Mathematical Programming Study 17 (1982) 77–90.
North-Holland Publishing Company

A MODIFICATION AND AN EXTENSION OF LEMARECHAL'S ALGORITHM FOR NONSMOOTH MINIMIZATION*

Robert MIFFLIN

Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164, U.S.A.

Received 20 November 1980
Revised manuscript received 14 September 1981

An algorithm is given for finding stationary points for constrained minimization problems having locally Lipschitz problem functions that are not necessarily convex or differentiable but are semismooth.

Key words: Nonsmooth Optimization, Nondifferentiable Programming, Constrained Minimization, Semismooth Functions.

1. Introduction

We consider the problem of minimizing f on $S=\{x\in\mathbf{R}^n: h(x)\le 0\}$ where f and h are real-valued locally Lipschitz continuous functions defined on $\mathbf{R}^n$. We give a modification and an extension of an algorithm due to Lemarechal [2] and show convergence to a stationary point of the problem if f and h also satisfy a weak 'semismoothness' [5, 6] hypothesis that is most likely satisfied by continuous functions arising in practical applications. The method is a feasible point descent method which combines a generalized cutting plane idea with quadratic approximation of some Lagrangian function. Even for the special case of no constraint function h (i.e., $S=\mathbf{R}^n$) and a convex objective function f, as considered in [2], this version differs from the original method, because of its rules for line search termination and the associated updating of the search direction finding subproblem. More specifically, our version does not require a user-specified uniform lower bound on the line search stepsizes. Instead, it uses a 'two-point' line search related to the one introduced in [5]. In this paper, we introduce a general 'α-function' whose values appear both in the subproblem constraints and in the stopping test for the line search. The line search stopping rules provide a direct generalization of those rules found useful for unconstrained minimization of a smooth function. The α-function concept along with the idea of bundling generalized gradients of f from points in S and those of h from points outside S is what deals with nonsmoothness arising either from nondifferentiability of the problem functions or from the presence of a constraint boundary.

*This material is based upon work supported by the National Science Foundation under Grant No. MCS 78–06716.

The algorithm requires a feasible starting point, i.e., an $x_1 \in S$, but requires no knowledge of f at infeasible points as do exact penalty function methods. If no such x_1 is available, then the algorithm can be used to minimize h starting from any point and if h is semiconvex [6] and there exists an $\hat{x}$ such that $h(\hat{x})<0$ then, by the convergence theorem in Section 4, the algorithm will find a feasible point in a finite number of iterations.

Also at feasible points the method requires no knowledge of h (other than h being nonpositive) as is required by many feasible point methods. In fact, since $h(x)$ may be replaced by $\max(h(x), 0)$ without changing the feasible set S, the constraint function may be assumed to be zero throughout S. The important question of scaling h (or its components if it is a maximum of several constraint functions) by a positive multiple is not discussed in this paper.

For a locally Lipschitz function F on $\mathbf{R}^n$ let ∂F denote the generalized gradient [1] of F, i.e., for $x \in \mathbf{R}^n$, $\partial F(x)$ is the convex hull (conv) of all limits of sequences of the form $\{\nabla F(x_k)\colon \{x_k\} \to x$ and F is differentiable at each $x_k\}$. Important properties of the point-to-convex set mapping $\partial F(\cdot)$ are uppersemicontinuity and local boundedness. If F is convex ∂F equals the subdifferential, i.e., for each $x \in \mathbf{R}^n$

$$g \in \partial F(x) \quad \text{if and only if} \quad F(y) \geq F(x) + \langle g, y-x\rangle \quad \text{for all } y \in \mathbf{R}^n. \tag{1}$$

If F is continuously differentiable (C^1) ∂F equals the (ordinary) gradient $\{\nabla F\}$. Furthermore, for many other functions F, such as those that are pieced together from C^1 functions, it is possible to determine ∂F or at least to give one element of $\partial F(x)$ at each x. For example, if $F(x) = \max[F_1(x)\ F_2(x), \dots, F_m(x)]$ where each F_i is C^1, then $\partial F(x) = \operatorname{conv}\{\nabla F_i(x)\colon F_i(x) = F(x)\}$. Such examples occur in decomposition, relaxation, duality, and/or exact penalty approaches to solving optimization problems. For further details, generalizations and related results see the references in the comprehensive nonsmooth optimization bibliography in [3] .

We say point $\bar{x} \in S$ is *stationary* for f on S if $0 \in M(\bar{x})$ where

$$M(x) = \begin{cases} \partial f(x), & \text{if } h(x) < 0, \\ \operatorname{conv}(\partial f(x) \cup \partial h(x)), & \text{if } h(x) = 0, \\ \partial h(x), & \text{if } h(x) > 0, \end{cases}$$

because $0 \in M(x^*)$ is a necessary condition for $x^* \in S$ to minimize f on S. However, as shown in [9] using an example of a nonsmooth nonmax function also given in [6], there may exist feasible directions of strict descent at a stationary point. As indicated in [5], it can be shown that the point-to-convex set mapping $M(\cdot)$ inherits the generalized gradient properties of uppersemicontinuity and local boundedness, which are required for showing convergence of the algorithm.

In order to implement the algorithm, we suppose that we have a subroutine that can evaluate a function $g(x) \in M(x)$ for each $x \in \mathbf{R}^n$. For ease of implementation and exposition, we also suppose that $g(x)\in\partial f(x)$ if $h(x)=0$. Of course, we are especially interested in the nonsmooth case where g is discontinuous at stationary points of the constrained minimization problem. Because g may have discontinuities, the algorithm employs a two-point line search. For example, if a line search proceeds along a direction that goes outside S, it must obtain two points, one feasible and the other infeasible, in order to simultaneously maintain feasibility and take the constraint boundary into account properly without requiring knowledge of ∂h at feasible points. Even for line searches along feasible directions, the method may need g-values on both sides of a discontinuity in the gradient of f. For example, if $f(x)=|x|=\max[x, -x]$ for $x\in\mathbf{R}$, then $g(x)=1$ for $x>0$ and $g(x)=-1$ for $x<0$, and we need to know both of these derivatives in order to identify $x=0$ as a minimizing point of f, i.e., in order to conclude that $0\in\partial f(0)=\operatorname{conv}\{-1, 1\}$. To insure that g-values taken at points near a g-discontinuity are close to being generalized gradients at the discontinuity the algorithm makes use of an α-function that is defined as follows:

Definition. Associated with f, h and g let $\alpha : S\times\mathbf{R}^n\to\mathbf{R}_+$ be a nonnegative-valued function satisfying

$$\{\alpha(x_k, y_k)\}\to 0, \qquad \text{if } \{(x_k, y_k)\}\to(\bar{x}, \bar{x}), \tag{2a}$$

$$\{\alpha(w_k, y_k)-\alpha(x_k, y_k)\}\to 0, \quad \text{if } \{(w_k, x_k, y_k)\}\to(\bar{x}, \bar{x}, \bar{y}), \tag{2b}$$

and

$$\bar{g}\in M(\bar{x}), \text{if } \{(x_k, y_k, g(y_k))\}\to(\bar{x}, \bar{y}, \bar{g}) \quad \text{and} \quad \{\alpha(x_k, y_k)\}\to 0. \tag{2c}$$

$\alpha(x, y)$ is intended to be an indication of how much $g(y)\in M(y)$ deviates from being an element of M at x. However, this measure may be somewhat arbitrary, because a fixed positive multiple of an α-function is also an α-function.

For a convex problem we may take α as a measure of deviation from linearity as indicated in the following lemma. Let S^c be the complement of S and note that S is closed, because h is continuous.

Lemma 1. *Suppose f and h are convex on $\mathbf{R}^n$. Then*

$$\alpha(x, y)=\begin{cases} f(x)-f(y)-\langle g(y), x-y\rangle & \text{for } (x, y)\in S\times S, \\ -h(y)-\langle g(y), x-y\rangle & \text{for } (x, y)\in S\times S^c \end{cases}$$

is nonnegative and satisfies (2).

Proof. If $y\in S$, then $g(y)\in\partial f(y)$ and by the convexity of f, the definition of α and the subgradient inequality (1) we have $\alpha(x, y)\geq 0$. If $y\notin S$, then $g(y)\in\partial h(y)$ and by the convexity of h

$$h(z)\geq h(y)+\langle g(y), z-y\rangle \quad \text{for all } z\in\mathbf{R}^n.$$

Then for $z=x\in S$ we have

$$0\geq h(x)\geq h(y)+\langle g(y), x-y\rangle=-\alpha(x, y).$$

So $\alpha(x, y)$ is nonnegative in either case.

The proofs of properties (2a) to (2c) require separating any sequence $\{y_k\}$ into two subsequences, one in S and the other in S^c, and then making separate arguments for each case as was done above for the nonnegativity proof. We will consider the latter case and omit the former for it has very similar arguments. So, for each sequence $\{y_k\}\to\bar{y}$ considered below, suppose $y_k\notin S$ for all k. Then

$$h(y_k)>0, \qquad g(y_k)\in\partial h(y_k),$$
$$\alpha(x_k, y_k) = -h(y_k)-\langle g(y_k), x_k - y_k\rangle \quad \text{for } x_k\in S, \tag{3}$$

and, by the continuity of h,

$$h(\bar{y})\geq 0. \tag{4}$$

Furthermore, suppose for each sequence $\{x_k\}\to\bar{x}$ considered below that $\{x_k\}\subset S$, so that

$$h(\bar{x})\leq 0. \tag{5}$$

To show (2a) suppose that $\bar{y}=\bar{x}$. Then, by (4) and (5)

$$0\geq h(\bar{x}) = h(\bar{y})\geq 0, \tag{6}$$

The local boundedness of ∂h implies

$$\{-h(y_k)-\langle g(y_k), x_k-y_k\rangle\}\to -h(\bar{y}). \tag{7}$$

Then (2a) follows from (3), (7) and (6).

To show (2b) suppose $\{w_k\}\to\bar{x}$. Since $\alpha(w_k, y_k)-\alpha(x_k, y_k)=\langle g(y_k), x_k-w_k\rangle$, (2b) follows from the local boundedness of ∂h.

To show (2c) suppose $\{\alpha(x_k, y_k)\}\to 0$ and $\{g(y_k)\}\to\bar{g}$. Then

$$\{-h(y_k)-\langle g(y_k), x_k-y_k\rangle\}\to 0,$$

so, by continuity,

$$-h(\bar{y})-\langle \bar{g}, \bar{x}-\bar{y}\rangle=0. \tag{8}$$

Furthermore, by uppersemicontinuity of ∂h, $\bar{g}\in\partial h(\bar{y})$, so, by convexity of h,

$$h(z)\geq h(\bar{y})+\langle \bar{g}, z-\bar{y}\rangle \quad \text{for all } z\in\mathbf{R}^n,$$

which combined with (4) gives

$$h(z)\geq\langle \bar{g}, z-\bar{x}\rangle \quad \text{for all } z\in\mathbf{R}^n. \tag{9}$$

Setting $z=\bar{x}$ gives $h(\bar{x})\geq 0$, which combined with (5) yields

$$h(\bar{x})=0. \tag{10}$$

Finally, (9) and (10) imply

$$h(z) \geq h(\bar{x}) + \langle \bar{g}, z - x \rangle \quad \text{for all } z \in \mathbf{R}^n,$$

which, by convexity of h, *implies that* $\bar{g} \in \partial h(\bar{x})$ and together with (10) implies that $\bar{g} \in M(\bar{x})$ and completes the proof.

For general problems it can be shown that the following function is satisfactory:

$$\alpha(x, y) = \begin{cases} \max[f(x) - f(y) - \langle g(y), x - y \rangle, \beta_0 |x - y|^2] & \text{for } (x, y) \in S \times S, \\ \max[-h(y) - \langle g(y), x - y \rangle, \beta_1 |x - y|^2] & \text{for } (x, y) \in S \times S^c, \end{cases}$$

where β_0 and β_1 are positive parameters and $|\cdot|$ denotes Euclidean norm. Note that if either problem function is known to be convex, then the corresponding β-parameter may be set equal to zero.

To motivate the search direction finding subproblem employed by the algorithm consider the simplest case where f is convex and the problem is unconstrained, i.e., $S = \mathbf{R}^n$. Suppose that f and g have been evaluated at x and y. A polyhedral approximation to $f(x+d)$ which, by convexity, agrees with $f(x)$ and $f(y)$ when $d = 0$ and $d = y - x$, respectively, is given by

$$\max[f(x) + \langle g(x), d \rangle, f(y) + \langle g(y), d - y + x \rangle]$$

or, using the definition of α given in Lemma 1,

$$\max[f(x) + \langle g(x), d \rangle, f(x) + \langle g(y), d \rangle - \alpha(x, y)].$$

Minimizing this polyhedral function of d is equivalent to the problem

$$\begin{aligned} &\underset{(w,d) \in \mathbf{R}^{1+n}}{\text{minimize}} && w, \\ &\text{subject to} && w \geq f(x) + \langle g(x), d \rangle, \\ &&& w \geq f(x) + \langle g(y), d \rangle - \alpha(x, y). \end{aligned}$$

A quadratic approximation to $f(x+d)$ is given by

$$f(x) + \langle g(x), d \rangle + \tfrac{1}{2} \langle d, Gd \rangle$$

where G is an n by n matrix approximating the curvature of f, for example, satisfying $G(x - z) = g(x) - g(z)$ where z is some past point (not necessarily y) at which g has been evaluated and there may be some indication that f is smooth along the line segment from z to x. Minimizing the above quadratic function of d is equivalent to the problem

$$\begin{aligned} &\underset{(w,d) \in \mathbf{R}^{1+n}}{\text{minimize}} && w + \tfrac{1}{2} \langle d, Gd \rangle, \\ &\text{subject to} && w \geq f(x) + \langle g(x), d \rangle. \end{aligned}$$

Merging these two problems and letting $v = w - f(x)$ give the following combined polyhedral-quadratic approximation subproblem used by the algorithm:

$$\begin{aligned} &\underset{(v,d)\in\mathbf{R}^{1+n}}{\text{minimize}} && v+\tfrac{1}{2}\langle d, Gd\rangle+f(x),\\ &\text{subject to} && v\geq\langle g(x), d\rangle,\\ &&& v\geq\langle g(y), d\rangle-\alpha(x, y). \end{aligned}$$

2. The algorithm

Let m_L and m_R be fixed parameters satisfying $0<m_L<m_R<1$ and (in case the problem is smooth) $m_L<\frac{1}{2}$.

Suppose initially that $x_1\in S$ and let $y_1=x_1$ and G_1 be a positive definite $n\times n$ matrix, such as the identity matrix. Note that, by the definitions of g and M, $g(y_1)\in\partial f(x_1)$ and, by property (2a), $\alpha(x_1, y_1)=0$.

In general, given a positive iteration integer k, a feasible point $x_k\in S$, generalized gradients $g(y_i)\in M(y_i)$ and corresponding scalars $\alpha(x_k, y_i)$ where $y_i\in\mathbf{R}^n$ for $i=1,2,\dots,k$, and a positive definite $n\times n$ matrix G_k solve for $(d, v)=(d_k, v_k)\in\mathbf{R}^{n+1}$ the kth quadratic programming subproblem:

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}\langle d, G_k d\rangle + v,\\ &\text{subject to} && -\alpha(x_k, y_i)+\langle g(y_i), d\rangle\leq v \quad \text{for } i=1,2,\dots,k. \end{aligned}$$

If $v_k=0$ stop. Otherwise, by a line search procedure as given below, find (if possible) two stepsizes t_L and t_R such that $0\leq t_L\leq t_R$ and such that the two corresponding points defined by

$$x_{k+1}=x_k+t_L d_k \quad\text{and}\quad y_{k+1}=x_k+t_R d_k$$

satisfy

$$h(x_{k+1})\leq 0, \tag{11a}$$

$$f(x_{k+1})\leq f(x_k)+m_L t_L v_k, \tag{11b}$$

and

$$-\alpha(x_{k+1}, y_{k+1})+\langle g(y_{k+1}), d_k\rangle\geq m_R v_k. \tag{11c}$$

If the line search procedure is successful, define the $(k+1)$st subproblem by replacing in the subproblem constraints $\alpha(x_k, y_i)$ by $\alpha(x_{k+1}, y_i)$ for $i=1,2,\dots,k$, appending the constraint

$$-\alpha(x_{k+1}, y_{k+1})+\langle g(y_{k+1}), d\rangle\leq v$$

and replacing G_k by a positive definite matrix G_{k+1}.

3. Remarks on the algorithm

If (d_k, v_k) solves the kth subproblem, then, necessarily, there exist multipliers λ_{ik} for $i=1,2,\dots,k$ such that

$$\lambda_{ik}\geq 0, \tag{12a}$$

$$\sum_{i=1}^{k} \lambda_{ik} = 1, \tag{12b}$$

$$G_k d_k = -\sum_{i=1}^{k} \lambda_{ik} g(y_i) \tag{12c}$$

and

$$v_k = \langle g(y_i), d_k \rangle - \alpha(x_k, y_i) \quad \text{if } \lambda_{ik} > 0. \tag{12d}$$

Combining (12a) to (12d) gives

$$v_k = -\langle G_k d_k, d_k \rangle - \sum_{i=1}^{k} \lambda_{ik} \alpha(x_k, y_i) \tag{12e}$$

or, since G_k is positive semidefinite and λ_{ik} and α are nonnegative

$$v_k = -\langle G_k^{1/2} d_k, G_k^{1/2} d_k \rangle - \sum_{i=1}^{k} \lambda_{ik} \alpha(x_k, y_i) \leq 0 \tag{13}$$

where $G_k^{1/2} G_k^{1/2} = G_k$.

From (12c) and (12e), we conclude that v_k can be thought of as an approximate directional derivative of some Lagrangian function at x_k in the direction d_k. From (13) we conclude that if $v_k \neq 0$ (i.e., the algorithm does not terminate) then the line search procedure is entered with $v_k < 0$. The next result justifies termination when $v_k = 0$.

Lemma 2. *If $v_k = 0$, then x_k is stationary for f on S.*

Proof. If $v_k = 0$, then, from (13) and the nonnegativity of λ_{ik} and α, we have

$$\lambda_{ik} \alpha(x_k, y_i) = 0 \quad \text{for } i = 1, 2, \dots, k \tag{14a}$$

and

$$G_k^{1/2} d_k = 0. \tag{14b}$$

Multiplying (14b) by $G_k^{1/2}$ and combining the result with (12c) gives

$$\sum_{i=1}^{k} \lambda_{ik} g(y_i) = 0. \tag{14c}$$

Furthermore, (14a) implies that $\alpha(x_k, y_i) = 0$ if $\lambda_{ik} > 0$. Thus, by property (2c), $g(y_i)$ is an element of the convex set $M(x_k)$ for each i such that $\lambda_{ik} > 0$. Now, stationary follows from (14c) and the fact that the λ_{ik} form a convex combination.

Note that the above discussion only requires G_k to be positive semidefinite. When G_k is positive semidefinite the subproblem has a bounded solution if there exist d_k and λ_{ik} for $i = 1, 2, \dots, k$ satisfying (12c). Assuming G_k to be positive definite guarantees that (12c) can be satisfied and that the subproblem has a dual, as developed in [2], given by

$$\text{minimize} \quad \frac{1}{2}\left\langle \sum_{i=1}^{k} \lambda_i g(y_i), \sum_{i=1}^{k} \lambda_i G_k^{-1} g(y_i) \right\rangle + \sum_{i=1}^{k} \lambda_i \alpha(x_k, y_i),$$

$$\text{subject to} \quad \sum_{i=1}^{k} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0 \quad \text{for } i = 1, 2, \ldots, k.$$

Elsewhere, we will report on how to extend the numerically stable constrained least squares algorithm in [7] for solving more general quadratic programming problems. A specialization of this new method will result in a reliable method for solving subproblem dual given here.

The constraints in the subproblem are present to determine the active generalized gradients (and their multipliers) at optimiality and the objective matrix G_k is present to represent the curvature (if any) of a Lagrangian associated with f and h with respect to directions that are orthogonal to all of the active gradients. Current research centers on finding out when and how

(a) to make some type of variable metric update of G_k based on differencing nonnegative combinations of gradients, i.e., differencing approximate gradients of a Lagrangian;

(b) to aggregate, reduce, or restart the constraint 'bundle'.

We speculate that the above combination of cutting-plane and quasi-Newton aspects are needed for the algorithm to have the potential for some kind of superlinear convergence. The details for bounding the number of subproblem constraints and for updating G_k in order to obtain superlinear convergence in the one-dimensional unconstrained convex case where $n=1$ and f is convex on $S=\mathbf{R}$ are given in [4]. This research (incomplete in the higher dimensional and constrained cases) is not discussed here, for the purpose of this paper is to establish some general convergence theory.

Line search stopping criteria (11a) and (11b) guarantee that x_{k+1} is feasible with a sufficiently smaller objective value than x_k and (11c) makes (d_k, v_k) sufficiently infeasible in subproblem $k+1$, because $v_k<0$ and $m_R<1$ imply that

$$-\alpha(x_{k+1}, y_{k+1}) + \langle g(y_{k+1}), d_k\rangle \geq m_R v_k > v_k. \tag{15}$$

Criteria (11b) and (11c) are generalizations of the line search stopping rules found useful in smooth minimization. (See [8], for example.) The smooth case appears when $g(x_k+td_k)=\nabla f(x_k+td_k)$ for all t, $v_k=\langle\nabla f(x_k), d_k\rangle$ (the exact directional derivative of f at x_k in direction d_k) and $t_R=t_L>0$, for then $y_{k+1}=x_{k+1}\neq x_k$ and, by (2a), $\alpha(x_{k+1}, y_{k+1})=0$. In general, discontinuities in g are detected and dealt with by allowing $y_{k+1}\neq x_{k+1}$, but imposing an upper bound on $\alpha(x_{k+1}, y_{k+1})$ via (15) which converges to zero if d_k and v_k converge to zero.

If f and h are weakly upper semismooth on $\mathbf{R}^n$ (see Appendix for the definition), then the following line search procedure either finds t_L and t_R or generates an increasing sequence $\{t_L\}$ such that $\{x_L+t_Ld_k\}\subset S$ and $\{f(x_k+t_Ld_k)\}\to -\infty$.

Line Search Procedure

Let ρ be a fixed parameter satisfying $0<\rho<\frac{1}{2}$.

Set $t=0$, $t_L=0$, and $t_u=+\infty$.

Loop: If $h(x_k+td_k)\le 0$ and $f(x_k+td_k)\le f(x_k)+m_L t v_k$
set $t_L=t$. Otherwise set $t_u=t$.
If $-\alpha(x_k+t_L d_k, x_k+td_k)+\langle g(x_k+td_k), d_k\rangle\ge m_R v_k$
set $t_R=t$ and return. Otherwise
if $t_u=+\infty$, choose $t\ge t_L+\rho$ by some extrapolation procedure,
or if t_u is finite choose $t\in[t_L+\rho(t_u-t_L), t_u-\rho(t_u-t_L)]$
by some interpolation procedure.
Go to Loop.

If the return never occurs, then the successive choices of t either cause $t_L\uparrow+\infty$ if t_u remains $+\infty$ or cause $(t_u-t_L)\downarrow 0$ if t_u ever becomes finite. In the former case, f is unbounded on the feasible set, because $m_L>0$ and $v_k<0$ imply that $m_L t_L v_k\downarrow-\infty$. In the latter case, a contradiction can be established as in [5] if f and h are weakly upper-semismooth. Hence, if the return does not occur after a finite number of loops, then the unbounded case obtains when the problem functions satisfy this additional semismoothness hypothesis. This result depends on the α-function property (2a) and the parameter inequality $m_L<m_R$ as well as the line search procedure rules.

The possibility where $\langle g(x_k), d_k\rangle\ge m_R v_k$ requires special consideration. This case occurs if and only if the loop is executed only once and $t_R=t_L=0$. Then $y_{k+1}=x_{k+1}=x_k$ and, by property (2a), the appended subproblem constraint is $\langle g(x_{k+1}), d\rangle\le v$. So, if at the next iteration $v_{k+1}<0$, then, since $m_R<1$, we have

$$\langle g(x_{k+1}), d_{k+1}\rangle\le v_{k+1}<m_R v_{k+1}.$$

Thus, defining $K_0=\{l: \langle g(x_l), d_l\rangle<m_R v_l\}$ we have that if $k\notin K_0$, then $x_{k+1}=x_k$ and either $v_{k+1}=0$ or $k+1\in K_0$. Therefore, if the algorithm does not terminate K_0 is an infinite set.

The final result of this section shows that in the convex case if the first positive t-value chosen does not exceed 1 and if it fails to become t_L, then the line search stops after only one evaluation of (h, f, g) with t_R equal to this t-value. On the other hand, if this first positive t-value does become t_L, then this value is a positive lower bound on the terminal value of t_L.

Lemma 3. *Suppose f and h are convex on $\mathbf{R}^n$ and the form of α is that given in Lemma 1. Also, suppose that if the line search does not terminate with $t_R=t_L=0$, then t is set equal to $t_1\in(0, 1]$ for the next loop.*

(a) *If t_u is set equal to t_1 at the next loop, then termination occurs with $t_R=t_1$ and $t_L=0$.*

(b) *Otherwise (i.e., t_L is set equal to t_1 at the next loop), then at termination $t_1\le t_L\le t_R$.*

Proof. (a) If the next loop execution leaves $t_L=0$ and sets $t_u=t_1$, then either

$$h(x_k+t_1d_k)>0 \tag{16a}$$

or

$$f(x_k+t_1d_k)>f(x_k)+m_Lt_1v_k.$$

We give the proof in the former case, for the latter case is similar and essentially contained in [2; p. 22]. So in the former case from the definition of α

$$-\alpha(x_k+t_Ld_k, x_k+t_1d_k)=h(x_k+t_1d_k)-t_1\langle g(x_k+t_1d_k), d_k\rangle$$

which combined with (16a) implies

$$-\alpha(x_k+t_Ld_k, x_k+t_1d_k)+\langle g(x_k+t_1d_k), d_k\rangle>(1-t_1)\langle g(x_k+t_1d_k), d_k\rangle. \tag{16b}$$

Now, by the convexity of h,

$$h(x_k+t_1d_k)-t_1\langle g(x_k+t_1d_k), d_k\rangle\leq h(x_k),$$

and, since $t_1>0$, $h(x_k)\leq 0$, and $h(x_k+t_1d_k)>0$, we have

$$0<[h(x_k+t_1d_k)-h(x_k)]/t_1\leq\langle g(x_k+t_1d_k), d_k\rangle. \tag{16c}$$

The assumption that $t_1\leq 1$ combined with (16b) and (16c) gives

$$-\alpha(x_k+t_Ld_k, x_k+t_1d_k)+\langle g(x_k+t_1d_k), d_k\rangle>0$$

which implies termination with $t_R=t_1$, because $m_Rv_k<0$.

(b) The conclusion follows from the fact that the successive values for t_L are increasing.

4. Convergence

Suppose that each execution of the line search procedure is finite and that the algorithm does not terminate. Then $v_k<0$ for all k and K_0 is an infinite set. The three lemmas that we prove in this section establish part (a) of the following convergence theorem. Part (b) follows from part (a), because, if f and h are semiconvex (see Appendix for the definition) and a constraint qualification is satisfied, then any stationary point is a minimizing point [6] and because every accumulation point of $\{x_k\}\subset S$ has the same f-value due to the monotonicity of $\{f(x_k)\}$.

Theorem 1. *Suppose $\{x_k\}$, $\{y_k\}$ and $\{G_k\}$ are bounded with $\{G_k\}$ uniformly positive definite. Then*

(a) *at least one of the accumulation points of $\{x_k\}$ is stationary for f on S; and*

(b) *if f and h are semiconvex on $\mathbf{R}^n$ and there exists an $\hat{x}\in\mathbf{R}^n$ such that $h(\hat{x})<0$, then every accumulation point of $\{x_k\}$ minimizes f on S.*

Remark. If $\{x\in\mathbf{R}^n: x\in S, f(x)\le f(x_1)\}$ is bounded, then $\{x_k\}$ is bounded and $\{y_k\}$ can be made bounded by choosing an additional parameter $\epsilon>0$ and imposing the additional line search requirement that

$$|y_{k+1}-x_{k+1}|=(t_R-t_L)|d_k|\le\epsilon.$$

For weakly uppersemismooth functions, it is possible to satisfy this condition and (11) simultaneously after a finite number of line search loops.

Let $\gamma_k>0$ be the smallest eigenvalue of the positive definite matrix G_k. Consider the following assumption that is trivially satisfied if the matrix sequence $\{G_k\}$ is uniformly positive definite:

If $\{d_k\}_{k\in K_0}$ has no zero accumulation point, then

$$\{\gamma_k\}_{k\in K_0} \text{ has no zero accumulation point.} \tag{17}$$

Lemma 4. *Suppose* (17) *holds and* $\{x_k\}$ *and* $\{y_k\}$ *are uniformly bounded. Then* $\{d_k\}_{k\in K_0}$ *has at least one zero accumulation point.*

Proof. Suppose for purposes of a proof by contradiction that there exists a positive number δ such that

$$|d_k|\ge\delta>0 \quad \text{for all } k\in K_0. \tag{18}$$

Then, by (17), there exists a positive number γ such that

$$\gamma_k\ge\gamma>0 \quad \text{for all } k\in K_0. \tag{19}$$

Combining (13), the nonnegativity of λ_{ik} and α, (19) and (18) gives

$$v_k\le-\gamma|d_k|^2\le-\gamma\delta|d_k|\le-\gamma\delta^2<0 \quad \text{for all } k\in K_0. \tag{20}$$

By the Cauchy–Schwarz inequality and the definition of K_0

$$-|g(x_k)||d_k| \le \langle g(x_k), d_k\rangle < m_R v_k \quad \text{for all } k \in K_0.$$

which combined with the left-most inequality of (20) gives

$$|d_k|<|g(x_k)|/(m_R\gamma) \quad \text{for all } k\in K_0.$$

Thus, since $\{x_k\}$ is assumed bounded and $M(\cdot)$ is locally bounded, $\{g(x_k)\}$ and, hence $\{d_k\}_{k\in K_0}$ and $\{v_k\}_{k\in K_0}$ are bounded. Let $\bar{v}$ and $\bar{d}$ be accumulation points of $\{v_k\}_{k\in K_0}$ and $\{d_k\}_{k\in K_0}$, respectively. Then, by (20),

$$\bar{v}\le-\gamma\delta^2<0. \tag{21}$$

By (11b) and by (20) for $k\in K_0$

$$f(x_{k+1})-f(x_k)\le m_L t_L v_k\le-m_L t_L\gamma\delta|d_k|$$

or, since $x_{k+1}=x_k+t_L d_k$,

$$f(x_{k+1})-f(x_k)\le-m_L\gamma\delta|x_{k+1}-x_k|. \tag{22}$$

Note that (22) also holds for $k \notin K_0$, because in this case $x_{k+1}=x_k$.

For any $p>k+1$, (22) and the triangle inequality imply

$$f(x_p)-f(x_{k+1})=\sum_{j=k+1}^{p-1} f(x_{j+1})-f(x_j) \leq -m_L\gamma\delta \sum_{j=k+1}^{p-1} |x_{j+1}-x_j|$$
$$\leq -m_L\gamma\delta|x_p-x_{k+1}|.$$

As f is continuous and $\{x_k\}$ is assumed bounded, the monotone nonincreasing sequence $\{f(x_k)\}$ is bounded from below and, hence, there exists an $\bar{x} \in S$ such that

$$\{x_k\} \to \bar{x}. \tag{23}$$

Also, for any $p \geq k+1$ we have, by the pth subproblem feasibility, that

$$-\alpha(x_p, y_{k+1})+\langle g(y_{k+1}), d_p\rangle \geq v_p. \tag{24}$$

Subtracting (11c) from (24) gives

$$\alpha(x_{k+1}, y_{k+1})-\alpha(x_p, y_{k+1})+\langle g(y_{k+1}, d_p-d_k\rangle \leq v_p-m_R v_k. \tag{25}$$

Now choose p and k in K_1, an infinite subset of K_0, where $\{y_{l+1}\}_{l\in K_1}\to \bar{y}$ $\{d_l\}_{l\in K_1} \to \bar{d}$ and $\{v_l\}_{l\in K_1} \to \bar{v}$, so that from (23), (25), property (2b) and the boundedness of $\{g(y_{k+1})\}$ we have

$$0 \leq \bar{v}-m_R\bar{v}=(1-m_R)\bar{v}.$$

Since $m_R<1$, this implies that $\bar{v}\geq 0$, which contradicts (21) and completes the proof.

Lemma 5. *Suppose that* $K_1 \subseteq K_0$ *is such that* $\{d_k\}_{k\in K_1}\to 0$, $\{G_k d_k\}_{k\in K_1}\to 0$ *and* $\{\langle g(x_k), d_k\rangle\}_{k\in K_1}\to 0$. *Then* $\{v_k\}_{k\in K_1}\to 0$ *and*

$$\left\{\sum_{i=1}^{k} \lambda_{ik}\begin{pmatrix}\alpha(x_k, y_i)\\ g(y_i)\end{pmatrix}\right\}_{k\in K_1} \to \begin{pmatrix}0\\0\end{pmatrix}$$

where $\lambda_{ik} \geq 0$ *for* $i=1,2,\ldots,k$ *and* $\sum_{i=1}^{k} \lambda_{ik}=1$.

Proof. The conclusions follow from the hypotheses, the definition of K_0, (12) and (13).

Lemma 6. *In addition to the hypotheses of Lemma 5, suppose that* $\{x_k\}_{k\in K_1}$ *and* $\{y_k\}_{k\in K_1}$ *are bounded and let* $\bar{x}$ *be any accumulation point of* $\{x_k\}_{k\in K_1}$. *Then* $\bar{x}$ *is stationary for* f *on* S.

Proof. As in the proof of Theorem 5.2 in [5], depending on the local boundedness and uppersemicontinuity of M and the fact that a convex combination of vectors in $\mathbf{R}^{n+1}$ can be expressed as a convex combination of $n+2$ or fewer of the vectors, Lemma 5 implies the existence of a positive integer $m \leq n+2$, an infinite subset $K_2 \subset K_1$ and convergent subsequences

$$\{(y_k^i, g(y_k^i))\}_{k\in K_2} \to (y^i, g^i) \in \mathbf{R}^n \times M(y^i) \quad \text{and} \quad \{\mu_k^i\}_{k\in K_2} \to \mu^i \geq 0$$

for $i = 1, \dots, m$ such that

$$\sum_{i=1}^{m} \mu^i = 1, \qquad \sum_{i=1}^{m} \mu^i g^i = 0,$$

and for $i = 1, 2, \dots, m$ $\{\alpha(x_k, y_k^i)\}_{k\in K_2} \to 0$ if $\mu^i > 0$. Now, because $\{x_k\}_{k\in K_2} \to \bar{x} \in S$, stationarity of $\bar{x}$ follows from property (2c) as in the proof of Lemma 2.

5. A generalization of the α-function class

Consider $\alpha(x, y)$ defined in Lemma 1 and note that if $y \in S$ and $f(x) \leq f(y)$, then

$$\alpha(x, y) = f(x) - f(y) - \langle g(y), x - y\rangle \leq |g(y)||x - y|$$

or if $y \notin S$, then

$$\alpha(x, y) = -h(y) - \langle g(y), x - y\rangle < |g(y)||x - y|,$$

so we may want the general class of α-functions to include $\beta|g(y)||x - y|$ where β is a positive parameter. Although such a function satisfies (2a) and (2b), it does not satisfy (2c), since when $\bar{x} \neq \bar{y}$ we cannot conclude that $\bar{g} \in M(\bar{x})$ if $\{\beta|g(y_k)||x_k - y_k|\} \to 0$, because $\{g(y_k)\} \to \bar{g} = 0$. However, when $\bar{g} = 0$, $\bar{y}$ is stationary if $\bar{y} \in S$. To force feasibility of $\bar{y}$, we could consider functions such as $\max[\beta_1|g(y)||x - y|, \beta_2 h(y)]$ where β_1 and β_2 are positive parameters. Then, in order to include such functions in the α-class, weaken (2c) as follows:

$$\text{either } \bar{g} \in M(\bar{x}) \text{ or } \bar{y} \text{ is stationary}$$
$$\text{if } \{(x_k, y_k, g(y_k))\} \to (\bar{x}, \bar{y}, \bar{g}) \quad \text{and} \quad \{\alpha(x_k, y_k)\} \to 0. \tag{2c)$'$}$$

If the definition of the α-class is so generalized, then it is easy to show that the following weakened version of Lemma 2 holds.

Lemma 2'. *If $v_k = 0$, then either x_k or y_i for some $i \in \{1, 2, \dots, k\}$ is stationary for f on S.*

Furthermore, Lemma 6 may be modified in a similar manner so that the following convergence result holds.

Theorem 1'. *Suppose $\{x_k\}$, $\{y_k\}$ and $\{G_k\}$ are bounded with $\{G_k\}$ uniformly positive definite. Then either $\{x_k\}$ or $\{y_k\}$ has at least one accumulation point that is stationary for f on S.*

Appendix

A function $F : \mathbf{R}^n \to \mathbf{R}$ is *weakly uppersemismooth* [5] *at* $x \in \mathbf{R}^n$ if

(a) F is Lipschitz continuous on a ball about x;

(b) for each $d \in \mathbf{R}^n$ and for any sequences $\{t_k\} \subset \mathbf{R}_+$ and $\{g_k\} \subset \mathbf{R}^n$ such that $\{t_k\} \downarrow 0$ and $g_k \in \partial F(x + t_k d)$ it follows that

$$\liminf_{k\to\infty} \langle g_k, d\rangle \geqq \limsup_{t \downarrow 0} [F(x+td)-F(x)]/t.$$

It can be shown that the right hand side of the above inequality is in fact equal to

$$F'(x; d) = \lim_{t \downarrow 0} [F(x + td) - F(x)]/t,$$

the *directional derivative* of F at x in the direction d.

The class of weakly uppersemismooth functions strictly contains the class of semismooth [6] functions. This latter class is closed under composition and contains convex, concave, C^1 and many other locally Lipschitz functions such as ones that result from piecing together C^1 functions as in [9].

A function $F : \mathbf{R}^n \to \mathbf{R}$ is *semiconvex* [6] at $x \in \mathbf{R}^n$ if

(a) F is Lipschitz continuous on a ball about x and for each $d \in \mathbf{R}^n$, $F'(x; d)$ exists and satisfies

(b) $F'(x; d) = \max[\langle g, d\rangle : g \in \partial F(x)]$,

(c) $F'(x; d) \geq 0$ implies $F(x+d) \geq F(x)$.

An example of a nondifferentiable nonconvex function that is both semismooth and semiconvex is $\log(1+|x|)$ for $x \in \mathbf{R}^n$.

References

[1] F.H. Clarke, "Generalized gradients and applications", *Transactions of the American Mathematical Society* 205 (1975) 247–262.

[2] C. Lemarechal, "Nonsmooth optimization and descent methods", RR-78-4, International Institute for Applied Systems Analysis, Laxenburg, Austria (1978).

[3] C. Lemarechal and R. Mifflin, eds., *Nonsmooth optimization* (Pergamon Press, Oxford, 1978).

[4] C. Lemarechal and R. Mifflin, "Global and superlinear convergence of an algorithm for one-dimensional minimization of convex functions", TR-81-3, Department of Pure and Applied Mathematics, Washington State University, Pullman, WA (1981).

[5] R. Mifflin, "An algorithm for constrained optimization with semismooth functions", *Mathematics of Operations Research* 2 (1977) 191–207.

[6] R. Mifflin, "Semismooth and semiconvex functions in constrained optimization", *SIAM Journal of Control and Optimization* 15 (1977) 959–972.

[7] R. Mifflin, "A stable method for solving certain constrained least squares problems", *Mathematical Programming* 16 (1979) 141–158.

[8] M.J.D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches", In: R. Cottle and C.E. Lemke, eds., *Nonlinear programming* (American Mathematical Society, Providence, RI, 1976) pp. 53–72.

[9] R.S. Womersley, "Optimality conditions for piecewise smooth functions", *Mathematical Programming Study* 17 (1982) 13–27 [This Volume.].