MARYLAND

TECHNICAL RESEARCH REPORT

SRC TR 87-28

Globally Convergent Algorithms for Semi-Infinite Optimization Problems Arising in Engineering Design

by

E.R. Panier, and A.L. Tits

SYSTEMS RESEARCH CENTER

UNIVERSITY OF MARYLAND

COLLEGE PARK, MARYLAND 20742

Globally Convergent Algorithms for Semi-Infinite Optimization Problems Arising in Engineering Design

Eliane R. Panier *

André L. Tits +

Electrical Engineering Department and Systems Research Center

University of Maryland, College Park, MD 20742.

Abstract. Optimization problems arising in engineering design often exhibit specific features which, in the interest of computational efficiency, ought to be exploited. Such is the possible presence of 'functional' specifications, i.e., specifications that are to be met over an interval of values of an independent parameter such as time or frequency. While problems involving such specifications could be handled by general purpose nondifferentiable optimization algorithms, the particular structure of functional constraints calls for specific techniques. Suitable schemes have been proposed in the literature. Global convergence is typically achieved by making use of some kind of adaptively refined discretization of the interval of variation of the independent parameter. One previously proposed algorithm exploits the regularity properties of the functions involved to dramatically reduce the computational overhead incurred once the discretization mesh becomes small. In this paper examples are given that show however that, if the initial discretization is coarse, convergence to a nonstationary point may occur. The cause of such failure is investigated and a class of algorithms is proposed that circumvent this difficulty.

Key words. Semi-infinite optimization, nonsmooth optimization, nondifferentiable optimization, engineering design

Please address all correspondence to A.L. Tits.

This research was supported by the National Science Foundation under grants No. DMC-84-20740 and CDR-85-00108.

* Phone number: 301-454-8832

+ Phone number: 301-454-6861

1. Introduction

Optimization problems arising in engineering design often exhibit specific features which, in the interest of computational efficiency, ought to be exploited. Such is the possible presence of 'functional' specifications, i.e., specifications that are to be met *over an interval* of values of an independent parameter such as time or frequency. Problems in circuit design or control system design often include many such specifications. While, at the expense of repeatedly solving univariate optimization subproblems, these constraints could be handled by general purpose nondifferentiable optimization algorithms (see, e.g., [1,2,3]), the particular structure of functional constraints calls for specific techniques. In this paper, such techniques are identified, leading to a class of computationally efficient algorithms enjoying global convergence properties.

For the sake of exposition, we consider the simple problem

$$(P) \qquad \text{minimize } f(x) \text{ s.t. } \phi(x,\omega) \leq 0 \; \forall \, \omega \in \Omega$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $\phi : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is continuous and is continuously differentiable with respect to its first argument, and $\Omega \subset \mathbb{R}$ is a compact interval. Polak *et al.* [4,5] have proposed two globally convergent algorithms for solving such problems, both using an adaptively refined discretization of Ω. In the first algorithm [4], a feasible direction scheme is used that yields an approximate solution to a problem P_q obtained by replacing Ω by a finite subset Ω_q. The search direction is based on gradients of ϕ at all ϵ-active discretization points. The discretization is then progressively refined and the corresponding problem is solved to a progressively better accuracy. Convergence of the overall algorithm to stationary points of P is proven. The second algorithm [5] achieves substantial computational savings over the first one by better exploiting the regularity properties of ϕ as a function of ω. The key observation is that, when the discretization is fine enough, the critical sensitivity information is essentially carried by the gradients of ϕ at the ϵ-active *local maximizers*. The algorithm proposed in [5] makes use of the latter only, and convergence to stationary points is preserved provided the initial discretization is fine enough.

In this paper we first note that, while the latter algorithm performs well with a small discretization mesh, convergence problems may arise in the early stages, when the discretization is still coarse. The reason is that the gradients at the local maximizers for the current iterate may not carry enough information on the local behavior of the constraints. This is illustrated by two examples where the discretization is never refined and the sequence of iterates converges to a nonstationary point. It is then shown how this problem can be avoided by using a standard idea of nondifferentiable optimization (see, e.g., [3,6,7,8]) that consists in using a search direction based not only on function and gradient information evaluated at the current iterate, but also on significant function and gradient information (i) 'remembered' from past iterates or (ii) gathered during suitably devised line searches.

The balance of this paper is organized as follows. In Section 2, the adaptive discretization scheme proposed in [5] is outlined and examples are exhibited where this scheme fails to converge when the initial discretization is too coarse. In Section 3, a class of algorithms is presented. There, the line search procedure, left undefined, is merely required to satisfy a certain condition motivated by the failure just observed. In Section 4 it is proven that all algorithms in the given class exhibit global convergence properties, irrespective of the size of the initial discretization mesh. Two types of line search satisfying the required condition are suggested in Section 5. Finally, Section 6 is dedicated to concluding remarks.

2. Preliminaries

In this section, after briefly outlining the algorithm of [5], we show by two examples that difficulties can arise when the initial discretization is coarse. The notations, terminology and assumptions to be used throughout this paper, essentially borrowed from [5], are introduced first.

For ease of reference, we restate the regularity assumptions for problem P.

Assumption 1.[1] $\Omega = [\omega_0, \omega_c]$ is a compact interval of $\mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $\phi : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is continuous and is differentiable with respect to its first argument x, and $\nabla_x \phi$ is continuous.

Assumption 1 implies that the function $\psi : \mathbb{R}^n \to \mathbb{R}$ given by

$$\psi(x) = \max_{\omega \in \Omega} \phi(x, \omega)$$

is well defined, so that problem P can be reformulated as

$$\text{minimize } f(x) \text{ s.t. } \psi(x) \leq 0.$$

Given $x \in \mathbb{R}^n$, the set of indices of active constraints at x is defined by

$$\Omega(x) = \{\omega \in \Omega \mid \phi(x, \omega) = \psi(x)\}$$

[1] The regularity assumption on ϕ can be weakened so as to allow functions that are merely piecewise continuous in ω (a frequent occurrence in engineering design problems.)

and the set of gradients of active constraints at x by

$$S(x) = \{\nabla_x \phi(x, \omega) \mid \omega \in \Omega(x)\}.$$

The convex hull of the latter is denoted by co$S(x)$. One more assumption will be made throughout.

Assumption 2. There exists no $x \in \mathbb{R}^n$ satisfying $\psi(x) \geq 0$ and $0 \in$co$S(x)$.

A point $x^* \in \mathbb{R}^n$ is called a *Kuhn-Tucker point* for P if $\psi(x^*) \leq 0$ and there exist a finite number l of points $\omega_j^* \in \Omega, \quad j = 1, \dots, l$ and some coefficients $\lambda_j^* \geq 0, \quad j = 1, \dots, l$ satisfying

$$\nabla f(x^*) + \sum_{j=1}^{l} \lambda_j^* \nabla \phi(x^*, \omega_j^*) = 0$$

and

$$\lambda_j^* \phi(x^*, \omega_j^*) = 0, \qquad j = 1, \dots l.$$

It is easily checked that, under the stated assumptions, every local solution of P is a Kuhn-Tucker point for P.

Given any $q \in \mathbb{N} \backslash \{0\}$, referred to below as *discretization index*, the set Ω is now discretized into

$$\Omega_q = \{\omega \in \Omega \mid \omega = \omega_o + l \frac{(\omega_c - \omega_o)}{q}, \quad l = 0, 1, \dots, q\}.$$

The constraint function is approximated accordingly by

$$\psi_q(x) = \max_{\omega \in \Omega_q} \phi(x, \omega)$$

and we define

$$\psi_q^+(x) = \max\{0, \psi_q(x)\}.$$

For any $\epsilon > 0$, the set of *ϵ-active points* of the discretization Ω_q is defined by

$$\Omega_{q,\epsilon}(x) = \{\omega \in \Omega_q \mid \phi(x,\omega) \geq \psi_q^+(x) - \epsilon\}.$$

A *left local maximizer* of ϕ over Ω_q at x is a point $\omega \in \Omega_q$ satisfying one of the three following properties.

i. $\omega_o < \omega < \omega_c$ and the two inequalities

$$\phi(x,\omega) \geq \phi(x,\omega + \frac{\omega_c - \omega_o}{q}) \tag{2.1}$$

and

$$\phi(x,\omega) > \phi(x,\omega - \frac{\omega_c - \omega_o}{q}) \tag{2.2}$$

hold.

ii. $\omega = \omega_o$ and (2.1)holds.

iii. $\omega = \omega_c$ and (2.2) holds.

The set of *ϵ-active left local maximizers* associated with the discretization is given by

$$\tilde{\Omega}_{q,\epsilon}(x) = \{\omega \in \Omega_{q,\epsilon}(x) \mid \omega \text{ is a left local maximizer of } \phi \text{ over } \Omega_q \text{ at } x\}$$

and we define

$$\hat{\Omega}_{q,\epsilon}(x) = \tilde{\Omega}_{q,\epsilon}(x) \cup \Omega_{q,0}(x).$$

Finally, the discretized problem is

$$(P_q) \qquad \text{minimize } f(x) \text{ s.t. } \psi_q(x) \leq 0.$$

For a given discretization, an initial precision ϵ and a current iterate x_i, the algorithm in [5] computes a search direction d_i by solving the quadratic program

$$\begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } < \nabla f(x_i), d > -\gamma\psi_q^+(x_i) \leq v \\ < \nabla_x \phi(x_i, \omega), d > \leq v \quad \forall\, \omega \in \tilde{\Omega}_{q,\epsilon}(x_i). \end{cases} \tag{2.3}$$

where $\gamma > 0$ is given. If the optimal value τ_i of this problem satisfies $\tau_i \geq -\delta\epsilon$, $\delta > 0$ given, ϵ is halfed and the search direction is recomputed accordingly. This process is repeated until the condition $\tau_i < -\delta\epsilon$ holds. If ϵ is decreased below a given threshold, the discretization is refined. The stepsize t_i along d_i is then determined by the following Armijo-like rule, which makes use of two scalars α, $\beta \in (0,1)$. If $\psi_q^+(x_i) > 0$ (Phase 1), t_i is the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$\psi_q(x_i + td_i) - \psi_q(x_i) \leq -\alpha t \delta \epsilon. \tag{2.4}$$

If $\psi_q^+(x_i) = 0$ (Phase 2), t_i is the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$f(x_i + td_i) - f(x_i) \leq -\alpha t \delta \epsilon$$

$$\psi_q(x_i + td_i) \leq 0.$$

Consider now the following problem.

Problem 1.

Problem P with $\Omega = [0,1]$ and functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $\phi : \mathbb{R}^2 \times [0,1] \to \mathbb{R}$ respectively given by

$$f(x) = \xi$$

and

$$\phi(x,\omega) = (2\omega - 1)\eta + \omega(1-\omega)(1-\eta) - \xi$$

where ξ and η are the components of x.

□

It can be checked that

$$\psi(x) = \begin{cases} \eta - \xi & \text{if } \eta \geq 1/3 \\ \frac{5\eta^2 - 2\eta + 1}{4(1-\eta)} - \xi & \text{otherwise} \end{cases}$$

and that the solution of Problem 1 is given by $(\sqrt{5}-2, 1-2\frac{\sqrt{5}}{5})^T$, the only Kuhn-Tucker point for this problem. On the other hand, for $q = 1$ (i.e., two points in the discretization),

$$\psi_q(x) = |\eta| - \xi$$

and the solution of the corresponding discretized problem is $(0,0)^T$. Consider now attempting to solve Problem 1 using the algorithm just outlined and suppose that $q = 1$ initially. Since two adjacent mesh points cannot both be left local maximizers, it is clear that for any iteration i, irrespective of the value of ϵ, $\tilde{\Omega}_{q,\epsilon}(x_i)$ will be $\{0\}$, $\{1\}$ or the empty set. It is then easily checked that, if $\xi_0 \geq |\eta_0|$ (so that $\psi_q(x_0) \leq 0$), τ_i will always be $-1/10$ or $-1/2$. Once $\delta\epsilon < 1/10$, assuming that q has remained unchanged, ϵ will never be further decreased, the discretization will never be refined, and convergence to $(0,0)^T$, the solution of the discretized problem, will occur. Note that the distinctive feature of Problem 1 is that the solution of the discretized problem is located on a 'corner' corresponding to two adjacent mesh points. Since such occurrences are fairly common when the discretization is coarse, failures such as the above should be expected to frequently take place.

Difficulty of the type just discussed can easily be overcome by refining the discretization whenever the step length $||x_{i+1} - x_i||$ falls below a given threshold. However in many cases such refinements may be wasteful as short steps can occur away from the solution of the discretized problem when the step is truncated due to the presence of a constraint that was not taken into account in the direction computation. A particularly acute such case is illustrated by the next example, where arbitrarily small steps are taken away from any stationary point of the discretized problem. For this example the algorithm outlined above fails again.

Problem 2.

Problem P with $\Omega = [0,1]$ and functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $\phi : \mathbb{R}^2 \times [0,1] \to \mathbb{R}$ respectively given by

$$f(x) = -\frac{3}{4}\xi$$

and[2]

$$\phi(x,\omega) = \omega(\omega-1) + (1-\omega)(-\frac{3}{4}\xi + \frac{7}{4}) + \omega(\xi+\eta).$$

Here again, ξ and η are the components of x.

$\square$

Note that

$$\psi(x) = \max(\phi(x,0), \phi(x,1)) = \max(-\frac{3}{4}\xi + \frac{7}{4}, \xi+\eta) = \psi_q(x)$$

for any $q \in \mathbb{N}\backslash\{0\}$, and thus that P_q is equivalent to P for any $q \in \mathbb{N}\backslash\{0\}$. Also, this problem is convex and satisfies Assumptions 1 and 2. Since (ξ,η) is feasible for $\xi \geq 7/3$ with $\xi+\eta \leq 0$

[2] The first term in $\phi(x,\omega)$ is introduced so as to satisfy assumption made in [5] that $\forall\ x \in \mathbb{R}^n$, $\Omega(x)$ is finite.

and since $f(x) \to -\infty$ as $\xi \to \infty$, there is no Kuhn-Tucker point. However, the following result holds.

Proposition 2.1.

Suppose the algorithm just outlined is used on Problem 2. Suppose that $q = 1$ initially and that $x_0 = (0,0)^T$. Also suppose that the parameters values α, β, δ and ϵ satisfy the relationships $\alpha = 1/2$, $\beta = 1/4$, $\delta\epsilon = 1/4$ and that γ is any positive number. Then ϵ is never decreased (so that the discretization is never refined) and the successive iterates are given by $x_i = (1 - 4^{-i}, 0)^T$. The limit point $(1,0)^T$ is not feasible for P (or P_q).

□

Proof

By induction. Clearly the result holds for $i = 0$. Thus suppose that $x_i = (1 - 4^{-i}, 0)^T$ and that ϵ has not been decreased in any of the first i iterations. Since $q = 1$, there are two mesh points, with corresponding constraint values

$$\phi(x_i, 0) = 1 + 3 \times 4^{-(i+1)}$$

and

$$\phi(x_i, 1) = 1 - 4^{-i}$$

so that

$$\psi_q(x_i) = 1 + 3 \times 4^{-(i+1)}.$$

Thus clearly $\omega = 0$ is the only left local maximizer and, irrespective of the value of ϵ, the only

ϵ-active left local maximizer. Since $\nabla_x \phi(x_i, 0) = \nabla f(x_i) = (-3/4, 0)^T$, the search direction is

$$d_i = (3/4, 0)^T$$

and the optimal value of problem (2.3) is $-9/32$. Thus, since $\delta\epsilon = 1/4$, ϵ is not decreased at Step i. Since the iterate x_i is not feasible, the line search performs a Phase 1 iteration and the next iterate is given by

$$x_{i+1} = x_i + \beta^{j_i} d_i$$

where j_i is the smallest nonnegative integer j such that

$$\psi_q(x_i + \beta^j d_i) \leq \psi_q(x_i) - \alpha\beta^j \delta\epsilon$$

It is easily checked that $j_i = i$. Thus,

$$\begin{aligned} x_{i+1} &= (1 - 4^{-i}, 0)^T + 4^{-i}(3/4, 0)^T \\ &= (1 - 4^{-(i+1)}, 0)^T. \end{aligned}$$

$\square$

Thus convergence occurs to a point that is not even a solution for the discretized problem. This failure can be explained as follows (see Fig. 2.1). At $x^* = (1, 0)^T$ the set of indices of active constraints is $\Omega(x^*) = \{0, 1\}$. However, along the constructed sequence, $\omega = 1$ is never a local maximizer for $\phi(x_i, \cdot)$ over Ω_q (although it is one over Ω), so that $\tilde{\Omega}_{q,\epsilon}(x_i) = \{0\}$ for all i. Thus, although the step performed by the line search is always truncated due to the presence of the constraint at $\omega = 1$, this constraint is never taken into account in the search direction computation, and consequently the sequence $\{x_i\}$ never leaves the subspace $\{x \mid \eta = 0\}$.

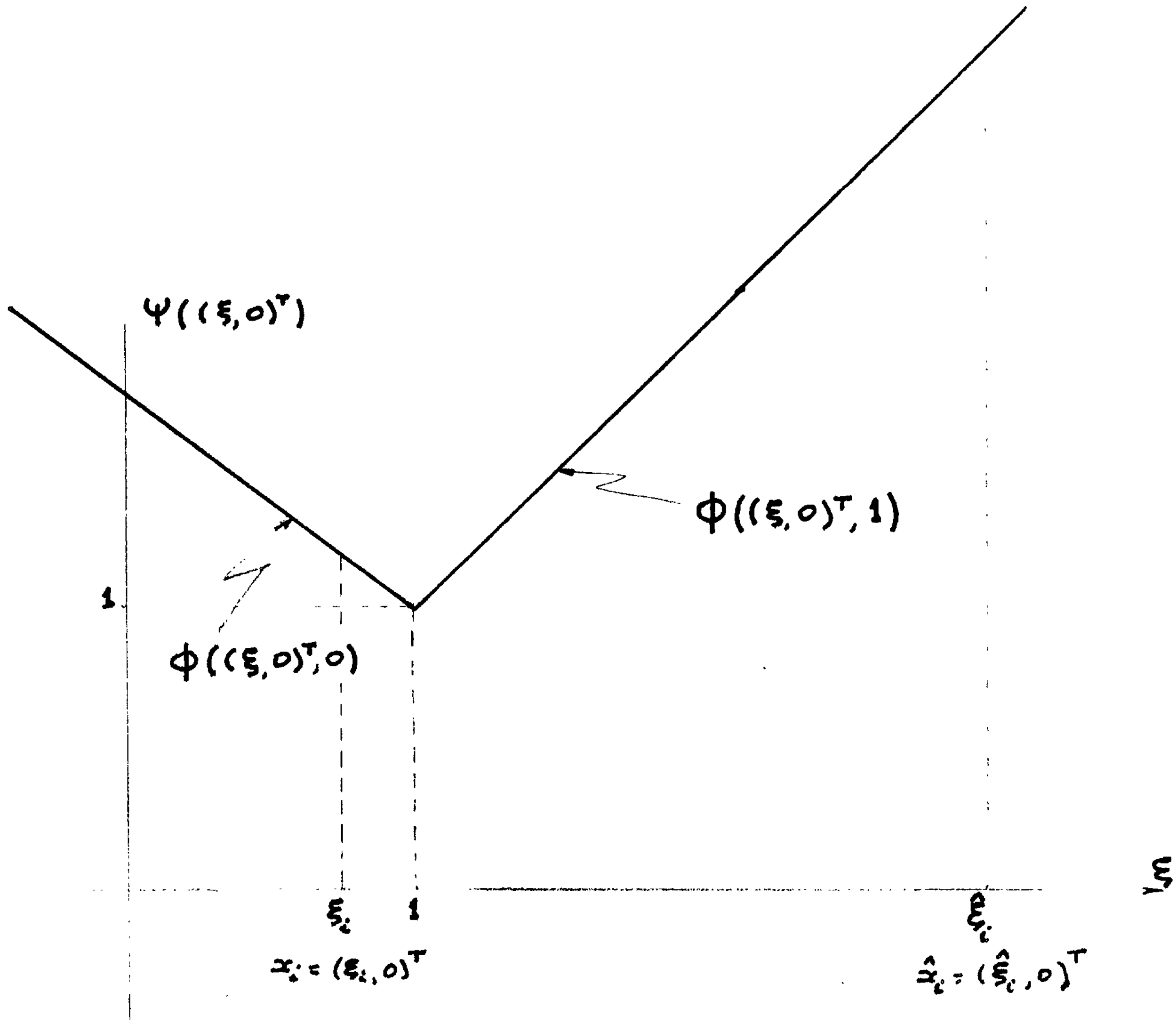

Figure 2.1. $\psi(x)$ from Problem 2 with $\eta = 0$

Obviously, in Problems 1 and 2, satisfactory performance will not be achieved unless computation of the search direction takes into account *simultaneously* the gradients $\nabla_x \phi(x,\omega)$ at both $\omega = 0$ and $\omega = 1$. It should be clear that similar considerations apply to a large class of problems. Thus one could use, in addition to the gradients at the left local maximizers for the current iterate, the gradient at the last point rejected by the Armijo rule (point $\hat{x}_i$ in Fig. 2.1 for Problem 2). Alternatively, one could combine a memory mechanism with the requirement that the line search yield a point not only satisfying (2.4) but also resulting in a directional derivative $\psi_q'(x_{i+1}, d_i)$ sufficiently larger than $\psi_q'(x_i, d_i)$, indicating that the gradients of active constraints at the new iterate are significantly different from those at the previous iterate. Such considerations motivate the line search framework (Condition C) and the updating scheme (Step 6) associated with the class of algorithms described in the next section. In Section 5 line search procedures of the types just suggested will be considered.

3. A Class of Algorithms

The idea of making use, in the computation of the search direction, of gradients evaluated at points other than the current iterate has been widely used in the nonsmooth optimization literature (see, e.g., [1,3,8]). For our purpose, a set J of pairs $(y, \omega) \in \mathbb{R}^n \times \Omega$ that carry relevant information will be maintained and updated at each iteration. For a current iterate x and a discretization index q, a 'weight' $W_q(x, y, \omega)$ will be associated to each pair $(y, \omega) \in J$, given by [3]

$$W_q(x, y, \omega) = \max\{||x - y||;\ \psi_q^+(y) - \phi(y, \omega);\ ||x - y||\ ||\nabla_x \phi(y, \omega)||\} \geq 0.$$

A *low* value of this weight indicates that $\nabla_x \phi(y, \omega)$ carries information that is *highly* relevant at x. The search direction $d_q(x, J)$ and an 'optimality function' $v_q(x, J)$ will then be obtained by solving the quadratic program (cf. (2.3))

$$(QP_q(x, J)) \quad \begin{cases} \min\ \frac{1}{2}||d||^2 + v & \\ \text{s.t.}\ < \nabla f(x), d > -\gamma \psi_q^+(x) \leq v & (3.1) \\ < \nabla_x \phi(y, \omega), d > -W_q(x, y, \omega) \leq v \quad \forall\ (y, \omega) \in J. & (3.2) \end{cases}$$

[3] The last term in the weight is introduced for purely technical reasons. It does not introduce any computational overhead since $\nabla_x \phi(y, \omega)$ already appears in $QP_q(x, J)$ below.

The optimal value of $QP_q(x, J)$ is

$$\tau_q(x, J) = \frac{1}{2}||d_q(x, J)||^2 + v_q(x, J).$$

The multipliers associated with the constraints of $QP_q(x, J)$ of type (3.1) and (3.2) will be represented by $\lambda_{f,q}(x, J)$ and $\lambda_{y,\omega,q}(x, J)$, respectively. These multipliers, which are not necessarily unique, are chosen in such a way that at most $n+1$ multipliers of the second type are different from 0.

In Algorithm A below, the line search is only partly specified. It will be required to satisfy a certain Condition C, which expresses that it should tend to gather significant information on the potentially binding constraints. The algorithm is inspired from the one in [5].[4]

Algorithm A.

Parameters. $\delta > 0$; $\gamma > 0$.

Data. $\epsilon_0 > 0$; $M_0 > 0$; $N_0 > 0$; $q_0 \in \mathbb{N}\backslash\{0\}$; $X_0 \in \mathbb{R}^n$.

Step 0. Initialization of the outer loop. Set $k = 0$.

Step 1. Initialization of the inner loop. Set $i = 0$, $x_0 = X_k$ and

$$J_0 = \{(X_k, \omega) \mid \omega \in \tilde{\Omega}_{q_k,\epsilon_k}(X_k)\}.$$

Step 2. Search direction computation. Compute $d_i = d_{q_k}(x_i, J_i)$ and $v_i = v_{q_k}(x_i, J_i)$ by solving $QP_{q_k}(x_i, J_i)$.

[4] Note however that, as $W_q(x_i, x_i, \omega)$ is positive for $\omega \notin \tilde{\Omega}_{q,0}(x_i)$, it becomes unnecessary to iteratively reduce ϵ. In order to limit the number of gradient evaluations, such reduction is nevertheless performed when the discretization is refined.

Step 3. Optimality test. If $v_i \geq -\delta\epsilon_k$ or $\|x_i\| > N_k$, go to Step 4. If $\psi_{q_k}(x_i) \leq 0$ and $f(x_i) < -M_k$, go to Step 4. Otherwise, go to Step 5.

Step 4. Discretization update. If $v_i \geq -\delta\epsilon_k$, set $\epsilon_{k+1} = \epsilon_k/2$. Otherwise, set $\epsilon_{k+1} = \epsilon_k$. If $\|x_i\| > N_k$, set $N_{k+1} = 2\|x_i\|$. Otherwise, set $N_{k+1} = N_k$. If $\psi_{q_k}(x_i) \leq 0$ and $f(x_i) < -M_k$, set $M_{k+1} = -2f(x_i)$. Otherwise, set $M_{k+1} = M_k$. Set $q_{k+1} = 2q_k$; $X_{k+1} = x_i$; $D_{k+1} = d_i$; $V_{k+1} = v_i$; $I_{k+1} = J_i$. Set $k = k+1$ and go back to Step 1.

Step 5. Line search. Find a stepsize $t_i \equiv t_{q_k,M_k}(x_i, d_i, v_i)$ and a set $Y_i \equiv Y_{q_k,M_k}(x_i, d_i, v_i)$ using the line search model given below.

Step 6. Updates. Set $x_{i+1} = x_i + t_i d_i$ and

$$J_{i+1} = J_i \backslash \{(y,\omega) \in J_i \mid \lambda_{y,\omega,q_k}(x_i, J_i) = 0\} \cup \{(x_{i+1}, \omega) \mid \omega \in \bar{\Omega}_{q_k,\epsilon_k}(x_{i+1})\} \cup Y_i.$$

Set $i = i+1$ and go back to Step 2.

$\square$

Line search model.

The following model is inspired from the line search procedure used in [5]. It depends on a parameter $\alpha \in (0,1)$. Suppose the current discretization index is q and let $M > 0$. Then given a point x, a direction d and a scalar $v < 0$ satisfying

$$< \nabla f(x), d > -\gamma\psi_q^+(x) \leq v \tag{3.3}$$

and

$$< \nabla_x \phi(x,\omega), d > \leq v \quad \forall\, \omega \in \Omega_{q,0}(x), \tag{3.4}$$

the line search yields a step $t \equiv t_{q,M}(x,d,v)$ satisfying the following conditions (essentially those in [5]). If $\psi_q^+(x) > 0$ (Phase 1), either

$$\psi_q(x+td) \leq 0 \tag{3.5}$$

or

$$\psi_q(x+td) - \psi_q(x) \leq \alpha t v \tag{3.6}$$

holds and if $\psi_q^+(x) = 0$ (Phase 2), either

$$f(x+td) < -M;\ \psi_q(x+td) \leq 0 \tag{3.7}$$

or both

$$f(x+td) - f(x) \leq \alpha t v \tag{3.8.a}$$

and

$$\psi_q(x+td) \leq 0 \tag{3.8.b}$$

hold. In addition, the line search must return a set $Y \equiv Y_{q,M}(x,d,v) \subset \mathbb{R}^n \times \Omega_q$ (possibly empty) and it is required that Condition C below be satisfied.

Condition C. There exists a scalar $\theta < 1$ such that the following holds for every discretization index q and for every positive value M: for every compact set $S \subset \mathbb{R}^n$ and for every $\bar{v} < 0$, there exist $\underline{t} > 0$ and $\bar{s} > 0$ such that for every $x,\ d \in S$ and $v < \bar{v}$ at which the line search is performed, the step $t \equiv t_{q,M}(x,d,v)$ and the set $Y \equiv Y_{q,M}(x,d,v)$ are such that

$$\|y\| \leq \bar{s} \quad \forall\ (y,\omega) \in Y;$$

and such that either

$$t \geq \underline{t}$$

or

$$< \nabla_x \phi(y,\omega), d > -W_q(x + td, y, \omega) \geq \theta v \tag{3.9}$$

for some $(y,\omega) \in Y \cup \{(x+td,\omega) \mid \omega \in \Omega_{q,0}(x+td)\}$, unless one of the following two conditions holds:

(i) $\psi^+(x) > 0$ and $\psi(x + td) \leq 0$ (transition from Phase 1 to Phase 2)

(ii) (3.7) (detection of a feasible point where the objective function value is very small).

□

Condition C and the updating scheme in Step 6 may be interpreted as follows. Suppose that at some point x_i in Algorithm A the step t_i becomes very small due to the presence of a constraint not taken into account in the search direction. Since (3.9) must then hold, the updated set J_{i+1} will contain a new pair (y,ω) significantly different from the pairs used in computing the current search direction (compare (3.9) with (3.2)), and the new search direction d_{i+1} will be significantly different from d_i, thus avoiding the problem encountered in Problem 2. By keeping in J_{i+1} the constraints that affected the current search direction (those with nonzero multipliers), the updated scheme of Step 6 will then ensure that zigzagging (as encountered in Problem 1) is avoided.

4. Convergence Analysis.

If the solution $(d_q(x,J), v_q(x,J))$ of $QP_q(x,J)$ is always well defined, Algorithm A will be well defined, provided the line search is. The latter will be discussed in Section 5. The following proposition establishes the former and gives a characterization of $(d_q(x,J), v_q(x,J))$ that will be used extensively below.

Proposition 4.1.

Problem $QP_q(x,J)$ has a unique solution $(d,v) \equiv (d_q(x,J), v_q(x,J))$, with $v \leq 0$. Moreover, (d,v) is the unique pair satisfying Relationships (4.1)-(4.4):

$$d + \lambda_f \nabla f(x) + \sum_{(y,\omega)\in J} \lambda_{y,\omega} \nabla_x \phi(y,\omega) = 0 \tag{4.1}$$

$$< \nabla f(x), d > -\gamma \psi_q^+(x) - v \leq 0 \tag{4.2}$$

(with equality if $\lambda_f \neq 0$)

$$< \nabla_x \phi(y,\omega), d > -W_q(x,y,\omega) - v \leq 0 \quad \forall\ (y,\omega) \in J \tag{4.3}$$

(with equality for (y,ω) such that $\lambda_{y,\omega} \neq 0$)

$$\lambda_f \geq 0,\ \lambda_{y,\omega} \geq 0 \quad \forall\ (y,\omega) \in J,\ \lambda_f + \sum_{(y,\omega)\in J} \lambda_{y,\omega} = 1, \tag{4.4}$$

where the multipliers $\lambda_f \equiv \lambda_{f,q}(x,J)$ and $\lambda_{y,\omega} \equiv \lambda_{y,\omega,q}(x,J)$, for all $(y,\omega) \in J$, are as defined in Section 3. Finally

$$v = -\|d\|^2 - \gamma \lambda_f \psi_q^+(x) - \sum_{(y,\omega)\in J} \lambda_{y,\omega} W_q(x,y,\omega) \tag{4.5}$$

Proof.

(d,v) solves $QP_q(x,J)$ if and only if d is solution of the problem in d

$$\text{minimize } \frac{1}{2}\|d\|^2 + \max\{< \nabla f(x), d > -\gamma\psi_q^+(x);\ < \nabla_x \phi(y,\omega), d > -W_q(x,y,\omega) \mid (y,\omega) \in J\}$$

and

$$v = \max\{< \nabla f(x), d > -\gamma\psi_q^+(x);\ < \nabla_x \phi(y,\omega), d > -W_q(x,y,\omega) \mid (y,\omega) \in J\}.$$

Since the objective function associated with this problem is strictly convex and radially unbounded, d and v are always uniquely defined. Moreover, since the optimal value may not exceed the value obtained for $d = 0$, the value v is nonpositive. Next, Relationships (4.1)-(4.4) come directly from the optimality conditions associated with the solution of $QP_q(x,J)$. Since, from convexity, those optimality conditions are also sufficient, it follows, from the uniqueness of the solution of $QP_q(x,J)$ that (d,v) is the unique pair satisfying (4.1)-(4.4). Finally, Relationships (4.2) and (4.3) yield

$$\lambda_f(< \nabla f(x), d > -\gamma\psi_q^+(x) - v) + \sum_{(y,\omega)\in J} \lambda_{y,\omega}(< \nabla_x \phi(y,\omega), d > -W_q(x,y,\omega) - v) = 0.$$

Using (4.4), we get

$$v = \lambda_f < \nabla f(x), d > + \sum_{(y,\omega)\in J} \lambda_{y,\omega} < \nabla_x \phi(y,\omega), d > -\gamma\lambda_f\psi_q^+(x) - \sum_{(y,\omega)\in J} \lambda_{y,\omega} W_q(x,y,\omega)$$

and since (4.1) implies

$$\lambda_f < \nabla f(x), d > + \sum_{(y,\omega)\in J} \lambda_{y,\omega} < \nabla_x \phi(y,\omega), d >= -\|d\|^2,$$

(4.5) holds.

□

In the remainder of this section, it is shown that Algorithm A is convergent in the sense that it generates an infinite sequence $\{X_k\}$ and that every accumulation point of that sequence is a Kuhn-Tucker point for P. We first prove the latter.

Proposition 4.2.

If the algorithm generates an infinite sequence $\{X_k\}$, then, every accumulation point of that sequence is a Kuhn-Tucker point for P.

Proof.

Consider an accumulation point X^* and a subsequence $\{X_k\}_{k\in K}$ converging to X^*. The inequalities $\|X_k\| > N_{k-1}$ and $f(X_k) < -M_{k-1}$ may only occur finitely many times on K. So, the inequality $V_k \geq -\delta\epsilon_{k-1}$ is always satisfied for $k \in K$ large enough. Since the numbers ϵ_{k-1} converge to zero, the numbers V_k, which are nonpositive in view of Proposition 4.1, must also tend to zero. It will be assumed, without loss of generality that I_k contains exactly $l \leq n+1$ elements (l possibly zero, indicating that I_k is empty). Let us now denote those elements by $(y_{j,k}, \omega_{j,k})$, $\quad j = 1, \ldots, l$. Let the multiplier associated with the first constraint of $QP_{q_k}(X_k, I_k)$ and the multipliers associated with the other constraints be respectively denoted by $\lambda_{f,k} \equiv \lambda_{f,q_k}(X_k, I_k)$ and $\lambda_{j,k} \equiv \lambda_{y_{j,k},\omega_{j,k},q_k}(X_k, I_k)$, $\quad j = 1, \ldots, l$. Using these notations, Relationships (4.1), (4.4) and (4.5) yield

$$D_k + \lambda_{f,k}\nabla f(X_k) + \sum_{j=1}^{l} \lambda_{j,k}\nabla_x\phi(y_{j,k}, \omega_{j,k}) = 0 \tag{4.6}$$

$$\lambda_{f,k} \geq 0, \ \lambda_{j,k} \geq 0 \ \ j = 1, \ldots, l, \ \lambda_{f,k} + \sum_{j=1}^{l} \lambda_{j,k} = 1 \tag{4.7}$$

and

$$V_k = -\|D_k\|^2 - \gamma\lambda_{f,k}\psi_{q_k}^+(X_k) - \sum_{j=1}^{l} \lambda_{j,k} W_{q_k}(X_k, y_{j,k}, \omega_{j,k}). \tag{4.8}$$

Reducing K, if necessary, to a subset of indices, it may be assumed, in view of (4.7), that the subsequences of coefficients $\lambda_{f,k}$ and $\lambda_{j,k}$, $\quad j = 1, \ldots, l$, converge on K to some numbers λ_f^*

and $\lambda_j^*, \quad j = 1, \ldots, l$ satisfying

$$\lambda_f^* \geq 0, \ \lambda_j^* \geq 0, \ \lambda_f^* + \sum_{j=1}^{l} \lambda_j^* = 1. \tag{4.9}$$

Since Ω is compact, it may also be assumed that the numbers $\omega_{j,k}, \quad j = 1, \ldots, l$, converge on K to some numbers $\omega_j^* \in \Omega, \quad j = 1, \ldots, l$. Convergence to zero on K of the sequence $\{V_k\}$ implies, in view of (4.8), the definition of the weights $W_{q_k}(X_k, y_{j,k}, \omega_{j,k})$, and continuity of $\nabla_x \phi$,

$$D_k \rightarrow 0, \ k \in K, \ k \rightarrow \infty \tag{4.10}$$

$$\lambda_{f,k} \psi_{q_k}^+(X_k) \rightarrow 0, \ k \in K, \ k \rightarrow \infty \tag{4.11}$$

$$\lambda_{j,k} \|\nabla_x \phi(y_{j,k}, \omega_{j,k})\| \rightarrow 0, \ k \in K, \ k \rightarrow \infty \tag{4.12}$$

for all j such that $\lambda_j^* = 0$ and

$$\|X_k - y_{j,k}\| \rightarrow 0, \ \psi_{q_k}^+(y_{j,k}) - \phi(y_{j,k}, \omega_{j,k}) \rightarrow 0, \ k \in K, \ k \rightarrow \infty \tag{4.13}$$

for all j such that $\lambda_j^* \neq 0$. In particular, from (4.13), the vectors $y_{j,k}$ corresponding to nonzero multipliers, converge to X^* on K so that, taking the limit in (4.6) we obtain, in view of regularity Assumption 1, and Relationships (4.10) and (4.12),

$$\lambda_f^* \nabla f(X^*) + \sum_{j=1}^{l} \lambda_j^* \nabla_x \phi(X^*, \omega_j^*) = 0. \tag{4.14}$$

Now, regularity Assumption 1 implies that

$$\psi_{q_k}^+(X_k) \rightarrow \psi^+(X^*), \ k \in K, \ k \rightarrow \infty$$

so that (4.11) yields

$$\lambda_f^* \psi^+(X^*) = 0.$$

Now, in view of (4.9), (4.14) and Assumption 2, $\psi(X^*) \leq 0$. Finally, using (4.13), the feasibility of X^* and regularity Assumption 1, we obtain,

$$\lambda_j^* \phi(X^*, \omega_j^*) = 0, \ \forall \, j. \tag{4.15}$$

Finally, remark that λ_f^* cannot be 0. Indeed, if it were the case, there would exist, in view of (4.9), an index j for which $\lambda_j^* \neq 0$. This would imply, from (4.15) and the feasibility of X^*, $\psi(X^*) = 0$. But, it would also hold, in view of (4.14), $0 \in \text{co}S(X^*)$, in contradiction with Assumption 2.

$\square$

It now remains to show that Algorithm A generates an infinite sequence $\{X_k\}$. We will suppose by contradiction that there exists an iteration k such that, for the corresponding discretized problem, the algorithm performs an infinite number of steps, thus generating an infinite sequence $\{x_i\}$. If the discretization is never changed, the following three conditions hold for all i:

$$\text{either } \psi_{q_k}(x_i) > 0 \text{ or } f(x_i) \geq -M_k \tag{4.16}$$

$$||x_i|| \leq N_k \tag{4.17}$$

and

$$v_i < -\delta \epsilon_k. \tag{4.18}$$

Lemma 4.3.

If there is an index k for which the discretization is never refined, then all the directions d_i constructed on the corresponding discretization satisfy $||d_i|| \geq \underline{d}$, for some $\underline{d} > 0$.

Proof.

The solutions (d_i, v_i) of $QP_{q_k}(x_i, J_i)$ satisfy, in view of (4.18), $< \nabla f(x_i), d_i > \leq v_i < -\delta\epsilon_k$ if $\psi_{q_k}(x_i) \leq 0$ and $< \nabla_x \phi(x_i, \omega_i), d_i > \leq v_i < -\delta\epsilon_k$ for some $\omega_i \in \Omega_{q_k,0}(x_i)$ otherwise. The result then follows from Relationship (4.17), compactness of Ω, and regularity Assumption 1.

□

Lemma 4.4.

If there is an index k for which the discretization is never refined, then the sequence $\{x_i\}$ of iterates constructed on the corresponding discretization converges to some vector x^*.

Proof.

In case $\psi_{q_k}(x_i) > 0 \; \forall \; i$, Relationships (3.6) and (4.5) imply that

$$\psi_{q_k}(x_{i+1}) \leq \psi_{q_k}(x_i) - \alpha t_i ||d_i||^2$$

and, from Lemma 4.3,

$$\psi_{q_k}(x_{i+1}) \leq \psi_{q_k}(x_i) - \alpha\underline{d}||x_{i+1} - x_i||.$$

Thus for any given $s \in \mathbb{N}$, we have,

$$\psi_{q_k}(x_{s+1}) - \psi_{q_k}(x_0) = \sum_{i=0}^{s}(\psi_{q_k}(x_{i+1}) - \psi_{q_k}(x_i)) \leq -\alpha\underline{d}\sum_{i=0}^{s}||x_{i+1} - x_i||.$$

Therefore

$$\sum_{i=0}^{s}||x_{i+1} - x_i|| \leq \frac{\psi_{q_k}(x_0) - \psi_{q_k}(x_{s+1})}{\alpha\underline{d}}$$

so that,

$$\sum_{i=0}^{s} \|x_{i+1} - x_i\| \leq \frac{\psi_{q_k}(x_0)}{\alpha \underline{d}}$$

so that $\{x_i\}$ is a Cauchy sequence, which proves the claim. If x_i eventually becomes feasible, the result follows similarly from Relationships (3.8.a) and (4.5).

□

Lemma 4.5.

If there is an index k for which the discretization is never refined then the sequence $\{t_i\}$ of stepsizes constructed on the corresponding discretization converges to zero.

Proof.

The result follows directly from Lemmas 4.3 and 4.4.

□

Proposition 4.6. Suppose Condition C holds. Then Algorithm A constructs an infinite sequence $\{X_k\}$.

Proof.

Let us show that we get a contradiction if we assume that k remains fixed. Let us denote by $\tau_i \equiv \tau_{q_k}(x_i, J_i)$ the optimal values of the quadratic problems $QP_{q_k}(x_i, J_i)$ and let

$$\tau^* = \lim\ \sup\ \tau_i.$$

In view of Lemma 4.4, the sequence $\{x_i\}$ is bounded and it is easily shown, using the optimality conditions (4.1)-(4.5) that the sequence of directions $\{d_i\}$ and the sequence of values $\{v_i\}$ are also bounded. Indeed, the values d_i and v_i satisfy, in view of (4.2),

$$< \nabla f(x_i), d_i > -\gamma\psi_q^+(x_i) - v_i \leq 0 \tag{4.19}$$

and, from (4.5),

$$v_i \leq -||d_i||^2.$$

These two inequalities yield

$$||d_i||(||d_i|| - ||\nabla f(x_i)||) \leq \gamma\psi^+(x_i)$$

which, in view of the convergence of the sequence $\{x_i\}$, ensures the boundedness of the sequence $\{d_i\}$. Boundedness of the sequence of negative values $\{v_i\}$ follows immediately from (4.19). From Condition C and compactness of Ω, the sequence $\{J_i\}$ is bounded as well. A subset of indices $I \subset \mathbb{N}$ can thus be extracted such that $\{\tau_i\}_{i\in I} \to \tau^*$, $\{d_i\}_{i\in I} \to d^*$, $\{v_i\}_{i\in I} \to v^*$, $\{d_{i+1}\}_{i\in I} \to d^{1*}$, $\{v_{i+1}\}_{i\in I} \to v^{1*}$ for some d^*, v^*, d^{1*}, v^{1*} and such that, for some $l \leq n+1$, possibly zero, exactly l constraints corresponding to points of J_i, $i \in I$ have a multiplier different from zero in $QP_{q_k}(x_i, J_i)$. Let us denote by $(y_{j,i}, \omega_{j,i})$, $\quad j = 1, \ldots, l$ the points of J_i corresponding to the nonzero multipliers. It may be assumed, without loss of generality, that $\{y_{j,i}\}_{i\in I} \to y_j^*$ and, $\{\omega_{j,i}\}_{i\in I} \to \omega_j^*$, for some y_j^*, ω_j^*, $\quad j = 1, \ldots, l$. Then, for all $i \in I$, (d_i, v_i) solves,

$$(QP_i) \quad \begin{cases} \text{minimize } \frac{1}{2}||d||^2 + v \\ \text{s.t. } < \nabla f(x_i), d > -\gamma\psi_{q_k}^+(x_i) \leq v \\ < \nabla_x \phi(y_{j,i}, \omega_{j,i}), d > -W_{q_k}(x_i, y_{j,i}, \omega_{j,i}) \leq v, \quad j = 1, \ldots, l \end{cases}$$

where only those constraints of nonzero multipliers appear, since clearly (d_i, v_i) satisfies the optimality conditions associated with QP_i and since, in view of Proposition 4.1, the solution to those optimality conditions is unique. Now, since $v_i < -\delta\epsilon_k$ for all i, in view of Condition C and Lemma 4.5, it may be assumed, without loss of generality, that there exists a sequence $\{(\tilde{y}_i, \tilde{\omega}_i)\}$ of points in J_{i+1} converging on I to some vector (y^*, ω^*) and satisfying,

$$< \nabla_x \phi(\tilde{y}_i, \tilde{\omega}_i), d_i > -W_{q_k}(x_{i+1}, \tilde{y}_i, \tilde{\omega}_i) \geq \theta v_i. \tag{4.20}$$

Next, (d_{i+1}, v_{i+1}) solves the quadratic problem $QP_{i+1} \equiv QP_{q_k}(x_{i+1}, J_{i+1})$,

$$(QP_{i+1}) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } < \nabla f(x_{i+1}), d > -\gamma\psi_{q_k}^+(x_{i+1}) \leq v \\ < \nabla_x \phi(y_{j,i}, \omega_{j,i}), d > -W_{q_k}(x_{i+1}, y_{j,i}, \omega_{j,i}) \leq v, \quad j = 1, \ldots, l \\ < \nabla_x \phi(\tilde{y}_i, \tilde{\omega}_i), d > -W_{q_k}(x_{i+1}, \tilde{y}_i, \tilde{\omega}_i) \leq v \\ + \text{ other inequalities} \end{cases}$$

and we suppose, without loss of generality that, for $i \in I$, the number of 'other inequalities' is fixed. The limit pair (d^*, v^*) is solution of the limit problem

$$(QP^*) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } < \nabla f(x^*), d > -\gamma\psi_{q_k}^+(x^*) \leq v \\ < \nabla_x \phi(y_j^*, \omega_j^*), d > -W_{q_k}(x^*, y_j^*, \omega_j^*) \leq v, \quad j = 1, \ldots, l. \end{cases}$$

This is because, for all i in I, (d_i, v_i) is solution of QP_i and thus satisfies the optimality conditions associated with QP_i. The limit d^* therefore satisfies the optimality conditions associated with QP^* and, in view of Proposition 4.1, (d^*, v^*) is the unique solution of QP^*. Similarly, (d^{1*}, v^{1*}) is solution of QP^{1*},

$$(QP^{1*}) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } < \nabla f(x^*), d > -\gamma\psi_{q_k}^+(x^*) \leq v \\ < \nabla_x \phi(y_j^*, \omega_j^*), d > -W_{q_k}(x^*, y_j^*, \omega_j^*) \leq v, \quad j = 1, \ldots, l \\ < \nabla_x \phi(y^*, \omega^*), d > -W_{q_k}(x^*, y^*, \omega^*) \leq v \\ +\text{other limit inequalities} \end{cases}$$

where the 'limit' inequalities correspond to a suitable subsequence. In view of (4.20), the unique solution (d^*, v^*) of QP^* satisfies

$$< \nabla_x \phi(y^*, \omega^*), d^* > - W_{q_k}(x^*, y^*, \omega^*) \geq \theta v^*.$$

Therefore, one constraint in QP^{1*} is not satisfied by (d^*, v^*). Thus, since all constraints in QP^* are included in QP^{1*}, $\tau^{1*} = \frac{1}{2}||d^{1*}||^2 + v^{1*}$ satisfies $\tau^{1*} > \tau^*$, in contradiction with the definition of τ^*.

$\square$

We conclude this section with a theorem that combines Propositions 4.2 and 4.6.

Theorem 4.7.

Suppose Condition C holds. Then Algorithm A constructs an infinite sequence $\{X_k\}$ and every accumulation point of this sequence is a Kuhn-Tucker point for P.

5. Two Line Search Procedures

In this section, two line search procedures satisfying Condition C are introduced and discussed. They are described for a given discretization index q, a scalar $M < 0$ and for some x, $d \in \mathbb{R}^n$ and $v < 0$ satisfying Relationships (3.3) and (3.4).

Line search LS1. (Armijo-like)

The following is a simple modification of the line search in [5]. Here, at each iteration, the last *unsuccessful* trial point in the Armijo-like test is 'remembered'. The step $t \equiv t_{q,M}(x, d, v)$ is computed through the following procedure for some α, $\beta \in (0, 1)$. If $\psi_q^+(x) > 0$ (Phase 1),

$t = \beta^{\underline{j}}$ where $\underline{j}$ is the smallest nonnegative integer j satisfying either

$$\psi_q(x + \beta^j d) \leq 0$$

or

$$\psi_q(x + \beta^j d) - \psi_q(x) \leq \alpha \beta^j v.$$

If $\psi_q^+(x) = 0$ (Phase 2), $t_i = \beta^{\underline{j}}$ where $\underline{j}$ is the smallest nonnegative integer j satisfying

$$\psi_q(x + \beta^j d) \leq 0$$

and either

$$f(x + \beta^j d) \leq -M$$

or

$$f(x + \beta^j d) - f(x) \leq \alpha \beta^j v$$

and the set $Y \equiv Y_{q,M}(x, d, v)$ is taken as the empty set if $\underline{j} = 0$ and as a singleton

$$Y = \{(x + \beta^{\underline{j}-1} d, \omega) \text{ where } \omega \text{ is any element in } \Omega_{q,0}(x + \beta^{\underline{j}-1} d)\}$$

otherwise.

Proposition 5.1.

Line search LS1 is well defined and satisfies Condition C.

Proof.

The proof of the first statement, can be found elsewhere (see, e.g., [4]). It thus remains to show that Condition C is satisfied by the line search. We show by contradiction that Condition C is satisfied with $\theta = 2\alpha$. It is easily checked that the sets Y are bounded on compact sets. Therefore, if Condition C does not hold, there is a fixed discretization index q, a compact set S, a scalar $\bar{v} < 0$, some sequences $\{x_i\}$ and $\{d_i\}$ with elements in S, and some numbers $v_i < \bar{v}$ such that the corresponding steps $t_i \equiv t_{q,M}(x_i, d_i, v_i)$ converge to zero and such that

$$< \nabla f(x_i), d_i > -\gamma\psi_q^+(x_i) \leq v_i \tag{5.1}$$

and

$$< \nabla_x \phi(y_i, \omega_i), d_i > -W_q(x_i + t_i d_i, y_i, \omega_i) < 2\alpha v_i. \tag{5.2}$$

for some $\omega_i \in \Omega_{q,0}(y_i)$ (the one chosen by the line search), with $y_i = x_i + \frac{t_i}{\beta} d_i$. The negative numbers v_i are bounded, in view of (5.1), the boundedness of the sequences $\{x_i\}$ and $\{d_i\}$ and regularity Assumption 1. Therefore, it is possible to extract a subset I of indices such that, on that subset, the sequences $\{x_i\}$, $\{d_i\}$, $\{v_i\}$, and $\{\omega_i\}$ converge respectively to some values x^*, d^*, v^*, and ω^*. Then, clearly, the sequence $\{y_i\}$ converges, on I, to the point x^* and $\omega^* \in \Omega_{q,0}(x^*)$. On the other hand, since the step $\frac{t_i}{\beta}$ was not accepted by the line search, a subset of indices $I' \subset I$ can be extracted so that, on that subset, one of the following three inequalities is always satisfied:

$$\psi_q(x_i) > 0 \text{ and } \psi_q(x_i + \frac{t_i}{\beta} d_i) - \psi_q(x_i) > \alpha \frac{t_i}{\beta} v_i \tag{5.3}$$

$$\psi_q(x_i) \leq 0 \text{ and } f(x_i + \frac{t_i}{\beta} d_i) - f(x_i) > \alpha \frac{t_i}{\beta} v_i \tag{5.4}$$

$$\psi_q(x_i) \leq 0 \text{ and } \psi_q(x_i + \frac{t_i}{\beta} d_i) > 0. \tag{5.5}$$

Relationship (5.4) cannot occur infinitely many times on I'. Otherwise, taking the limit of (5.4) on I' would yield $< \nabla f(x^*), d^* > \geq \alpha v^* > v^*$, in contradiction with the fact that, in view of (5.1), since $\psi_q(x_i) \leq 0$, the pair (d_i, v_i) satisfies $< \nabla f(x_i), d_i > \leq v_i$. If Relationship (5.3) is always satisfied, it also holds, since $\omega_i \in \Omega_{q,0}(y_i)$,

$$\phi(x_i + \frac{t_i}{\beta} d_i, \omega_i) - \phi(x_i, \omega_i) > \alpha \frac{t_i}{\beta} v_i.$$

Taking the limit on I' yields $< \nabla_x \phi(x^*, \omega^*), d^* > \geq \alpha v^*$. Now, since the iterates x_i are not feasible and ω^* belongs to $\Omega_{q,0}(x^*)$, the limit weight $W_q(x^*, x^*, \omega^*)$ is zero. So that, for $i \in I'$ large enough, (5.2) is never satisfied.[5] In case Relationship (5.5) always holds, $< \nabla \phi(x^*, \omega^*), d^* > \geq 0$ so that, again, inequality (5.2) never holds for $i \in I'$ large enough.

□

Line search LS2. (Wolfe-like)

The second line search to be introduced is of the Wolfe type. This kind of line search is widely used in nondifferentiable optimization. Besides $\alpha \in (0,1)$, a number $\bar{\alpha} \in (\alpha, 1)$ is given. The set $Y \equiv Y_{q,M}(x,d,v)$ is taken as the empty set and a stepsize $t \equiv t_{q,M}(x,d,v)$ satisfying the following conditions (see below for an explicit procedure for computing such t) is obtained.

(a. Phase 1.) If $\psi_q^+(x) > 0$, either (a1) or (a2) holds.

(a1)

$$\psi_q(x + td) \leq 0 \tag{5.6}$$

[5] Note that, possibly, $\omega_i \notin \Omega_{q,0}(x_i) \quad \forall\ i$, so that the arguments used in connection with (5.4) cannot be repeated for (5.3).

(a2)

$$\psi_q(x+td) - \psi_q(x) < \alpha t v \tag{5.7}$$

and

$$< \nabla_x \phi(x+td, \omega), d > \geq \bar{\alpha} v \tag{5.8}$$

for some $\omega \in \Omega_{q,0}(x+td)$.

(b. Phase 2.) If $\psi_q^+(x) = 0$, either (b1) or (b2) holds.

(b1)

$$f(x+td) < -M;\ \psi_q(x+td) \leq 0 \tag{5.9}$$

(b2) (5.10.a-b) holds and either (5.11) or (5.12) holds.

$$f(x+td) - f(x) \leq \alpha t v \tag{5.10.a}$$

$$\psi_q(x+td) < 0 \tag{5.10.b}$$

$$< \nabla f(x+td), d > \geq \bar{\alpha} v \tag{5.11}$$

$$< \nabla_x \phi(x+td, \omega), d > + \phi(x+td, \omega) \geq \bar{\alpha} v \tag{5.12}$$

for some $\omega \in \Omega_{q,0}(x+td)$.

Let us now describe a procedure for finding such stepsize. The method presented below is a modification of a line search due to Mifflin [1]. To limit the number of gradient evaluations, (5.8) and (5.12) are tested for only one arbitrarily selected $\omega \in \Omega_{q,0}(x+td)$. It is shown below that this will not hinder termination of the procedure. The exit of the line search depends on whether or not the current iterate is feasible. The method is described for a point x satisfying

$\psi_q^+(x) > 0$ (resp. $\psi_q^+(x) = 0$). The procedure is as follows. Set $t^l = 0$, $t^u = +\infty$. Pick $t^t > 0$.

Loop.

If t^t satisfies all required conditions, set $t = t^t$ and stop.

If t^t satisfies (5.7) (resp.(5.10)), set $t^l = t^t$. Otherwise, set $t^u = t^t$.

If $t^u = +\infty$, set $t^t = 2t^t$.

If $t^u < +\infty$, set $t^t = \frac{1}{2}(t^l + t^u)$.

Go back to loop.

$\square$

Proposition 5.2.

Line search LS2 is well defined and Condition C is satisfied.

Proof.

We show by contradiction that Condition C is satisfied with $\theta = \bar{\alpha}$. If Condition C does not hold, there is a fixed discretization index q, a compact set S, a scalar $\bar{v} < 0$, some sequences $\{x_i\}$ and $\{d_i\}$ with elements in S, and some numbers $v_i < \bar{v}$ such that the corresponding steps $t_i \equiv t_{q,M}(x_i, d_i, v_i)$ converge to zero and such that

$$< \nabla f(x_i), d_i > \leq v_i \tag{5.13}$$

and

$$< \nabla f(x_i + t_i d_i), d_i > \geq \bar{\alpha} v_i. \tag{5.14}$$

In view of (5.13) and regularity Assumption 1, the values v_i are bounded. Therefore, it is possible to extract a subset I of indices such that, on that subset, the sequences $\{x_i\}$, $\{d_i\}$

and $\{v_i\}$ converge respectively to some values x^*, d^* and $v^* \leq \bar{v}$. Taking the limit in (5.13) and (5.14) gives

$$< \nabla f(x^*), d^* > \leq v^*$$

and

$$< \nabla f(x^*), d^* > \geq \bar{\alpha} v^*.$$

Those last two inequalities yield $(1 - \bar{\alpha})v^* > 0$, a contradiction with the relationships $\bar{\alpha} < 1$ and $v^* \leq \bar{v} < 0$. Let us now show that the line search is always well defined. We will suppose that x is feasible for the current discretization Ω_q, i.e., that $\psi_q^+(x) = 0$. The proof for the case when x is not feasible is very similar and is thus omitted. Assuming, by contradiction, that the procedure does not terminate, two cases are to be considered. In the first case, (5.10) is always satisfied (so that t^u remains infinite). In that case, the step t^t generated by the line search procedure satisfies

$$f(x + t^t d) - f(x) \leq \alpha t^t v$$

and

$$\psi_q(x + t^t d) < 0$$

and t^t keeps increasing. Since $v < 0$, (5.9) will eventually hold, and the procedure will terminate, a contradiction. In the second case, (5.10) is not always satisfied, so that t^u becomes finite and a binary search takes place. Suppose by contradiction that the line search does not terminate. In that case, the entire interval $[t^l, t^u]$ converges to some nonnegative value t^*. Let us first show that, in that case, the value t^l does not remain equal to 0. Indeed, if t^l is always 0, the values t^u converge to 0 and satisfy either $f(x+t^u d) - f(x) > \alpha t^u v$ or $\psi_q(x+t^u d) > 0$. This contradicts the fact that x is feasible and that, whenever the line search is performed, in view of (3.3) and (3.4), the direction d satisfies $< \nabla f(x), d > \leq v$ if $\psi_q^+(x) = 0$ and $< \nabla_x \phi(x,\omega), d >< 0 \quad \forall\, \omega \in \Omega_{q,0}(x)$. Therefore t^l becomes different from 0, and, since the line search does not terminate, t^l always satisfies

$$f(x+t^l d) - f(x) \leq \alpha t^l v, \tag{5.15}$$

$$\psi_q(x+t^l d) < 0, \tag{5.16}$$

and

$$< \nabla f(x+t^l d), d > \, < \bar{\alpha} v, \tag{5.17}$$

and there exists a point ω of the discretization Ω_q, depending on t^l, satisfying

$$\omega \in \Omega_{q,0}(x+t^l d); \ < \nabla_x \phi(x+t^l d, \omega), d > + \phi(x+t^l d, \omega) < \bar{\alpha} v \tag{5.18}$$

and, for t^u, it holds either

$$f(x+t^u d) - f(x) > \alpha t^u v \tag{5.19}$$

or

$$\psi_q(x+t^u d) > 0. \tag{5.20}$$

If (5.19) is satisfied infinitely often, it holds, in view of (5.15),

$$f(x+t^u d) - f(x+t^l d) > \alpha(t^u - t^l)v$$

and therefore, $< \nabla f(x+t^*d), d > \geq \alpha v$, in contradiction with (5.17) and the inequality $\alpha \leq \bar{\alpha}$. On the other hand, if (5.20) is always satisfied, it follows from (5.16) that

$$\psi_q(x + t^*d) = 0 \tag{5.21}$$

and, since Ω_q is finite, there exist a mesh point $\hat{\omega} \in \Omega_q$ and a subsequence of points t^l such that $\psi_q(x + t^l d)$ is strictly increasing and (5.18) is always satisfied with $\hat{\omega}$. In particular, $\hat{\omega} \in \Omega_{q,0}(x + t^l d)$ so that

$$\psi_q(x + t^l d) = \phi(x + t^l d, \hat{\omega})$$

on the entire subsequence and thus

$$< \nabla_x \phi(x + t^* d, \hat{\omega}), d > \geq 0.$$

However, in view of (5.21)

$$\phi(x + t^* d, \hat{\omega}) = 0$$

so that (5.18) implies

$$< \nabla_x \phi(x + t^* d, \hat{\omega}), d > \leq \bar{\alpha} v,$$

a contradiction.

$\square$

6. Discussion

A class of globally convergent algorithms has been presented and two corresponding line search procedures have been discussed. These two procedures, while conceptually simple, may be impeded by an excessive number of function evaluations per iteration due to repeated violation of a constraint not taken into account in the search direction. Such a problem will likely not occur when the discretization is fine, as the global maximizers of $\phi(x,\cdot)$ are typically well approximated by points in $\bar{\Omega}_{q,\epsilon}(x)$. When the discretization is coarse, it may be appropriate to use a more sophisticated line search such as the one proposed by Mifflin in [10] in the context univariate nonsmooth optimization. This could essentially amount to replacing the binary search in Line search LS2 by an interpolation scheme based on derivative information. Yet another possibility would be that of not requiring that (5.8) (or (5.12)) be satisfied by the next iterate provided it is satisfied at some pair (y,ω) to be included in the set Y. A similar idea has been widely used in nondifferentiable optimization (see, e.g., [1,3]).

Other refinements may be appropriate in the interest of computational efficiency. First, in Step 1 of Algorithm A, initialization of the set J_0 for a given discretization level k (inner loop) could take into account information contained in I_{k-1}, collected at the previous discretization level. Also, inside the inner loop, it may be desirable to systematically drop from the set J_i pairs (y,ω) with a high value $W_{q_k}(x_i,y,\omega)$. Second, although we have assumed throughout that the discretization was uniformed, all our convergence results still hold if it is merely assumed that the partition 'grows dense' as k goes to infinity. Correspondingly, nonuniform discretization patterns could be used to take advantage of a priori (or acquired) information on the 'shape' of $\phi(x,\cdot)$. Finally, it is clear that proper scaling should be introduced at various places in Algorithm A.

References

[1] R. Mifflin, "An Algorithm for Constrained Optimization with Semismooth Functions," *Math. Oper. Res.*, 2, no. 2, pp. 191–207, 1977.

[2] E. Polak, D. Q. Mayne and Y. Wardi, "On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems," *SIAM J. Control Optim.*, 21, no. 2, pp. 179–203, 1983.

[3] K. C. Kiwiel, *Methods of Descent in Nondifferentiable Optimization* (Lecture Notes in Mathematics). Berlin, Heidelberg, Springer-Verlag, 1985.

[4] E. Polak and D. Q. Mayne, "An Algorithm for Optimization Problems with Functional Inequality Constraints," *IEEE Trans. Automat. Control*, 21, no. 2, pp. 184–193, 1976.

[5] C. Gonzaga, E. Polak and R. Trahan, "An Improved Algorithm for Optimization Problems with Functional Inequality Constraints," *IEEE Trans. Automat. Control*, AC-25, no. 1, pp. 49–54, 1980.

[6] K. C. Kiwiel, *An Aggregate Subgradient Method for Nonsmooth Convex Minimization.* System Research Institute of the Polish Academy of Sciences, 1983.

[7] ———, *A Linearization Algorithm for Nonsmooth Minimization.* System Research Institute of the Polish Academy of Sciences.

[8] C. Lemaréchal, J. J. Strodiot and A. Bihain, "On a Bundle Algorithm for Nonsmooth Optimization," in *Nonlinear Programming 4* . Academic Press, New York-London , pp. 245–282, 1981.

[9] P. Wolfe, "On the Convergence of Gradient Methods Under Constraints," IBM Yorktown Research Facility, Yorktown Heights, N.Y., IBM Research Report RC 1752, 1967.

[10] R. Mifflin, "Stationarity and Superlinear Convergence of an Algorithm for Univariate Locally Lipschitz Constrained Minimization," *Math. Programming*, 28, pp. 50–71, 1984.