ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

LINE SEARCH ALGORITHMS WITH GUARANTEED SUFFICIENT DECREASE

Jorge J. Moré and David J. Thuente*

Mathematics and Computer Science Division

Preprint MCS-P330-1092

October 1992

Work supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

* Permanent address: Department of Mathematical Sciences, Indiana-Purdue University, Fort Wayne, Indiana 46805

ABSTRACT

The problem of finding a point that satisfies the sufficient decrease and curvature condition is formulated in terms of finding a point in a set $T(\mu)$. We describe a search algorithms for this problem that produces a sequence of iterates that converge to a point in $T(\mu)$ and that, except for pathological cases, terminates in a finite number of steps. Numerical results for an implementation of the search algorithm on a set of test functions show that the algorithm terminates within a small number of iterations.

LINE SEARCH ALGORITHMS WITH GUARANTEED SUFFICIENT DECREASE

Jorge J. Moré and David J. Thuente*

1 Introduction

Given a continuously differentiable function $\phi : \mathbb{R} \to \mathbb{R}$ defined on $[0, \infty)$ with $\phi'(0) < 0$, and constants μ and η in $(0, 1)$, we are interested in finding an $\alpha > 0$ such that

$$\phi(\alpha) \le \phi(0) + \mu \phi'(0) \alpha \tag{1.1}$$

and

$$|\phi'(\alpha)| \le \eta |\phi'(0)|. \tag{1.2}$$

The development of a search procedure that satisfies these conditions is a crucial ingredient in a line search method for minimization. The search algorithm described in this paper has been used by several authors, for example, Liu and Nocedal [10], O'Leary [12], Schlick and Fogelson [14, 15], and Gilbert and Nocedal [7]. This paper describes this search procedure and the associated convergence theory.

In a line search method we are given a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a descent direction p for f at a given point $x \in \mathbb{R}^n$. Thus, if

$$\phi(\alpha) \equiv f(x + \alpha p), \qquad \alpha \ge 0, \tag{1.3}$$

then (1.1) and (1.2) define an acceptable step. The motivation for requiring conditions (1.1) and (1.2) in a line search method should be clear. If α is not too small, condition (1.1) forces a *sufficient decrease* in the function. However, this condition is not sufficient to guarantee convergence, because it allows arbitrarily small choices of $\alpha > 0$. Condition (1.2) rules out arbitrarily small choices of α and usually guarantees that α is near a local minimizer of ϕ. Condition (1.2) is a *curvature condition* because it implies that

$$\phi'(\alpha) - \phi'(0) > (1 - \eta)|\phi'(0)|,$$

and thus the average curvature of ϕ on $(0, \alpha)$ is positive. The curvature condition (1.2) is particularly important in a quasi-Newton method because it guarantees that a positive definite quasi-Newton update is possible. See, for example, Dennis and Schnabel [4] and Fletcher [6].

Work supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the U.S. Department of Energy under Contract W-31-109-Eng-38.

*Permanent address: Department of Mathematical Sciences, Indiana-Purdue University, Fort Wayne, Indiana 46805.

As final motivation for the solution of (1.1) and (1.2), we mention that if a step satisfies these conditions, then the line search method is convergent for reasonable choices of direction. See, for example, Dennis and Schnabel [4] and Fletcher [6] for gradient-related methods; Powell [13] and Byrd, Nocedal, and Yuan [3] for quasi-Newton methods; and Al-Baali [1], Liu and Nocedal [10], and Gilbert and Nocedal [7] for conjugate gradient methods.

In most practical situations it is important to impose additional requirements on α. In particular, it is natural to require that α satisfy the bounds

$$0 \le \alpha_{min} \le \alpha \le \alpha_{max}. \tag{1.4}$$

The main reason for requiring a lower bound α_{min} is to terminate the iteration, while the upper bound α_{max} is needed when the search is used for linearly constrained optimization problems or when the function ϕ is unbounded below. In linearly constrained optimization problems the parameter α_{max} is a function of the distance to the nearest active constraint. An unbounded problem can be approached by accepting any α in $[\alpha_{min}, \alpha_{max}]$ such that $\phi(\alpha) \le \phi_{min}$, where $\phi_{min} < \phi(0)$ is a lower bound specified by the user of the search. In this case

$$\alpha_{max} = \frac{1}{\mu}\left(\frac{\phi(0) - \phi_{min}}{-\phi'(0)}\right) \tag{1.5}$$

is a reasonable setting because if α_{max} satisfies the sufficient decrease condition (1.1), then $\phi(\alpha_{max}) \le \phi_{min}$. On the other hand, if α_{max} does not satisfy the sufficient decrease condition, then we will show that it is possible to determine an acceptable α.

The main problem that we consider in this paper is to find an acceptable α in the sense that α belongs to the set

$$T(\mu) \equiv \Big\{ \alpha > 0 : \phi(\alpha) \le \phi(0) + \alpha\mu\phi'(0),\ |\phi'(\alpha)| \le \mu|\phi'(0)| \Big\}.$$

By phrasing our results in terms of $T(\mu)$ we make it clear that the search algorithm is independent of η; the parameter η is used only in the termination test of the algorithm. Another advantage of phrasing the results in terms of $T(\mu)$ is that $T(\mu)$ is usually not empty. For example, $T(\mu)$ is not empty when ϕ is bounded below.

Several authors have done related work on the solution of (1.1) and (1.2). For example, Gill and Murray [8] attacked (1.1) and (1.2) by using a univariate minimization algorithm for ϕ to find a solution α^* to (1.2). If α^* did not satisfy (1.1), then α^* was repeatedly halved in order to obtain a solution β^* to (1.1). Of course, β^* did not necessarily satisfy (1.2); but if μ was sufficiently small, then it was argued that this was an unlikely event. In a similar vein, we mention that the search algorithm of Shanno and Phua [16, 17] is not guaranteed to work in all cases. In particular, the sufficient decrease condition (1.1) can rule out many of the points that satisfy (1.2), and then the algorithm is not guaranteed to converge.

Gill, Murray, Saunders, and Wright [9] proposed an interesting variation on (1.1) and (1.2) when they argued that if there is no solution to (1.1) and (1.2), then it was necessary to compute a point such that

$$\phi(\alpha) = \phi(0) + \mu\phi'(0)\alpha. \tag{1.6}$$

If (1.6) has a solution, then their algorithm computes a sequence of nested intervals such that each interval contains points that satisfy (1.1) and (1.2), or just (1.6). Their algorithm, however, is not guaranteed to produce a point that satisfies (1.1) and (1.2).

Fletcher [5] suggested that it is possible to compute a sequence of nested intervals that contain points that satisfy (1.1) and (1.2), but he did not prove any result along these lines. This suggestion led to the algorithms developed by Al-Baali and Fletcher [2] and Moré and Sorensen [11]. In this paper we provide a convergence analysis, implementation details, and numerical results for the algorithm of Moré and Sorensen [11].

The search algorithm for $T(\mu)$ is defined in Section 2. We show that the search algorithm produces a sequence of iterates that converge to a point in $T(\mu)$ and that, except for pathological cases, the search algorithm produces a finite sequence $\alpha_0, \ldots, \alpha_m$ of trial values in $[\alpha_{min}, \alpha_{max}]$, where $\alpha_m \in T(\mu)$ or is one of the bounds. Termination at one of the bounds can be avoided by a suitable choice of bounds. For example, if $\alpha_{min} = 0$ and α_{max} is defined by (1.5), then either α_m lies in $T(\mu)$ or $\phi(\alpha_m) \leq \phi_{min}$.

The results of Section 2 show that the search algorithm can be used to find an α that satisfies (1.1) and (1.2) when $\mu \leq \eta$. A result for an arbitrary $\eta \in (0, \mu)$ requires additional assumptions because there may not be an α that satisfies (1.1) and (1.2) even if ϕ is bounded below. In Section 3 we show that if the search algorithm generates an iterate α_k that satisfies the sufficient decrease condition and $\phi'(\alpha_k) > 0$, then the search algorithm terminates at an α_k that satisfies (1.1) and (1.2).

Given α_0 in $[\alpha_{min}, \alpha_{max}]$, the search algorithm generates a sequence of nested intervals $\{I_k\}$ and a sequence of iterates $\alpha_k \in I_k \cap [\alpha_{min}, \alpha_{max}]$. Section 4 describes the specific choices for the trial values α_k that are used in our algorithm. Our numerical results indicate that these choices lead to fast termination.

Section 5 describes a set of test problems and numerical results for the search procedure. The first three functions in the test set have regions of concavity, while the last three functions are convex. In all cases the functions have a unique minimizer. The emphasis in the numerical results is to explain the qualitative behavior of the algorithm for a wide range of values of μ and η.

2 The Search Algorithm for $T(\mu)$

In this section we present the search algorithm for determining an α in $T(\mu)$. We assume that ϕ is continuously differentiable on $[0, \alpha_{max}]$ with $\phi'(0) < 0$. Most work on line searches assumes that $\mu < \frac{1}{2}$, because if ϕ is a quadratic with $\phi'(0) < 0$ and $\phi''(0) > 0$, then the global minimizer α^* of ϕ satisfies

$$\phi(\alpha^*) = \phi(0) + \frac{1}{2}\alpha^* \phi'(0),$$

and thus α^* satisfies (1.1) only if $\mu \le \frac{1}{2}$. The restriction $\mu < \frac{1}{2}$ also allows $\alpha = 1$ to be ultimately acceptable to Newton and quasi-Newton methods. In this section we need only assume that μ lies in $(0, 1)$.

Given α_0 in $[\alpha_{min}, \alpha_{max}]$, the search algorithm generates a sequence of nested intervals $\{I_k\}$ and a sequence of iterates $\alpha_k \in I_k \cap [\alpha_{min}, \alpha_{max}]$ according to the following procedure.

Search Algorithm. Set $I_0 = [0, \infty]$.

For $k = 0, 1, \ldots$

Choose a safeguarded $\alpha_k \in I_k \cap [\alpha_{min}, \alpha_{max}]$.

Test for convergence.

Update the interval I_k.

In this description the term *safeguarded* α_k refers to the rules that force convergence of the algorithm. For the moment we assume that a safeguarded choice is made, and discuss the updating of I_k.

The aim of the updating process for the intervals I_k is to identify and generate an interval I_k such that $T(\mu) \cap I_k$ is not empty, and then refine the interval so that $T(\mu) \cap I_k$ remains not empty. We now specify conditions on the endpoints of an interval I that guarantee that I has a nonempty intersection with $T(\mu)$. The conditions on the endpoints α_l and α_u are specified in terms of the auxiliary function ψ defined by

$$\psi(\alpha) \equiv \phi(\alpha) - \phi(0) - \mu\phi'(0)\alpha.$$

We assume that $\alpha_l \ne \alpha_u$ but do not assume that α_l and α_u are ordered.

Theorem 2.1 *Let I be a closed interval with endpoints α_l and α_u. If the endpoints satisfy*

$$\psi(\alpha_l) \le \psi(\alpha_u), \quad \psi(\alpha_l) \le 0, \quad \psi'(\alpha_l)(\alpha_u - \alpha_l) < 0, \tag{2.1}$$

then there is an α^* in I with $\psi(\alpha^*) \le \psi(\alpha_l)$ and $\psi'(\alpha^*) = 0$. In particular, $\alpha^* \in (T(\mu) \cap I)$.

Proof. Assume that $\alpha_u > \alpha_l$; the proof in the other case is similar. Define

$$\alpha_m = \sup\Big\{\alpha \in [\alpha_l, \alpha_u] : \psi(\beta) \le 0,\ \beta \in [\alpha_l, \alpha]\Big\}.$$

Then $\alpha_m > \alpha_l$, because $\psi'(\alpha_l) < 0$. We first claim that $\psi(\alpha_m) \ge \psi(\alpha_l)$. The assumption on α_u shows that this is certainly the case if $\alpha_m = \alpha_u$. This also holds if $\alpha_m < \alpha_u$, because in this case the definition of α_m implies that $\psi(\alpha_m) = 0$, and thus $\psi(\alpha_m) = 0 \ge \psi(\alpha_l)$.

Define α^* to be a global minimizer of ψ on $[\alpha_l, \alpha_m]$. We claim that $\alpha^* \in T(\mu)$. The global minimum cannot be achieved at α_l because $\psi'(\alpha_l) < 0$; and since we have established that $\psi(\alpha_m) \geq \psi(\alpha_l)$, the global minimum cannot be achieved at α_m. Hence, α^* is in the interior of $[\alpha_l, \alpha_m]$. In particular, $\psi'(\alpha^*) = 0$, and thus $|\phi'(\alpha^*)| = \mu|\phi'(0)|$. We also know that α^* satisfies (1.1) because $\psi(\alpha) \leq 0$ for all α in $[\alpha_l, \alpha_m]$. Hence, $\alpha^* \in T(\mu)$, as desired. ■

Theorem 2.1 provides the motivation for the search algorithm by showing that if the endpoints of I satisfy (2.1), then ψ has a minimizer α^* in the interior of I and, moreover, that α^* belongs to $T(\mu)$. Thus the search algorithm can be viewed as a procedure for locating a minimizer of ψ.

The assumptions that Theorem 2.1 imposes on the endpoints α_l and α_u cannot be relaxed because if we fix α_l and α_u by the assumption $\psi(\alpha_l) \leq \psi(\alpha_u)$, then the result fails to hold if either of the other two assumptions are violated. The assumptions (2.1) can be paraphrased by saying that α_l is the endpoint with lowest ψ value, that α_l satisfies the sufficient decrease condition (1.1), and that $\alpha_u - \alpha_l$ is a descent direction for ψ at α_l so that $\psi(\alpha) < \psi(\alpha_l)$ for all α in I sufficiently close to α_l. In particular, this last assumption guarantees that ψ can be decreased by searching near α_l.

We now describe an algorithm for updating the interval I, and then show how to use this algorithm to obtain an interval that satisfies the conditions of Theorem 2.1.

Updating Algorithm. Given a trial value α_t in I, the endpoints α_l^+ and α_u^+ of the updated interval I_+ are determined as follows:

Case U1: If $\psi(\alpha_t) > \psi(\alpha_l)$, then $\alpha_l^+ = \alpha_l$ and $\alpha_u^+ = \alpha_t$.

Case U2: If $\psi(\alpha_t) \leq \psi(\alpha_l)$ and $\psi'(\alpha_t)(\alpha_l - \alpha_t) > 0$, then $\alpha_l^+ = \alpha_t$ and $\alpha_u^+ = \alpha_u$.

Case U3: If $\psi(\alpha_t) \leq \psi(\alpha_l)$ and $\psi'(\alpha_t)(\alpha_l - \alpha_t) < 0$, then $\alpha_l^+ = \alpha_t$ and $\alpha_u^+ = \alpha_l$.

It is straightforward to show that if the endpoints α_l and α_u satisfy (2.1), then the updated endpoints α_l^+ and α_u^+ also satisfy (2.1) unless $\psi'(\alpha_t) = 0$ and $\psi(\alpha_t) \leq \psi(\alpha_l)$. Of course, in this case there is no need to update I because α_t belongs to $T(\mu)$.

Al-Baali and Fletcher [2] present two updating schemes. The aim of scheme S1 is to identify a point that satisfies (1.1) and $\phi'(\alpha) \geq \eta\phi'(0)$, while scheme S2 seeks a point that satisfies (1.1) and (1.2). In scheme S2 the endpoints α_l^+ and α_u^+ are determined as follows:

If $\psi(\alpha_t) > 0$ or if $\phi(\alpha_t) > \phi(\alpha_l)$, then

$$\alpha_l^+ = \alpha_l \text{ and } \alpha_u^+ = \alpha_t,$$

else if $\phi'(\alpha_t)(\alpha_l - \alpha_u) > 0$, then

$$\alpha_l^+ = \alpha_t \text{ and } \alpha_u^+ = \alpha_u,$$

else $\alpha_l^+ = \alpha_t$ and $\alpha_u^+ = \alpha_l$.

The two updating algorithms produce the same iterates as long as

$$\psi(\alpha_t) > 0, \qquad \text{or} \qquad \psi(\alpha_t) \le 0, \quad \psi(\alpha_t) \le \psi(\alpha_l),$$

but differ in their treatment of the situation where

$$\psi(\alpha_t) \le 0, \quad \psi(\alpha_t) > \psi(\alpha_l), \quad \phi(\alpha_t) < \phi(\alpha_l), \quad \alpha_l < \alpha_t < \alpha_u.$$

In this case, our updating algorithm chooses $I_+ = [\alpha_l, \alpha_t]$, while scheme S2 sets $I_+ = [\alpha_t, \alpha_u]$ if $\phi'(\alpha_t) < 0$. Our algorithm seems to be preferable in this situation because the interval I_+ contains an acceptable point, while the interval generated by scheme S2 is not guaranteed to contain an acceptable point.

We now show how the updating algorithm can be used to determine an interval I in $[0, \alpha_{max}]$ with endpoints that satisfy (2.1). Initially $\alpha_l = 0$ and $\alpha_u = \infty$. Consider any trial value α_t in $[\alpha_{min}, \alpha_{max}]$. If cases U1 or U3 hold, then we have determined an interval with endpoints α_l and α_u that satisfy the conditions of Theorem 2.1. Otherwise case U2 holds, and we can repeat the process for some α_t^+ in $[\alpha_t, \alpha_{max}]$. We continue generating trial values in $[\alpha_l, \alpha_{max}]$ as long as case U2 holds, but require that eventually α_{max} be used as a trial value. This is done by choosing

$$\alpha_t^+ \in \Big[\min\{\delta_{max}\alpha_t, \alpha_{max}\}, \alpha_{max}\Big] \tag{2.2}$$

for some factor $\delta_{max} > 1$. In our implementation we use

$$\alpha_t^+ = \min\{\alpha_t + \delta(\alpha_t - \alpha_l), \alpha_{max}\}, \qquad \delta \in [1.1, 4],$$

and thus an induction argument shows that (2.2) holds with $\delta_{max} = 1.1$.

Since $\alpha_l = 0$ initially and $\psi(0) = 0$, the sequence $\alpha_0, \alpha_1, \ldots$ of trial values is increasing with

$$\psi(\alpha_k) \le 0 \text{ and } \psi'(\alpha_k) < 0, \qquad k = 0, 1, \ldots, \tag{2.3}$$

as long as case U2 holds. The search algorithm terminates at α_{max} if $\psi(\alpha_{max}) \le 0$ and $\psi'(\alpha_{max}) < 0$. This is a reasonable termination criterion because Theorem 2.1 shows that when these conditions do not hold there is an $\alpha^* \in T(\mu)$ with $\alpha^* \le \alpha_{max}$. Thus, after a finite number of trial values, either the search algorithm terminates at α_{max}, or the search algorithm generates an interval with endpoints that satisfy conditions (2.1).

Given an interval that satisfies conditions (2.1), the search algorithm uses the updating algorithm to refine I. We claim that if the search algorithm does not generate an interval I in $[\alpha_{min}, \alpha_{max}]$ that satisfies conditions (2.1), then the sequence $\{\alpha_k\}$ of trial values is decreasing with

$$\psi(\alpha_k) > 0 \text{ or } \psi'(\alpha_k) \ge 0, \qquad k = 0, 1, \ldots. \tag{2.4}$$

This claim is established by considering all three cases of the updating algorithm. If we use an α_t with $\psi(\alpha_t) \le 0$ and $\psi'(\alpha_t) < 0$, and case U2 or U3 holds, then the updating algorithm shows that the interval I_+ lies to the right of α_t. Since $\alpha_t \ge \alpha_{min}$, the interval I_+ contains $[\alpha_{min}, \alpha_{max}]$. If case U1 holds, then $\psi(\alpha_l) < 0$, and thus $\alpha_l \ge \alpha_{min}$. Hence, the updated interval contains $[\alpha_{min}, \alpha_{max}]$.

We force the search algorithm to use α_{min} as a trial value when (2.4) holds and $\alpha_{min} > 0$. This is done by choosing

$$\alpha_t^+ \in \Big[\alpha_{min}, \max\{\delta_{min}\alpha_t, \alpha_{min}\}\Big] \tag{2.5}$$

for some factor $\delta_{min} < 1$. In our implementation (2.5) holds with $\delta_{min} = \frac{7}{12}$. For more details, see Section 4.

The search algorithm terminates at α_{min} if $\psi(\alpha_{min}) > 0$ or $\psi'(\alpha_{min}) \ge 0$. This is a reasonable termination criterion because Theorem 2.1 shows that there is an $\alpha^* \in T(\mu)$ with $\alpha^* \le \alpha_{min}$ when these conditions hold. Thus, after a finite number of trial values, either the search algorithm terminates at α_{min}, or the search algorithm generates an interval I in $[\alpha_{min}, \alpha_{max}]$ with endpoints that satisfy conditions (2.1).

The requirements (2.2) and (2.5) are two of the safeguarding rules. Note that (2.2) is enforced only when (2.3) holds, while (2.5) is used when (2.4) holds. If the search algorithm generates an interval I in $[\alpha_{min}, \alpha_{max}]$, then we need a third rule to guarantee that the choice of α_t forces the length of I to zero. In our implementation this is done by monitoring the length of I; if the length of I does not decrease by a factor of $\delta < 1$ after two trials, then a bisection step is used for the next trial α_t. In our implementation we use $\delta = 0.66$.

Theorem 2.2 *The search algorithm produces a sequence $\{\alpha_k\}$ in $[\alpha_{min}, \alpha_{max}]$ such that after a finite number of trial values one of the following conditions hold.*

The search terminates at α_{max}, the sequence of trial values is increasing, and (2.3) *holds.*

The search terminates at α_{min}, the sequence of trial values is decreasing, and (2.4) *holds.*

An interval $I_k \subset [\alpha_{min}, \alpha_{max}]$ is generated.

Proof. In this proof we essentially summarize the arguments presented above. Let $\alpha_l^{(k)}$ and $\alpha_u^{(k)}$ be the endpoints of I_k, and define

$$\beta_l^{(k)} = \min\Big\{\alpha_l^{(k)}, \alpha_u^{(k)}\Big\}, \quad \beta_u^{(k)} = \max\Big\{\alpha_l^{(k)}, \alpha_u^{(k)}\Big\}.$$

The left endpoint $\beta_l^{(k)}$of I_k is nondecreasing, while the right endpoint is nonincreasing.

We first show that $\beta_u^{(k)} = \infty$ cannot hold for all $k \ge 0$. If $\beta_u^{(k)} = \infty$, then only case U2 of the updating algorithm holds because in the other two cases both endpoints are set to finite values. Since only case U2 holds, it is clear that (2.3) holds, and thus the safeguarding rule (2.2) shows that the bound α_{max} is eventually used as a trial value. If the search does not terminate at α_{max}, then $\beta_u^{(k)} = \alpha_{max}$.

A similar argument shows that $\beta_l^{(k)} = 0$ cannot hold for all $k \geq 0$. If $\beta_l^{(k)} = 0$, then only case U1 or U3 of the updating algorithm holds because in case U2 both endpoints are set to positive values. Moreover, in this case (2.4) holds. The safeguarding rule (2.5) shows that (2.4) cannot hold for all $k \geq 0$ when $\alpha_{min} = 0$, and that if $\alpha_{min} > 0$, then α_{min} is eventually used as a trial value. If the search does not terminate at α_{min}, then $\beta_u^{(k)} = \alpha_{min}$.

We have thus shown that after a finite number of trial values, either the search terminates at one of the two bounds α_{min} or α_{max}, or $\beta_l^{(k)} > 0$ and $\beta_u^{(k)} < \infty$. Of course, in this last case I_k is a subset of $[\alpha_{min}, \alpha_{max}]$. ■

The most interesting case of Theorem 2.2 occurs when an interval $I_k \subset [\alpha_{min}, \alpha_{max}]$ is generated. In this case the safeguarding rules guarantee that the length of the intervals $\{I_k\}$ converges to zero, and thus the sequence $\{\alpha_k\}$ converges to some α^* in $T(\mu)$.

We can rule out finite termination at one of the bounds by ruling out (2.3) and (2.4). The simplest way to do this is to assume that α_{min} satisfies

$$\psi(\alpha_{min}) \leq 0 \text{ and } \psi'(\alpha_{min}) < 0, \tag{2.6}$$

and that α_{max} satisfies

$$\psi(\alpha_{max}) > 0 \text{ or } \psi'(\alpha_{max}) \geq 0. \tag{2.7}$$

Under these assumptions, Theorem 2.2 shows that an interval $I_k \subset [\alpha_{min}, \alpha_{max}]$ is generated after a finite number of trial values.

Conditions (2.6) and (2.7) can be easily satisfied. For example, if $\alpha_{min} = 0$, then (2.6) holds. Condition (2.7) holds if α_{max} is defined by (1.5) and ϕ_{min} is a strict lower bound for ϕ. Condition (2.7) also holds if $\phi'(\alpha_{max}) \geq 0$.

Theorem 2.3 *If the bounds α_{min} and α_{max} satisfy* (2.6) *and* (2.7)*, then the search algorithm terminates in a finite number of steps with an $\alpha_k \in T(\mu)$, or the iterates $\{\alpha_k\}$ converge to some $\alpha^* \in T(\mu)$ with $\psi'(\alpha^*) = 0$. If the search algorithm does not terminate in a finite number of steps, then there is an index k_0 such that the endpoints $\alpha_l^{(k)}$, $\alpha_u^{(k)}$ of the interval I_k satisfy $\alpha_l^{(k)} < \alpha^* < \alpha_u^{(k)}$. Moreover, if $\psi(\alpha^*) = 0$, then ψ' changes sign on $[\alpha_l^{(k)}, \alpha^*]$ for all $k \geq k_0$, while if $\psi(\alpha^*) < 0$, then ψ' changes sign on $[\alpha_l^{(k)}, \alpha^*]$ or $[\alpha^*, \alpha_u^{(k)}]$ for all $k \geq k_0$.*

Proof. Assume that $\alpha_k \notin T(\mu)$ for all the iterates generated by the search algorithm. Since the intervals I_k are uniformly bounded and their lengths tends to zero, any sequence $\{\theta_k\}$ with $\theta_k \in I_k$ must converge to a common limit α^*. Theorem 2.1 guarantees that there is a $\theta_k \in (T(\mu) \cap I_k)$ with $\phi'(\theta_k) = \mu\phi'(0)$. This implies that $\alpha^* \in T(\mu)$ and that

$$\phi'(\alpha^*) = \mu\phi'(0).$$

In particular, $\psi'(\alpha^*) = 0$.

We define k_0 by noting that the continuity of ϕ' shows that there is a $k_0 > 0$ such that $\phi'(\alpha) < 0$ for all $\alpha \in I_k$ and all $k \geq k_0$. Since $\psi(\alpha_l^{(k)}) \leq 0$ and $\alpha_l^{(k)} \notin T(\mu)$, we must have $|\phi'(\alpha_l^{(k)})| > \mu|\phi'(0)|$. We also know that $\phi'(\alpha_l^{(k)}) < 0$ for $k \geq k_0$, and thus $\phi'(\alpha_l^{(k)}) < \mu\phi'(0)$. Hence, $\psi'(\alpha_l^{(k)}) < 0$. Condition (2.1) on the endpoints implies that $\alpha_l^{(k)} < \alpha_u^{(k)}$, and in particular, $\alpha_l^{(k)} < \alpha^* < \alpha_u^{(k)}$.

Now consider the case where $\psi(\alpha^*) = 0$. We cannot have $\psi'(\alpha) \leq 0$ on $[\alpha_l^{(k)}, \alpha^*]$ because this implies that $\psi(\alpha_l^{(k)}) > \psi(\alpha^*) = 0$. Thus $\psi'(\beta_k) > 0$ for some $\beta_k \in [\alpha_l^{(k)}, \alpha^*]$. Since $\psi'(\alpha_l^{(k)}) < 0$, we have shown that ψ' changes sign on $[\alpha_l^{(k)}, \alpha^*]$.

Finally, consider the case where $\psi(\alpha^*) < 0$. Assume that k_0 is such that $\psi(\alpha_u^{(k)}) < 0$ for all $k \geq k_0$. If $\psi'(\alpha_u^{(k)}) \geq 0$, then $\phi'(\alpha_u^{(k)}) \geq \mu\phi'(0)$, and since $\phi'(\alpha_u^{(k)}) < 0$, we have $\alpha_u^{(k)} \in T(\mu)$. This contradiction shows that $\psi'(\alpha_u^{(k)}) < 0$. We have already shown that $\psi'(\alpha_l^{(k)}) < 0$, so ψ' changes sign on $[\alpha_l^{(k)}, \alpha^*]$ or $[\alpha^*, \alpha_u^{(k)}]$ if $\psi'(\beta_k) > 0$ for some β_k in $[\alpha_l^{(k)}, \alpha_u^{(k)}]$. This is clear because if $\psi'(\alpha) \leq 0$ on $[\alpha_l^{(k)}, \alpha_u^{(k)}]$, then $\psi(\alpha_l^{(k)}) > \psi(\alpha_u^{(k)})$. ■

If the search algorithm does not terminate in a finite number of steps, then Theorem 2.3 implies that ψ' changes sign an infinite number of times in the sense that there is a monotone sequence $\{\beta_k\}$ that converges to α^* and such that $\psi'(\beta_k)\psi'(\beta_{k+1}) < 0$. Theorem 2.3 thus justifies our claim that, except for pathological cases, the search algorithm terminates in a finite number of iterations. Closely related results have been established by Al-Baali and Fletcher [2] and Moré and Sorensen [11]. In these results, however, the emphasis is on showing that the search algorithm eventually generates an α_k that satisfies (1.1) and (1.2) provided $\mu < \eta$.

3 Search for a Local Minimizer

Theorem 2.3 guarantees finite termination at an α_k that satisfies (1.1) and (1.2) provided $\eta > \mu$. In this section we modify the search algorithm and show that under reasonable conditions we can guarantee that the modified search algorithm generates an α_k that satisfies (1.1) and (1.2) for any $\eta > 0$.

A difficulty with setting $\eta < \mu$ is that, even if $T(\mu)$ is not empty, there may not be an $\alpha \geq 0$ that satisfies (1.1) and (1.2). We illustrate this point with a minor modification of an example of Al-Baali and Fletcher [2]. Define

$$\phi(\alpha) = \begin{cases} \frac{1}{2}(1-\sigma)\alpha^2 - \alpha, & 0 \leq \alpha \leq 1 \\ \frac{1}{2}(\sigma - 1) - \sigma\alpha, & 1 \leq \alpha, \end{cases} \tag{3.1}$$

where $\eta < \sigma < \mu$. The solid plot in Figure 3.1 is the function ϕ with $\sigma = 0.1$; the dashed plot is the function $l(\alpha) = \phi(0) + \mu\phi'(0)\alpha$ with $\mu = 0.25$. A computation shows that ϕ is

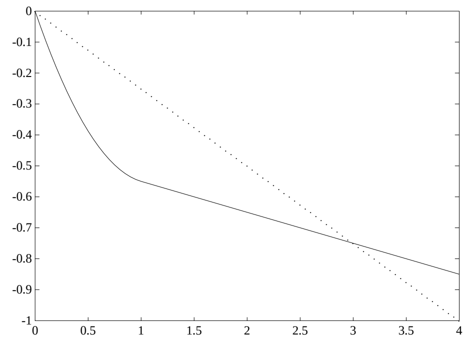

Figure 3.1: Solid plot is ϕ; dotted plot is $l(\alpha) = \phi(0) + \mu\phi'(0)\alpha$.

continuously differentiable and that

$$|\phi'(\alpha)| \geq \sigma > \eta$$

for all $\alpha \geq 0$. Moreover, if $\mu < \frac{1}{2}$, then

$$T(\mu) = \left[\frac{1-\mu}{1-\sigma}\ ,\ \frac{1-\sigma}{2(\mu-\sigma)}\right].$$

Thus $T(\mu)$ is a nonempty interval with $\alpha = 1$ in the interior. In Figure 3.1 we have set $\sigma = 0.1$ and $\mu = 0.25$ and thus $T(\mu) = [\frac{5}{6}, 3]$.

We now show that if during the search for $T(\mu)$ we compute a trial value α_k such that $\psi(\alpha_k) \leq 0$ and $\psi'(\alpha_k) \geq 0$, then α_k belongs to $T(\mu)$ or we have identified an interval that contains points that satisfy the sufficient decrease condition (1.1) and the curvature condition (1.2).

Theorem 3.1 *Assume that the bounds α_{min} and α_{max} satisfy* (2.6) *and* (2.7). *Let $\{\alpha_k\}$ be the sequence generated by the search algorithm, and let $\alpha_l^{(k)}$ and $\alpha_u^{(k)}$ be the endpoints of the interval I_k generated by the search algorithm. If α_k is the first iterate that satisfies*

$$\psi(\alpha_k) \leq 0, \quad \psi'(\alpha_k) \geq 0, \tag{3.2}$$

then $\alpha_l^{(k)} < \alpha_u^{(k)}$. Moreover, $\alpha_k \in T(\mu)$ or $\phi'(\alpha_k) > 0$. If $\phi'(\alpha_k) > 0$, then the interval

$$I^* \equiv \left[\alpha_l^{(k)}, \alpha_k\right],$$

contains an α^* that satisfies (1.1) *and $\phi'(\alpha^*) = 0$. Moreover, any $\alpha \in I^*$ with $\phi(\alpha) \leq \phi(\alpha_k)$ also satisfies* (1.1).

Proof. We first claim that $\psi'(\alpha_l^{(k)}) < 0$. If this is not the case, then $\psi'(\alpha_j) \geq 0$ for some index $j < k$ because $\alpha_l^{(k)}$ is a previous iterate. However, this contradicts the assumption that α_k is the first iterate that satisfies (3.2). This proves that $\psi'(\alpha_l^{(k)}) < 0$.

Since $\psi'(\alpha_l^{(k)}) < 0$, assumptions (2.1) imply that $\alpha_l^{(k)} < \alpha_u^{(k)}$. This implies, in particular, that $\alpha_k > \alpha_l^{(k)}$ so I^* is well defined.

If $\psi'(\alpha_k) \geq 0$ and $\phi'(\alpha_k) \leq 0$, then it is clear that $|\phi'(\alpha_k)| \leq \mu|\phi'(0)|$. Since $\psi(\alpha_k) \leq 0$, this implies that $\alpha_k \in T(\mu)$.

Now assume that $\phi'(\alpha_k) > 0$, and let α^* be a global minimizer of ϕ on I^*. Since $\phi'(\alpha_k) > 0$ and α^* is a minimizer, $\alpha^* \neq \alpha_k$. Similarly, since we proved above that $\psi'(\alpha_l^{(k)}) < 0$, and α^* is a minimizer, $\alpha^* \neq \alpha_l^{(k)}$.

We have shown that α^* is in the interior of I^*. Hence, $\phi'(\alpha^*) = 0$ as desired. We complete the proof by noting that if $\phi(\alpha) \leq \phi(\alpha_k)$ for some $\alpha \in I^*$, then

$$\phi(\alpha) \leq \phi(\alpha_k) \leq \phi(0) + \mu\phi'(0)\alpha_k \leq \phi(0) + \mu\phi'(0)\alpha.$$

The second inequality holds because α_k satisfies (1.1), while the third inequality holds because $\alpha \leq \alpha_k$. Hence, any $\alpha \in I^*$ with $\phi(\alpha) \leq \phi(\alpha_k)$ also satisfies (1.1). ■

There is no guarantee that the search algorithm will generate an iterate α_k such that $\psi(\alpha_k) \leq 0$ and $\phi'(\alpha_k) > 0$. For example, if ϕ is the function shown in Figure 3.1, then $\phi'(\alpha) < 0$ for all α. Even if ϕ has a minimizer α^* that satisfies the sufficient decrease condition, the search algorithm may be trapped in a region that contains points in $T(\mu)$, but where (1.1) and (1.2) are not satisfied.

Theorem 3.1 is one of the ingredients needed to develop a search algorithm for a minimizer that satisfies the sufficient decrease condition (1.1) and the curvature condition (1.2). We also need to show that the interval I^* specified by Theorem 3.1 satisfies the assumptions of the following result.

Theorem 3.2 *Let I be a closed interval with endpoints α_l and α_u. If the endpoints satisfy*

$$\phi(\alpha_l) \leq \phi(\alpha_u), \quad \phi'(\alpha_l)(\alpha_u - \alpha_l) < 0,$$

then there is an α^* in I with $\phi(\alpha^*) \leq \phi(\alpha_l)$ and $\phi'(\alpha^*) = 0$.

Proof. The proof of this result is almost immediate. If α^* is the global minimizer of ϕ on I, then the assumptions on α_l and α_u guarantee that α^* is in the interior of I and thus $\phi'(\alpha^*) = 0$. ■

The interval I^* specified by Theorem 3.1 satisfies the assumptions of Theorem 3.2 because the derivative of ϕ has the proper sign at the endpoints. We assumed that $\phi'(\alpha_k) > 0$. Moreover, in Theorem 3.1 we established that $\alpha_l^{(k)} < \alpha_u^{(k)}$, and thus assumptions (2.1) on the endpoints of I_k imply that $\psi'(\alpha_l^{(k)}) < 0$. Hence $\phi'(\alpha_l^{(k)}) < 0$. These two results show that I^* has the desired properties.

We now need to modify the updating algorithm so that we can guarantee finite termination at an iterate that satisfies the sufficient decrease condition (1.1) and the curvature condition (1.2). The modification is simple; we just replace ψ by ϕ in the updating algorithm.

Modified Updating Algorithm. Given a trial value α_t in I, the endpoints α_l^+ and α_u^+ of the updated interval I_+ are determined as follows:

Case a: If $\phi(\alpha_t) > \phi(\alpha_l)$, then $\alpha_l^+ = \alpha_l$ and $\alpha_u^+ = \alpha_t$.

Case b: If $\phi(\alpha_t) \le \phi(\alpha_l)$ and $\phi'(\alpha_t)(\alpha_l - \alpha_t) > 0$, then $\alpha_l^+ = \alpha_t$ and $\alpha_u^+ = \alpha_u$.

Case c: If $\phi(\alpha_t) \le \phi(\alpha_l)$ and $\phi'(\alpha_t)(\alpha_l - \alpha_t) < 0$, then $\alpha_l^+ = \alpha_t$ and $\alpha_u^+ = \alpha_l$.

We have shown that the interval I^* specified by Theorem 3.1 satisfies the assumptions of Theorem 3.2. Moreover, a short computation shows that if I is any interval that satisfies the assumptions of Theorem 3.1, then the modified updating algorithm preserves these assumptions.

Our implementation of the search algorithm of Section 2 uses the modified updating algorithm in an obvious manner: If some iterate α_k satisfies $\psi(\alpha_k) \le 0$ and $\phi'(\alpha_k) > 0$, then the modified updating algorithm is used on that iteration and all further iterations.

Theorem 3.3 *Assume that the bounds α_{min} and α_{max} satisfy* (2.6) *and* (2.7). *If the modified search algorithm generates an iterate such that $\psi(\alpha_k) \le 0$ and $\phi'(\alpha_k) > 0$, then the modified search terminates at an α_k that satisfies* (1.1) *and* (1.2).

Proof. If the search algorithm generates an α_k with $\psi(\alpha_k) \le 0$ and $\phi'(\alpha_k) > 0$, then Theorem 3.1 shows that $\alpha_k > \alpha_l^{(k)}$, and thus the modified updating algorithm sets

$$I_{k+1} = [\alpha_l^{(k)}, \alpha_k]$$

because case U2 does not hold. Moreover, Theorem 3.1 guarantees that any $\alpha \in I_{k+1}$ with $\phi(\alpha) \le \phi(\alpha_k)$ satisfies (1.1). This implies that for any iteration $j > k$ the endpoint $\alpha_l^{(j)}$ satisfies (1.1). We also know that any sequence $\{\theta_k\}$ with $\theta_k \in I_k$ must converge to a common limit α^*. Since Theorem 3.2 shows that there is a $\theta_k \in I_k$ such that $\phi'(\theta_k) = 0$, we obtain that $\phi'(\alpha^*) = 0$. Hence, $\alpha_l^{(j)}$ satisfies (1.2) for all $j > k$ sufficiently large. This proves that the modified search terminates at an iterate that satisfies (1.1) and (1.2). ■

4 Trial Value Selection

Given the endpoints α_l and α_u of the interval I, and a trial value α_t in I, the updating algorithm described in the preceding section produces an interval I_+ that contains acceptable points. We now specify the new trial value α_t^+ in I_+.

We assume that in addition to the endpoints α_l and α_u, and the trial point α_t, we have function values f_l, f_u, f_t and derivatives g_l, g_u, g_t. The function values f_l, f_u, f_t and derivatives g_l, g_u, g_t can be obtained from either the function ϕ or the auxiliary function ψ. The function and derivative values are obtained from the auxiliary function ψ until some iterate satisfies the test $\psi(\alpha_k) \le 0$ and $\psi'(\alpha_k) \ge 0$. Once this test is satisfied, ϕ is used.

We have divided the trial value selection in four cases. In the first two cases we choose α_t^+ by interpolating the function values at α_l and α_t so that the trial value α_t^+ lies between α_l and α_t. We define α_t^+ in terms of α_c (the minimizer of the cubic that interpolates f_l, f_t, g_l, and g_t), α_q (the minimizer of the quadratic that interpolates f_l, f_t, and g_l), and α_s (the minimizer of the quadratic that interpolates g_l and g_t).

Case 1: $f_t > f_l$. In this case compute α_c, α_q, and set

$$\alpha_t^+ = \begin{cases} \alpha_c & \text{if } |\alpha_c - \alpha_l| < |\alpha_q - \alpha_l| \\ \frac{1}{2}(\alpha_q + \alpha_c) & \text{otherwise.} \end{cases}$$

Both α_c and α_q lie in I_+ so they are both candidates for α_t^+. We desire a choice that is close to α_l since this is the point with the lowest function value. Both α_q and α_c are relatively close to α_l because

$$|\alpha_c - \alpha_l| \le \frac{2}{3}|\alpha_u - \alpha_l|, \qquad |\alpha_q - \alpha_l| \le \frac{1}{2}|\alpha_u - \alpha_l|.$$

Thus, for the above choice of α_t^+,

$$|\alpha_t^+ - \alpha_l| \le \frac{7}{12}|\alpha_u - \alpha_l|.$$

A choice close to α_l is clearly desirable when f_t is much larger than f_l. In this case the quadratic step is closer to α_l than α_c, but usually abnormally so. Indeed, if $\alpha_q(f_t)$ is the value of α_q as a function of f_t, then

$$\lim_{f_t \to \infty} \alpha_q(f_t) = \alpha_l.$$

On the other hand, a computation shows that

$$\lim_{f_t \to \infty} \alpha_c(f_t) = \alpha_l + \frac{2}{3}(\alpha_u - \alpha_l).$$

Thus, the midpoint of α_c and α_t is a reasonable compromise.

Case 2: $f_t \le f_l$ and $g_t g_l < 0$. In this case compute α_c, α_s, and set

$$\alpha_t^+ = \begin{cases} \alpha_c & \text{if } |\alpha_c - \alpha_t| \ge |\alpha_s - \alpha_t| \\ \alpha_s & \text{otherwise.} \end{cases}$$

Both α_c and α_s lie in I_+ so they are both candidates for α_t^+. Since $g_t g_l < 0$, a minimizer lies between α_l and α_t. Choosing the step that is farthest from α_t tends to generate a step that straddles a minimizer, and thus the next step is also likely to fall in this case.

In the next case we choose α_t^+ by extrapolating the function values at α_l and α_t, so the trial value α_t^+ lies outside of the interval with α_t and α_l as endpoints. We define α_t^+ in terms of α_c (the minimizer of the cubic that interpolates f_l, f_t, g_l, and g_t) and α_s (the minimizer of the quadratic that interpolates g_l and g_t).

Case 3: $f_t \le f_l$, $g_t g_l \ge 0$, and $|g_t| \le |g_l|$. In this case the cubic that interpolates the function values f_l and f_t and the derivatives g_l and g_t may not have a minimizer. Moreover, even if the minimizer α_c exists, it may be in the wrong direction. For example, we may have $\alpha_t > \alpha_l$ but $\alpha_c < \alpha_t$. On the other hand, the secant step α_s always exists and is in the right direction.

If the cubic tends to infinity in the direction of the step and the minimum of the cubic is beyond α_t, set

$$\alpha_t^+ = \begin{cases} \alpha_c & \text{if } |\alpha_c - \alpha_t| < |\alpha_s - \alpha_t| \\ \alpha_s & \text{otherwise.} \end{cases}$$

Otherwise, set $\alpha_t^+ = \alpha_s$. This choice is based on the observation that during extrapolation it is sensible to be cautious and choose the step closest to α_t.

The trial value α_t^+ defined above may be outside of the interval with α_t and α_u as endpoints, or it may be in this interval but close to α_u. Either situation is undesirable, so we redefine α_t^+ by setting

$$\alpha_t^+ = \begin{cases} \min\{\alpha_t + \delta(\alpha_u - \alpha_t), \alpha_t^+\} & \text{if } \alpha_t > \alpha_l \\ \max\{\alpha_t + \delta(\alpha_u - \alpha_t), \alpha_t^+\} & \text{otherwise} \end{cases}$$

for some $\delta < 1$. In our algorithm we use $\delta = 0.66$.

In the last case the information available at α_l and α_t indicates that the function is decreasing rapidly in the direction of the step, but there does not seem to be a good way to choose α_t^+ from the available information.

Case 4: $f_t \le f_l$, $g_t g_l \ge 0$, and $|g_t| > |g_l|$. In this case we choose α_t^+ as the minimizer of the cubic that interpolates f_u, f_t, g_u, and g_t,

5 Numerical Results

The set of test problems that we use to illustrate the behavior of the search algorithm includes convex and general functions. The first three functions have regions of concavity, while the last three functions are convex. In all cases the functions have a unique minimizer. Our numerical results were done in double precision on an IPX Sparcstation.

The region of concavity of the first function in the test set is to the right of the minimizer, while the second function is concave to the left of the minimizer. The first function is defined by

$$\phi(\alpha) = -\frac{\alpha}{(\alpha^2 + \beta)} \tag{5.1}$$

with $\beta = 2$, while the second function is defined by

$$\phi(\alpha) = (\alpha + \beta)^5 - 2(\alpha + \beta)^4 \tag{5.2}$$

with $\beta = 0.004$. Plots for these two functions appear in Figures 5.1 and 5.2.

The third function in the test set was suggested by Paul Plassmann. This function is defined in terms of the parameters l and β by

$$\phi(\alpha) = \phi_0(\alpha) + \frac{2(1-\beta)}{l\pi} \sin(\frac{l\pi}{2}\alpha), \tag{5.3}$$

where

$$\phi_0(\alpha) = \begin{cases} 1 - \alpha & \text{if } \alpha \le 1 - \beta \\ \alpha - 1 & \text{if } \alpha \ge 1 + \beta \\ \frac{1}{2\beta}(\alpha - 1)^2 + \frac{1}{2}\beta & \text{if } \alpha \in [1 - \beta, 1 + \beta]. \end{cases}$$

The parameter β controls the size of $\phi'(0) = -\beta$. This parameter also controls the size of the interval where (1.2) holds because $|\phi'(\alpha)| \ge \beta$ for $|\alpha - 1| \ge \beta$, and thus (1.2) can hold only for $|\alpha - 1| < \beta$. The parameter l controls the number of oscillations in the function for $|\alpha - 1| \ge \beta$ because in that interval $\phi''(\alpha)$ is a multiple of $\sin(\frac{l\pi}{2}\alpha)$. Note that if l is odd, then $\phi'(1) = 0$, and that if $l = 4k - 1$ for some integer $k \ge 1$, then $\phi''(1) > 0$. Also note that ϕ is convex for $|\alpha - 1| < \beta$ if

$$\beta(1 - \beta)\frac{l\pi}{2} \le 1.$$

We have chosen $\beta = 0.01$ and $l = 39$. A plot of this function with these parameter settings appears in Figure 5.3.

The other three functions in the test set are from the paper of Yanai, Ozawa, and Kaneko [18]. These functions are defined in terms of parameters β_1 and β_2 by

$$\phi(\alpha) = \gamma(\beta_1)[(1 - \alpha)^2 + \beta_2^2]^{\frac{1}{2}} + \gamma(\beta_2)[\alpha^2 + \beta_1^2]^{\frac{1}{2}}, \tag{5.4}$$

where

$$\gamma(\beta) = (1 + \beta^2)^{\frac{1}{2}} - \beta.$$

These functions are convex, but different choices of β_1 and β_2 lead to functions with quite different characteristics. This can be seen clearly in Figures 5.4, 5.5, and 5.6.

In the tables below we present numerical results for different values of α_0. We have used $\alpha_0 = 10^i$ for $i = \pm 1, \pm 3$. This illustrates the behavior of the algorithm from different starting points. We are particularly interested in the behavior from the remote starting points $\alpha_0 = 10^{\pm 3}$.

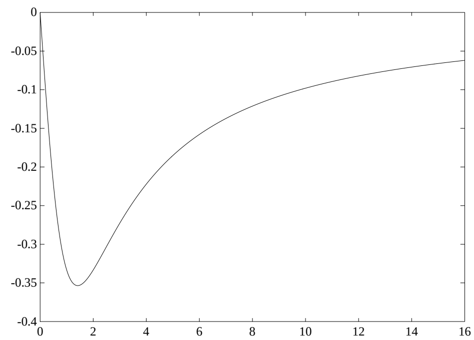

Figure 5.1: Plot of function (5.1) with $\beta = 2$

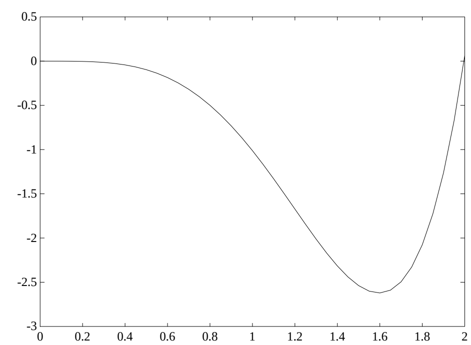

Figure 5.2: Plot of function (5.2) with $\beta = 0.004$

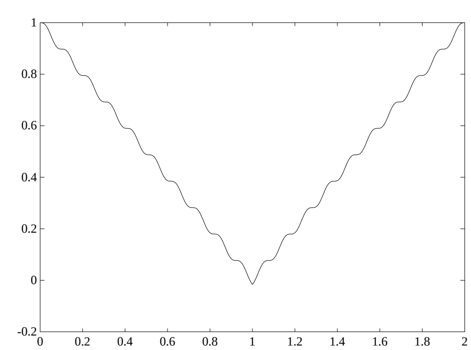

Figure 5.3: Plot of function (5.3) with $\beta = 0.01$ and $l = 39$

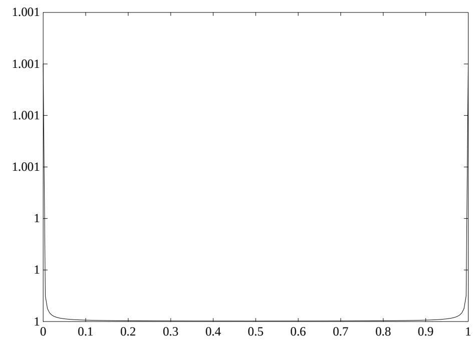

Figure 5.4: Plot of function (5.4) with $\beta_1 = 0.001$ and $\beta_2 = 0.001$

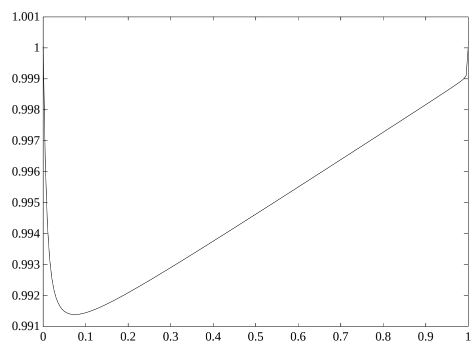

Figure 5.5: Plot of function (5.4) with $\beta_1 = 0.01$ and $\beta_2 = 0.001$

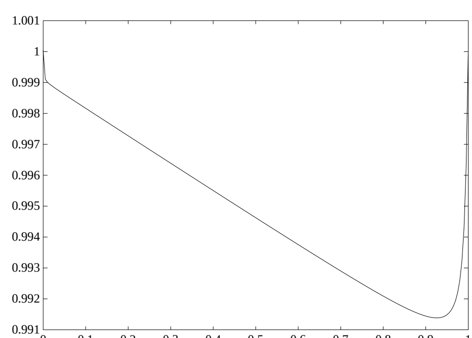

Figure 5.6: Plot of function (5.4) with $\beta_1 = 0.001$ and $\beta_2 = 0.01$

Table 5.1: Results for the function in Figure 5.1 with $\mu = 0.001$ and $\eta = 0.1$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	6	1.4	$-9.2\ 10^{-3}$
10^{-1}	1	3	1.4	$4.7\ 10^{-3}$
10^{+1}	1	1	10	$9.4\ 10^{-3}$
10^{+3}	1	4	37	$7.3\ 10^{-4}$

In our numerical results we have used different values of μ and η in order to illustrate different features of the problems and the search algorithm. In many problems we have used $\eta = 0.1$ because this value is typical of those used in an optimization setting. We comment on what happens for other values of μ and η. The general trend is for the number of function evaluations to decrease if μ is decreased or if η is increased. The reason for this trend is that as μ is decreased or η is increased, the measure of the set of acceptable values of α increases.

An interesting feature of the numerical results for the function in Figure 5.1 is that values of α much larger than $\alpha^* \approx 1.4$ can satisfy (1.1) and (1.2). This should be clear from Figure 5.1 and from the results in Table 5.1. These results show that if we use $\mu = 0.001$ and $\eta = 0.1$, then the starting point $\alpha_0 = 10$ satisfies (1.1) and (1.2), and thus the search algorithm exits with α_0. Similarly, the search algorithm exits with $\alpha_4 \approx 37$ when the starting point is $\alpha_0 = 10^{+3}$.

We can avoid termination at points far away from the minimizer α^* by increasing μ or decreasing η. If we increase μ and set $\mu = \eta = 0.1$, then the algorithm terminates with $\alpha_3 \approx 1.6$ when $\alpha_0 = 10$ and with $\alpha_7 \approx 1.6$ when $\alpha_0 = 10^{+3}$. There is no change in the behavior of the algorithm from the other two starting points. If we decrease η by setting $\eta = 0.001$ but leave μ unchanged at $\mu = 0.1$, then the final iterate α_m is near α^* for all starting points. For $\eta = 0.001$ the search algorithm needs six function evaluations for $\alpha_0 = 10$ and ten function evaluations for $\alpha_0 = 10^{+3}$. The number of function evaluations for $\alpha_0 = 10^{-3}$ and $\alpha_0 = 10^{-1}$ is, respectively, 8 and 4. This increase in the number of function evaluations is to be expected because now the set of acceptable α is smaller.

Another interesting feature of the results in Table 5.1 is that the six function evaluations needed for $\alpha_0 = 10^{-3}$ could have been predicted from the nature of the extrapolation process. This can be explained by noting that in a typical situation the extrapolation process generates iterates by setting $\alpha_t^+ = \alpha_t + \delta(\alpha_t - \alpha_l)$ with $\delta = 4$, and thus

$$\alpha_1 = 0.005, \quad \alpha_2 = 0.021, \quad \alpha_3 = 0.085, \quad \alpha_4 = 0.341, \quad \alpha_5 = 1.365,$$

Table 5.2: Results for the function in Figure 5.2 with $\mu = \eta = 0.1$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	12	1.6	$7.1\ 10^{-9}$
10^{-1}	1	8	1.6	$1.0\ 10^{-10}$
10^{+1}	1	8	1.6	$-5.0\ 10^{-9}$
10^{+3}	1	11	1.6	$-2.3\ 10^{-8}$

until the minimizer is bracketed, or until one of these iterates satisfies the termination conditions. This implies, for example, that if the minimizer is $\alpha^* = 1.4$, then either one of the above iterates satisfies (1.1) and (1.2), or at least six functions are evaluations are required before the search algorithm exits.

The number of function evaluations needed to find an acceptable α is usually dependent on the measure of the set of acceptable α. From this point of view, the only difficult test problems are those based on the functions in Figures 5.2 and 5.3, because for these functions the set of acceptable α is small. The choice of $\beta = 0.004$ for the function in Figure 5.2 guarantees that this function has a large region of concavity, but also forces $\phi'(0)$ to be quite small (approximately $-5\ 10^{-7}$). As a consequence (1.2) is quite restrictive for any $\eta < 1$. Similar remarks apply to the numerical results for the function in Figure 5.3. This is a difficult test problem because information based on derivatives is unreliable as a result of the oscillations in the function. Moreover, as already noted, (1.2) can hold only for $|\alpha - 1| < \beta$.

In Table 5.2 we present the numerical results for the function in Figure 5.2. In this table we have used $\mu = \eta = 0.1$, but these results remain unchanged if we set $\eta = 0.1$ and choose any $\mu < \eta$.

The number of function evaluations in Table 5.2 compares favorably with a search algorithm based on bisection. Given the starting value $\alpha_0 = 10$, a search algorithm based on bisection requires 48 function evaluations to determine an acceptable α because in this problem the set of acceptable α is an interval of approximate length $2.5\ 10^{-9}$. The comparison is even more favorable for the starting point $\alpha_0 = 10^{+3}$ because in this case a bisection algorithm requires 107 function evaluations.

For the function in Figure 5.3 the set of acceptable α is an interval of length 10^{-3}, so a bisection algorithm requires 10 function evaluations for the starting value $\alpha_0 = 10$, and 30 function evaluations for $\alpha_0 = 10^{+3}$. If we now compare this information with the numerical results in Table 5.3, we see that the search algorithm of this paper performs better than an algorithm based on bisection. This is surprising because for this function the information

Table 5.3: Results for the function in Figure 5.3 with $\mu = \eta = 0.1$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	12	1.0	$-5.1\ 10^{-5}$
10^{-1}	1	12	1.0	$-1.9\ 10^{-4}$
10^{+1}	1	10	1.0	$-2.0\ 10^{-6}$
10^{+3}	1	13	1.0	$-1.6\ 10^{-5}$

Table 5.4: Results for the function in Figure 5.4 with $\mu = \eta = 0.001$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	4	0.08	$-6.9\ 10^{-5}$
10^{-1}	1	1	0.10	$-4.9\ 10^{-5}$
10^{+1}	1	3	0.35	$-2.9\ 10^{-6}$
10^{+3}	1	4	0.83	$1.6\ 10^{-5}$

provided by ϕ' is unreliable.

The numerical results for the problems based on function (5.4) appear in Tables 5.4, 5.5, and 5.6. In all these tables we have chosen $\mu = \eta = 0.001$. Although these choices are not typical of those found in an optimization environment, they lead to more interesting results.

If we compare the results in these three tables, we notice that for a given starting point, the number of function evaluations sometimes differs considerably. The results in Tables 5.5 are typical of those that occur for $\eta = 0.001$. In examining the results in Table 5.5, allowances must be made for the fact that the starting points are not distributed symmetrically around the minimizer $\alpha^* \approx 0.074$. In particular, the small number of function evaluations for $\alpha_0 = 0.1$ is mainly due to the fact that in this case α_0 is close to α^*.

The number of function evaluations in Table 5.4 is lower because the set of acceptable α is unusually large. In particular, note that the value α_m returned by the search algorithm is not close to the minimizer $\alpha^* = \frac{1}{2}$ of the function in Figure 5.4.

The number of function evaluations in Table 5.6 is higher because in this problem it is difficult to determine an iterate α_k such that $\phi'(\alpha_k) > 0$ and α_k satisfies the sufficient decrease condition. Recall that once such an iterate is determined, we know that the problem has a minimizer that satisfies the sufficient decrease condition.

Table 5.5: Results for the function in Figure 5.5 with $\mu = \eta = 0.001$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	6	0.075	$1.9\ 10^{-4}$
10^{-1}	1	3	0.078	$7.4\ 10^{-4}$
10^{+1}	1	7	0.073	$-2.6\ 10^{-4}$
10^{+3}	1	8	0.076	$4.5\ 10^{-4}$

Table 5.6: Results for the function in Figure 5.6 with $\mu = \eta = 0.001$

α_0	info	m	α_m	$\phi'(\alpha_m)$
10^{-3}	1	13	0.93	$5.2\ 10^{-4}$
10^{-1}	1	11	0.93	$8.4\ 10^{-5}$
10^{+1}	1	8	0.92	$-2.4\ 10^{-4}$
10^{+3}	1	11	0.92	$-3.2\ 10^{-4}$

In an optimization setting, we would not tend to use $\eta = 0.001$, and then the number of function evaluations needed to obtain an acceptable α would decrease considerably. Consider, for example, the results for the function in Figure 5.6 with $\mu = 0.001$ and $\eta = 0.1$. For these settings, the number of function evaluations needed to obtain an acceptable α from the starting points $\alpha_0 = 10^i$ for $i = -3, -1, 1, 3$ would be, respectively, $2, 1, 3, 4$. Similar results would be obtained for the functions in Figures 5.4 and 5.5.

Acknowledgments. This work benefited from discussions and suggestions from many sources. We thank, in particular, Bill Davidon, Roger Fletcher, and Michael Powell. We are also grateful to Jorge Nocedal for his encouragement during the final phases of this work, to Paul Plassmann for providing the monster function of Section 5, and to Gail Pieper for her careful reading of the manuscript.

References

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line searches*, IMA J. Numer. Anal., 5 (1985), pp. 121–124.

[2] M. Al-Baali and R. Fletcher, *An efficient line search for nonlinear least squares*, J. Optim. Theory Appl., 48 (1984), pp. 359–377.

[3] R. H. Byrd, J. Nocedal, and Y. Yuan, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal., 24 (1987), pp. 1171–1190.

[4] J. E. Dennis and R. E. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, 1983.

[5] R. Fletcher, *Practical Methods of Optimization Volume 1: Unconstrained Optimization*, John Wiley & Sons, 1980.

[6] ——, *Practical Methods of Optimization*, John Wiley & Sons, Second ed., 1987.

[7] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optimization, 2 (1992), pp. 21–42.

[8] P. E. Gill and W. Murray, *Safeguarded steplength algorithms for optimization using descent methods*, Report NAC 37, National Physical Laboratory, Teddington, England, 1974.

[9] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *Two steplength algorithms for numerical optimization*, Report SOL 79-25, Stanford University, Systems Optimization Laboratory, Stanford, California, 1979.

[10] D. C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Programming, 45 (1989), pp. 503–528.

[11] J. J. Moré and D. C. Sorensen, *Newton's method*, in Studies in Numerical Analysis, G. H. Golub, ed., The Mathematical Association of America, 1984, pp. 29–82.

[12] D. P. O'Leary, *Robust regression computation using iteratively reweighted least squares*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 466–480.

[13] M. J. D. Powell, *Some global convergence properties of a variable metric method without line searches*, in Nonlinear Programming, R. W. Cottle and C. E. Lemke, eds., vol. IX of SIAM-AMS Proceedings, American Mathematical Society, 1976, pp. 53–72.

[14] T. Schlick and A. Fogelson, *TNPACK – A truncated Newton minimization package for large-scale problemns: I. Algorithms and usage*, ACM Trans. Math. Software, 18 (1992), pp. 46–70.

[15] ——, *TNPACK – A truncated Newton minimization package for large-scale problemns: II. Implementations examples*, ACM Trans. Math. Software, 18 (1992), pp. 71–111.

[16] D. F. Shanno and K. H. Phua, *Minimization of unconstrained multivariate functions*, ACM Trans. Math. Software, 2 (1976), pp. 87–94.

[17] ——, *Remark on Algorithm 500*, ACM Trans. Math. Software, 6 (1980), pp. 618–622.

[18] H. Yanai, M. Ozawa, and S. Kaneko, *Interpolation methods in one dimensional optimization*, Computing, 27 (1981), pp. 155–163.