SIAM J. Optimization
Vol. x, No. x, pp. x–xx, xxx 19xx

AN SQP ALGORITHM FOR FINELY DISCRETIZED CONTINUOUS MINIMAX PROBLEMS AND OTHER MINIMAX PROBLEMS WITH MANY OBJECTIVE FUNCTIONS*

JIAN L. ZHOU† AND ANDRÉ L. TITS‡

Abstract. A common strategy for achieving global convergence in the solution of semi-infinite programming (SIP) problems, and in particular of continuous minimax problems, is to (approximately) solve a sequence of discretized problems, with a progressively finer discretization meshes. Finely discretized minimax and SIP problems, as well as other problems with many more objectives/constraints than variables, call for algorithms in which successive search directions are computed based on a small but significant subset of the objectives/constraints, with ensuing reduced computing cost per iteration and decreased risk of numerical difficulties. In this paper, an SQP-type algorithm is proposed that incorporates this idea in the particular case of minimax problems. The general case will be considered in a separate paper. The quadratic programming subproblem that yields the search direction involves only a small subset of the objective functions. This subset is updated at each iteration in such a way that global convergence is insured. Heuristics are suggested that take advantage of a possible close relationship between "adjacent" objective functions. Numerical results demonstrate the efficiency of the proposed algorithm.

Key words. continuous minimax, semi-infinite programming, many constraints, sequential quadratic programming, discretization, global convergence.

AMS(MOS) subject classifications. 49M07, 49M37, 49M39, 65K05, 90C06, 90C30, 90C34

1. Introduction.

Optimization problems that arise in engineering design often belong to the class of Semi-Infinite Programming (SIP) problems, *i.e.*, they involve specifications that are to be satisfied over an interval of values of an independent parameter such as time, frequency, temperature or modeling error (see, *e.g.*, [2], [3], [30], [33]). A simple example is given by the problem

$$\text{(SI)} \qquad \text{minimize} \quad f(x) \quad \text{s.t.} \quad \Phi_{[0,1]}(x) \leq 0,\ x \in R^n,$$

with

$$\Phi_{[0,1]}(x) := \sup_{\omega \in [0,1]} \phi(x,\omega).$$

An important class of SIP problems is that of "continuous" minimax problems such as

$$\text{(CMM)} \qquad \text{minimize} \sup_{\omega \in [0,1]} \phi(x,\omega).$$

Note that (CMM) can be equivalently formulated in the form of (SI) as

$$\text{minimize } x^0 \quad \text{s.t.} \sup_{\omega \in [0,1]} \phi(x,\omega) \leq x^0, (x,x^0) \in R^{n+1}.$$

*Received by the editors xxx xx, 19xx; accepted for publication (in revised form) xxx xx, 19xx. This research was supported in part by NSF's Engineering Research Centers Program No. NSFD-CDR-88-03012, by NSF grant No. DMC-88-15996 and by a grant from the Westinghouse Corporation.

†Institute for Systems Research, University of Maryland, College Park, MD 20742.

‡Institute for Systems Research and Electrical Engineering Department, University of Maryland, College Park, MD 20742; all correspondence should be addressed to this author.

The difficulties in solving (SI), and in particular (CMM), stem mostly from the facts that (i) the accurate evaluation of $\Phi_{[0,1]}$ for each x involves a potentially time consuming global maximization, and (ii) $\Phi_{[0,1]}$ is nondifferentiable in general, even when ϕ is smooth. Various approaches have been proposed to circumvent these difficulties (see [18] for a recent survey). Some algorithms are based on the characterization of maximizers of $\phi(x, \cdot)$ over $[0, 1]$ in the neighborhood of a local solution of (SI) (see, *e.g.*, [12], [17], [20], [34]). Under mild assumptions, the set of such maximizers contains a "small" number of points (for small n). The solution of the original problem can then be reduced to the solution of a problem involving approximations to these maximizers $\omega_i(x)$. Application of Newton's method, or of a Sequential Quadratic Programming (SQP) method to the reduced problem (with constraints $\phi(x, \omega_i(x)) \leq 0$) brings about a fast local rate of convergence. However global convergence, when insured at all, involves a potentially very costly line search ([5], [42]). A large class of globally convergent algorithms, on the other hand, is based on approximating $\Phi_{[0,1]}$ by means of progressively finer discretizations of $[0, 1]$, *i.e.*, substituting for (SI) the problems

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

with, for instance,

$$\Omega = \{0, \frac{1}{q}, \frac{2}{q}, \cdots, \frac{(q-1)}{q}, 1\},$$

where q, a positive integer, is progressively increased (see, *e.g.*, [10], [13], [16], [27], [31], [32], [34], [38]). The overall performance of these algorithms depends heavily on the performance at each discretization level, especially when q becomes large.

Problem (DSI) involves finitely many smooth constraints and thus in principle can be solved by classical constrained optimization techniques. Yet typically, if q is large compared to the number n of variables, only a small portion of the constraints are active at the solution. Suitably taking advantage of this situation may lead to substantial computational savings. Similar considerations arise in connection with inequality constrained optimization problems of the form

$$\text{(MC)} \qquad \text{minimize} \quad f(x) \quad \text{s.t.} \quad \phi_i(x) \leq 0 \quad i = 0, \ldots, \ell,$$

in which $\ell \gg n$, *i.e.*, in which constraints far outnumber variables. The minimax problem (here, with finitely many objective functions) is an important special case of this problem. Examples of (MC) include mechanical design problems involving trusses (see, *e.g.*, [37], [43] or papers in [6], [25]). Note that there is no essential difference between (DSI) and (MC). Their similarity is particularly strong if the constraints in (MC) are "sequentially related" in the sense that the values taken by ϕ_i are typically close to those taken by ϕ_{i+1}.

In [32], [27], (DSI) is solved by means of first order (thus, slow) methods. In [32], based on ideas of Zoutendijk [46] and Polak [29, Section 4.3], the construction of the search direction at iteration k makes use of the gradients $\nabla_x \phi(x_k, \omega)$ at all points $\omega \in \Omega$ at which $\phi(x_k, \omega) \geq -\epsilon$ ("ϵ-active" constraints), where $\epsilon > 0$ is appropriately small. When the discretization is fine, however, the set of such points is often unduly large as it contains entire neighborhoods of local maximizers. In [27], it is shown that only a small subset of these points need be used, by suitably detecting "critical" values of ω and "remembering" them from iteration to iteration in a manner reminiscent of bundle type methods in nonsmooth optimization (see, *e.g.*, [21], [23]). Specifically, at iteration k, a first order direction d_k is computed using a certain subset Ω_k of Ω. After a new iterate x_{k+1} has been obtained, a new set Ω_{k+1} is constructed by including (i) all ω's that globally maximize $\phi(x_{k+1},\cdot)$ over Ω; (ii) all ω's that globally maximize $\phi(\bar{x}_{k+1},\cdot)$, where $\bar{x}_{k+1}$ is a trial point that was rejected in the previous line search; and (iii) all ω's in Ω_k that affected direction d_k. This scheme is shown in [27] to induce global convergence. It is efficient because, under mild assumptions, the dimension of the quadratic programming problem that yields d_k is moderate, and gradient evaluations are only required at a few grid points. However, at each level of discretization (*i.e.*, for each fixed q), the algorithm proposed in [27] (like that proposed in [32]) exhibits at best a linear rate of convergence.

SQP-type algorithms, while often impractical for problems with large numbers of variables, are particularly suited to various classes of engineering applications where the number of variables is not too large but evaluations of objective/constraint functions and of their gradients are highly time consuming. Indeed, as these algorithms use quadratic programs as successive models, progress between (expensive) function evaluations is typically significantly better than with algorithms making use of mere linear systems of equations as models. In the context of SQP-type algorithms for the solution of problems with many constraints, Biggs [1] proposed to replace with equality constraints the active inequality constraints and to ignore all other inequality constraints in the computation of the search direction. Much later, Polak and Tits [34] and Mine et al. [24] adapted the "ϵ-active" idea to the SQP context, and Powell [36] proposed a "tolerant" algorithm for linearly constrained problems, which also borrows from the "ϵ-active" concept. Again, however, in the case of finely discretized SIP problems, the number of constraints may be unduly large. Recently, Conn and Li [4] proposed a working set scheme for the minimax problem and obtained promising numerical results. Finally, in [41], Schittkowski proposes modifications of standard SQP methods for the solution of problems with many constraints. However, no convergence analysis is provided; in practice global convergence may or may not take place, depending on the heuristics used to update an active working set of constraints.

In this paper, we propose and analyze an SQP-type algorithm based on the scheme introduced in [27] for the special case of the discretized minimax problem

$$(P) \qquad \text{minimize} \quad \max_{\omega\in\Omega} \phi(x,\omega),$$

where Ω is again a finite set. The general discretized SIP case involves additional intrinsic difficulties and will be considered in a separate paper. We define

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

At iteration k, given an iterate x_k and a subset Ω_k of Ω, a search direction d_k is obtained as the solution of the "quadratic program" $QP(x_k, H_k, \Omega_k)$.[2] Here, for any $x \in \mathbb{R}^n$, $H \in \mathbb{R}^{n\times n}$ symmetric positive definite, and $\hat{\Omega} \subseteq \Omega$, $QP(x, H, \hat{\Omega})$ is defined by

$$QP(x,H,\hat{\Omega}) \qquad \text{minimize} \ \ \frac{1}{2}\langle d, Hd\rangle + \Phi'_{\hat{\Omega}}(x,d), \quad \text{s.t.} \ \ d \in R^n$$

[2] $QP(x,H,\hat{\Omega})$ is equivalent to the true quadratic program (over R^{n+1})

$$\text{minimize} \ \ \frac{1}{2}\langle d, Hd\rangle + d^0 \quad \text{s.t.} \ \ \phi(x,\omega) + \langle \nabla_x \phi(x,\omega), d\rangle - \Phi_{\hat{\Omega}}(x) - d^0 \le 0 \ \ \forall \omega \in \hat{\Omega}.$$

.

where

$$\Phi'_{\hat{\Omega}}(x,d) = \max_{\omega\in\hat{\Omega}}\{\phi(x,\omega) + \langle\nabla_x\phi(x,\omega), d\rangle\} - \Phi_{\hat{\Omega}}(x) \tag{1.1}$$

is a first order approximation to $\Phi_{\hat{\Omega}}(x+d) - \Phi_{\hat{\Omega}}(x)$, with

$$\Phi_{\hat{\Omega}}(x) = \max_{\omega\in\hat{\Omega}} \phi(x,\omega).$$

A line search (*e.g.*, of Armijo type such as that suggested by Han [14], [15]) is performed along direction d_k to obtain a next iterate $x_{k+1} = x_k + t_k d_k$, with $t_k \in (0,1]$; H_k is updated to H_{k+1}; and a new subset Ω_{k+1} of Ω is constructed according to a scheme inspired from that used in [27]. In particular, if $t_k < 1$, Ω_{k+1} includes a point $\bar{\omega}_k$ that caused the last trial point to be rejected by the line search. However, in the present context, a difficulty arises. Suppose $\bar{\omega}_k$ was not in Ω_k. The rationale for including it in Ω_{k+1} is that, had it been included in Ω_k, a larger step would likely have been accepted (since $\bar{\omega}_k$ is now preventing a larger step). In the context of [27] where a first order search direction is used (*i.e.*, $H_k = I$ for all k), it follows that d_{k+1} will likely allow a larger step to be taken. In the current framework however it is unclear whether $\bar{\omega}_k$ is of any help in the new metric H_{k+1}, and global convergence may not occur. One remedy would be to renounce updating H_k whenever $t_k < 1$ and $\bar{\omega}_k \notin \Omega_k$ is picked by the algorithm. As it will be proved that, eventually, $\bar{\omega}_k$ can only be picked from Ω_k (Lemma 3.14), such scheme will not prevent normal updating from eventually taking place (thus will not jeopardize the anticipated superlinear rate of convergence). Yet, disallowing normal updating of H_k in early iterations can hinder the algorithm's effectiveness. To obviate this effect we will disallow normal updating of H_k only if the additional condition $t_k < \delta$ is satisfied, where δ is a small positive number. Indeed, if t_k stays bounded away from zero, then $\{d_k\}$ must go to zero (Lemma 3.3(iv)) and global convergence takes place in any case (Lemma 3.4(ii)). It is shown below that this overall algorithm indeed achieves global convergence and maintains a fast rate of local convergence.

A well known possible adverse effect is that the line search may truncate the unit step even arbitrarily close to a solution, thus preventing superlinear convergence (Maratos effect). It will be shown that this can be avoided by incorporating in the basic algorithm standard techniques such as a second order correction (see, *e.g.*, [28], [44]).

The algorithm stated and analyzed below (Algorithm 2.1) allows that additional ω's be included in Ω_k at each iteration. Clever heuristics may significantly speed up the algorithm, especially in early iterations. In our implementation (discussed in a subsequent section) we paid special attention to finely discretized SIP problems in which $\phi(x,\omega)$ is continuous in ω and to other problems in which "adjacent" objectives are closely related.

The remainder of paper is organized as follows. The basic algorithm is stated in §2. A complete convergence analysis is presented in §3. In §4, implementation issues are discussed, and numerical results are reported. §5 is devoted to final remarks. The paper ends with an appendix, §6, containing proofs of results in §3.

2. Preliminaries and Algorithm Statement. The following assumption is made throughout.

Assumption 1. For every $\omega \in \Omega$, $\phi(\cdot,\omega) : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable.

Let x^* be a local minimizer for (P). Then (see, *e.g.*, [15]) it is a *KKT point* for (P), *i.e.*, there exist *KKT multipliers* μ_ω^*, $\omega \in \Omega$ such that

$$\begin{cases} \sum_{\omega\in\Omega} \mu_\omega^* \nabla_x \phi(x^*,\omega) = 0 \\ \mu_\omega^* \geq 0 \ \forall\omega\in\Omega \ \text{ and } \sum_{\omega\in\Omega}\mu_\omega^* = 1 \\ \mu_\omega^* = 0 \ \forall\omega\in\Omega \ \text{ s.t. } \ \phi(x^*,\omega) < \Phi(x^*). \end{cases} \tag{2.1}$$

It is readily verified that there is a natural correspondence between the KKT points of (P) and those of the equivalent constrained minimization problem

$$(P_{eq}) \qquad \text{minimize} \ \ x^0 \ \text{ s.t. } \ \Phi(x) \leq x^0, \ \ x \in R^n, \ \ x^0 \in R.$$

Specifically, the following holds.

LEMMA 2.1. *A point x^* is a KKT point for (P) if and only if $(x^*, \Phi(x^*))$ is a KKT point for (P_{eq}). The associated KKT multipliers are identical.*

Similarly, given $x \in R^n$, $H = H^T > 0$ and $\hat{\Omega} \subseteq \Omega$, if d solves $QP(x, H, \hat{\Omega})$, then it is a KKT point for $QP(x, H, \hat{\Omega})$, *i.e.*, there exist μ_ω, $\omega \in \hat{\Omega}$ such that

$$\begin{cases} Hd + \sum_{\omega\in\hat{\Omega}} \mu_\omega \nabla_x \phi(x,\omega) = 0 \\ \mu_\omega \geq 0 \ \forall\omega\in\hat{\Omega} \ \text{ and } \sum_{\omega\in\hat{\Omega}}\mu_\omega = 1 \\ \mu_\omega = 0 \ \forall\omega\in\hat{\Omega} \ \text{ s.t. } \ \phi(x,\omega) + \langle\nabla_x\phi(x,\omega), d\rangle - \Phi_{\hat{\Omega}}(x) < \Phi'_{\hat{\Omega}}(x,d). \end{cases} \tag{2.2}$$

Moreover, since $\varphi(d) := \frac{1}{2}\langle d, Hd\rangle + \Phi'_{\hat{\Omega}}(x,d)$ is strictly convex in d (sum of a strictly convex function and of a convex function), it has a unique minimizer d^*. It follows that the equivalent quadratic program in R^{n+1} has a unique minimizer $(d^*, \varphi(d^*))$ and thus that $QP(x, H, \hat{\Omega})$ has d^*, its global minimizer, as its only KKT point.

We are now ready to make precise the rule for updating Ω_k. Following [27], Ω_{k+1} contains the union of three sets.[3] Given $x \in R^n$, let

$$\Omega_{max}(x) = \{\omega \in \Omega : \phi(x,\omega) = \Phi(x)\}$$

be the set of maximizers of $\phi(x,\cdot)$. The first component of Ω_{k+1} is $\Omega_{max}(x_{k+1})$. Indeed if $\Omega_{max}(x_{k+1})$ were not included, d_{k+1} might not be a direction of descent for Φ at x_{k+1}. The second component of Ω_{k+1} is obtained from the line search. While the essence of the ideas put forth in this paper is independent of the specifics of this line search, for the sake of exposition, we will consider the case of an Armijo-type line search inspired from the line search used by Han [14], [15]. Thus $x_{k+1} = x_k + t_k d_k$, where t_k is the largest number t in $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$\Phi(x_k + td_k) \leq \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle, \tag{2.3}$$

where $\alpha \in (0, 1/2)$ and $\beta \in (0,1)$ are fixed. Suppose the line search at iteration k results in $t_k < 1$, implying that the line search test (2.3) is violated at $x_k + \frac{t_k}{\beta} d_k$. A

[3]In [27], Ω_{k+1} is set to be *equal* to this union.

next search direction taking this into account is called for. Thus, Ω_{k+1} will include some $\bar{\omega}_k$ such that

$$\phi(x_k + \frac{t_k}{\beta} d_k, \bar{\omega}_k) > \Phi(x_k) - \alpha \frac{t_k}{\beta} \langle d_k, H_k d_k \rangle.$$

Finally, to avoid zigzagging (which could prevent global convergence; see the example in [27]) it is important that key elements in Ω_k be kept in Ω_{k+1}. A natural choice is to preserve all $\omega \in \Omega_k$ that are binding at the solution of $QP(x_k, H_k, \Omega_k)$, *i.e.*, those ω for which the corresponding multiplier $\mu_{k,\omega}$ is strictly positive[4] (clearly, this is also needed for fast local convergence). Thus, the third component of Ω_{k+1} is

$$\Omega_k^b = \{\omega \in \Omega_k : \mu_{k,\omega} > 0\}.$$

Thus the overall algorithm for the solution of (P) is as follows.

Algorithm 2.1.
Parameters. $\alpha \in (0, \frac{1}{2})$, $\beta \in (0,1)$, $0 < \delta \ll 1$.
Data. $x_0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n \times n}$ with $H_0 = H_0^T > 0$.
Step 0. Initialization. Set $k = 0$ and pick $\Omega_0 \supseteq \Omega_{max}(x_0)$.
Step 1. Computation of search direction and step length.
(i). Compute d_k by solving $QP(x_k, H_k, \Omega_k)$. If $d_k = 0$, stop.
(ii). Compute t_k, the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$\Phi(x_k + t d_k) \leq \Phi(x_k) - \alpha t \langle d_k, H_k d_k \rangle. \tag{2.4}$$

Step 2. Updates. Set $x_{k+1} = x_k + t_k d_k$. If $t_k < 1$, pick $\bar{\omega}_k$ such that

$$\phi(x_k + \frac{t_k}{\beta} d_k, \bar{\omega}_k) > \Phi(x_k) - \alpha \frac{t_k}{\beta} \langle d_k, H_k d_k \rangle.$$

Pick

$$\Omega_{k+1} \supseteq \begin{cases} \Omega_{max}(x_{k+1}) \cup \Omega_k^b & \text{if } t_k = 1 \\ \Omega_{max}(x_{k+1}) \cup \Omega_k^b \cup \{\bar{\omega}_k\} & \text{if } t_k < 1. \end{cases} \tag{2.5}$$

If $t_k \leq \delta$ and $\bar{\omega}_k \notin \Omega_k$, set $H_{k+1} = H_k$; otherwise, compute a new positive definite approximation H_{k+1} to the Hessian of the Lagrangian of (P) at the solution. Set $k = k + 1$. Go back to *Step 1.*

□

3. Convergence Analysis. Although (P) takes the form of an ordinary minimax problem, the classical convergence analysis for such problems (e.g., [14], [15]) cannot be directly applied to the present situation since, at each iteration, only a subset of the discretized set Ω is employed to construct a search direction.

3.1. Global convergence. The following additional standard assumptions are made.

Assumption 2. For any $x_0 \in \mathbb{R}^n$, the level set $\{x \in \mathbb{R}^n : \Phi(x) \leq \Phi(x_0)\}$ is compact.

---

[4]If the multiplier vector associated with $QP(x_k, H_k, \Omega_k)$ is not unique, any of the choices is appropriate.

Assumption 3. There exist σ_1, $\sigma_2 > 0$ such that

$$\sigma_1\|d\|^2 \leq \langle d, H_k d\rangle \leq \sigma_2\|d\|^2 \quad \forall d \in \mathbb{R}^n, \forall k.$$

We first show that, owing to the fact that Ω_k always contains $\Omega_{max}(x_k)$, Algorithm 2.1 is well defined.

LEMMA 3.1. *At any iteration k there exists $\bar{t}_k > 0$ such that, for all $t \in [0, \bar{t}_k]$,*

$$\Phi(x_k + td_k) \leq \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle.$$

Proof. Since d_k solves $QP(x_k, H_k, \Omega_k)$, it yields an objective value no larger than that at $d = 0$ for that problem, and thus

$$\Phi'_{\Omega_k}(x_k, d_k) \leq -\frac{1}{2}\langle d_k, H_k d_k\rangle.$$

In view of (1.1) and Assumption 3, it follows that

$$\langle\nabla\phi(x_k, \omega), d_k\rangle < -\frac{1}{2}\langle d_k, H_k d_k\rangle \quad \forall\omega \in \Omega_k\ .$$

Since $\alpha < 1/2$ and since $\Omega_k \supseteq \Omega_{max}(x_k)$, it then follows that there exists $\hat{t}_k > 0$ such that, for all $t \in [0, \hat{t}_k]$,

$$\Phi_{\Omega_k}(x_k + td_k) \leq \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle.$$

On the other hand, since $\Omega_{max}(x_k) \subseteq \Omega_k$, $\phi(x_k, \omega) < \Phi(x_k)$ for all $\omega \notin \Omega_k$. In view of the continuity assumptions, this implies that there exists $\bar{t}_k > 0$ such that, for all $t \in [0, \bar{t}_k]$,

$$\Phi(x_k + td_k) \leq \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle,$$

proving the claim. □

Thus the line search is always well defined and Algorithm 2.1 stops only when $d_k = 0$. The following lemma implies that, if this occurs, the last point x_k must be a KKT point.

LEMMA 3.2. *Let $H > 0$, $x \in \mathbb{R}^n$, and $\hat{\Omega} \subseteq \Omega$ with $\Phi_{\hat{\Omega}}(x) = \Phi(x)$. Then the unique KKT point d of $QP(x, H, \hat{\Omega})$ is zero if and only if x is a KKT point for (P).*

Proof. Suppose the unique KKT point of $QP(x, H, \hat{\Omega})$ is $d = 0$ and let $\{\hat{\mu}_\omega : \omega \in \hat{\Omega}\}$ be the associated KKT multipliers. In view of (2.2) and since $\Phi_{\hat{\Omega}}(x) = \Phi(x)$, the KKT condition (2.1) for (P) holds at x with multipliers $\mu_\omega = \hat{\mu}_\omega$ for $\omega \in \hat{\Omega}$ and $\mu_\omega = 0$ for $\omega \in \Omega\backslash\hat{\Omega}$. Thus, x is a KKT point for (P). The converse is proved similarly. □

We now assume that an infinite sequence $\{x_k\}$ is generated by Algorithm 2.1. The following facts are direct consequences of Lemma 3.1 and of our assumptions.

LEMMA 3.3. *(i) The sequence $\{x_k\}$ is bounded; (ii) the sequence $\{d_k\}$ is bounded; (iii) the sequence $\{\Phi(x_k)\}$ converges; and (iv) the sequence $\{t_k d_k\}$ converges to zero.*

Proof. The claims follow directly from Assumptions 1, 2, 3 and the fact that $\Phi(x_{k+1}) \leq \Phi(x_k) - \alpha t_k\langle d_k, H_k d_k\rangle$ (since $|t_k| \leq 1$, convergence to zero of $\{t_k\langle d_k, H_k d_k\rangle\}$ implies convergence to zero of $\{\langle(t_k d_k), H_k(t_k d_k)\rangle\}$). □

Now, let v_k denote the optimal value of $QP(x_k, H_k, \Omega_k)$, i.e.,

$$v_k = \frac{1}{2}\langle d_k, H_k d_k\rangle + \Phi'_{\Omega_k}(x_k, d_k). \tag{3.1}$$

Since $d = 0$ is feasible for $QP(x_k, H_k, \Omega_k)$, v_k is nonpositive for all k. It turns out that convergence of $\{d_k\}$ to zero is equivalent to convergence of $\{v_k\}$ to zero and implies that accumulation points of $\{x_k\}$ are KKT points. More generally, the following holds.

LEMMA 3.4. *Let $K \subseteq I\!N$ be an infinite index set. Then, (i) $\{d_k\}$ converges to zero on K if and only if $\{v_k\}$ converges to zero on K; (ii) if $\{d_k\}$ converges to zero on K, then all accumulation points of $\{x_k\}_{k\in K}$ are KKT points for (P).*

Proof. Since, for all k, $\Phi_{\Omega_k}(x_k) = \Phi(x_k)$, it follows from (1.1) and (2.2) with $x = x_k$, $H = H_k$, $\hat{\Omega} = \Omega_k$, $d = d_k$, that, for all k and some $\mu_{k,\omega} \geq 0, \omega \in \Omega_k$, with $\sum_{\omega\in\Omega_k} \mu_{k,\omega} = 1$,

$$\begin{aligned}\Phi'_{\Omega_k}(x_k, d_k) &= \max_{\omega\in\Omega_k}\{\phi(x_k,\omega) + \langle\nabla_x\phi(x_k,\omega), d_k\rangle\} - \Phi(x_k)\\ &= \sum_{\omega\in\Omega_k} \mu_{k,\omega}\{\phi(x_k,\omega) + \langle\nabla_x\phi(x_k,\omega), d_k\rangle\} - \Phi(x_k)\\ &\leq \sum_{\omega\in\Omega_k} \mu_{k,\omega}\langle\nabla_x\phi(x_k,\omega), d_k\rangle,\end{aligned}$$

yielding, again from (2.2),

$$\Phi'_{\Omega_k}(x_k, d_k) \leq -\langle d_k, H_k d_k\rangle. \tag{3.2}$$

In view of (3.1), it follows that

$$v_k \leq -\frac{1}{2}\langle d_k, H_k d_k\rangle.$$

Thus, the "if" part of (i) follows directly from Assumption 3. On the other hand, if $\{d_k\}$ goes to zero on K, since $\{x_k\}$ is bounded, it follows from (1.1) that

$$\lim_{k\in K, k\to\infty} \Phi'_{\Omega_k}(x_k, d_k) = 0.$$

The "only if" part of (i) then follows from (3.1).

To prove (ii), suppose $\{d_k\}$ goes to zero on K and let $K' \subseteq K$ be any infinite index set such that $\{x_k\}$ converges to some $\hat{x}$ on K'. Without loss of generality, assume $\Omega_k = \hat{\Omega}$ for all $k \in K'$, for some $\hat{\Omega} \subseteq \Omega$. Then $\Phi_{\hat{\Omega}}(x_k) = \Phi(x_k)$ for all $k \in K'$ and thus $\Phi_{\hat{\Omega}}(x^*) = \Phi(x^*)$. In view of Assumption 3 and of the boundedness of $\{\mu_{k,\omega}\}$ for all $\omega \in \Omega$, there exists $K'' \subseteq K'$ such that $\{H_k\}$ converges to some H^* on K'' and, for each $\omega \in \hat{\Omega}$, there exists $\hat{\mu}_\omega$ such that $\{\mu_{k,\omega}\}$ converges to $\hat{\mu}_\omega$ on K''. Letting $\hat{\mu}_\omega = 0$ for $\omega \in \Omega\backslash\hat{\Omega}$, taking limits for $k \in K''$ in the optimality condition (2.2) associated with $QP(x_k, H_k, \hat{\Omega})$ and comparing with (2.1) shows that $\hat{x}$ is a KKT point for (P). □

The next lemma, which is the same as Lemma 4.7 in [21, Chapter 3], is central to the proof of global convergence.

LEMMA 3.5. *Let $x^* \in I\!R^n$ be such that*

$$\liminf_{k\to\infty}\ \max\{|v_k|,\ \|x_k - x^*\|\} = 0.$$

Then, x^* is a KKT point for (P).

Proof. The assumption implies that there exists an infinite index set K such that $\{x_k\}$ converges to x^* and $\{v_k\}$ converges to zero, both on K. Thus, the conclusion follows from Lemma 3.4. □

The establishment of the global convergence of Algorithm 2.1 employs a contradiction argument inspired from [21, Chapter 3]. If $\{x_k\}$ has a limit point x^* that is not a KKT point, v_k is bounded away from zero on the corresponding subsequence (Lemma 3.5), with a uniform lower bound $\epsilon^* > 0$ for *all* subsequences over which $\{x_k\}$ converges to x^*. It is shown below (Lemma 3.6) that in such case $|v_{k+1}|$ is significantly smaller than $|v_k|$ on any such subsequence K. Since, in view of Lemma 3.3(iv), $\{x_{k+1}\}$ also converges to x^*, $|v_{k+2}|$ is also significantly smaller than $|v_{k+1}|$. A careful repeated application of this argument shows that $|v_k|$ becomes smaller than ϵ^* on a sequence at "finite distance" from K, a contradiction.

The proof of the following lemma is inspired from that of Lemma 4.11 in [21, Chapter 3] (see also the proof of Lemma 3.15 in [44]) and is given in the appendix. It relies crucially on the assumption that $\{\|H_{k+1} - H_k\|\} \to 0$ whenever $\{t_k\} \to 0$, which is insured in Algorithm 2.1 by setting $H_{k+1} = H_k$ when t_k is small and $\bar{\omega}_k \notin \Omega_k$; it also relies on the inclusion in Ω_{k+1} of the second and third subsets in (2.5).

LEMMA 3.6. *There exists $c > 0$ such that, if K is an infinite index set on which $\{d_k\}$ is bounded away from zero, then there exists an integer N such that*

$$|v_{k+1}| \le |v_k| - c|v_k|^2, \quad \forall k \ge N, k \in K. \tag{3.3}$$

Repeated application of this results yields the following.

LEMMA 3.7. *There exists $c > 0$ such that, if K is an infinite index set on which $\{x_k\}$ is bounded away from KKT points, then, given any positive integer i_0, there exists an integer N such that*

$$|v_{k+i+1}| \le |v_{k+i}| - c|v_{k+i}|^2 \quad \forall k \ge N, k \in K, \forall i \in [0, i_0].$$

□

Proof. For any integer i, in view of Lemma 3.3(iv), $\{x_{k+i}\}$ is bounded away from KKT points for $k \in K$ and this in turn implies, in view of Lemma 3.4, that $\{d_{k+i}\}$ is bounded away from zero for $k \in K$. Therefore, in view of Lemma 3.6, for each i there exists N_i such that

$$|v_{k+i+1}| \le |v_{k+i}| - c|v_{k+i}|^2, \quad \forall k \ge N_i, k \in K.$$

Choosing $N = \max_{0 \le i \le i_0} \{N_i\}$ proves the claim. □

LEMMA 3.8. *Given $\eta > 0$ and $\epsilon > 0$, there exists an integer i_0 depending only on η and ϵ such that, for any sequence $\{z_i\}$ of real numbers satisfying*

$$0 \le z_{i+1} \le z_i - \eta z_i^2 \quad \forall i \in \mathbb{N},$$

$z_i < \epsilon$ for all $i \ge i_0$.

Proof. See the appendix. □

We are now ready to establish the global convergence of Algorithm 2.1. The facts that c in Lemma 3.7 is independent of K and that i_0 in Lemma 3.8 is independent of z_0 play a crucial role in the proof.

THEOREM 3.9. *Let $\{x_k\}$ be the sequence generated by Algorithm 2.1. Then, every accumulation point of $\{x_k\}$ is a KKT point.*

Proof. Let x^* and K be such that $\{x_k\}$ converges to x^* on K, an infinite index set. Proceeding by contradiction, we assume x^* is not a KKT point. It follows from Lemma 3.5 that there exists $\epsilon^* > 0$ such that

$$\liminf_{k \to \infty} \max\{|v_k|, \|x_k - x^*\|\} > \epsilon^*. \tag{3.4}$$

Thus, $\{v_k\}$ is bounded away from zero on K. In view of Lemma 3.4, $\{d_k\}$ is also bounded away from zero on K and, in view of Lemma 3.3(iv), $\{t_k\}$ converges to zero on K. Let c be as given by Lemma 3.7. Let i_0 be as given in Lemma 3.8 with $\epsilon = \epsilon^*$ and $\eta = c$. In view of Lemma 3.7, there exists an integer N such that

$$|v_{k+i+1}| \leq |v_{k+i}| - c|v_{k+i}|^2, \quad \forall k \geq N, k \in K, \forall i \in [0, i_0].$$

From the definition of i_0, it follows from Lemma 3.8 with $z_i = |v_{k+i}|$, $k \in K$, $i = 0, \ldots, i_0$ and $z_i = 0$, $k \in K$, $i > i_0$, that

$$|v_{k+i_0}| < \epsilon^* \quad \forall k \geq N, k \in K.$$

On the other hand, since by assumption x^* is not a KKT point and since in view of Lemma 3.3(iv) $\{x_{k+i_0}\}$ also converges to x^* on K, it follows from (3.4) that

$$\mathop{lim\,inf}_{k \in K, k \to \infty} |v_{k+i_0}| > \epsilon^*,$$

a contradiction. □

3.2. Local convergence. Under additional regularity conditions, it is shown that, close to a strong local minimizer x^*, the right hand side of (2.5) becomes equal to $\Omega_{max}(x^*)$ for all k and no $\omega \in \Omega \backslash \Omega_{max}(x^*)$ hinders the line search, so that Algorithm 2.1 behaves as if solving (P) with Ω replaced with $\Omega_{max}(x^*)$. Further, it is shown that H_k will be updated normally, thus will not be prevented from asymptotically suitably approximating $\nabla^2 L(x^*, \mu^*)$, the Hessian of the Lagrangian at the limit KKT pair. If H_k does become a suitable approximation to $\nabla^2 L(x^*, \mu^*)$ and if the full step of one is eventually accepted by the line search, 2-step superlinear convergence will result.

Assumption 1 is replaced by the following.

Assumption 1'. For every $\omega \in \Omega$, the function $\phi(\cdot, \omega) : \mathbb{R}^n \to \mathbb{R}$ is three times continuously differentiable.

Let x^* be an accumulation point of $\{x_k\}$ (thus a KKT point for (P)).

Assumption 4. Any scalars λ_ω, $\omega \in \Omega_{max}(x^*)$, satisfying

$$\sum_{\omega \in \Omega_{max}(x^*)} \lambda_\omega \nabla_x \phi(x^*, \omega) = 0 \text{ and } \sum_{\omega \in \Omega_{max}(x^*)} \lambda_\omega = 0$$

must be all zero.

Thus the KKT multipliers μ^*_ω, $\omega \in \Omega$, corresponding to x^*, for problem (P), are unique.

Assumption 5. The second order sufficiency conditions with strict complementary slackness are satisfied at x^*, *i.e.* (see, *e.g.*, [14]), $\mu^*_\omega > 0$ for all $\omega \in \Omega_{max}(x^*)$ and

$$\langle d, \nabla^2_{xx} L(x^*, \mu^*) d \rangle > 0, \quad \forall d \in \mathcal{S}^*, d \neq 0,$$

with

$$\nabla^2_{xx} L(x^*, \mu^*) = \sum_{\omega \in \Omega_{max}(x^*)} \mu^*_\omega \nabla^2_{xx} \phi(x^*, \omega)$$

and

$$\mathcal{S}^* = \{d : \langle d, \nabla_x \phi(x^*, \omega) \rangle = 0 \;\; \forall \omega \in \Omega_{max}(x^*)\}.$$

The following result is standard for ordinary constrained problems (see, e.g., [8, Theorem 2.3.2]). A proof in the minimax case is given in the appendix for sake of completeness.

LEMMA 3.10. *The point x^* is an isolated KKT point for (P).*

PROPOSITION 3.11. *The entire sequence $\{x_k\}$ converges to x^*.*

Proof. The claim follows from Theorem 3.9, Lemma 3.10, and Lemma 3.3(iv). □

Much of the remainder of this section is devoted to showing that, close to x^*, the right hand side of (2.5) remains constant and equal to $\Omega_{max}(x^*)$, and no $\omega \notin \Omega\backslash\Omega_{max}(x^*)$ hinders the line search, *i.e.*, Algorithm 2.1 eventually behaves as if solving a minimax problem of the type

$$\text{minimize} \quad \max_{\omega\in\hat{\Omega}} \ \phi(x,\omega)$$

(with $\hat{\Omega} = \Omega_{max}(x^*)$) with Ω_k always set to be equal to $\hat{\Omega}$ instead of updated according to the rule in Step 2. A consequence of this is that rate of convergence results obtained for such standard "constant Ω_k" algorithms hold for Algorithm 2.1. The first step is to show that, for k large enough, $\Omega_{max}(x^*) \subseteq \Omega_k$ (Lemma 3.14). This is first proved on a subsequence using Lemma 3.7(ii).

LEMMA 3.12. *There exists an infinite index set K such that*

$$\Omega_{max}(x^*) \subseteq \Omega_k^b \quad \forall k \in K. \tag{3.5}$$

Proof. First, there exists an infinite index set K such that $\{d_k\}$ converges to zero on K. Indeed, if $\{d_k\}$ were bounded away from zero, it would follow from Lemma 3.6 that

$$|v_{k+1}| \leq |v_k| - c|v_k|^2, \quad \forall k \geq N$$

for some $c > 0$ and some integer N, implying that $\{v_k\}$ converges to zero, in violation of Lemma 3.4(i). Next, in view of the finite cardinality of Ω, without loss of generality, we may assume that $\Omega_k^b = \hat{\Omega}$ for all $k \in K$ for some constant set $\hat{\Omega}$. Let $\mu_k \in R^{|\Omega|}$ be a vector with components $\{\mu_{k,\omega}\}$ such that $\mu_{k,\omega}$, $\omega \in \hat{\Omega}$, are the KKT multipliers associated with $QP(x_k, H_k, \Omega_k)$ and $\mu_{k,\omega} = 0$, $\omega \in \Omega\backslash\Omega_k$. Without loss of generality, $\{\mu_k\} \to \hat{\mu}$ as $k \to \infty$, $k \in K$, for some $\hat{\mu}$. We show that $\hat{\mu}$ together with x^* satisfies the KKT conditions (2.1) of the original problem. In view of Proposition 3.11, since $\{d_k\}$ converges to zero on K, taking limits in the optimality condition (2.2) associated with $QP(x_k, H_k, \Omega_k)$, $k \in K$, yields, since $\hat{\mu}_\omega = 0$ for all $\omega \in \Omega\backslash\Omega_k$,

$$\sum_{\omega\in\Omega} \hat{\mu}_\omega \nabla_x \phi(x^*,\omega) = 0,$$

$$\hat{\mu}_\omega \geq 0 \ \forall \omega \in \Omega \ \text{ and } \sum_{\omega\in\Omega} \hat{\mu}_\omega = 1,$$

$$\hat{\mu}_\omega = 0 \quad \forall \omega \in \Omega \text{ s.t. } \phi(x^*,\omega) < \Phi(x^*).$$

Therefore, x^* with $\{\hat{\mu}_\omega, \omega \in \hat{\Omega}; \hat{\mu}_\omega = 0, \omega \in \Omega\backslash\hat{\Omega}\}$ satisfies (2.1). Uniqueness of the multipliers for (P) at x^* and strict complementarity (Assumptions 4 and 5) imply that $\omega \in \hat{\Omega}$ for all ω such that $\phi(x^*,\omega) = \Phi(x^*)$, *i.e.*, (3.5) holds. □

The following lemma, on the other hand, establishes that d_k is small whenever (3.5) holds.

LEMMA 3.13. *Let K be an infinite index set such that $\Omega_{max}(x^*) \subseteq \Omega_k$ for all $k \in K$. Then, $\{d_k\}$ converges to zero on K.*

Proof. Given $\hat{\Omega} \subseteq \Omega$, let $K_{\hat{\Omega}} = \{k \in K : \Omega_k = \hat{\Omega}\}$. For any $\hat{\Omega} \subseteq \Omega$ such that $K_{\hat{\Omega}}$ is an infinite set, we prove by contradiction that $\{d_k\}$ converges to zero on $K_{\hat{\Omega}}$. Since Ω has only finitely many subsets, the lemma will follow. Thus suppose that for some infinite index set $K' \subseteq K_{\hat{\Omega}}$, $\{d_k\}$ is bounded away from zero on K' and let $K'' \subseteq K'$ be such that $\{H_k\}$ converges to H^* on K'' for some $H^* > 0$ (such K'' exists in view of Assumption 3). Since $\Omega_{max}(x^*) \subseteq \hat{\Omega}$, $QP(x^*, H^*, \hat{\Omega})$ has $d = 0$ as its unique solution (Lemma 3.2). It follows from [39, Theorem 2.1] that $\{d_k\} \to 0$ as $k \to \infty$, $k \in K''$, contradicting the fact that $\{d_k\}$ is bounded away from zero on K'. □

LEMMA 3.14. *For k large enough, $\Omega_{max}(x^*) \subseteq \Omega_k^b$.*

Proof. In view of Lemma 3.12, the claim holds on an infinite subsequence. To complete the proof we show that, given any infinite index set K such that $\Omega_{max}(x^*) \subseteq \Omega_k^b$ for all $k \in K$, it holds that $\Omega_{max}(x^*) \subseteq \Omega_{k+1}^b$ for all $k \in K$, k large enough. In view of the construction of Ω_{k+1}, it is enough to show that $\mu_{k+1,\omega} > 0$ for all $\omega \in \Omega_{max}(x^*)$, $k \in K$, k large enough, where $\mu_{k+1,\omega}$, $\omega \in \Omega_{k+1}$, are the KKT mutipliers associated with $QP(x_{k+1}, H_{k+1}, \Omega_{k+1})$. Thus let K be an infinite index set such that $\Omega_{max}(x^*) \subseteq \Omega_k^b$ for all $k \in K$ (so that $\Omega_{max}(x^*) \subseteq \Omega_{k+1}$ for all $k \in K$). Lemma 3.13 implies that $\{d_{k+1}\}$ converges to zero on K. Suppose by contradiction that there exists $\omega^* \in \Omega_{max}(x^*)$ and an infinite index set $K' \subseteq K$ such that $\mu_{k+1,\omega^*} = 0$ for all $k \in K'$ (note that $\Omega_{max}(x^*)$ is a finite set). An argument similar to that used in the proof of Lemma 3.12 shows that, in view of Assumption 4, $\mu^*_{\omega^*} = 0$, contradicting strict complementarity (Assumption 5). □

The following result directly follows from Lemmas 3.13 and 3.14.

LEMMA 3.15. *The entire sequence $\{d_k\}$ converges to zero.* □

This leads to the main result of this section

PROPOSITION 3.16. *For k large enough,*

$$\Omega_k^b = \Omega_{max}(x^*),$$

$$\phi(x_k + td_k, \omega) \leq \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle \quad \forall t \in [0,1],\ \omega \in \Omega\backslash\Omega_{max}(x^*), \tag{3.6}$$

and

$$\phi(x_{k+1}, \omega) < \Phi(x_{k+1}) \quad \forall \omega \in \Omega\backslash\Omega_{max}(x^*).$$

Proof. To prove the first claim, in view of Lemma 3.14, it suffices to show that, for k large enough, $\Omega_k^b \subseteq \Omega_{max}(x^*)$. To this end, let $\hat{\omega} \in \Omega\backslash\Omega_{max}(x^*)$, *i.e.*, suppose that $\phi(x^*, \hat{\omega}) < \Phi(x^*)$. Our continuity assumption and Proposition 3.11 then imply that, for k large enough, $\phi(x_k, \hat{\omega}) < \Phi(x_k)$, or equivalently $\phi(x_k, \hat{\omega}) < \max_{\omega \in \Omega_k} \phi(x_k, \omega)$, since by construction $\Omega_{max}(x^*) \subseteq \Omega_k$. This, together with Proposition 3.11 and Lemma 3.15 and the continuity assumption, implies that, for k large enough,

$$\phi(x_k, \omega) + \langle \nabla_x \phi(x_k, \hat{\omega}), d_k\rangle < \max_{\omega \in \Omega_k}\{\phi(x_k, \omega) + \langle \nabla_x \phi(x_k, \omega), d_k\rangle\},$$

so that $\mu_{k,\hat{\omega}} = 0$ and $\hat{\omega} \notin \Omega_k^b$, proving the first claim. The second and third claims directly follow from Proposition 3.11, Lemma 3.15, and the continuity assumption. □

In view of Lemma 3.15, it follows from Proposition 3.16 that, if Ω_{k+1} is always picked to be equal to the right-hand side of (2.4) (rather than to merely contain it), then, for k large enough, $\Omega_k = \Omega_{max}(x^*)$. Whether or not this is the case, for k large enough, Algorithm 2.1 will behave exactly as if solving the problem

$$\text{minimize} \quad \max_{\omega \in \Omega_{max}(x^*)} \phi(x,\omega) \tag{3.7}$$

with $\Omega_k = \Omega_{max}(x^*)$ selected at each iteration. (Indeed any ω not in Ω_k^b does not affect direction d_k and, when d_k is small enough, any ω not in $\Omega_{max}(x^*)$ does not affect the line search.) In particular, for k large enough, if some $\bar{\omega}_k$ is picked by Algorithm 2.1, it must already be in Ω_k. Thus H_k is eventually updated at every iteration and the local behavior of Algorithm 2.1 becomes identical to that of the algorithm proposed by Han [14], [15] (except for a different rule for selecting t_k satisfying (2.4)).

Suppose that, as a result of the updating rule, H_k approaches the Hessian of the Lagrangian in the sense that

$$\lim_{k\to\infty} \frac{\|P_k\{H_k - \nabla^2_{xx}L(x^*,\mu^*)\}P_k d_k\|}{\|d_k\|} = 0 \tag{3.8}$$

where the matrices P_k are defined by

$$P_k = I - R_k(R_k^T R_k)^{-1}R_k^T$$

with $R_k = [\nabla_x\phi(x,\omega_i) - \nabla_x\phi(x,\omega_1) : i = 2,\ldots,s]$,[5] $\omega_1,\ldots,\omega_s$ being the elements of $\Omega_{max}(x^*)$; and suppose moreover that $t_k = 1$ for k large enough. Then (see [14]), the convergence rate is two-step superlinear, *i.e.*,

$$\lim_{k\to\infty} \frac{\|x_{k+2} - x^*\|}{\|x_k - x^*\|} = 0.$$

To achieve $t_k = 1$ for k large enough, it is necessary to introduce a scheme to avoid the Maratos effect. One option is to adopt a second order correction such as that used in [28] and [45] (in the latter, it is combined with a "nonmonotone line search"; using such line search here would entail a more complicated analysis). Specifically, Step 1(ii) in Algorithm 2.1 is replaced with the following.

Step 1(ii$_1$). If $\Phi(x_k+d_k) \le \Phi(x_k) - \alpha\langle d_k, H_k d_k\rangle$, set $\tilde{d}_k = 0$. Otherwise, compute a correction $\tilde{d}_k$ solution of the problem[6] in $\tilde{d}$

$$\text{minimize} \quad \frac{1}{2}\langle d_k + \tilde{d}, H_k(d_k+\tilde{d})\rangle + \tilde{\Phi}'(x_k+d_k, x_k, \tilde{d})$$

where

$$\tilde{\Phi}'(x_k+d_k, x_k, \tilde{d}) = \max_{\omega\in\Omega_k}\{\phi(x_k+d_k,\omega) + \langle\nabla\phi(x_k,\omega),\tilde{d}\rangle\} - \Phi(x_k+d_k). \tag{3.9}$$

If $\|\tilde{d}_k\| > \|d_k\|$, set $\tilde{d}_k = 0$.

Step 1(ii$_2$). Compute t_k, the first number t in the sequence $\{1,\beta,\beta^2,\ldots\}$ satisfying

$$\Phi(x_k + td_k + t^2\tilde{d}_k) \le \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle.$$

[5] Note that P_k remains invariant if in the definition of R_k the role of ω_1 is played by any other ω_i.

[6] Alternatively, $\tilde{d}$ could be selected as the solution of a linear least squares problem, see [28].

Also, in Step 2 of Algorithm 2.1, x_{k+1} is set to $x_k + t_k d_k + t_k^2 \tilde{d}_k$ and, if $t_k < 1$, $\bar{\omega}_k$ is any ω satisfying

$$\phi(x_k + \frac{t_k}{\beta} d_k + \left(\frac{t_k}{\beta}\right)^2 \tilde{d}_k, \omega) > \Phi(x_k) - \alpha \frac{t_k}{\beta} \langle d_k, H_k d_k \rangle.$$

It is readily checked that such modification does not affect the analysis carried out in this section, the only necessary changes being to substitute, in the statements of Lemma 3.1 and of Lemma 6.1 and at various places in the proofs, all instances of $x_k + td_k$ with $x_k + td_k + t^2 \tilde{d}_k$ (and similarly, *mutatis mutandis*, when t_k or t_k/β is present instead of t; the fact that $\|\tilde{d}_k\| \leq \|d_k\|$ is enforced in the modified algorithm is a key to the validity of the modified proofs). Thus the modified algorithm will eventually behave as if solving (3.7) with, at each iteration, Ω_k selected to be equal to $\Omega_{max}(x^*)$ and H_k normally updated. It is shown in [45] that, if (3.8) holds, the step $t_k = 1$ will always be accepted for k large enough (in fact the proof in [45] makes use of the weaker assumption

$$\lim_{k \to \infty} \frac{\|d_k^T P_k \{H_k - \nabla_{xx}^2 L(x^*, \mu^*)\} P_k d_k\|}{\|d_k\|^2} = 0.$$

It is also shown in [45] that $\tilde{d}_k = O(\|d_k\|^2)$. It follows that two-step superlinear convergence is preserved when (3.8) holds.

4. Implementation and numerical results. An efficient implementation of Algorithm 2.1, including the Maratos effect avoidance scheme described at the end of §3, has been developed as part of a C code dubbed CFSQP [22].[7] Version 2.0 of CFSQP was used to perform the numerical tests described below.

The specifics of the CFSQP implementation are as follows. In Algorithm 2.1, the rule for updating Ω_k only specifies that it must contain a certain subset of "critical" points of Ω. In practice, initial convergence is often sped up if additional "potentially critical" elements of Ω are also included. On the other hand, it is clear that increasing the size of Ω_k increases the number of gradient evaluations per iteration and makes $QP(x_k, H_k, \Omega_k)$ more complex to solve. Thus a compromise must be struck. Various heuristics come to mind (see, *e.g.*, [41]). The current CFSQP implementation focusses on the frequent case where "adjacent" objectives are closely related (objectives are "sequentially related"). It follows the idea, used in [11], [27], to include in Ω_k the set $\Omega_\epsilon^{\ell\ell m}(x_k)$ of "ϵ-active left local maximizers" at x_k, for some $\epsilon > 0$. A point $\omega_i \in \Omega := \{\omega_0, \ldots, \omega_q\}$ is ϵ-active if it belongs to

$$\Omega_\epsilon(x) = \{\omega_i : \phi(x, \omega_i) > \Phi(x) - \epsilon\}.$$

It is a left local maximizer of ϕ over Ω at x if one of the following three conditions holds: (i) $i \in \{1, \ldots, q-1\}$ and

$$\phi(x, \omega_i) > \phi(x, \omega_{i-1}) \tag{4.1}$$

and

$$\phi(x, \omega_i) \geq \phi(x, \omega_{i+1}); \tag{4.2}$$

[7] CFSQP is available from the authors.

(ii) $i = 0$ and (4.2); (iii) $i = q$ and (4.1). We also found that using $\Omega_0 = \Omega_{max}(x_0)$ often gave a poor initial search direction and performance could be improved if additional points were heuristically selected for the first iteration. For many problems, the performance was improved if the end points ω_0 and ω_q were included in Ω_0. Thus, for Ω_0 and Ω_{k+1} (in Steps 0 and 2 of Algorithm 2.1), CFSQP selects respectively

$$\Omega_0 = \Omega_{max}(x_0) \cup \Omega_\epsilon^{\ell\ell m}(x_0) \cup \{\omega_0\} \cup \{\omega_q\}$$

and

$$\Omega_{k+1} = \begin{cases} \Omega_{max}(x_{k+1}) \cup \Omega_k^b \cup \Omega_\epsilon^{\ell\ell m}(x_{k+1}) & \text{if } t_k = 1 \\ \Omega_{max}(x_{k+1}) \cup \Omega_k^b \cup \{\bar{\omega}_k\} \cup \Omega_\epsilon^{\ell\ell m}(x_{k+1}) & \text{if } t_k < 1. \end{cases}$$

Many problems encountered in practice involve more than one set of "sequentially related" objectives, e.g., a finely discretized version of the problem

$$\text{minimize } \max\{\sup_\omega \phi^1(x,\omega), \ldots, \sup_\omega \phi^\ell(x,\omega)\}, \tag{4.3}$$

where ω ranges over some interval. An important example of this type of problem is Chebyshev approximation, which has the form

$$\text{minimize } \sup_\omega |\phi(x,\omega)|,$$

or, equivalently

$$\text{minimize } \max\{\sup_\omega \phi(x,\omega), \sup_\omega -\phi(x,\omega)\}.$$

All but the last problem in the numerical tests discussed below are of this type. Note that Algorithm 2.1 and the analysis of §3 apply without modification to (4.3) by "linearly ordering" the discrete objectives as, say, $\phi^1(\cdot,\omega_0^1)$, …, $\phi^1(\cdot,\omega_{q_1}^1)$, $\phi^2(\cdot,\omega_0^2)$, …, $\phi^\ell(\cdot,\omega_{q_\ell}^\ell)$, where it is assumed that the ith set contains of q_i objective functions. However, in its selection of Ω_k, the CFSQP implementation takes into account the grouping of the objectives into subsets. Specifically, Ω_0 and Ω_k include the global maximizers and the ϵ-active left local maximizers for each of the ϕ^i's considered independently; and Ω_0 includes the "end-points" for each of the ϕ^i's.

The following parameter values are used in CFSQP: $\alpha = 0.1$, $\beta = 0.5$, δ is the square root of the machine precision, and $\epsilon = 1$ (in $\Omega_\epsilon^{\ell\ell m}(x)$). For the solution of the QP subproblems, CFSQP invokes QLD, a code due to Powell and Schittkowski [40]. H_k is updated using the BFGS formula with Powell's modification [35] with the following stipulations: the evaluation of the gradient of the Lagrangian function is based on the KKT multipliers corresponding to the QP subproblem and multipliers associated with values of ω not used in the QP are set to 0. Assigning the value 0 to multipliers associated with constraints not considered in the current QP subproblem is equivalent to considering them inactive, which is consistent with the intuition underlying the selection of Ω_k.

The numerical results reported below were obtained on discretized versions of nine test problems borrowed from the literature. Problems `OET 1` through `OET 7` are taken from [26], `HET-Z` from [19], and `PT` from [34]. Problems `OET 1` through `OET 7` and `HET-Z` are of the form

$$\text{minimize } \max_{\omega \in I} |\phi(x,\omega)|,$$

with ϕ and I as follows (the ξ's are the components of x):

OET 1: $\phi(x,\omega) = \omega^2 - (\xi_1\omega + \xi_2\exp(\omega))$, $I = [0,2]$, $x_0 = (1,1)$.
OET 2: $\phi(x,\omega) = \frac{1}{1+\omega} - \xi_1\exp(\xi_2\omega)$, $I = [-0.5, 0.5]$, $x_0 = (1,-1)$.
OET 3: $\phi(x,\omega) = \sin(\omega) - (\xi_1 + \xi_2\omega + \xi_3\omega^2)$, $I = [0,1]$, $x_0 = (1,1,1)$.
OET 4: $\phi(x,\omega) = \exp(\omega) - \frac{\xi_1+\xi_2\omega}{1+\xi_3\omega}$, $I = [0,1]$, $x_0 = (1,1,1)$.
OET 5: $\phi(x,\omega) = \sqrt{\omega} - (\xi_4 - (\xi_1\omega^2 + \xi_2\omega + \xi_3)^2)$, $I = [0.25, 1]$, $x_0 = (1,1,1,1)$.
OET 6: $\phi(x,\omega) = \frac{1}{1+\omega} - (\xi_1\exp(\xi_3\omega) + \xi_2\exp(\xi_4\omega))$, $I = [-0.5, 0.5]$, $x_0 = (1,1,-3,-1)$.
OET 7: $\phi(x,\omega) = \frac{1}{1+\omega} - (\xi_1\exp(\xi_4\omega) + \xi_2\exp(\xi_5\omega) + \xi_3\exp(\xi_6\omega))$, $I = [-0.5, 0.5]$, $x_0 = (1,1,1,-7,-3,-1)$.
HET-Z: $\phi(x,\omega) = (1-\omega^2) - (0.5x^2 - 2x\omega)$, $I = [-1,1]$, $x_0 = 1$.

Problem PT is of the form

$$\text{minimize } \max_{\omega \in I} \phi(x,\omega),$$

with $\phi(x,\omega) = (2\omega^2 - 1)x + \omega(1-\omega)(1-x)$, $I = [0,1]$ and $x_0 = 5$.

To assess the efficiency of the scheme proposed in this paper, we compared the CFSQP implementation of Algorithm 2.1 with two algorithms differing from it only in the selection of Ω_k at each iteration. In algorithm FULL, $\Omega_k = \Omega$ at each iteration, which essentially corresponds to Han's algorithm [14], [15]. In algorithm ϵ-ACT, a simple "ϵ-active" scheme is used, specifically, $\Omega_k = \Omega_\epsilon(x_k)$ for all k, with $\epsilon = 0.1$ (both for $+\phi$ and $-\phi$ in the case of the first 8 problems). For all three algorithms, the optimization process was terminated whenever $\|d_k\| \leq 1.\text{E-}4$ was achieved.

In Tables 1 and 2, results are reported for 101 and 501 uniformly spaced mesh points, respectively (for a total of, respectively, 202 and 1002 "discrete objectives" in the case of the first 8 problems); specifically,

$$\Omega = \{a, a + \frac{b-a}{q}, a + \frac{2(b-a)}{q}, \ldots, b\},$$

with $q = 100$ and 500, respectively, where a and b are the end points of the interval of variation of ω for the problem under consideration. In the tables, NF is the number of evaluations of objective function ϕ,[8,9] IT is the total number of iterations, $\sum |\Omega_k|$ is the sum over all iterations k of the cardinality of Ω_k (in case of NEW and FULL, it is equal to the total number of gradient evaluations), and $|\Omega^*|$ is the number of points in $\Omega_k = \Omega^*$ at the stopping point $x_k = x^*$. TIME indicates the execution time in seconds, and OBJECTIVE the value of the objective function at x^*. All tests were conducted on a SUN/SPARC 1 workstation.

The following observations may be made. In most cases, the number of iterations and the total number of function evaluations are lowest for FULL and highest for NEW. This is expected though since the search directions in NEW are computed based on a much simpler QP model. Note, however, that the increase in the number of iterations and function evaluations when using NEW instead of FULL is typically moderate. In contrast, NEW provides dramatic savings in terms of number of gradient evaluations and of size of the QP subproblems (whereas the savings achieved by ϵ-ACT are modest). Note, in particular, that $|\Omega^*|$ remains essentially unchanged when the number of mesh points is increased from 101 to 501. The decrease in computational effort achieved by NEW is clearly evident in the dramatically lower TIME of execution.

[8] For the first eight problems, all numbers in this column are pessimistic by a factor of about 2: evaluation of $+\phi$ and $-\phi$ at a given point counts as two function evaluations.

[9] Note that NF is in general not a multiple of the cardinality of Ω. Indeed computation of $\tilde{d}_k$ involves evaluating $\phi(x_k + d_k, \omega)$ only for $\omega \in \Omega_k$ (see (3.9)).

Finally note that, in the implementation used for these numerical tests, the QP solver does not take into account information from the solution of the previous QP subproblem when starting a new one (QLD does not allow for such "crash start"). One could argue that a crash start may significantly speed up the solution of QP subproblems in algorithm FULL, its effect being akin to drastically reducing the number of constraints to be dealt with by the QP solver; and that, as a result, if a crash start were used, the computational cost of solving QP subproblems in FULL might be comparable to that of solving QP subproblems in NEW. To investigate this issue, we conducted additional tests with QPSOL [9] (which allows for crash starts) replacing QLD. It was observed that a crash start is helpful only in the final iterations, when the active set is correctly identified. Our interpretation of this phenomenon is that, in early iterations, while as evidenced by the good behavior of NEW the crash set is a reasonable approximation for the active set in the sense that there are values of ω in the crash set close to most values of ω in the true active set, there nevertheless may be very little (or none at all) overlap between these two sets. As a result a crash set may be of no use to an off-the-shelf QP solver. Overall, the QP-solving time in NEW is still significantly lower then the QP solving time in FULL.

PROB	n	ALGO	NF	IT	$\sum \lvert\Omega_k\rvert$	$\lvert\Omega^*\rvert$	TIME	OBJECTIVE	$\lVert d^*\rVert$
OET 1	2	NEW	2546	10	56	6	0.88	0.53819574	0.38E-16
		FULL	1445	6	1212	202	1.54	0.53813894	0.97E-14
		ϵ-ACT	2444	8	560	92	1.63	0.53819574	0.38E-16
OET 2	2	NEW	861	4	22	6	0.53	0.08715336	0.12E-05
		FULL	642	3	606	202	1.05	0.08715640	0.47E-05
		ϵ-ACT	842	4	448	202	1.00	0.08716226	0.35E-04
OET 3	3	NEW	1805	7	47	8	0.96	0.00450481	0.51E-15
		FULL	1387	5	1010	202	2.31	0.00450481	0.11E-16
		ϵ-ACT	1905	7	988	202	1.81	0.00450481	0.19E-15
OET 4	3	NEW	2805	10	68	8	1.34	0.00429463	0.27E-11
		FULL	2472	9	1818	202	4.50	0.00429463	0.76E-11
		ϵ-ACT	3525	12	1494	202	4.51	0.00429463	0.37E-10
OET 5	4	NEW	6727	19	152	8	3.26	0.00264951	0.23E-05
		FULL	5533	18	3636	202	12.2	0.00264951	0.29E-06
		ϵ-ACT	7407	22	3360	202	13.0	0.00264951	0.24E-05
OET 6	4	NEW	4314	14	128	10	2.54	0.00206863	0.18E-09
		FULL	3765	12	2424	202	9.47	0.00206878	0.22E-06
		ϵ-ACT	4035	13	2376	202	10.2	0.00206880	0.39E-06
OET 7	6	NEW	38106	97	1186	12	28.6	0.00006644	0.98E-04
		FULL	11887	31	6262	202	55.8	0.00004432	0.18E-12
		ϵ-ACT	25974	68	13274	202	124.	0.00004432	0.32E-12
HET-Z	1	NEW	606	2	7	3	0.29	1.00000000	0.22E-15
		FULL	1437	7	1414	202	1.63	0.99995000	0
		ϵ-ACT	1010	4	194	64	0.56	0.99995000	0
PT	1	NEW	1224	7	18	2	0.35	0.23605381	0
		FULL	602	5	505	101	0.55	0.23605381	0
		ϵ-ACT	986	5	155	66	0.34	0.23605381	0

Table 1: Numerical Results with Discretization $|\Omega| = 101$

PROB	n	ALGO	NF	IT	$\sum \lvert\Omega_k\rvert$	$\lvert\Omega^*\rvert$	TIME	OBJECTIVE	$\lVert d^*\rVert$
OET 1	2	NEW	13420	11	62	6	4.17	0.53824312	0.38E-16
		FULL	7153	6	6012	1002	10.6	0.53824312	0.97E-13
		ϵ-ACT	11919	8	2748	448	7.46	0.53824312	0.48E-13
OET 2	2	NEW	4207	4	23	6	1.61	0.08716106	0.15E-05
		FULL	3155	3	3006	1002	4.40	0.08716395	0.47E-05
		ϵ-ACT	4128	4	2212	1002	3.89	0.08716768	0.31E-04
OET 3	3	NEW	8920	7	50	9	3.25	0.00450552	0.25E-05
		FULL	6873	5	5010	1002	9.95	0.00450505	0.46E-17
		ϵ-ACT	9829	7	4904	1002	12.2	0.00450505	0.17E-15
OET 4	3	NEW	13886	10	71	9	4.78	0.00429567	0.12E-06
		FULL	11998	9	9018	1002	21.6	0.00429543	0.76E-11
		ϵ-ACT	17161	12	7404	1002	21.5	0.00429543	0.32E-10
OET 5	4	NEW	33441	19	158	8	12.2	0.00265008	0.27E-5
		FULL	27460	18	18036	1002	59.6	0.00265008	0.30E-06
		ϵ-ACT	42085	25	19694	1002	57.4	0.00265008	0.16E-04
OET 6	4	NEW	21345	14	131	11	8.63	0.00206998	0.16E-05
		FULL	18625	12	12024	1002	50.8	0.00206989	0.22E-06
		ϵ-ACT	19995	13	11856	1002	52.2	0.00206996	0.35E-06
OET 7	6	NEW	54584	30	355	15	27.2	0.00013273	0.32E-04
		FULL	60521	32	32064	1002	273.	0.00004446	0.11E-12
		ϵ-ACT	127096	67	64806	1002	383.	0.00006876	0.77E-13
HET-Z	1	NEW	3006	2	7	3	0.89	1.00000000	0.22E-13
		FULL	10062	10	10020	1002	10.9	0.99999800	0.14E-13
		ϵ-ACT	7092	7	1910	316	3.35	0.99999800	0.14E-13
PT	1	NEW	7337	8	22	2	1.03	0.23606791	0
		FULL	2991	5	2505	501	1.84	0.23606792	0
		ϵ-ACT	3895	5	799	334	0.90	0.23606792	0

Table 2: Numerical Results with Discretization $|\Omega| = 501$

5. Conclusion. An SQP-type algorithm has been proposed and analyzed for the solution of minimax optimization problems with many more objective functions than variables, in particular, of finely discretized continuous minimax problems. (It has been argued that SQP-type algorithms are particularly suited to certain classes of such problems.) At each iteration, a quadratic programming problem involving only a small set of constraints is solved and, correspondingly, only a few gradients are evaluated. Numerical results indicate that the proposed scheme is efficient.

There is no conceptual difficulty in extending the algorithm to tackle discretized versions of continuous minimax problems where the maximization is with respect to several free variables ranging over arbitrary compact sets. The proposed algorithm, with appropriate modifications, has been implemented in an optimization-based design package [7] and has proven very successful in solving various types of engineering design problems.

6. Appendix: Proofs.

6.1. Proof of Lemma 3.6. We denote by $H_k^{\frac{1}{2}}$ the symmetric positive definite square root of H_k, by $H_k^{-\frac{1}{2}}$ its inverse, and we make use of the following notation:

$$\begin{aligned}
\gamma_k(\omega) &= \Phi(x_k) - \phi(x_k,\omega)\\
\pi_k &= \sum_{\omega\in\Omega_k} \mu_{k,\omega}\gamma_k(\omega)\\
\pi_k^+ &= \sum_{\omega\in\Omega_k} \mu_{k,\omega}\gamma_{k+1}(\omega)\\
g_k(\omega) &= H_k^{-\frac{1}{2}}\nabla_x\phi(x_k,\omega)\\
p_k &= \sum_{\omega\in\Omega_k} \mu_{k,\omega}g_k(\omega) = -H_k^{\frac{1}{2}}d_k\\
p_k^+ &= \sum_{\omega\in\Omega_k} \mu_{k,\omega}g_{k+1}(\omega).
\end{aligned}$$

It follows from (3.1), boundedness of $\{\mu_k\}$ and Assumptions 1-3 that, for some $M > 1$,

$$max\{|v_k|, \|p_k^+\|, \|g_{k+1}(\bar\omega)\|\} \le M, \quad \forall k\in \mathbb{N}, \quad \forall \omega\in\Omega. \tag{6.1}$$

We will show that (3.3) holds with

$$c = \frac{(1-2\tilde\alpha)^2}{16M^2}, \tag{6.2}$$

where $\tilde\alpha$ is any number in $(\alpha, 1/2)$.
A few more lemmas are first established.

Lemma 6.1. *Let K be an infinite index set such that $\{\|H_{k+1} - H_k\|\}$ converges to zero on K and $\{d_k\}$ is bounded away from zero on K. Then, given any $\tilde\alpha > \alpha$,*

$$\phi(x_{k+1},\omega) + \langle\nabla_x\phi(x_{k+1},\omega), H_{k+1}^{-\frac{1}{2}}H_k^{\frac{1}{2}}d_k\rangle - \Phi(x_{k+1}) \ge -\tilde\alpha\langle d_k, H_kd_k\rangle$$

whenever $t_k < 1$, for $k\in K$, k large enough, and for all $\omega\in\Omega$ such that

$$\phi(x_k + \frac{t_k}{\beta}d_k,\omega) > \Phi(x_k) - \alpha\frac{t_k}{\beta}\langle d_k, H_kd_k\rangle.$$

Proof. In view of Lemma 3.3(iv), $\{t_k\}$ converges to zero on K. Proceeding by contradiction, suppose the claim does not hold, *i.e.*, there exists an infinite index set $K' \subseteq K$ such that, for all $k\in K'$, $t_k < 1$ and for some $\bar\omega_k\in\Omega$

$$\phi(x_k + \frac{t_k}{\beta}d_k,\bar\omega_k) > \Phi(x_k) - \alpha\frac{t_k}{\beta}\langle d_k, H_kd_k\rangle \tag{6.3}$$

and

$$\phi(x_{k+1},\bar\omega_k) + \langle\nabla_x\phi(x_{k+1},\bar\omega_k), H_{k+1}^{-\frac{1}{2}}H_k^{\frac{1}{2}}d_k\rangle - \Phi(x_{k+1}) < -\tilde\alpha\langle d_k, H_kd_k\rangle. \tag{6.4}$$

In view of Lemma 3.3(i,ii) and Assumption 3, there exists an infinite index set $K'' \subseteq K'$ such that the sequences $\{x_k\}$, $\{d_k\}$, $\{H_k\}$ and $\{\bar\omega_k\}$ converge on K'' respectively to some x^*, d^*, H^* and ω^*. In view of Lemma 3.3(iv), $\{x_{k+1}\}$ also converges to x^* on K''. Furthermore, since $\{t_k\}$ goes to zero on K'', it follows from (6.3) and our continuity assumption that $\omega^*\in\Omega_{max}(x^*)$. Also, a simple contradiction argument using Assumption 3 shows that $\{H_{k+1}^{-\frac{1}{2}} H_k^{\frac{1}{2}}\} \to I$ on K. Thus taking the limit of (6.4) on K'' yields

$$\langle \nabla_x \phi(x^*, \omega^*), d^* \rangle \leq -\tilde{\alpha} \langle d^*, H^* d^* \rangle < -\alpha \langle d^*, H^* d^* \rangle. \tag{6.5}$$

On the other hand, (6.3) implies that

$$\phi(x_k + \frac{t_k}{\beta} d_k, \bar{\omega}_k) > \phi(x_k, \bar{\omega}_k) - \alpha \frac{t_k}{\beta} \langle d_k, H_k d_k \rangle.$$

Thus,

$$\frac{\phi(x_k + \frac{t_k}{\beta} d_k, \bar{\omega}_k) - \phi(x_k, \bar{\omega}_k)}{\frac{t_k}{\beta}} > -\alpha \langle d_k, H_k d_k \rangle. \tag{6.6}$$

Taking the limit of (6.6) as $k \to \infty$ on K'' yields

$$\langle \nabla_x \phi(x^*, \omega^*), d^* \rangle \geq -\alpha \langle d^*, H^* d^* \rangle,$$

which contradicts (6.5). □

As in [21, Chapter 3], using the dual of $QP(x_k, H_k, \Omega_k)$ facilitates the analysis.

LEMMA 6.2. *Given any $x \in R^n$, $H = H^T > 0$, and $\hat{\Omega} \subseteq \Omega$, the dual quadratic program $\overline{QP}(x, H, \hat{\Omega})$ of $QP(x, H, \hat{\Omega})$ is given by*

$$\overline{QP}(x, H, \hat{\Omega}) \qquad \text{maximize} \quad -\left(\frac{1}{2} \| \sum_{\omega \in \hat{\Omega}} \mu_\omega g(\omega) \|^2 + \sum_{\omega \in \hat{\Omega}} \mu_\omega \gamma(\omega) \right) \text{ s.t. } \mu \in U,$$

where $\gamma(\omega) = \Phi(x) - \phi(x, \omega)$, $g(\omega) = H^{-\frac{1}{2}} \nabla_x \phi(x, \omega)$, and

$$U = \{\mu \in I\!R^{|\hat{\Omega}|} : \sum_{\omega \in \hat{\Omega}} \mu_\omega = 1 \quad \textit{and} \quad \mu_\omega \geq 0 \ \forall \omega \in \hat{\Omega}\}.$$

Proof. The dual is given by

$$\text{maximize } \varphi(\mu) \quad \text{s.t. } \mu \in U,$$

where φ is the dual functional, *i.e.*,

$$\varphi(\mu) = \min_d \left\{ \frac{1}{2} \langle d, Hd \rangle + \sum_{\omega \in \hat{\Omega}} \mu_\omega (\phi(x, \omega) + \langle \nabla_x \phi(x, \omega), d \rangle) - \Phi(x) \right\}. \tag{6.7}$$

In view of Assumption 3, the unique minimizer d^* in (6.7) is given by

$$d^* = -H^{-1} \sum_{\omega \in \hat{\Omega}} \mu_\omega \nabla_x \phi(x, \omega) = -H^{-\frac{1}{2}} \sum_{\omega \in \hat{\Omega}} \mu_\omega g(\omega)$$

yielding

$$\sum_{\omega \in \hat{\Omega}} \mu_\omega \langle \nabla_x \phi(x, \omega), d^* \rangle = -\langle d^*, Hd^* \rangle.$$

Therefore,

$$\varphi(\mu) = -\frac{1}{2}\langle d^*, Hd^*\rangle - \sum_{\omega\in\hat{\Omega}} \mu_\omega\{\Phi(x) - \phi(x,\omega)\}$$
$$= -\left(\frac{1}{2}\|\sum_{\omega\in\hat{\Omega}} \mu_\omega g(\omega)\|^2 + \sum_{\omega\in\hat{\Omega}} \mu_\omega \gamma(\omega)\right)$$

and the result follows. □

LEMMA 6.3. *There exists* $\underline{t} > 0$ *such that*

$$\phi(x_k + td_k, \omega) \le \Phi(x_k) - \alpha t\langle d_k, H_k d_k\rangle$$

for all k*, all* $t \in [0, \underline{t}]$*, and all* $\omega \in \Omega_k$*.*

Proof. In view of Assumption 1' and boundedness of $\{d_k\}$, there exist $c_1 > 0$ and $c_2 > 0$ such that, for all $\omega \in \Omega$, all $t \in [0,1]$ and all k,

$$\phi(x_k + td_k, \omega) \le \phi(x_k, \omega) + c_1 t\|d_k\|$$

and

$$\phi(x_k + td_k, \omega) \le \phi(x_k, \omega) + t\langle\nabla_x\phi(x_k,\omega), d_k\rangle + c_2t^2\|d_k\|^2.$$

Thus, it follows from (2.2) applied to $QP(x_k, H_k, \Omega_k)$ that, for all $\omega \in \Omega_k$, all $t \in [0,1]$ and all k,

$$\begin{aligned}
&\phi(x_k{+}td_k, \omega)\\
&\quad\le (1-t)\phi(x_k,\omega) + t\{\phi(x_k,\omega) + \langle\nabla_x\phi(x_k,\omega), d_k\rangle\} + c_2t^2\|d_k\|^2\\
&\quad\le (1-t)\phi(x_k,\omega) + t\max_{\omega\in\Omega_k}\{\phi(x_k,\omega) + \langle\nabla_x\phi(x_k,\omega), d_k\rangle\} + c_2t^2\|d_k\|^2\\
&\quad= (1-t)\phi(x_k,\omega) + t\sum_{\omega\in\Omega_k}\mu_{k,\omega}\{\phi(x_k,\omega) + \langle\nabla_x\phi(x_k,\omega), d_k\rangle\} + c_2t^2\|d_k\|^2\\
&\quad\le (1-t)\Phi(x_k) + t\Phi(x_k)\sum_{\omega\in\Omega_k}\mu_{k,\omega} + t\sum_{\omega\in\Omega_k}\mu_{k,\omega}\langle\nabla_x\phi(x_k,\omega), d_k\rangle + c_2t^2\|d_k\|^2\\
&\quad= \Phi(x_k) - t\langle d_k, H_kd_k\rangle + c_2t^2\|d_k\|^2,
\end{aligned}$$

where, again, $\mu_{k,\omega}$, $\omega \in \Omega_k$ are the KKT multipliers associated with $QP(x_k, H_k, \Omega_k)$. Thus, in view of Assumption 3, since $\alpha \in (0, 1/2)$,

$$\begin{aligned}
\phi(x_k + td_k, \omega) &\le \Phi(x_k) - \alpha t\langle d_k, H_kd_k\rangle + t(\alpha-1)\langle d_k, H_kd_k\rangle + c_2t^2\|d_k\|^2\\
&\le \Phi(x_k) - \alpha t\langle d_k, H_kd_k\rangle + t\|d_k\|^2\{(\alpha-1)\sigma_1 + c_2t\}\\
&\le \Phi(x_k) - \alpha t\langle d_k, H_kd_k\rangle, \quad \forall t \in [0, \bar{t}],
\end{aligned}$$

for all $\omega \in \Omega_k$, with $\bar{t} = (1-\alpha)\sigma_1/c_2 > 0$. □

Proof of Lemma 3.6. Since $\{d_k\}$ is bounded away from zero on K, it follows from Lemma 3.3(iv) that $\{t_k\}$ goes to zero on K. Without loss of generality assume that $t_k < 1$ for all $k \in K$. For each $k \in K$, let $\bar{\omega}_k \in \Omega_{k+1}$ be the value picked in Step 2 of Algorithm 2.1; $\bar{\omega}_k$ satisfies

$$\phi(x_k + \frac{t_k}{\beta}d_k, \bar{\omega}_k) > \Phi(x_k) - \alpha\frac{t_k}{\beta}\langle d_k, H_kd_k\rangle.$$

In view of Lemma 6.3, $\bar{\omega}_k \notin \Omega_k$ for all $k \in K$. Define $\Omega'_k = \Omega^b_k \cup \{\bar{\omega}_k\}$. Let v'_{k+1} denote the optimal value of $QP(x_{k+1}, H_{k+1}, \Omega'_k)$. In view of the construction of Ω_{k+1}, $\Omega'_k \subseteq \Omega_{k+1}$. Thus, $|v_{k+1}| \le |v'_{k+1}|$. Therefore, it suffices to prove (3.3) with the left hand side replaced by $|v'_{k+1}|$. Define the quadratic function in ν

$$\begin{aligned} Q(\nu) &= \frac{1}{2}\|\nu g_{k+1}(\bar{\omega}_k) + (1-\nu)\sum_{\omega\in\Omega_k}\mu_{k,\omega} g_{k+1}(\omega)\|^2 + \nu\gamma_{k+1}(\bar{\omega}_k) \\ &\quad + (1-\nu)\sum_{\omega\in\Omega_k}\mu_{k,\omega}\gamma_{k+1}(\omega) \\ &= \frac{1}{2}\|\nu g_{k+1}(\bar{\omega}_k) + (1-\nu)p^+_k\|^2 + \nu\gamma_{k+1}(\bar{\omega}_k) + (1-\nu)\pi^+_k. \end{aligned}$$

Let $\nu \in [0,1]$. Let $\mu_{k,\omega}$, $\omega \in \Omega_k$, be the KKT multipliers associated to $QP(x_k, H_k, \Omega_k)$. With the (dual feasible) choice $\mu_{\bar{\omega}_k} = \nu$, $\mu_\omega = (1-\nu)\mu_{k,\omega}$, for all $\omega \in \Omega^b_k$, and $\mu_\omega = 0$ for all $\omega \in \Omega\backslash\Omega'_k$, the objective of the dual quadratic program $\overline{QP}(x_{k+1}, H_{k+1}, \Omega'_k)$ takes value $-Q(\nu)$. By duality, v'_{k+1} is the optimal objective value for both $QP(x_{k+1}, H_{k+1}, \Omega'_k)$ and $\overline{QP}(x_{k+1}, H_{k+1}, \Omega'_k)$. Thus,

$$|v'_{k+1}| \le Q(\nu), \quad \forall \nu \in [0,1].$$

Thus, it suffices to prove (3.3) with the left hand side replaced by $min_{\nu\in[0,1]}\, Q(\nu)$. Expanding the quadratic term of $Q(\nu)$ yields

$$\begin{aligned} Q(\nu) &= \frac{1}{2}\nu^2\|g_{k+1}(\bar{\omega}_k)\|^2 + \frac{1}{2}(1-\nu)^2\|p^+_k\|^2 + \nu(1-\nu)\langle g_{k+1}(\bar{\omega}_k), p^+_k\rangle \\ &\quad + \nu\gamma_{k+1}(\bar{\omega}_k) + (1-\nu)\pi^+_k \\ &= \frac{1}{2}\|p^+_k\|^2 + \frac{\nu^2}{2}\|g_{k+1}(\bar{\omega}_k) - p^+_k\|^2 + \nu\langle g_{k+1}(\bar{\omega}_k), p^+_k\rangle + \nu\gamma_{k+1}(\bar{\omega}_k) \\ &\quad - \nu\|p^+_k\|^2 + (1-\nu)\pi^+_k. \end{aligned} \tag{6.8}$$

Note that

$$\langle g_{k+1}(\bar{\omega}_k), p_k\rangle = -\langle \nabla_x\phi(x_{k+1}, \bar{\omega}_k), H^{-\frac{1}{2}}_{k+1} H^{\frac{1}{2}}_k d_k\rangle.$$

Since $\{t_k\}$ converges to zero on K, $t_k < \delta$ for $k \in K$, k large enough. Since $\bar{\omega}_k \notin \Omega_k$, it follows from *Step 2* in Algorithm 2.1 that $H_{k+1} = H_k$ for $k \in K$, k large enough. Thus, assumptions of Lemma 6.1 are all satisfied. Given $\tilde{\alpha} \in (\alpha, 1/2)$, in view of Lemma 6.1 with $\omega = \bar{\omega}_k$, there exists an integer k_1 such that, for all $k \ge k_1$, $k \in K$,

$$\langle g_{k+1}(\bar{\omega}_k), p_k\rangle + \gamma_{k+1}(\bar{\omega}_k) \le \gamma_{k+1}(\bar{\omega}_k) + \phi(x_{k+1}, \bar{\omega}_k) - \Phi(x_{k+1}) + \tilde{\alpha}\langle d_k, H_k d_k\rangle.$$

In view of the definition of γ_k and of relationships (3.1) and (3.2), it follows that

$$\langle g_{k+1}(\bar{\omega}_k), p_k\rangle + \gamma_{k+1}(\bar{\omega}_k) \le -2\tilde{\alpha} v_k, \quad \forall k \in K, k \ge k_1.$$

Hence, for all $k \ge k_1$, $k \in K$,

$$\langle g_{k+1}(\bar{\omega}_k), p^+_k\rangle + \gamma_{k+1}(\bar{\omega}_k) \le -2\tilde{\alpha} v_k - \langle g_{k+1}(\bar{\omega}_k), p_k - p^+_k\rangle.$$

Also,

$$\begin{aligned} \|p^+_k\|^2 &= \|p_k - p_k + p^+_k\|^2 \\ &= \|p_k\|^2 + \|p_k - p^+_k\|^2 - 2\langle p_k, p_k - p^+_k\rangle \\ &= \|p_k\|^2 + O(\|p_k - p^+_k\|). \end{aligned}$$

On the other hand, since $M > 1$, inequality (6.1) implies that

$$\|g_{k+1}(\bar{\omega}_k) - p_k^+\|^2 \leq 4M^2.$$

Substituting all these into (6.8) yields, for all $k \geq k_1$, $k \in K$, $\nu \in [0,1]$,

$$\begin{aligned} Q(\nu) &\leq \frac{1}{2}\|p_k\|^2 + 2M^2\nu^2 - \nu(2\tilde{\alpha}v_k + \|p_k\|^2) + (1-\nu)\pi_k^+ + O(\|p_k - p_k^+\|) \\ &= \frac{1}{2}\|p_k\|^2 + \pi_k + 2M^2\nu^2 - \nu(2\tilde{\alpha}v_k + \|p_k\|^2 + \pi_k) - (1-\nu)(\pi_k - \pi_k^+) \\ &\quad + O(\|p_k - p_k^+\|). \end{aligned}$$

In view of Lemma 6.2, since duality holds,

$$|v_k| = -v_k = \frac{1}{2}\|p_k\|^2 + \pi_k.$$

Thus, for all $k \geq k_1$, $k \in K$,

$$\begin{aligned} Q(\nu) &\leq |v_k| + 2M^2\nu^2 - \nu(2\tilde{\alpha}v_k + |v_k|) - \frac{\nu}{2}\|p_k\|^2 \\ &\quad + O(|\pi_k - \pi_k^+|) + O(\|p_k - p_k^+\|) \\ &\leq |v_k| + 2M^2\nu^2 - \nu(1-2\tilde{\alpha})|v_k| + O(|\pi_k - \pi_k^+|) + O(\|p_k - p_k^+\|). \end{aligned}$$

The minimum of the right hand side is achieved at $\bar{\nu}_k = 4c|v_k|$, with $c = (1 - 2\tilde{\alpha})/16M^2$. Since $\tilde{\alpha} < 1/2$, $|v_k| \leq M$ and $M > 1$, it follows that $\bar{\nu}_k \in [0,1]$ and thus, for all $k \geq k_1$, $k \in K$,

$$\min_{\nu\in[0,1]} Q(\nu) \leq Q(\bar{\nu}_k) \leq |v_k| - 32M^2c^2|v_k|^2 + O(|\pi_k - \pi_{k^+}|) + O(\|p_k - p_k^+\|).$$

Now, in view of Lemma 3.3(i,iv) and Assumption 1, and since $H_{k+1} = H_k$ for k large enough, $k \in K$, $\{\pi_k - \pi_k^+\}$ and $\{p_k - p_k^+\}$ both tend to zero as k goes to infinity, $k \in K$. Thus, since v_k is bounded away from zero on K (Lemma 3.3(i)), there exists a positive integer N such that, for all $k \geq N$, $k \in K$

$$\min_{\nu\in[0,1]} Q(\nu) \leq |v_k| - c|v_k|^2.$$

Therefore (3.3) follows from the inequality

$$|v_{k+1}| \leq |v'_{k+1}| \leq \min_{\nu\in[0,1]} Q(\nu).$$

□

6.2. Proof of Lemma 3.8. $z_0 - \eta z_0^2$ achieves its largest value with $z_0 = \frac{1}{2\eta}$, yielding a largest possible value for z_1 given by

$$z_1 = \frac{1}{2\eta} - \eta(\frac{1}{2\eta})^2 = \frac{1}{4\eta}.$$

The mapping

$$z \mapsto z - \eta z^2$$

is monotonic increasing over $[0, \frac{1}{2\eta}]$. Thus, given any z_0, the sequence defined by

$$z_1 = \frac{1}{4\eta}$$
$$z_{i+1} = z_i - \eta z_i^2, \quad i = 1, 2, \ldots$$

is the largest of all nonnegative sequences satisfying the given inequality condition, in the sense that given any such sequence $\{y_i\}$,

$$y_i \leq z_i, \quad i = 1, 2, \ldots$$

Let now i_0 be such that $z_i < \epsilon$ for all $i \geq i_0$. It follows that $y_i < \epsilon$ for all $i \geq i_0$. □

6.3. Proof of Lemma 3.10. We first show that $(x^*, \Phi(x^*), \mu^*)$ is an isolated solution of the nonlinear system of equations in (x, x^0, μ) (see (2.1))

$$\begin{cases} \sum_{\omega\in\Omega} \mu_\omega \nabla_x \phi(x, \omega) = 0 \\ \sum_{\omega\in\Omega} \mu_\omega = 1 \\ \mu_\omega(\phi(x, \omega) - x^0) = 0 \;\; \forall \omega \in \Omega. \end{cases} \tag{6.9}$$

This will result from the Inverse Function Theorem if the Jacobian of the left-hand side of this system of equations is nonsingular. To show that this is indeed the case, let $(d, d^0, \lambda) \in R^n \times R \times R^{|\Omega|}$ be such that

$$\begin{bmatrix} \nabla^2_{xx} L(x^*, \mu^*) & 0 & \nabla_x \phi(x^*, \omega_1) & \ldots & \nabla_x \phi(x^*, \omega_\ell) \\ 0 & 0 & 1 & \ldots & 1 \\ \mu^*_{\omega_1} \nabla_x \phi(x^*, \omega_1)^T & -\mu^*_{\omega_1} & \phi(x^*, \omega_1) - \Phi(x^*) & & \\ \vdots & \vdots & & \ddots & \\ \mu^*_{\omega_\ell} \nabla_x \phi(x^*, \omega_\ell)^T & -\mu^*_{\omega_\ell} & & & \phi(x^*, \omega_\ell) - \Phi(x^*) \end{bmatrix} \begin{bmatrix} d \\ d^0 \\ \\ \lambda \\ \end{bmatrix} = 0,$$

where $\omega_1, \ldots, \omega_\ell$ are the elements of Ω, *i.e.*, suppose that

$$\nabla^2_{xx} L(x^*, \mu^*) d + \sum_{\omega\in\Omega} \lambda_\omega \nabla_x \phi(x^*, \omega) = 0, \tag{6.10}$$

$$\sum_{\omega\in\Omega} \lambda_\omega = 0, \tag{6.11}$$

$$\mu^*_\omega \langle \nabla_x \phi(x^*, \omega), d\rangle - \mu^*_\omega d^0 + \lambda_\omega(\phi(x^*, \omega) - \Phi(x^*)) = 0 \;\forall\; \omega \in \Omega. \tag{6.12}$$

From (6.12) and the last line in (2.1) it follows that

$$\lambda_\omega = 0 \quad \forall \omega \quad \text{s.t.} \quad \phi(x^*, \omega) < \Phi(x^*). \tag{6.13}$$

Together with (6.12) this implies that

$$\mu^*_\omega(\langle \nabla_x \phi(x^*, \omega), d\rangle - d^0) = 0 \quad \forall \omega \in \Omega$$

and, in view of the strict complementarity assumption

$$\langle \nabla_x \phi(x^*, \omega), d \rangle = d^0 \quad \forall \omega \in \Omega_{max}(x^*).$$

From the first two lines in (2.1), it follows that $d^0 = 0$, thus

$$\langle \nabla_x \phi(x^*, \omega), d \rangle = 0 \quad \forall \omega \in \Omega_{max}(x^*). \tag{6.14}$$

Performing the inner product with d of both sides of (6.10) and using (6.13) and (6.14) yields

$$\langle d, \nabla^2_{xx} L(x^*, \mu^*) d \rangle = 0.$$

In view of Assumption 5, this together with (6.14) implies that $d = 0$. Substituting this in (6.10) and using (6.11)-(6.13) and Assumption 4, we conclude that $\lambda_\omega = 0$ for all $\omega \in \Omega_{max}(x^*)$. Thus $(d, d^0, \lambda) = 0$ and the Jacobian of the left-hand side of (6.9) at $(x^*, \Phi(x^*), \mu^*)$ is nonsingular. Thus $(x^*, \Phi(x^*), \mu^*)$ is an isolated solution of (6.9) and therefore (x^*, μ^*) is an isolated KKT pair for (P). To complete the proof, we now proceed by contradiction. Thus suppose that there exists a sequence $\{x^*_k\}$ of KKT points for (P) that converges to x^* as k goes to infinity, with $x^*_k \neq x^*$ for all k, and let μ^*_k be a KKT multiplier vector associated with x^*_k. Since (x^*, μ^*) is an isolated KKT pair for (P), $\{\mu^*_k\}$ must be bounded away from μ^*. Since $\{\mu^*_k\}$ is bounded, we may assume without loss of generality that $\{\mu^*_k\}$ converges to $\hat{\mu}$ as k goes to infinity, with $\hat{\mu} \neq \mu^*$. In view of our continuity assumption, taking the limit at k goes to infinity of both sides of (2.1) shows that $(x^*, \hat{\mu})$ is a KKT pair for (P), in contradiction with uniqueness of the KKT multiplier vector associated to x^*, which is guaranteed by Assumption 4.

Acknowledgment. The authors wish to thank Craig T. Lawrence for putting together the CFSQP implementation of the scheme proposed in this paper and for running the numerical tests; and two referees for their insightful criticism and suggestions.

REFERENCES

[1] M. C. Biggs, *Constrained Minimization Using Recursive Equality Quadratic Programming*, in Numerical Methods for Non-Linear Optimization, F. A. Lootsma, ed., Academic Press, London, New York, 1972, pp. 411–428.

[2] S. P. Boyd and C. H. Barratt, *Linear Controller Design: Limits of Performance*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

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

[4] A. R. Conn and Y. Li, *A Structure-Exploiting Algorithm for Nonlinear Minimax Problems*, SIAM J. Optimization, 2 (May, 1992), pp. 242–263.

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

[6] H. A. Eschenauer, C. Mattheck, and N. Olhoff, eds., *Engineering Optimization in Design Processes,*, Springer-Verlag, Berlin, 1991.

[7] M. K. H. FAN, A. L. TITS, J. L. ZHOU, L. S. WANG, and J. KONINCKX, *CONSOL-OPTCAD User's Manual (Version c1.2-s1.7, Released 8/1992)*, Systems Research Center, University of Maryland, Technical Report TR-87-212r4, College Park, Maryland 20742, 1992.

[8] A. V. FIACCO, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, in Academic Press, R. Bellman, ed., Mathematics in Science and Engineering, 165, 1983.

[9] P. E. GILL, W. MURRAY, M. A. SAUNDERS, and M. H. WRIGHT, *User's Guide for QPSOL (Version 3.2): A Fortran Package for Quadratic Programming*, Systems Optimization Laboratory, Stanford University, Technical Report SOL 84-6, Stanford, CA 94305, 1984.

[10] C. GONZAGA AND E. POLAK, *On Constraint Dropping Schemes and Optimality Functions for a Class of Outer Approximations Algorithms*, SIAM J. Control Optim., 17 (1979), pp. 477–493.

[11] C. GONZAGA, E. POLAK, and R. TRAHAN, *An Improved Algorithm for Optimization Problems with Functional Inequality Constraints*, IEEE Trans. Automat. Control, AC-25 (1980), pp. 49–54.

[12] G. GRAMLICH, R. HETTICH, and E. W. SACHS, *Local Convergence of SQP-Methods in Semi-Infinite Programming*, SIAM J. Optimization, 1993 (to appear).

[13] S. A. GUSTAFSON, *A Three-Phase Algorithm for Semi-Infinite Programs*, in Semi-Infinite Programming and Applications, A. V. Fiacco and K. O. Kortanek, eds., Lecture Notes in Economics and Mathematical Systems, 215, Springer-Verlag, New York–Heidelberg–Berlin, 1983, pp. 138–157.

[14] S. P. HAN, *Superlinear Convergence of a Minimax Method*, Department of Computer Science, Cornell University, TR78-336, 1978.

[15] ———, *Variable Metric Methods for Minimizing a Class of Nondifferentiable Functions*, Math. Programming, 20 (1981), pp. 1–13.

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

[17] R. HETTICH AND 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, pp. 97–111.

[18] R. HETTICH AND K. O. KORTANEK, *Semi-Infinite Programming: Theory, Methods, and Applications*, SIAM Rev., 35 (1993), pp. 380–429.

[19] R. HETTICH AND P. ZENCKE, *Numerische Methoden der Approximation und Semi-Infiniten Optimierung*, Teubner Studienbücher, 1982.

[20] W. van HONSTEDE, *An Approximation Method for Semi-Infinite Problems*, in Semi-Infinite Programming, R. Hettich, ed., Lecture Notes in Control and Information Sciences, 15, Springer Verlag, 1979, pp. 126–136.

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

[22] C. T. LAWRENCE, J. L. ZHOU, and A. L. TITS, *User's Guide for CFSQP Version 2.0: A C Code for Solving Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality and Linear Constraints.*, Institute for Systems Research, University of Maryland, TR-94-16, College Park, MD 20742, 1994.

[23] C. LEMARÉCHAL, *Nondifferentiable Optimization*, in Optimization, G. Nemhauser, A. Rinnooy-Kan, and M. Todd, eds., Handbooks in Operations Research and Management Science, North Holland, 1989.

[24] H. MINE, M. FUKUSHIMA, and Y. TANAKA, *On the Use of ϵ-Most-Active Constraints in an Exact Penalty Function Method for Nonlinear Optimization*, IEEE Trans. Automat. Control, AC-29 (1984), pp. 1040–1042.

[25] C. A. MOTA SOARES, ed., *Computer-Aided Optimal Design: Structural and Mechanical Systems*, NATO ASI Series, Series F , 27 , Springer-Verlag, Berlin, 1987.

[26] K. OETTERSHAGEN, *Ein Superlinear Konvergenter Algorithmus zur Losung Semi-Infiniter Optimierungsprobleme*, Ph.D. Thesis, Bonn University, 1982.

[27] E. R. PANIER AND A. L. TITS, *A Globally Convergent Algorithm with Adaptively Refined Discretization for Semi-Infinite Optimization Problems Arising in Engineering Design*, IEEE Trans. Automat. Control, AC-34 (1989), pp. 903–908.

[28] ———, *A Superlinearly Convergent Method of Feasible Directions for Optimization Problems Arising in the Design of Engineering Systems*, in Proc. of the Seventh International Conference on Analysis and Optimization of Systems — Antibes, June 25-27, 1986, A. Bensoussan and J. L. Lions, eds., Lecture Notes in Control and Information Sciences, 83, Springer Verlag, Berlin-Heidelberg-New York-Tokyo, June 1986, pp. 65–73.

[29] E. POLAK, *Computational Methods in Optimization*, Academic Press, New York, N.Y., 1971.

[30] ———, *On the Mathematical Foundations of Nondifferentiable Optimization*, SIAM Rev., 29 (1987), pp. 21–89.

[31] E. POLAK AND L. HE, *Rate Preserving Discretization Strategies for Semi-Infinite Programming and Optimal Control*, in Proceedings of the 29th IEEE Conference on Decision and Control, December 1990, pp. 2444–2449.

[32] E. POLAK AND D. Q. MAYNE, *An Algorithm for Optimization Problems with Functional Inequality Constraints*, IEEE Trans. Automat. Control, 21 (1976), pp. 184–193.

[33] E. POLAK, D. Q. MAYNE, and D. M. STIMLER, *Control System Design via Semi-Infinite Optimization: A Review*, IEEE Proc., 72 (1984), pp. 1777–1794.

[34] E. POLAK AND A. L. TITS, *A Recursive Quadratic Programming Algorithm for Semi-Infinite Optimization Problems*, Appl. Math. Optim., 8 (1982), pp. 325–349.

[35] M. J. D. POWELL, *A Fast Algorithm for Nonlinearly Constrained Optimization Calculations*, in Numerical Analysis, Dundee, 1977, Lecture Notes in Mathematics 630, G. A. Watson, ed., Springer-Verlag, 1978, pp. 144–157.

[36] ———, *A Tolerant Algorithm for Linearly Constrained Optimization Calculations*, Math. Programming, 45 (1989), pp. 547–566.

[37] J. RAKOWSKA, R. T. HAFTKA, and L. T. WATSON, *Multi-Objective Control-Structure Optimization Via Homotopy Methods*, SIAM J. Optimization, 3 (1993), pp. 654–667.

[38] R. REEMTSEN, *Discretization methods for the solution of semi-infinite programming problems*, J. Optim. Theory Appl., 71 (1991), pp. 85–103.

[39] S. M. ROBINSON, *Perturbed Kuhn-Tucker Points and Rates of Convergence for a Class of Nonlinear-Programming Algorithms*, Math. Programming, 7 (1974), pp. 1–16.

[40] K. Schittkowski, *QLD : A FORTRAN Code for Quadratic Programming, User's Guide*, Mathematisches Institut, Universität Bayreuth, Germany, 1986.

[41] ———, *Solving Nonlinear Programming Problems with Very Many Constraints*, Mathematisches Institut, Universität Bayreuth, Report No. 294, Bayreuth, Germany, 1991.

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

[43] W. Zhao and S. Azarm, *A Cross-Sectional Shape Multiplier Method for Two-Level Optimum Design of Frames*, in Advances in Design Automation, Volume II - Optimal Design and Mechanical Systems Analysis, ASME Publication DE, 23-2, September 1990, pp. 197–205.

[44] J. L. Zhou, *Fast, Globally Convergent Optimization Algorithms, with Application to Engineering System Design*, University of Maryland, PhD Thesis, College Park, MD 20742, 1992.

[45] J. L. Zhou and A. L. Tits, *Nonmonotone Line Search for Minimax Problems*, J. Optim. Theory Appl., 76 (1993), pp. 455–476.

[46] G. Zoutendijk, *Methods of Feasible Directions*, Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, 1960.