible point. The choice of neighborhood and the acceptance or rejection of a trial point are determined by the ratio of the improvement in the original objective function over the improvement in the approximating function. Under weak assumptions it is shown that the objective value of the iterates converges to the optimal value. The proof extends some of the techniques of [Zhang, et al 84] , which considered an algorithm based on linear approximation.

2. Piecewise-linear approximation

We first consider the construction of a separable piecewise-linear approximation "centered" at a feasible point $\bar{x} \epsilon X$. We define a "shifted" function $h(d)$, where d denotes the difference between $\bar{x}$ and its potential successor and h is the difference in their function values:

$$h(d) := f(d + \bar{x}) - f(\bar{x}).$$

Note that $d = 0$ thus corresponds to $\bar{x}$, with $h(0) = 0$, and $h(d) \le 0$ if $d + \bar{x}$ has a lower objective function value than $\bar{x}$. We now define a *separable* approximation of h by first considering the *single-variable functions* obtained by fixing (at 0) all variables but one (in this definition, e^j is the j^{th} unit vector):

$$s_j(d_j) := h(d_j e^j) \qquad (j = 1, \ldots, n).$$

Note that s_j thus corresponds to the restriction of h to the d_j-axis, and the convexity of f implies the convexity of d_j. A separable approximation of h is then given by summing these single-variable functions:

$$s(d) := \sum_{j=1}^{n} s_j(d_j).$$

For computational purposes, we carry out one final approximation step by first defining $\tilde{s}_j(d_j)$ $(j = 1, \ldots, n)$ to be the convex *piecewise-linear* function obtained by linearly interpolating between the values of s_j at a set of points (to be defined below) on the d_j-axis, and then letting

$$\tilde{h}(d) := \sum_{j=1}^{n} \tilde{s}_j(d_j).$$

Note that $\tilde{h}$ is a separable, convex, piecewise-linear function with $\tilde{h}(0) = 0$, and that $\tilde{h}$ may be thought of as an approximation of the separable function s.

This type of approximation may be contrasted with the linear approximation trust region algorithm of [Zhang, et al 85] and the quadratic approximation techniques of [Fletcher 81] and [Yuan 85]. In terms of accuracy and complexity of the approximation, it lies in between these two approaches. Its main disadvantage, as with the linearization approach of [Zhang, et al 85] is that it is separable and therefore does not contain any "cross terms" that take into account linkage between variables in the objective. The principal advantage is that it is generally more accurate (particularly on the translated axes, where it even enjoys an accuracy advantage over quadratic approximations) than linear approximation, and yet still yields subproblems that may be solved by linear programming. It should also be noted that the step-bound constraints of trust region algorithms may easily be incorporated into the piecewise-linear approach via bounds on the size and number of segments of the approximation, whereas these bounds lead to additional computational overhead with respect to the other approximations. Computational experience cited in [Feijoo 85] and [Feijoo and Meyer, 84,85] indicates that algorithms utilizing piecewise-linear approximation converge in many fewer iterations than those based on linear approximation.

At each iteration of the algorithm to be defined below, we solve a *piecewise-linear approximating problem* of the form:

$$PL(\bar{x}, \alpha) \qquad \begin{array}{ll} \min\limits_{d} & \tilde{h}(d) \\ s.t. & Ad \leq \bar{b}, \quad \|d\| \leq \alpha, \end{array}$$

where α is the so-called *trust region parameter* (α generally changes at each iteration), $\bar{b} := b - A\bar{x}$, and $\|d\|$ represents the ℓ_∞ norm. Observe that $Ad \leq \bar{b}$ is equivalent to $A(\bar{x} + d) \leq b$, so that a feasible solution of $PL(\bar{x}, \alpha)$ will also be *feasible* for the original problem (NLP). The problem $PL(\bar{x}, \alpha)$ may be converted into an equivalent *linear program* by the standard substitutions associated with separable programming. In this regard, the augmentation of the original constraints by the norm constraint $\|d\| \leq \alpha$ actually serves to reduce computational effort by restricting the number of segments employed in the piecewise-linear approximation of the objective function. This norm constraint in fact need not be explicitly considered, since the use of the δ-formulation of separable programming decomposes each variable d_j into a set of of bounded variables (each of which corresponds to a segment of the piecewise-linear approximation $\tilde{s}_j$), and the bounds on these variables may be chosen so as to guarantee satisfaction of the norm constraint. See section 6 for an explicit construction of the corresponding linear program.

The *grid points* defining the piecewise-linear functions $\tilde{s}_j$ will include the point 0 as well as additional grid points spaced not more than α apart (see Figure 1). A key property that will be used in the validity proof of the algorithm is that for $\bar{x}$ *non-optimal* for (NLP), the choice of an α *sufficiently small* will ensure that an optimal solution of $PL(\bar{x}, \alpha)$ will have a value for f smaller than $f(\bar{x})$. (Since $d = 0$ is feasible for $PL(\bar{x}, \alpha)$, it follows that the *optimal value* of $PL(\bar{x}, \alpha)$, denoted by $\tilde{h}^*(\alpha)$, satisfies $\tilde{h}^*(\alpha) \leq 0$, with $\tilde{h}^*(\alpha) = 0$ if and only if $d = 0$ is optimal for that problem. We will show that an improvement in the value of the approximation $\tilde{h}$ translates into a improvement in the "true" objective f for sufficiently small α). More precisely, for an optimal solution $d^*(\alpha)$ of $PL(\bar{x}, \alpha)$ such that $\tilde{h}(d^*(\alpha)) < 0$, we will be concerned with the *ratio* of the improvement of the "true" objective function h to the improvement in the approximating function $\tilde{h}$, i.e., for $\tilde{h}^*(\alpha) < 0$, we define $h^*(\alpha) := h(d^*(\alpha))$ and $\tilde{h}^*(\alpha) := \tilde{h}(d^*(\alpha))$ and consider $r^*(\alpha) := h^*(\alpha)/\tilde{h}^*(\alpha)$ (the notation here conceals the fact that the value of $h^*(\alpha)$ may vary among alternative optima for $PL(\bar{x}, \alpha)$, but this factor will be accounted for since the bounds to be derived are independent of the particular optimal solution considered). In order to develop conditions that will ensure that the *improvement ratio* $r^*(\alpha)$ is at least ρ_0 (a positive parameter to be specified in the algorithm), we will now consider optimality conditions for (NLP) .

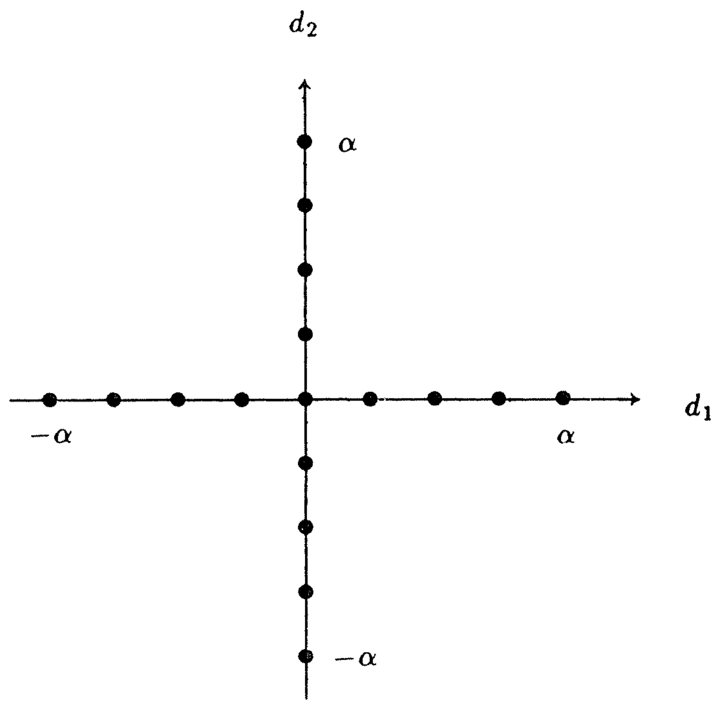

Figure 1: Grid points in two-dimensional case

First-order optimality conditions for (NLP) may be developed in terms of the problem obtained by linearizing the objective at $\bar{x}$:

$$L(\bar{x}, \alpha) \qquad \begin{array}{ll} \min\limits_{d} & \nabla f(\bar{x}) d \\ s.t. & Ad \le \bar{b}, \quad \|d\| \le \alpha. \end{array}$$

By convexity, $\bar{x}$ is an *optimal* solution of (NLP) if and only if $d = 0$ is an optimal solution of $L(\bar{x}, \alpha)$ for every $\alpha \ge 0$. An *optimal* solution of $L(\bar{x}, \alpha)$ will be denoted as $d_L^*(\alpha)$. For notational convenience we define $h_L(d) := \nabla f(\bar{x}) d$ and the *optimal value* of $L(\bar{x}, \alpha)$ as $h_L^*(\alpha)$.

Lemma 1: Let the grid size for $\tilde{h}(d)$ be no greater than α and let $\|d\| \le \alpha$. Then the following *error bounds* hold (and are independent of the choice of $\bar{x} \epsilon X$ in any fixed bounded set):

(a) $h(d) - h_L(d) = o(\|d\|) = o(\alpha)$
(b) $s(d) - h_L(d) = o(\alpha)$
(c) $s(d) - \tilde{h}(d) = O(\alpha^2)$
(d) $h_L(d) - \tilde{h}(d) = o(\alpha)$
(e) $h(d) - \tilde{h}(d) = o(\alpha)$
(f) $h_L^*(\alpha) - \tilde{h}^*(\alpha) = o(\alpha)$

Pf: (a) Follows from differentiability of f.
(b) Follows from $\nabla s(0) = \nabla h(0)$.
(c) This is a property of piecewise-linear approximations of separable functions; see, e.g., [Feijoo 85].
(d) Follows from (b) and (c).
(e) Follows from (a) and (d).
(f) By part (d) and the definitions of the terms,

$$\tilde{h}^*(\alpha) \le \tilde{h}(d_L^*(\alpha)) = h_L(d_L^*(\alpha)) + o(\alpha) = h_L^*(\alpha) + o(\alpha).$$

However, since $h_L(d) \le \tilde{h}(d)$, it follows that $h_L^*(\alpha) \le \tilde{h}^*(\alpha)$, and the combined inequalities $h_L^*(\alpha) \le \tilde{h}^*(\alpha) \le h_L^*(\alpha) + o(\alpha)$ yield (f). ∎

The following lemma establishes some useful convexity properties of the optimal value functions.

Lemma 2: For $\lambda \epsilon [0,1]$,

$$h_L(d_L^*(\lambda\alpha)) \leq h_L(\lambda d_L^*(\alpha)) \leq \lambda h_L(d_L^*(\alpha))$$

and

$$\tilde{h}(d^*(\lambda\alpha)) \leq \tilde{h}(\lambda d^*(\alpha)) \leq \lambda \tilde{h}(d^*(\alpha)).$$

Pf: The left inequality for $\tilde{h}$ follows from the feasibility of $\lambda d^*(\alpha)$ for the trust region corresponding to $\lambda\alpha$. The right inequality follows from $d^*(0) = 0$ and the convexity of $\tilde{h}$. The arguments for h_L are analogous. ∎

We now show that an improved value for the approximation $\tilde{h}$ may be obtained by solving $PL(\bar{x}, \alpha)$.

Lemma 3: If $\bar{x}$ is *not* an optimal solution of (NLP), then for all α sufficiently small, $\tilde{h}^*(\alpha) < 0$.

Pf: Suppose there exists a positive sequence $\alpha_i \to 0$ such that $\tilde{h}^*(\alpha_i) = 0$ for all i. Since $\tilde{h}^*(\alpha) = h_L^*(\alpha) + o(\alpha)$ by Lemma 1, setting $\alpha = \alpha_i$, dividing through by α_i, and taking limits, yields $h_L^*(\alpha_i)/\alpha_i \to 0$. However, since $\bar{x}$ is non-optimal, for any fixed $\alpha \geq \max_i \alpha_i$ we have $h_L^*(\alpha) < 0$ and, by Lemma 2, $h_L^*(\frac{\alpha_i}{\alpha} \cdot \alpha) \leq \frac{\alpha_i}{\alpha} h_L^*(\alpha)$. Dividing this inequality by α_i yields $h_L^*(\alpha_i)/\alpha_i \leq h_L^*(\alpha)/\alpha$, contradicting the convergence of the left-hand-side terms to 0. ∎

By using convexity properties and error bounds, improvement in the approximating function $\tilde{h}$ may be related to improvement in the true objective h.

Lemma 4: If for some $\tilde{\alpha}$, $\tilde{h}^*(\tilde{\alpha}) < 0$, then for any $\rho < 1$, the inequality $r^*(\lambda\tilde{\alpha}) \geq \rho$ holds for all sufficiently small $\lambda > 0$.

Pf:

$$\begin{aligned} r^*(\lambda\tilde{\alpha}) = h^*(\lambda\alpha)/\tilde{h}^*(\lambda\alpha) = & \\ \frac{\tilde{h}^*(\lambda\tilde{\alpha}) + o(\lambda\tilde{\alpha})}{\tilde{h}^*(\lambda\tilde{\alpha})} &= 1 + o(\lambda\tilde{\alpha})/\tilde{h}^*(\lambda\tilde{\alpha}) \\ &\geq 1 + o(\lambda\tilde{\alpha})/\lambda\tilde{h}^*(\tilde{\alpha}). \end{aligned}$$

The result follows by noting that the last ratio tends to 0 as $\lambda \to 0$. ∎

In order to guarantee that the improvement ratio behaves properly in the neighborhood of $\bar{x}$, we now prove that $\tilde{h}^*$ is continuous.

Lemma 5: $\tilde{h}^*$ is a continuous function of x and α for $x \epsilon X$ and $\alpha \geq 0$.

Pf: Suppose $(y^i, \alpha^i) \to (\bar{x}, \alpha)$ where $y^i \epsilon X$ and $\alpha^i \geq 0$, and let $\tilde{h}_i^*$ denote the optimal value of the problem corresponding to (y^i, α^i). By considering an appropriate convergent subsequence of optimal solutions it follows that $\liminf \tilde{h}_i^* \geq \tilde{h}^*(\alpha)$. To establish the other required inequality, $\limsup \tilde{h}_i^* \leq \tilde{h}^*(\alpha)$, we construct a sequence of feasible solutions for the sequence of problems $PL(y^i, \alpha^i)$ as follows: let $\bar{z} := \bar{x} + d^*(\alpha)$, where $d^*(\alpha)$ is any optimal solution of $PL(\bar{x}, \alpha)$, (we assume $d^*(\alpha) \neq 0$, since the result is trivial if $\tilde{h}^*(\alpha) = 0$) and define λ_i and d^i such that $y^i + d^i = \bar{z}$ and such that λ_i is the largest scalar in $[0,1]$ such that $\|\lambda_i d^i\| \leq \alpha^i$. Since $d^i \to d^*(\alpha)$ and $\alpha^i \to \alpha$, it follows that $\lambda_i \to 1$. Since $\bar{z} \epsilon X$, the convexity of the feasible sets of the problems $PL(y^i, \alpha^i)$ implies that the $\lambda_i d^i$ form a sequence of feasible solutions for those problems. Therefore,

$$\limsup \tilde{h}_i^* \leq \lim h_i^*(\lambda_i d^i) = \tilde{h}(d^*(\alpha)) = \tilde{h}^*(\alpha). \quad \blacksquare$$

The key factor in the validity proof of the algorithm is the guarantee of a minimum improvement ratio in a neighborhood of non-optimal points. In the following, ρ_0 is a parameter in $(0,1), x^i \epsilon X$, and r_i^* denotes the improvement ratio corresponding to x^i.

Lemma 6: If $x^i \to \bar{x}$, where $\bar{x}$ is non-optimal for (NLP), then there exists an $\bar{\alpha} > 0$ such that $r_i^*(\alpha) \geq \rho_0$ for all $\alpha \varepsilon (0, \bar{\alpha})$ and all x^i sufficiently close to $\bar{x}$.

Pf: By using the uniformity of the approximation error and the continuity of $\tilde{h}^*$, the proof used for Lemma 4 may be extended as needed for this result. $\blacksquare$

3. A piecewise-linear trust region algorithm

The algorithm to be described below determines for each *distinct* non-optimal x^i a value of the trust region parameter that provides at least a value of ρ_0 for the improvement ratio. This is accomplished by starting with a value for α that is at least a threshhold value, and decreasing this value as needed to achieve the required improvement ratio. This algorithm is modelled after the trust region algorithm based on linear approximation in [Zhang, et al 84]. Two versions of the algorithm are presented. The second version is slightly more complex , but in practice eliminates the need to compute the optimal value of the LP $L(x^i, \alpha_i)$ (this quantity is denoted by $h_i^L(\alpha_i)$).

Algorithm PLTR

(0) Let $x^1 \epsilon X$ be given. Choose positive scalars $m > 1$, $\alpha' \leq \alpha_1$, and $\rho_0 < \rho_1 < \rho_2 < 1$. Set $i = 1$.

(1) Solve $PL(x^i, \alpha_i)$, obtaining an optimal solution d_i^*.

(2) If $\tilde{h}_i(d_i^*)/h_i^L(\alpha_i) \leq \rho_0$, or if $h_i(d_i^*)/\tilde{h}_i(d_i^*) \leq \rho_0$,
then set $x^{i+1} = x^i$, $\alpha_{i+1} = \alpha_i/m$, $i \leftarrow i+1$, and return to step (1).

(3) Otherwise, let $r_i^* = h_i(d_i^*)/\tilde{h}_i(d_i^*)$, $x^{i+1} = x^i + d_i^*$ and

$$\alpha_{i+1} = \begin{cases} \max\{\alpha_i/m, \alpha'\} & \text{if } r_i^* \leq \rho_1; \\ \max\{m\alpha_i, \alpha'\} & \text{if } r_i^* \geq \rho_2; \\ \max\{\alpha_i, \alpha'\} & \text{otherwise.} \end{cases}$$

Set $i \leftarrow i+1$ and return to (1). ∎

Note that the *initial* value of the trust region parameter for each *distinct* x^i is at least α', since the value of x^i changes only in step (3), where the new value of α is set to at least α'. The validity of the algorithm now follows in a straightforward manner.

Theorem: If $\bar{x}$ is an accumulation point of a sequence generated by Algorithm PLTR, then $\bar{x}$ is an optimal solution of (NLP).

Pf: Assume that the result is false, and let K be a subsequence such that $x^i \xrightarrow{K} \bar{x}$. Using the preceding lemma and its analog for $\tilde{h}/h^L$, we consider those $i \epsilon K$ sufficiently large such that the ratio conditions in step (2) of the algorithm are satisfied for all $\alpha \epsilon (0, \bar{\alpha})$. Moreover, since the initial value of α for each distinct x^i is at least α', it is the case that for arbitrarily large $i \epsilon K$ that $\alpha_i \geq \alpha^* := \min\{\bar{\alpha}/m, \alpha'\}$ (since the trust region parameter is not reduced below this quantity to order to achieve the required ratios) and $r_i^* > \rho_0$, so that $h_i(d_i^*) < \rho_0 \cdot \tilde{h}_i(d_i^*) < \rho_0^2 \cdot h_L^*(\alpha^*)/2$, where the latter inequality follows from $h_i^L(\alpha_i) < h_L^*(\alpha^*)/2$ for $i \epsilon K$ sufficiently large. However, for x^i sufficiently close to $\bar{x}$, the relations $h_i(d_i^*) = f(x^{i+1}) - f(x^i) < \rho_0^2 \cdot h_L^*(\alpha^*)/2$ contradict $f(x^i) \rightarrow f(\bar{x})$. ∎

For computational efficiency, the algorithm PLTR may be modified by bypassing the initial ratio condition in step (2) whenever $\tilde{h}_i(d_i^*) < -\tau$, where τ is a positive tolerance. (It is easily seen that a slight revision in the proof establishes convergence for this modification.) In this revised algorithm, the solution of a piecewise-linear problem is thus accepted if ($\tilde{h}_i(d_i^*) < -\tau$ or $\tilde{h}_i(d_i^*)/h_i^L(\alpha_i) > \rho_0$) and $h_i(d_i^*)/\tilde{h}_i(d_i^*) > \rho_0$. By choosing τ sufficiently small, $\tilde{h}_i(d_i^*) < -\tau$ will be satisfied whenever $\tilde{h}_i$ is non-zero relative to machine precision, and it thus will not be necessary to solve $L(x^i, \alpha_i)$.

4. Non-differentiable, convex objective functions

The results of the previous section may be extended to allow non-differentiable convex objective function terms of the form $g(x) := \max\{c(x), 0\}$, where c is convex and differentiable on X. The key to this extension is the observation that the results of Lemma 1 may be generalized to approximations of the form

$$\tilde{g}(x) := \max\{c(\bar{x}) + a(x), 0\},$$

where $a(x)$ is a linear, separable, or piecewise-linear separable approximation of $[c(x) - c(\bar{x})]$. For example, letting $d := x - \bar{x}$ as before, if the approximation satisfies $c(x) - [c(\bar{x}) + a(x)] = o(d)$, then it follows from the relation $\max\{c(x) - o(d), 0\} = \max\{c(x), 0\} - o(d)$ that $g(x) - \tilde{g}(x) = o(d)$. Analogous results hold for objective function terms of the form $\max\{c_1(x), \ldots, c_m(x)\}$, provided that each c_i is differentiable and convex. Note that first-order optimality conditions that are sufficient for optimality may be obtained in a manner analogous to the conditions involving $L(\bar{x}, \alpha)$ by considering in place of $L(\bar{x}, \alpha)$ the problem

$$\begin{aligned} \min_{d} \quad & \max\{c(\bar{x}) + \nabla c(x) d, 0\} \\ s.t. \quad & Ad \leq \bar{b}, \quad \|d\| \leq \alpha. \end{aligned}$$

In the case that the original objective function is a convex *separable* non-differentiable function, the convergence of local piecewise-linear approximation methods was established in [Meyer 79]. For such problems, the separable approximation dominates the true objective function, so that the improvement ratio is at least 1 whenever an improvement is obtained in the approximating problem, regardless of the size of the trust region.

5. Objective functions with absolute value terms

For objective function terms of the form $|t(x)|$, where t is differentiable on X, the preceding error bounds may be extended to approximations of the form $\tilde{t}(x) := |t(\bar{x}) + k(x)|$, where k is a linear, separable, or piecewise-linear separable approximation of $t(x) - t(\bar{x})$. However, even if t is convex, $|t|$ will generally be *non-convex,* and two modifications in the preceding results are required. First, when the "inner" approximation k is other than linear, the resulting approximation $\tilde{t}$ will generally be *non-convex,* and therefore difficult to utilize computationally (one could employ integer programming to deal with non-convex piecewise-linear objectives, but the resulting problem is still very difficult computationally). Second, because of the non-convexity, the first-order optimality conditions (associated with the problem analogous to $L(\bar{x}, \alpha)$) are necessary but *not sufficient* for optimality.

Therefore, to obtain linear programming subproblems, it is necessary to let $k(x)$ be a *linear* approximation (in this case, since the linear term $k(x)$ appears within an absolute value term, the corresponding objective function term is piecewise-linear and convex), and from a theoretical viewpoint, we can only establish that the accumulation points of the corresponding algorithm will be *stationary* points that satisfy first-order necessary conditions. This is the result described in [Zhang, et al, 84].

6. Computational Aspects

The piecewise-linear trust region algorithm has a number of nice computational properties. Although the approximation that it generates is generally more accurate than a linear approximation (and is significantly more accurate for points near the translated axes), the subproblems remain linear programs. Moreover, since the approximation is also separable, it is easily seen that in the differentiable case, the additional constraints needed to model the piecewise-linear function are simply bounds on some additional variables. For example, in a two-segment approximation, the original vector variable x is replaced by substituting $x = \bar{x} + x^+ - x^-$ with $0 \leq x^+ \leq u^+$ and $0 \leq x^- \leq u^-$, yielding the approximating problem

$$\begin{array}{ll} \min\limits_{x} & c^+x^+ - c^-x^- \\ s.t. & A(x^+ - x^-) \leq \bar{b}, \quad 0 \leq x^+ \leq u^+ \text{ and } 0 \leq x^- \leq u^- \ , \end{array}$$

where c^+ and c^- are vectors of slopes, and u^+ and u^- are used to enforce the trust region constraints.

Preservation of constraint structure is particularly important in problem classes such as nonlinear networks, where it is crucial for efficiency to maintain the original network constraints . The separability of the approximation also makes possible a decomposition of the subproblem into independent subproblems in the case of those problem classes (such as traffic assignment problems, see, e.g., [Feijoo and Meyer 84, 85]) in which the critical coupling between variables takes place in the objective function. Such decompositions permit the use of parallel computing for the solution of the subproblems.

References

B. Feijoo: "Piecewise-linear approximation methods and parallel algorithms in optimization", University Wisconsin Computer Sciences Department Tech. Rpt. #598, 1985.

B. Feijoo and R. R. Meyer: "Piecewise-linear approximation methods for nonseparable convex optimization", University of Wisconsin Computer Sciences Department Technical Report #521, 1984.

B. Feijoo and R. R. Meyer: "Optimization on the Crystal multicomputer", in *Computing 85,* G. Bucci and G. Valle eds., Elsevier Science Publishers, 1985.

R. Fletcher: *Practical Methods of Optimization,* John Wiley, 1981.

R. R. Meyer: " Two-segment separable programming", *Management Science* , 25, 385-395, 1979.

Y.Yuan: "Conditions for convergence of trust region algorithms for nonsmooth optimization", *Mathematical Programming,* 31, 220-228, 1985.

J. Zhang, N. H. Kim, and L. Lasdon: "An improved successive linear programming algorithm", The University of Texas Graduate School of Business Working paper 84/85-3-2, Austin, 1984.