SRC TR 88-16

A Globally Convergent Algorithm with Adaptively Refined Discretization for Semi-Infinite Optimization Problems Arising in Engineering Design

by

E.R. Panier and A.L. Tits

A Globally Convergent Algorithm with Adaptively Refined Discretization for Semi-Infinite Optimization Problems Arising in Engineering Design

Eliane R. Panier *
André L. Tits +
Systems Research Center and Electrical Engineering Department
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. Such problems pertain to semi-infinite optimization. While most of the algorithms that have been proposed for the solution of these problems make use, at each iteration, of a set of local maximizers over the range of the independent parameter, the question of suitably approximating such maximizers is generally left aside. It has been suggested that this issue can be addressed by means of an adaptively refined discretization of the interval of variation of the independent parameter. The algorithm proposed in this paper makes use of such a technique and, by means of a certain memory mechanism, avoids the potential lack of convergence suffered by an existing algorithm.

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 of this type pertain to *semi-infinite optimization.* They are typical, among other places, in the context of circuit design or control system design (see, e.g., [1,2] and references therein). While, at the expense of repeatedly computing global solutions of univariate optimization subproblems, semi-infinite optimization problems could be handled by general purpose nondifferentiable optimization algorithms (see, e.g., [3–6]), their particular structure calls for specific techniques. Such techniques have been investigated by many authors (see, e.g., [7–14]). While most approaches make use, at each iteration, of a set of local maximizers over the range of the independent parameter, the question of suitably approximating such maximizers is generally left aside. Notable exceptions are found in [15] for linear problems and in the work of Polak *et al.* [7, 10] for the general nonlinear case. The proposed schemes are based on an adaptively refined discretization of the interval of variation of the independent parameter. This is also the focus of this paper.

For the sake of exposition, we consider the simple problem

$$(P) \qquad \text{minimize } f(x) \text{ s.t. } \phi(x,\omega) \leq 0 \; \forall \; \omega \in [0,1]$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $\phi : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ is continuous and is continuously differentiable with respect to its first argument.[1] Polak *et al.* [7,10] have proposed two algorithms for solving problem (P), both using an adaptively refined discretization of the interval [0,1]. In the first algorithm [7], a feasible direction scheme is used that yields an approximate solution to a problem (P_q) obtained by replacing $\Omega = [0,1]$ by a finite subset Ω_q. The search direction computation makes use of the gradients of ϕ at all ϵ-active discretization points (see below for a precise definition). The discretization is then progressively refined and the corresponding problem is solved to a progressively better accuracy. Convergence

[1] Extension of the ideas discussed in this paper to problems including multiple functional constraints over different compact intervals, together with multiple 'ordinary' constraints, does not present any conceptual difficulties.

of the overall algorithm to stationary points of (P) is proven. The second algorithm [10] 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 [10] makes use of the latter only.

In this paper it is shown that, while the algorithm of [10] usually performs quite well, there are cases where convergence to stationary points is lost. The reason is that, for a given mesh size, the gradients at the local maximizers for the current iterate do not always 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 that convergence can be recovered via a simple modification of the algorithm in [10].

The balance of the paper is organized as follows. In Section 2, the adaptive discretization scheme proposed in [10] is outlined and examples are exhibited where the corresponding algorithm fails to converge. In Section 3, a modified scheme is proposed. Numerical results are presented in Section 4. Concluding remarks are given in Section 5. A proof of convergence for the modified algorithm is sketched in the Appendix.

2. Preliminaries

In this section, after briefly outlining the algorithm of [10], we show by two examples that difficulties can arise, preventing convergence to a stationary point. The notations, terminology and assumptions to be used throughout this paper, essentially borrowed from [10], are introduced first.

For ease of reference, we restate the regularity assumptions for Problem (P).

Assumption 1.[2] $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, $\phi : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ 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

$$\psi(x) = \max_{\omega \in [0,1]} \phi(x, \omega)$$

[2] 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.)

is well defined, so that Problem (P) can be reformulated as

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

Note that ψ is not everywhere differentiable and that, due to the infinite cardinality of $[0,1]$, its value at a given point x cannot in general be computed exactly in finitely many operations. Below, we will make use of finite discretizations of $[0,1]$.

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

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

and the set of corresponding gradients by

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

The convex hull of the latter is denoted by $\mathrm{co}S(x)$. The following standard assumption will be made throughout.

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

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 $\omega_j^* \in [0,1], \quad j = 1,\ldots,l$ and some coefficients $\lambda_j^* \geq 0, \quad j = 1,\ldots,l$ satisfying

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

and

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

Under Assumptions 1 and 2, every local minimizer 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 $\Omega = [0,1]$ is now discretized into

$$\Omega_q = \{\omega = \frac{l}{q}, \quad l = 0,1,\ldots,q\}.$$

The constraint function is approximated accordingly by

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

and we define

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

The discretized problem is

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

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. $0 < \omega < 1$ and the two inequalities

$$\phi(x,\omega) \geq \phi(x,\omega + \frac{1}{q}) \tag{2.1}$$

and

$$\phi(x,\omega) > \phi(x,\omega - \frac{1}{q}) \tag{2.2}$$

hold.

ii. $\omega = 0$ and (2.1)holds.

iii. $\omega = 1$ 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

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

For a given discretization, an initial precision ϵ and a current iterate x_i, the algorithm in [10] computes a search direction d_i by solving the quadratic program in (d, v)

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

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 $\alpha,\ \beta \in (0,1)$. 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

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

and

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

Notice that the computation of the search direction takes into account constraint information only at local maximizers ω at the current iterate x_i. While this information often provides a suitable local representation of ϕ around x_i, such is not always the case. Below, we illustrate this fact by means of two examples where Assumptions 1 and 2 hold but where convergence fails to occur.[3] In both cases the constructed sequence cannot overtake a 'corner' in $\psi(x)$, as this corner is never 'seen' at current local maximizers.

Problem 1. [4]

Problem (P) with $n = 2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ 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

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

[3] In fact, both examples violate an assumption implicitly used in [10] that there is no $x \in \mathbb{R}^n$ at which, for some discretization mesh, $\phi(x,\cdot)$ takes identical values at two successive mesh points. This strong assumption is made explicit by Polak in [16] (Assumption 7.2).

[4] This problem is, in disguised form, the minimax problem $\min_\eta \max_{\omega\in[0,1]} (2\omega-1)\eta+\omega(1-\omega)(1-\eta)$.

and that the solution of Problem 1 is given by $(\sqrt{5}-2, 1-2\frac{\sqrt{5}}{5})$, 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 $(0,0)$. 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 $\psi_q(x_0) \leq 0$) for $q = 1$, problem (2.3) will always be one of three fixed problems with corresponding optimal values τ_i of $-1/10$, $-1/2$ and $-1/2$ respectively. Thus, if $\delta\epsilon < 1/10$ initially, ϵ will never be decreased and the discretization will never be refined. It can be checked that convergence to $(0,0)$, 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, failures such as the above should be expected to frequently take place.

Difficulties 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 refinement 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 $n = 2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ and $\phi : \mathbb{R}^2 \times [0,1] \to \mathbb{R}$ respectively given by

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

and[5]

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

[5] The first term in $\phi(x,\omega)$ is introduced so as to satisfy assumption made in [10] that for all $x \in \mathbb{R}^n$, $\Omega(x)$ is finite.

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

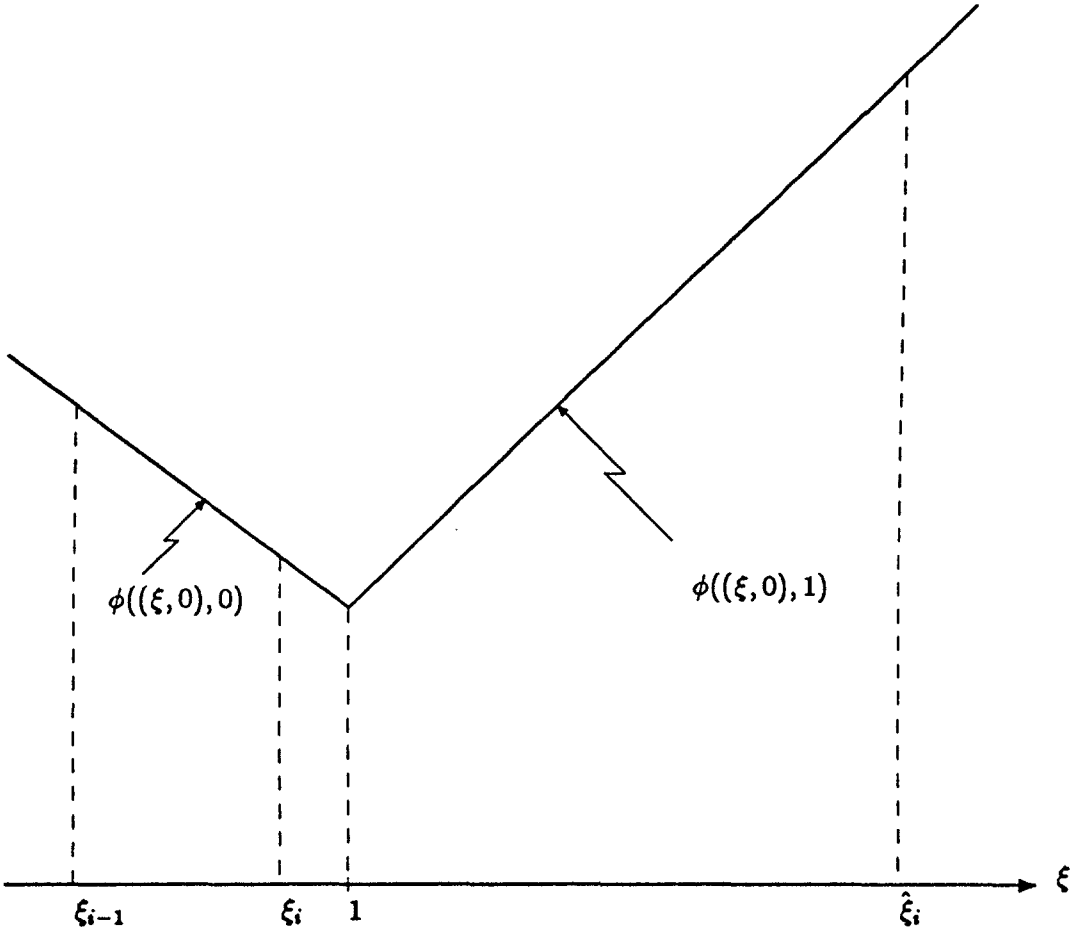
□

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 note that this problem is convex. Since (ξ,η) is feasible for $\xi \geq 7/3$ with $\xi + \eta \leq 0$ and since $f(x) \to -\infty$ as $\xi \to +\infty$, there is no Kuhn-Tucker point. However, it is easily checked that if the algorithm just outlined is used on Problem 2 with parameter values α, β, δ and ϵ satisfying the relationships $\alpha = 1/2$, $\beta = 1/4$, $\delta\epsilon = 1/4$ and γ being any positive number, then, if $q = 1$ initially and $x_0 = (0,0)$, ϵ is never decreased (so that the discretization is never refined) and the successive iterates are given by $x_i = (1 - 4^{-i}, 0)$, $i = 1, 2, \cdots$. The limit point $(1,0)$ is not feasible for (P) (or (P_q)). Thus convergence occurs to a point that is not even a solution of the discretized problem. This failure can be explained as follows (see Fig. 2.1). At $x^* = (1,0)$ 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 $\{x \mid \eta = 0\}$.

Obviously, in Problems 1 and 2, satisfactory performance will not be achieved unless computation of the search direction takes into account *simultaneously* the gradients $\nabla_x \phi(x_i,\omega)$ at both $\omega = 0$ and $\omega = 1$. More generally, it should be clear that, in a large class of problems, difficulties may arise unless one makes use in the search direction computation of gradient information at values of ω that maximize $\phi(x,\cdot)$ at suitably selected nearby points x, rather than exclusively at values of ω that are local maximizers at the iterate x_i. Indeed, as in Problems 1 and 2, the line search may be impeded by large values of ϕ at the former values of ω. In the context of the Armijo line search used in [10], a simple idea would be to use as additional values of ω the maximizers of $\phi(\hat{x}_i,\cdot)$ at the last *unsuccessful* trial point $\hat{x}_i$ of the previous line search (see Fig. 2.1; this would prevent the type of jamming observed in Problem 2), as well as certain values of ω used at recent iterations (to avoid the type of jamming observed in Problem 1). These ideas are made precise in the next section. Other



$x_{i-1} = (\xi_{i-1}, 0)$; $x_i = (\xi_i, 0)$; $x^* = (1, 0)$; $\hat{x}_i = (\hat{\xi}_i, 0)$

Fig. 2.1. $\psi(x)$ for Problem 2 with $\eta = 0$

schemes, possibly more efficient, are briefly discussed in Section 5.

3. An Algorithm

Based on the previous observations, we propose the following simple modification of the algorithm in [10]. Here the iterate X_k is updated every time the discretization is refined (outer loop) while x_i denotes subiterates, at a given discretization level (inner loop).

Algorithm A.

Parameters. $\delta > 0$; $\gamma > 0$; α, $\beta \in (0,1)$; $\epsilon_0 > 0$; $N_0 > 0$; $q_0 \in \mathbb{N}\backslash\{0\}$.

Data. $X_0 \in \mathbb{R}^n$.

Step 0. Initialization of the outer loop. Set $k = 0$, $\epsilon = \epsilon_0$, $N = N_0$, $q = q_0$.

Step 1. Initialization of the inner loop. Set $i = 0$, $x_0 = X_k$, $\hat{\Omega}_0 = \bar{\Omega}_{q,\epsilon}(x_0)$.

Step 2. Search direction computation. Compute (d_i, v_i) solution of the quadratic program in (d, v)

$$\begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } \langle \nabla f(x_i), d\rangle - \gamma\psi_q^+(x_i) \leq v \\ \quad \langle \nabla_x \phi(x_i, \omega), d\rangle + \phi(x_i, \omega) - \psi_q^+(x_i) \leq v \quad \forall\, \omega \in \hat{\Omega}_i. \end{cases} \tag{3.1}$$

Let τ_i be the optimal value of this quadratic program and let $\lambda_i(\omega)$, $\omega \in \hat{\Omega}_i$ be the multipliers corresponding to the third line of (3.1). 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 (such a choice is always possible in view of Caratheodory's Theorem).

Step 3. Optimality test. If $\tau_i \geq -\delta\epsilon$ or $\|x_i\| > N$, go to Step 6. Otherwise, go to Step 4.

Step 4. Line search. If $\psi_q^+(x_i) > 0$ (Phase 1), let t_i be the first number t in the sequence $\{1,\ \beta,\ \beta^2,\ \ldots\}$ satisfying either

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

or

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

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

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

and

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

Step 5. Updates. Set $x_{i+1} = x_i + t_i d_i$ and $\hat{\Omega}_{i+1} =$

$$\begin{cases} \bar{\Omega}_{q,\epsilon}(x_{i+1}) \cup \{\omega \in \hat{\Omega}_i | \lambda_i(\omega) \neq 0\} \cup \Omega_{q,0}(\hat{x}_{i+1}), \text{with } \hat{x}_{i+1} = x_i + \beta^{-1} t_i d_i, & \text{if } t_i < 1 \text{ and } \psi_q^+(\hat{x}_{i+1}) > 0 \\ \bar{\Omega}_{q,\epsilon}(x_{i+1}) \cup \{\omega \in \hat{\Omega}_i | \lambda_i(\omega) \neq 0\} & \text{otherwise.} \end{cases} \tag{3.2}$$

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

Step 6. Discretization refinement. If $\tau_i \geq -\delta\epsilon$, set $\epsilon = \epsilon/2$; if $\|x_i\| > N$, set $N = 2\|x_i\|$. Set $q = 2q$, $X_{k+1} = x_i$, $k = k + 1$, and go back to Step 1.

□

The main difference between Algorithm A and the algorithm in [10] is the use of $\hat{\Omega}_i$ instead of $\tilde{\Omega}_{q,\epsilon}(x_i)$ in the search direction computation (compare (3.1) to (2.3)). As seen in (3.2), the former includes, besides $\bar{\Omega}_{q,\epsilon}(x_i)$, [6] all points from $\hat{\Omega}_{i-1}$ that affected the previous search direction (i.e., those with non zero multipliers $\lambda_{i-1}(\omega)$) as well as all global maximizers at $\hat{x}_i$, the last unsuccessful trial point in the previous line search. The other differences between Algorithm A and the algorithm in [10] are not essential: (i) A term $\phi(x_i,\omega) - \psi_q^+(x_i)$ is introduced in the last line of (3.1). The effect of this term is to de-emphasize the points $\omega \in \hat{\Omega}_i$ for which $\phi(x_i,\omega)$ is far from being critical (i.e., is significantly less than $\psi_q^+(x_i)$), as is clear from the equivalent formulation

$$\text{minimize } \frac{1}{2}\|d\|^2 + \max\{\langle \nabla f(x_i), d\rangle - \gamma\psi_q^+(x_i); \langle \nabla_x \phi(x_i,\omega), d\rangle + \phi(x_i,\omega) - \psi_q^+(x_i) \mid \omega \in \hat{\Omega}_i\}.$$

While this makes it unnecessary to adaptively reduce ϵ, ϵ is nethertheless halfed in Step 6, thus decreasing the number of gradients necessary in the quadratic program. (ii) The discretization is refined not only when τ_i is small[7] but also if $\|x_i\|$ becomes large. This ensures that k will be eventually incremented (and the discretization refined), even if $\{x_i\}$ tends to diverge at the current discretization level. (iii) In phase 1, the line search is terminated if $\psi_q(x_i + td_i) \leq 0$ (transition to phase 2).

We now state the main convergence theorem for Algorithm A. An outline of its proof is given in the appendix.

Theorem T (Global Convergence).
Algorithm A constructs an infinite sequence $\{X_k\}$ and every accumulation point of this sequence is a Kuhn-Tucker point for (P).

□

4. Numerical Considerations

To keep Algorithm A simple, we made no attempt at achieving high computational efficiency (but, see Section 5 below for some suggestions in that respect). Rather, our purpose was merely to devise an algorithm that would avoid the jamming phenomena observed

[6] $\bar{\Omega}_{q,\epsilon}(x_i)$ is used instead of $\tilde{\Omega}_{q,\epsilon}(x_i)$ to avoid the situation (occurring with probability 0 in exact arithmetic) where no positive step would satisfy the Armijo line search of Step 4.

[7] Note that refining the discretization everytime $\tau_i > -\delta\epsilon$ is compatible with the algorithm in [10].

in connection with Problems 1 and 2. This was achieved by making use of an iteratively updated set $\hat{\Omega}_i$ instead of $\tilde{\Omega}_{q,\epsilon}(x_i)$ in the search direction computation. To experimentally assess the validity of this scheme, we first tested Algorithm A on Problems 1 and 2, without including the nonessential modifications discussed above. Parameters values were $\delta = 0.09$, $\gamma = 2.$, $\alpha = 0.5$, $\beta = 0.25$, $\epsilon_0 = 1.$, $N_0 = +\infty$, $q_0 = 1$ for Problem 1 and $\delta = 0.25$, $\gamma = 2.$, $\alpha = 0.5$, $\beta = 0.25$, $\epsilon_0 = 1.$, $N_0 = +\infty$, $q_0 = 1$ for Problem 2. We also tried the same algorithm with $\hat{\Omega}_{i+1} = \bar{\Omega}_{q,\epsilon}(x_{i+1})$ in Step 5, i.e., the algorithm of [10], with the same parameter values. While the latter failed exactly as described in Section 2, at discretization level $q = 1$, the former behaved satisfactorily, converging fairly rapidly to $(\sqrt{5} - 2, 1 - 2\frac{\sqrt{5}}{5})$ on Problem 1 and diverging on Problem 2.

While our goal was to avoid possible jamming, it would be counter productive if this were achieved at the expense of a significantly reduced efficiency on many "well behaved" problems of interest. With this concern in mind we tested Algorithm A on the example problem used in [10]. This problem consists in the design of a PID controller for a third order SISO system, for which the integral of the square of the error in the output due to a step input is to be minimized, subject to a (functional) constraint on the Nyquist plot to ensure robust stability. We use as parameter values, $\delta = 10^{-3}$, $\gamma = 2.$, $\alpha = 0.2$, $\beta = 0.3$, $\epsilon_0 = 0.2$, $N_0 = +\infty$, $q_0 = 1$. Most of these values are identical to those used in [10]. The initial discretization level $q_0 = 1$ (compared to $q_0 = 128$ in [10]) was selected, to test more intensively the discretization refinement scheme. For this example the behavior of the algorithm was *identical* whether $\hat{\Omega}_i$ was updated as specified in Step 5 or whether it was set to $\bar{\Omega}_{q,\epsilon}(x_i)$. In fact, for every i, either this set was empty on it consisted of a single point, the global maximizer at x_i.

5. Discussion

The essential modification in Algorithm A with respect to the algorithm in [10] is that in the former the search direction computation makes use of gradients $\nabla_x \phi(x_i, \omega)$ at some values of ω that are not ϵ-active local maximizers at x_i but have proved critical at nearby points x. This technique has some similarity with that used in bundle algorithms, in nondifferentiable optimization, where the search direction computation makes use of elements of generalized gradients at other points than the current iterate x_i (see, e.g., [17]). The idea of null step used in that context could be adopted here as well: one would terminate the line search after a few unsuccessful trial points, add to $\hat{\Omega}_i$ the global maximizers ω at the closest to

x_i among these points and recompute a search direction at x_i. Another popular tool in nondifferentiable optimization is the Wolfe-type line search, which yields a new iterate $x_{i+1} = x_i + t_i d_i$ that, besides achieving a sufficient decrease of the objective function, results in a directional derivative $\psi'(x_{i+1},\ d_i)$ of the potentially binding constraint sufficiently larger then $\psi'(x_i,\ d_i)$ (see, e.g., [3]). A similar idea can be used here, with ψ replaced by ψ_q (see [18] for details). It would be of interest to investigate the use in the present context of other nondifferentiable optimization techniques (see, e.g., [5,6,19]).

The set $\bar{\Omega}_{q,\epsilon}(x_i)$, which includes all ϵ-active local maximizers at x_i could have been replaced in Steps 1 and 5 by the set of global maximizers $\Omega_{q,0}(x_i)$, without jeopardizing the convergence properties. As another alternative, as suggested at the end of Section 3, ϵ could have been kept constant throughout. The option selected here tends to allow good initial progress and carries relatively moderate computational overhead. Other refinements may be appropriate in the interest of computational efficiency. First, in Step 1 of Algorithm A, initialization of the set $\hat{\Omega}_0$ for a given discretization level k (inner loop) could take into account information collected at the previous discretization level. Also, inside the inner loop, it may be desirable to systematically drop from the set $\hat{\Omega}_i$ mesh points ω that are far from being ϵ-active at the current iterate. Second, although we have assumed throughout that the discretization was uniform, 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 $\phi(x,\cdot)$. Also, it is clear that proper scaling should be introduced at various places in Algorithm A. Finally, while conceptually simple, Algorithm A 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. While feasible direction methods are of special interest in engineering design, they may be very slow in our context. It may therefore be appropriate to allow, on a given discretization, a slight violation of the constraints and to introduce a penalty term in the objective function to force the iterate towards the feasible set.

Finally, it should be clear that, with minor modifications, Algorithm A could be used in cases when $\Omega \subset \mathbb{R}^p$, for some $p > 1$. The computational overhead, however, would soon become prohibitive.

Appendix.

Proof of Theorem T.

Most of the convergence results in [10] are still valid here. The key point in the proof of Theorem T is to show that the discretization mesh is refined infinitely many times. This is sketched below. Suppose by contradiction that there exits a discretization mesh q for which an infinite sequence $\{x_i\}$ is constructed. For simplicity, we will assume that all the iterates in the sequence are feasible, so that a phase 2 iteration is always performed by the line search. If the discretization is never changed, then, for all i, $\|x_i\| \leq N$ and

$$\tau_i < -\delta\epsilon. \tag{a.1}$$

Using standard arguments, it can be shown that under these conditions, the sequence of iterates $\{x_i\}$ converges to some vector, say x^*; the sequence of steps $\{t_i\}$ converges to zero; and the sequences $\{d_i\}$, $\{v_i\}$, $\{\tau_i\}$ are bounded. Let $\tau^* = \limsup \tau_i$. In view of the above, there exists an infinite subset of indices $I \subset \mathbb{N}$, a nonnegative integer $l \leq n+1$, numbers $\omega_1^*, \cdots, \omega_l^*$ in Ω_q, vectors d^* and d^{1*} and real numbers v^*, v^{1*}, τ^{1*} such that for all $i \in I$, $t_i < 1$ and exactly l constraints corresponding to $\omega_1^*, \cdots, \omega_l^*$, have nonzero multipliers in (3.1), and such that $\{d_i\}_{i\in I} \to d^*$, $\{d_{i+1}\}_{i\in I} \to d^{1*}$, $\{v_i\}_{i\in I} \to v^*$, $\{v_{i+1}\}_{i\in I} \to v^{1*}$, $\{\tau_i\}_{i\in I} \to \tau^*$, $\{\tau_{i+1}\}_{i\in I} \to \tau^{1*}$. In view of (a.1), $v^* \leq -\delta\epsilon < 0$. For all $i \in I$, (d_i, v_i) solves

$$(QP_i) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } \langle \nabla f(x_i), d\rangle \leq v \\ \qquad \langle \nabla_x \phi(x_i, \omega_j^*), d\rangle + \phi(x_i, \omega_j^*) \leq v, \quad j = 1, \ldots, l \end{cases}$$

where only those constraints of nonzero multipliers appear, since clearly (d_i, v_i) satisfies the optimality conditions associated with (QP_i) and it can be shown that the solution to those optimality conditions is unique. Also, since a phase 2 iteration is performed and, on I, the stepsize t_i is strictly less than one then, for all $i \in I$, either

$$f(\hat{x}_{i+1}) - f(x_i) > -\alpha\beta^{-1} t_i \delta\epsilon \tag{a.2}$$

or

$$\psi_q^+(\hat{x}_{i+1}) > 0$$

with $\hat{x}_{i+1} = x_i + \beta^{-1} t_i d_i$. Using standard arguments, one can show that $(a.2)$ cannot occur infinitely many times. Thus, without loss of generality, there exists a point $\omega^* \in \Omega_{q,0}(\hat{x}_{i+1})$ (ω^* independent of i) such that, for $i \in I$,

$$\phi(\hat{x}_{i+1}, \omega^*) > 0\ . \tag{a.3}$$

Next, (d_{i+1}, v_{i+1}) solves the quadratic program

$$(QP_{i+1}) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } \langle \nabla f(x_{i+1}), d\rangle \leq v \\ \quad \langle \nabla_x \phi(x_{i+1}, \omega_j^*), d\rangle + \phi(x_{i+1}, \omega_j^*) \leq v, \quad j = 1, \ldots, l \\ \quad \langle \nabla_x \phi(x_{i+1}, \omega^*), d\rangle + \phi(x_{i+1}, \omega^*) \leq v \\ \quad + \text{ other inequalities} \end{cases}$$

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

$$(QP^*) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } \langle \nabla f(x^*), d\rangle \leq v \\ \quad \langle \nabla_x \phi(x^*, \omega_j^*), d\rangle + \phi(x^*, \omega_j^*) \leq v, \quad j = 1, \ldots, l. \end{cases}$$

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^*, v^*) therefore satisfies the optimality conditions associated with (QP^*). It can be shown that the solution to these optimality conditions is unique, so that (d^*, v^*) must solve (QP^*). Similarly, (d^{1*}, v^{1*}) solves

$$(QP^{1*}) \quad \begin{cases} \text{minimize } \frac{1}{2}\|d\|^2 + v \\ \text{s.t. } \langle \nabla f(x^*), d\rangle \leq v \\ \quad \langle \nabla_x \phi(x^*, \omega_j^*), d\rangle + \phi(x^*, \omega_j^*) \leq v, \quad j = 1, \ldots, l \\ \quad \langle \nabla_x \phi(x^*, \omega^*), d\rangle + \phi(x^*, \omega^*) \leq v \\ \quad + \text{ other limit inequalities} \end{cases}$$

where the 'other limit inequalities' correspond to a suitable subsequence. In view of $(a.3)$, and since $\psi_q^+(x_i) = 0$ for all i, it follows that $\phi(x^*, \omega^*) = 0$ and that the unique solution (d^*, v^*) of (QP^*) satisfies

$$\langle \nabla_x \phi(x^*, \omega^*), d^*\rangle \geq 0 > v^*.$$

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

□

Acknowledgement

The authors wish to thank Tam Q. Nguyen for implementing and testing Algorithm A.

References

[1] R.K. Brayton, G.D. Hachtel & A.L. Sangiovanni-Vincentelli, "A Survey of Optimization Techniques for Integrated Circuit Design," *IEEE Proc.* 69 (1981), 1334–1362.

[2] E. Polak, D.Q. Mayne & D.M. Stimler, "Control System Design via Semi-Infinite Optimization: A Review," *IEEE Proc.* 72 (1984), 1777–1794.

[3] R. Mifflin, "An Algorithm for Constrained Optimization with Semismooth Functions," *Math. Oper. Res.* 2 (1977), 191–207.

[4] E. Polak, D.Q. Mayne & Y. Wardi, "On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems," *SIAM J. Control Optim.* 21 (1983), 179–203.

[5] K.C. Kiwiel, *Methods of Descent in Nondifferentiable Optimization*, Lecture Notes in Mathematics #1133, Springer-Verlag, Berlin, Heidelberg, New-York, Tokyo, 1985.

[6] C. Lemaréchal, "Nondifferentiable Optimization," in *Handbook on Operations Research*, Nemhauser, Rinnooy Kan & Todd, eds., North Holland, 1989,to appear.

[7] E. Polak & D.Q. Mayne, "An Algorithm for Optimization Problems with Functional Inequality Constraints," *IEEE Trans. Automat. Control* 21 (1976), 184–193.

[8] D.Q. Mayne, E. Polak & R. Trahan, "An Outer Approximations Algorithm for Computer-Aided Design Problems," *J. Optim. Theory Appl.* 28 (1979), 331–351.

[9] R. Hettich & W. van Honstede, "On Quadratically Convergent Methods for Semi-Infinite Programming," in *Semi-Infinite Programming*, R. Hettich, ed., Lecture Notes in Control and Information Sciences #15, Springer Verlag, 1979, 97–111.

[10] C. Gonzaga, E. Polak & R. Trahan, "An Improved Algorithm for Optimization Problems with Functional Inequality Constraints," *IEEE Trans. Automat. Control* AC-25 (1980), 49–54.

[11] G.A. Watson, "Globally Convergent Methods for Semi-Infinite Programming," *BIT* 21 (1981), 362–373.

[12] S.A. Gustafson, "A Three-Phase Algorithm for Semi-Infinite Programs," in *Semi-Infinite Programming and Applications*, A.V. Fiacco & K.O. Kortanek, eds., Lecture Notes in Economics and Mathematical Systems #215, Springer Verlag, 1983, 138–157.

[13] E. Polak & A.L. Tits, "A Recursive Quadratic Programming Algorithm for Semi-Infinite Optimization Problems," *Appl. Math. Optim.* 8 (1982), 325–349.

[14] I.D. Coope & G.A. Watson, "A Projected Lagrangian Algorithm for Semi-Infinite Programming," *Math. Programming* 32 (1985), 337–356.

[15] R. Hettich, "An Implementation of a Discretization Method for Semi-Infinite Programming," *Math. Programming* 34 (1986), 354–361.

[16] E. Polak, "On the Mathematical Foundations of Nondifferentiable Optimization ," *SIAM Rev.* 29 (1987), 21–89.

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

[18] E.R. Panier & A.L. Tits, "Globally Convergent Algorithms for Semi-Infinite Optimization Problems Arising in Engineering Design," Systems Research Center, University of Maryland, Technical Report TR-87-28, College Park, MD 20742, 1987.

[19] V.F. Dem'yanov & L.V. Vasil'ev, *Nondifferentiable Optimization*, Translations Series in Mathematics and Engineering, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1985.