TECHNICAL
RESEARCH
REPORT

SYSTEMS
RESEARCH
CENTER

Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University

An SQP Algorithm for Finely Discretized SIP Problems and Other Problems with Many Constraints

by J.L. Zhou and A.L. Tits

TR 92-125

AN SQP ALGORITHM FOR FINELY DISCRETIZED SIP PROBLEMS AND OTHER PROBLEMS WITH MANY CONSTRAINTS*

JIAN L. ZHOU† AND ANDRÉ L. TITS‡

Abstract. A common strategy for achieving global convergence in the solution of semi-infinite programming (SIP) problems is to (approximately) solve a sequence of discretized problems, with a progressively finer discretization mesh. Finely discretized SIP problems, as well as other problems with many more constraints than variables, call for algorithms in which successive search directions are computed based on a small but significant subset of the 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. The quadratic programming subproblem that yields the search direction involves only a small subset of the constraints. This subset is updated at each iteration in such a way that global convergence is insured. Heuristics are suggested that take advantage of possible close relationship between "adjacent" constraints. Numerical results demonstrate the efficiency of the proposed algorithm.

Key words. Semi-Infinite Programming, Sequential Quadratic Programming, discretization, global convergence.

AMS(MOS) subject classifications. 65, 90

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 or temperature (see, *e.g.*, [1], [2], [17], [20]). As a simple example, consider the problem

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

with

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

where $f(x)$ and $\phi(x,\omega)$ depend smoothly on x. The difficulties in solving (SI) stem mostly from the facts that (i) the accurate evaluation of $\psi_{[0,1]}$ for each x involves a potentially time consuming global maximization, and (ii) $\psi_{[0,1]}$ is nondifferentiable in general, even when ϕ is smooth. Various approaches have been proposed to circumvent these difficulties. 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.*, [10], [11], [21]). 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 of these maximizers $\omega_i(x)$. Application of Newton's method, or of a Sequential Quadratic Programming (SQP) method to the reduced problem (with constraints $\phi(x,\omega_i(x)) \le 0$) brings about a fast rate of

*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.

convergence. However global convergence, when insured at all, involves a potentially very costly line search ([3], [27]). A large class of globally convergent algorithms, on the other hand, is based on approximating $\psi_{[0,1]}$ by means of a progressively finer discretization of $[0, 1]$, *i.e.*, substituting for (SI) the problems

$$(SI_d) \qquad \min_{x\in R^n} \; f(x) \quad \text{s.t.} \quad \phi(x,\omega) \le 0 \quad \forall \omega \in \Omega$$

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.*, [6], [8], [9], [15], [18], [19], [21], [24]). The overall performance of these algorithms largely depends on the the performance at each discretization level, especially when q becomes large.

Problem (SI_d) involves finitely many 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

$$(MC) \qquad \min_{x\in R^n} \; f(x) \quad \text{s.t.} \quad \phi_i(x) \le 0 \quad i = 0, \ldots, \ell,$$

in which $\ell \gg n$, *i.e.*, in which constraints far outnumber variables. These include, among others, mechanical design problems involving trusses (see, *e.g.*, [23], [28] or papers in [4], [14]). Note that there is no essential difference between (SI_d) and (MC).

In [19], [15], (SI_d) is solved by means of first order (thus, slow) methods. In [19], the construction of the search direction at iteration k is based on the gradients $\nabla_x \phi(x_k, \omega)$ at all points $\omega \in \Omega$ at which $\phi(x_k, \omega) \ge -\epsilon$. In [15], it is shown that a small subset of these points can be used instead 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.*, [12], [13]). 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} is 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 [15] 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 required at a few grid points only. However, at each level of discretization (*i.e.*, for each fixed q), the algorithm proposed in [15] (as well as that proposed in [19]) exhibits at best a linear rate of convergence.

In [26], modifications of standard SQP methods are proposed 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 a related SQP type algorithm based on the scheme introduced in [15]. To simplify the exposition, we focus instead on an unconstrained problem closely related to (SI_d), namely,

$$(P) \qquad \min_{x\in R^n} \; \max_{\omega\in\Omega} \phi(x,\omega),$$

and we define

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

where Ω is again a finite set (a corresponding algorithm for (SI_d) is stated in Appendix 1). 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 $QP(x_k, H_k, \Omega_k)$. Here, for any $x \in \mathbb{R}^n$, $H \in \mathbb{R}^{n \times n}$ symmetric positive definite, and $\hat{\Omega} \subset \Omega$, $QP(x, H, \hat{\Omega})$ is defined by

$$(QP(x,H,\hat{\Omega})) \qquad \min_{d \in R^n} \ \frac{1}{2}\langle d, Hd \rangle + \psi'_{\hat{\Omega}}(x,d)$$

where

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

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

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

A line search (*e.g.*, of Armijo type) is performed along direction d_k to obtain a next iterate x_{k+1}; H_k is updated to H_{k+1}; and a new subset Ω_{k+1} of Ω is constructed following the idea used in [15]. In particular, Ω_{k+1} includes the global maximizers $\bar{\omega}_k$ at the last trial point $\bar{x}_{k+1}$ at which a restriction of the step was forced. However, in the present context, a difficulty arises. The rationale for including $\bar{\omega}_k$ in Ω_{k+1} is that, had $\bar{\omega}_k$ 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 [15] where a first order search direction is used (*i.e.*, $H_k = I$ for all k), it follows that d_{k+1} will be a "better" direction than d_k. 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 be jeopardized. One remedy would be to renounce updating H_k when a small step is taken while x_k is away from a Karush-Kuhn-Tucker (KKT) point, while updating H_k normally if convergence to a KKT point appears to be taking place, so as not to jeopardize the anticipated superlinear rate of convergence. From the analysis for ordinary minimax problems (see, *e.g.*, [30]), it is known that, normally, $\{d_k\}$ goes to zero and $\{t_k\}$ is bounded away from zero (even if d_k is a first order direction). This is the case in the current context (Lemmas 3.13 and 3.15 below) as it can be proved that, eventually, all critical points of Ω are included in Ω_k (Lemma 3.12). Thus a possible criterion for not updating H_k would be

$$t_k \leq \|d_k\|.$$

However, under such a criterion, undesirable resetting may take place when t_k is not small, in particular when $\|d_k\| \geq 1$. To get around this effect, a small number $\delta \in (0,1)$ is prescribed so that $H_{k+1} = H_k$ is enforced if

$$t_k \leq \min\{\delta, \ \|d_k\|\}.$$

It is shown below that this overall algorithm indeed achieves global convergence and maintains a fast rate of local convergence.

As with conventional constrained optimization problems or minimax problems, a 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 can be shown that this can be avoided by incorporating in the basic algorithm standard techniques such as second order correction and nonmonotone line search (see, *e.g.*, [29], [30]). Finally, while the algorithm just outlined does have the desired theoretical convergence properties, it may entail too extreme a reduction in size of Ω to Ω_k, and the number of iterations required for reaching the solution may become unduly large. Heuristics (possibly inspired from those proposed in [26]) can be used to include additional constraints into Ω_k without jeopardizing the convergence properties. In particular, regularity properties of $\phi(x, \omega)$ as a function of ω can be put to advantage (this was done in [7] and a similar scheme was later adopted in [15]).

The paper is organized as follows. The basic algorithm is stated in § 2. The convergence analysis is presented in § 3. In § 4, implementation and extension issues are discussed, and numerical results are reported. § 5 is devoted to final remarks. The paper ends with two appendices: the first one, § 6, contains a complete algorithm for the solution of (SI_d); the second one, § 7, contains some proofs.

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 it is a *KKT point* for (P), *i.e.*, there exist *KKT multipliers* μ^*_ω, $\omega \in \Omega$ such that

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

Similarly, given $x \in R^n$, $H = H^T > 0$ and $\hat{\Omega} \subset \Omega$, 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

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

Moreover, since $QP(x, H, \hat{\Omega})$ is a strictly convex quadratic program, it has exactly one KKT point, which is the global minimizer.

We are now ready to make precise the rule for updating Ω_k. Following [15], Ω_{k+1} is constructed as the union of three sets. Given $x \in R^n$, let

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

be the set of maximizers of $\phi(x, \cdot)$. The first component of Ω_{k+1} is $\Omega_{max}(x_{k+1})$. The second component of Ω_{k+1} is obtained from the line search. While the ideas put forth in this paper are largely unaffected by the specifics of this line search, for the sake of exposition, we will consider the case of an Armijo type line search. 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

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

where $\alpha \in (0, 1/2)$ and $\beta \in (0, 1)$ are fixed. Suppose the line search at iteration k results in $t_k < 1$, *i.e.*, the line search test (2.3) is violated at $\bar{x}_{k+1} = x_k + \frac{t_k}{\beta} d_k$. A next search direction taking this into account is called for. Thus, $\Omega_{max}(\bar{x}_{k+1})$ is the second component of Ω_{k+1}. Finally, to avoid zigzagging it is important that key elements in Ω_k be preserved 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.† Thus, the third component of Ω_{k+1} is

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

so that, overall,

$$\Omega_{k+1} = \Omega_{max}(x_{k+1}) \cup \Omega_{max}(\bar{x}_{k+1}) \cup \Omega_k^b.$$

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 $\Omega_0 = \Omega_{max}(x_0)$.
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

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

Step 2. Updates. Set $x_{k+1} = x_k + t_k d_k$. If $t_k < 1$, set $\bar{x}_{k+1} = x_k + \frac{t_k}{\beta} d_k$. Set

$$\Omega_{k+1} = \begin{cases} \Omega_{max}(x_{k+1}) \cup \Omega_k^b & \text{if } t_k = 1 \\ \Omega_{max}(x_{k+1}) \cup \Omega_{max}(\bar{x}_{k+1}) \cup \Omega_k^b & \text{if } t_k < 1. \end{cases}$$

If $t_k \leq min\{\delta, \|d_k\|\}$, set $H_{k+1} = H_k$; otherwise, compute a new positive definite approximation H_{k+1} to the Hessian of the Lagrangian of (P). 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 problem 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.

---

†If the multiplier vector associated with $QP(x_k, H_k, \Omega_k)$ is not unique, any choice is appropriate.

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

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

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

The following lemma gives certain basic properties of Algorithm 2.1.

LEMMA 3.1. *(i). If d_k nonzero solves $QP(x_k, H_k, \Omega_k)$, then there exists $\underline{t}_k > 0$ such that, for all $t \in [0, \underline{t}_k]$,*

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

(ii). The sequences $\{x_k\}$ and $\{d_k\}$ are bounded. (iii). The sequence $\{\psi(x_k)\}$ is convergent and the sequence $\{t_k d_k\}$ converges to zero.

Proof. Since $\psi_{\Omega_k}(x_k) = \psi(x_k)$, it follows from (2.2) that, at iteration k with $\hat{\Omega} = \Omega_k$,

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

thus,

$$\psi'_{\Omega_k}(x_k, d_k) \le -\langle d_k, H_k d_k\rangle. \tag{3.1}$$

Thus, in view of Assumption 3, since $\alpha < 1$, there exists $\hat{t}_k > 0$ such that, for all $t \in [0, \hat{t}_k]$,

$$\max_{\omega\in\Omega_k}\phi(x_k + td_k, \omega) \le \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle.$$

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

$$\max_{\omega\in\Omega\setminus\Omega_k}\phi(x_k + td_k, \omega) \le \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle.$$

(i) thus follows by letting $\underline{t}_k = min\{\hat{t}_k, \tilde{t}_k\}$. (ii) and (iii) follows directly from Assumptions 1, 2, 3 and the fact that $\psi(x_{k+1}) \le \psi(x_k) - \alpha t_k\langle d_k, H_k d_k\rangle$ (since, if $\{t_k d_k\}$ does not go to zero, then $\{d_k\}$ does not go to zero). □

In view of statement (i), if d_k is nonzero the line search is well-defined, even though not all points of Ω are used in constructing d_k. The following lemma implies that, if the sequence $\{x_k\}$ generated by Algorithm 2.1 is finite, the last point must be a KKT point.

LEMMA 3.2. *Let $H > 0$, and $\hat{\Omega} \subset \Omega$ with $\hat{\Omega} \cap \Omega_{max}(x) \ne \emptyset$. If the unique KKT point d of $QP(x, H, \hat{\Omega})$ is zero, then 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 $\psi_{\hat{\Omega}}(x) = \psi(x)$, the KKT condition (2.1) for (P) holds at x with multipliers $\mu_\omega = \hat{\mu}_\omega$ for $\omega \in \hat{\Omega}$ and $\mu_\omega = 0$ otherwise. Thus, x is a KKT point for (P). □

We now assume that an infinite sequence $\{x_k\}$ is generated by Algorithm 2.1. Let v_k denote the optimal value of $QP(x_k, H_k, \Omega_k)$, *i.e.*,

$$v_k = \frac{1}{2}\langle d_k, H_k d_k\rangle + \psi'_{\Omega_k}(x_k, d_k). \tag{3.2}$$

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

LEMMA 3.3. *Let $K \subset I\!N$ 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. In view of (3.1) and (3.2),

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

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

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

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

To prove (ii), suppose $\{d_k\}$ goes to zero on K and let $K' \subset K$ 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 $\hat{\Omega} \subset \Omega$. In view of Lemma 3.1(ii), Assumption 3 and the boundedness of $\{\mu_{k,\omega}\}$ for all $\omega \in \Omega$, there exists $K'' \subset K'$ 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 [12, Chapter 3], is central to the proof of global convergence.

LEMMA 3.4. *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 3.3. □

The establishment of the global convergence of Algorithm 2.1 employs a contradiction argument inspired from [12, 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 3.4) (with a uniform lower bound ϵ_1 for *all* subsequences over which $\{x_k\}$ converges to x^*). It is shown below (Lemma 3.5) that in such case $|v_{k+1}|$ is significantly smaller than $|v_k|$. Since, in view of Lemma 3.1, $\{x_{k+1}\}$ 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_1 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 [12, Chapter 3] (see also the proof of Lemma 3.14 in [29]) and is given in Appendix 2. It relies crucially on the assumption that $H_{k+1}^{-1}H_k \to I$ (the identity matrix) when $t_k \to 0$, which is ensured in Algorithm 2.1 by setting $H_{k+1} = H_k$ when t_k is small.

LEMMA 3.5. *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 integer i_0, there exists an integer N such that*

$$|v_{k+i+1}| \leq |v_{k+i}| - c|v_{k+i}|^2 \quad \forall k \geq N, k \in K, \forall i \in [0, i_0]. \tag{3.3}$$

□

LEMMA 3.6. *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*

$$0 \leq z_{i+1} \leq z_i - \eta z_i^2 \quad \forall i \in I\!N,$$

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

Proof. See Appendix 2. □

We are now ready to establish the global convergence of Algorithm 2.1.

THEOREM 3.7. *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. Proceeding by contradiction, we assume x^* is not a KKT point. It follows from Lemma 3.4 that there exists $\epsilon_1 > 0$ such that

$$\liminf_{k \to \infty} \; max\{|v_k|, \; \|x_k - x^*\|\} > \epsilon_1. \tag{3.4}$$

Thus, $\{v_k\}$ is bounded away from zero on K. In view of Lemma 3.3, $\{d_k\}$ is also bounded away from zero on K and, in view of Lemma 3.1(iii), $\{t_k\}$ converges to zero on K. Let c be as given by Lemma 3.5. Let i_0 be as given in Lemma 3.6 with $\epsilon = \epsilon_1$ and $\eta = c$. In view of Lemma 3.5 and since $v_k \leq 0$ for all k, 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 3.6 with $z_i = |v_{k+i}|$, $k \in K$, that

$$|v_{k+i_0}| < \epsilon_1 \quad \forall k \geq N, k \in K.$$

On the other hand, since by assumption x^* is not a KKT point and since in view of Lemma 3.1(iii) $\{x_{k+i_0}\}$ also converges to x^* on K, it follows from (3.4) that

$$\liminf_{k \in K, k \to \infty} |v_{k+i_0}| > \epsilon_1,$$

a contradiction. □

3.2. Local convergence. Under additional regularity conditions, it is shown that, close to a strong local minimizer, Ω_k remains constant so that Algorithm 2.1 becomes a standard algorithm for finite minimax problems. Further, it is shown that H_k will be updated normally, thus will not be prevented from asymptotically suitably approximating the Hessian of the Lagrangian. If H_k does become a suitable approximation 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

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 μ_ω, $\omega \in \Omega_{max}(x^*)$, satisfying

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

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., $\mu^*_\omega > 0$ for all $\omega \in \Omega_{max}(x^*)$ and

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

with

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

and

$$\mathcal{S}^* = \{h : \langle h, \nabla_x\phi(x^*,\omega_i)\rangle = \langle h, \nabla_x\phi(x^*,\omega_j)\rangle, \ \ \forall \omega_i,\omega_j \in \Omega_{max}(x^*)\}.$$

PROPOSITION 3.8. *The entire sequence $\{x_k\}$ converges to x^*.*

Proof. Assumption 5 implies that x^* is an isolated KKT point. The claim then follows from Lemma 3.1(iii) and Theorem 3.7. □

Much of the remainder of this section is devoted to showing that, close to x^*, Ω_k does not change from iteration to iteration, *i.e.*, Algorithm 2.1 becomes identical to a standard SQP algorithm for a minimax problem of the type

$$\min_{x\in R^n} \ \max_{\omega\in\hat{\Omega}} \ \phi(x,\omega)$$

where $\hat{\Omega}$ is a subset of Ω (in fact, $\hat{\Omega} = \Omega_{max}(x^*)$). A consequence of this is that rate of convergence results obtained for such algorithms hold for Algorithm 2.1. The first step is to show that, for k large enough, $\Omega_{max}(x^*) \subset \Omega_k$ (Lemma 3.12). This is first proved on a subsequence (Lemma 3.10), using the following result.

LEMMA 3.9. *There exists an infinite index set K on which $\{d_k\}$ converges to zero.*

Proof. Suppose by contradiction that $\{d_k\}$ is bounded away from zero. Then, in view of the definition of v_k, of Lemma 3.1(ii) and of Assumptions 1 and 3, there exist $\underline{v}$ and $\overline{v}$ such that $0 < \underline{v} \le |v_k| \le \overline{v}$. An obvious modification of the proof of Lemma 3.5 can be used to show that there exists $c > 0$ such that, given any i_0, there exists an integer N such that

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

In particular, with $i_0 > (\overline{v} - \underline{v})/c\underline{v}^2$, there exist $N \in \mathbb{N}$ such that, for all $i \in [0, i_0]$,

$$\begin{aligned} |v_{N+i+1}| &\le |v_{N+i}| - c|v_{N+i}|^2 \\ &\le |v_{N+i}| - c\underline{v}^2 \\ &\le |v_N| - ci\underline{v}^2 \\ &\le \overline{v} - ci\underline{v}^2. \end{aligned}$$

Thus, $|v_{N+i_0+1}| < \underline{v}$, a contradiction. □

LEMMA 3.10. *There exists an infinite index set K such that*

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

Proof. Let K be an infinite index set such that $\{d_k\}$ converges to zero on K (in view of Lemma 3.9, such K exists). In view of the finite cardinality of Ω, without loss of generality, we assume $\Omega_k = \hat{\Omega}$ 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 $QP(x_k, H_k, \hat{\Omega})$ and $\mu_{k,\omega} = 0$, $\omega \in \Omega \backslash \hat{\Omega}$. 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 3.8, since $\{d_k\}$ converges to zero on K, taking limits in the optimality condition (2.2) associated with $QP(x_k, H_k, \hat{\Omega})$, $k \in K$, yields, since $\hat{\mu}_\omega = 0$ for all $\omega \in \Omega \backslash \hat{\Omega}$,

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

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

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

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 $\phi(x^*, \omega) = \psi(x^*)$, *i.e.*, (3.5) holds. □

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

LEMMA 3.11. *Let K be an infinite index set such that $\Omega_{max}(x^*) \subset \Omega_k$ for all $k \in K$. Then, $\{d_k\}$ converges to zero on K.*

Proof. Given $\hat{\Omega} \subset \Omega$, let $K_{\hat{\Omega}} = \{k \in K : \Omega_k = \hat{\Omega}\}$. For any $\hat{\Omega} \subset \Omega$ 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 $\hat{K} \subset K_{\hat{\Omega}}$, $\{d_k\}$ is bounded away from zero on $\hat{K}$ and let $\tilde{K} \subset \hat{K}$ be such that $\{H_k\}$ converges to H^* on $\tilde{K}$ for some $H^* > 0$ (such $\tilde{K}$ exists in view of Assumption 3). Since $QP(x^*, H^*, \hat{\Omega})$ has $d = 0$ as unique solution, it follows from Theorem 2.1 in [25] that $d_k \to 0$ as $k \to \infty$, $k \in \tilde{K}$, contradicting the fact that $\|d_k\|$ is bounded from below on $\hat{K}$. □

LEMMA 3.12. *For k large enough, $\Omega_{max}(x^*) \subset \Omega_k$.*

Proof. In view of Lemma 3.10, the claim holds on an infinite subsequence. To complete the proof, we show that given any infinite index set K such that $\Omega_{max}(x^*) \subset \Omega_k$ for all $k \in K$, it holds that $\Omega_{max}(x^*) \subset \Omega_{k+1}$ for all $k \in K$, k large enough. In view of the construction of Ω_{k+1}, it is enough to show that $\mu_{k,\omega} > 0$ for all $\omega \in \Omega_{max}(x^*)$,

$k \in K$, k large enough, where $\mu_{k,\omega}$, $\omega \in \Omega_k$, are the KKT mutipliers associated with $QP(x_k, H_k, \Omega_k)$. Thus let K be an infinite index set such that $\Omega_{max}(x^*) \subset \Omega_k$ for all $k \in K$. Lemma 3.11 implies $\{d_k\}$ converges to zero on K. Suppose by contradiction that there exists $\omega^* \in \Omega_{max}(x^*)$ and an infinite index set $K' \subset K$ such that $\mu_{k,\omega^*} = 0$ 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 3.10 shows that, in view of Assumption 4, $\mu^*_{\omega^*} = 0$, contradicting strict complementarity (Assumption 5). □

The following result directly follows from Lemmas 3.11 and 3.12.

LEMMA 3.13. *The entire sequence $\{d_k\}$ converges to zero.* □

This leads to the main result of this section

PROPOSITION 3.14. *For k large enough,*

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

Proof. In view of Proposition 3.8 and Lemma 3.13, it holds that, for k large enough,

$$\begin{aligned}\Omega_k^b &= \{\omega \in \Omega_k : \mu_{k,\omega} > 0\} \\ &\subset \{\omega \in \Omega_k : \phi(x_k, \omega) + \langle \nabla_x \phi(x_k, \omega), d_k \rangle = \max_{\omega \in \Omega_k} \{\phi(x_k, \omega) + \langle \nabla_x \phi(x_k, \omega), d_k \rangle\} \\ &\subset \Omega_{max}(x^*).\end{aligned}$$

Moreover, in view of Proposition 3.8 and Assumption 1', for k large enough,

$$\Omega_{max}(x_k) \subset \Omega_{max}(x^*),$$

and, in view of Lemma 3.1(iii), whenever $t_k < 1$, for k large enough,

$$\Omega_{max}(\bar{x}_k) \subset \Omega_{max}(x^*).$$

Thus, from the construction of Ω_{k+1} in Algorithm 2.1, $\Omega_k \subset \Omega_{max}(x^*)$ for k large enough. This together with Lemma 3.12 yields the claim. □

Thus, for k large enough, the behavior of Algorithm 2.1 is identical to that of standard SQP algorithms for minimax problems. However if superlinear convergence is to take place, it is essential that the Hessian matrix H_k be normally updated for k large enough, *i.e.*, in view of Step 2 in Algorithm 2.1, that $t_k > min\{\delta, \|d_k\|\}$ for k large enough. Since $\{d_k\}$ goes to zero as k goes to infinity, this will occur if t_k is bounded away from zero. This is indeed the case, as shown next.

LEMMA 3.15. *There exists $\underline{t} > 0$ such that, for k large enough, $t_k \geq \underline{t}$.*

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,

$$\phi(x_k + td_k, \omega) \leq \phi(x_k, \omega) + c_1 t \|d_k\|. \tag{3.6}$$

and

$$\phi(x_k + td_k, \omega) \leq \phi(x_k, \omega) + t\langle \nabla_x \phi(x_k, \omega), d_k \rangle + c_2 t^2 \|d_k\|^2.$$

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,

$$\begin{aligned}
\phi(x_k+td_k,\omega) & \\
&\leq (1-t)\phi(x_k,\omega) + t\{\phi(x_k,\omega) + \langle \nabla_x \phi(x_k,\omega), d_k\rangle\} + c_2 t^2 \|d_k\|^2 \\
&\leq (1-t)\phi(x_k,\omega) + t \max_{\omega\in\Omega_k}\{\phi(x_k,\omega) + \langle \nabla_x \phi(x_k,\omega), d_k\rangle\} + c_2 t^2 \|d_k\|^2 \\
&= (1-t)\phi(x_k,\omega) + t \sum_{\omega\in\Omega_k} \mu_{k,\omega}\{\phi(x_k,\omega) + \langle \nabla_x \phi(x_k,\omega), d_k\rangle\} + c_2 t^2 \|d_k\|^2 \\
&\leq (1-t)\psi(x_k) + t\psi(x_k)\sum_{\omega\in\Omega_k}\mu_{k,\omega} + t\sum_{\omega\in\Omega_k}\mu_{k,\omega}\langle \nabla_x \phi(x_k,\omega), d_k\rangle + c_2 t^2 \|d_k\|^2 \\
&= \psi(x_k) - t\langle d_k, H_k d_k\rangle + c_2 t^2\|d_k\|^2,
\end{aligned}$$

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 $\alpha \in (0, 1/2)$,

$$\begin{aligned}
\phi(x_k + td_k,\omega) &\leq \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle + t(\alpha-1)\langle d_k, H_k d_k\rangle + c_2 t^2\|d_k\|^2 \\
&\leq \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle + t\|d_k\|^2\{(\alpha-1)\sigma_1 + c_2 t\} \\
&\leq \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle, \quad \forall t \in [0,\hat{t}],
\end{aligned}$$

with $\hat{t} = (1-\alpha)\sigma_1/c_2 > 0$. On the other hand, since in view of Proposition 3.14 $\phi(x^*,\omega) < \psi(x^*)$ for all $\omega \in \Omega\backslash\Omega_k$ for k large enough, it follows from Propostion 3.8 that there exists $\epsilon > 0$ such that, for k large enough,

$$\phi(x_k,\omega) \leq \psi(x_k) - \epsilon, \quad \forall \omega \in \Omega\backslash\Omega_k. \tag{3.7}$$

Combining (3.6) and (3.7) yields that, for all $\omega \in \Omega\backslash\Omega_k$, all $t \geq 0$ and all k large enough

$$\phi(x_k + td_k,\omega) \leq \psi(x_k) - \epsilon + c_1 t\|d_k\|.$$

Since d_k converges to zero (Lemma 3.13), there exists $\tilde{t} > 0$, such that, for k large enough,

$$\phi(x_k + td_k,\omega) \leq \psi(x_k) - \alpha t\langle d_k, H_k d_k\rangle, \quad \forall t \in [0,\tilde{t}].$$

The claim follows readily. □

Thus H_k is eventually updated at every iteration and the local behavior of Algorithm 2.1 becomes identical to that of the algorithm in [30].

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

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

where the matrices P_k are defined by

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

with $R_k = [\nabla_x\phi(x,\omega_i) - \nabla_x\phi(x,\omega_1) : i = 2,\ldots,s]$,† $\omega_1,\ldots,\omega_s$ being the elements of $\Omega_{max}(x^*)$; and suppose moreover that $t_k = 1$ for k large enough. Then (see [30]), the convergence rate is two-step superlinear, *i.e.*,

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

†Note that P_k remain invariant if in the definition of R_k the role of ω_1 is played by any other ω_i.

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 the "nonmonotone line search" used in [29], [30]. It is a straightforward if tedious exercise to verify that such a scheme does not affect any of the results obtained in this section.

4. Implementation, extention and numerical results. In Algorithm 2.1, the search direction is computed based on, in some sense, the smallest subset Ω_k of Ω such that global convergence is guaranteed. In practice, faster initial convergence will often take place if additional "potentially critical" elements of Ω are included. Yet clearly, increasing the size of Ω_k will make $QP(x_k, H_k, \Omega_k)$ more complex to solve. Various heuristics come to mind (see, *e.g.*, [26]). For the frequent case where "adjacent" constraints are closely related, we follow the idea used in [7], [15], and 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 $\omega \in \Omega$ is ϵ-active if it belongs to

$$\Omega_\epsilon(x) = \{\omega \in \Omega : \phi(x,\omega) > \psi(x) - \epsilon\}$$

and it is a left local maximizer of ϕ over Ω at x if one of the following three conditions holds: (i) $\omega \in (0,1)$ and

$$\phi(x,\omega) > \phi(x, \omega - \frac{1}{q}) \tag{4.1}$$

and

$$\phi(x,\omega) \geq \phi(x, \omega + \frac{1}{q}); \tag{4.2}$$

(ii) $\omega = 0$ and (4.2); (iii) $\omega = 1$ and (4.1).

There is no conceptual difficulty in modifying Algorithm 2.1 so as to handle problem (SI_d) (or (MC)). Our numerical experiments were carried out on (SI_d)-type problems with such a modified algorithm incorporating the "feasible" SQP idea of [16]. A precise statement of this algorithm is given in Appendix 1 (Algorithm 6.1) and a complete analysis can be found in [29]. We only found a small number of SIP problems in the literature and, due to the insufficient reported information, comparison of our algorithm with existing algorithms is difficult. Thus, Table 1 and Table 2 only contain results obtained with Algorithm 6.1 (`NEW`) and with an algorithm identical to it except that $\Omega_k \equiv \Omega$ for all k (`OLD`). The test problems are discretized versions of Examples 2-6 from [27] (denoted by `EXPL2`, ..., `EXPL6` respectively) and of a problem from [19] (denoted by `P-M`). In the tables, n stands for the number of variables (n), `NF` for the number of evaluations of objective function f, `NC` for the number of evaluation of constraint ψ_Ω, `IT` for the total number of iterations and $|\Omega^*|$ for the number of points in $\Omega_k = \Omega^*$ at the stopping point $x_k = x^*$. `TIME` indicates the execution time in seconds on a SUN/SPARC 1 workstation, `OBJECTIVE` the value of the objective function at x^*, and ϵ_s is the stopping parameter (the algorithm is terminated as soon as $\|d^*\| \leq \epsilon_s$). Each problem is tested starting from a hand-picked feasible point: $x_0 = (1,2)^T$ for `EXPL2`, $x_0 = (-100,1,1)^T$ for `EXPL3`, $x_0 = (5,\ldots,5)^T$ for `EXPL4`, $x_0 = (1,0.5,0)^T$ for `EXPL5` and $x_0 = (0.5,-2)^T$ for `EXPL6`; `P-M` is solved using the same initial point as in [7].

It can be observed that $|\Omega^*|$ is much smaller than $|\Omega|$ for all problems. Also, comparison of Tables 1 and 2 shows that a finer discretization does not always result in a larger Ω^* and in most cases $|\Omega^*|$ remains very small.

PROB	n	ALGO	NF	NC	IT	$\|\Omega^*\|$	TIME	OBJECTIVE	$\|\|d^*\|\|$	ϵ_s
EXPL2	2	NEW	5	11	4	4	0.28	2.6180660	2E-05	1E-04
		OLD	4	4	3		0.50	2.6180860	3E-05	1E-04
EXPL3	3	NEW	14	27	13	2	0.74	5.3346873	5E-06	1E-04
		OLD	11	15	10		1.54	5.3346873	1E-05	1E-04
EXPL4	3	NEW	13	34	13	4	0.75	0.6490311	7E-10	1E-04
		OLD	11	11	10		1.98	0.6490311	4E-08	1E-04
EXPL4	6	NEW	23	49	18	5	3.70	0.6168213	9E-04	1E-02
		OLD	18	18	17		7.12	0.6169191	9E-04	1E-02
EXPL4	8	NEW	23	87	23	16	8.05	0.6165831	8E-03	2E-02
		OLD	14	14	14		15.48	0.6167966	2E-03	2E-02
EXPL5	3	NEW	5	7	4	4	0.27	4.3011578	2E-10	1E-04
		OLD	4	4	3		0.80	4.3012897	5E-05	1E-04
EXPL6	2	NEW	8	46	7	1	0.70	97.159034	1E-05	1E-04
		OLD	8	47	7		1.20	97.159034	1E-05	1E-04
P-M	3	NEW	59	69	50	8	6.78	0.1745942	4E-05	1E-04
		OLD	33	33	32		13.22	0.1846945	3E-05	1E-04

Table 1: Numerical Results with Discretization $|\Omega| = 100$

PROB	n	ALGO	NF	NC	IT	$\|\Omega^*\|$	TIME	OBJECTIVE	$\|\|d^*\|\|$	ϵ_s
EXPL2	2	NEW	9	13	5	2	0.73	2.4305360	3E-05	1E-04
		OLD	4	4	3		2.95	2.6189860	3E-05	1E-04
EXPL3	3	NEW	14	28	14	2	3.00	5.3346873	1E-05	1E-04
		OLD	10	15	9		6.44	5.3346873	1E-05	1E-04
EXPL4	3	NEW	12	33	11	4	2.55	0.6490417	2E-09	1E-04
		OLD	11	11	10		9.62	0.6490417	4E-06	1E-04
EXPL4	6	NEW	17	46	16	24	13.22	0.6168140	9E-03	1E-02
		OLD	14	14	13		35.26	0.6169080	9E-04	1E-02
EXPL4	8	NEW	28	158	29	9	56.70	0.6163828	5E-03	2E-02
		OLD	18	18	17		77.85	0.6167391	2E-03	2E-02
EXPL5	3	NEW	5	6	4	3	0.66	4.3011838	4E-05	1E-04
		OLD	5	5	4		4.75	4.3011838	6E-06	1E-04
EXPL6	2	NEW	9	46	7	1	3.00	97.159034	1E-05	1E-04
		OLD	8	27	7		5.45	97.159034	1E-05	1E-04
P-M	3	NEW	52	54	39	27	21.30	0.1746273	1E-05	1E-04
		OLD	50	50	39		83.50	0.1746567	5E-05	1E-04

Table 2: Numerical Results with Discretization $|\Omega| = 500$

5. Conclusion. An SQP-type algorithm has been proposed and analyzed for the solution of optimization problems with many more constraints than variables, in particular, of finely discretized semi-infinite programming problems. At each iteration, a quadratic problem involving only a small set of constraints is solved and correspondingly, only a few constraint gradients are evaluated. Numerical results indicate that the proposed algorithm is very efficient.

There is no conceptual difficulty in extending the algorithm to the solution of problems with more than one free independent parameters ranging over an arbitrary compact set. The proposed algorithm, with appropriate modifications, has been implemented in an optimization-based engineering system design tool [5] and has proven very successful in solving various types of engineering design problems.

6. Appendix 1. In this appendix, the algorithm for problem (SI_d) that was used to obtain the numerical results of § 5 is precisely described. This algorithm enforces feasibility of all iterates and avoids the Maratos effect.

At iteration k, a direction d_k^0 is first computed by solving the quadratic program

$$(QP_0(x_k, H_k, \Omega_k)) \quad \min_{d^0 \in R^n} \quad \frac{1}{2}\langle d^0, H_k d^0\rangle + \langle \nabla f(x_k), d^0\rangle$$
$$\text{s.t.} \quad \phi(x_k,\omega) + \langle \nabla_x \phi(x_k,\omega), d^0\rangle \le 0, \quad \forall \omega \in \Omega_k.$$

Then, a first order feasible descent direction d^1 is obtained as the solution of the quadratic program

$$(QP_1(x, d^0, \hat{\Omega})) \quad \min_{d^1 \in R^n, \gamma \in R} \quad \frac{\eta}{2}\langle d_k^0 - d^1, d_k^0 - d^1\rangle + \gamma$$
$$\text{s.t.} \quad \langle \nabla f(x_k), d^1\rangle \le \gamma$$
$$\phi(x_k,\omega) + \langle \nabla_x \phi(x_k,\omega), d^1\rangle \le \gamma, \quad \forall \omega \in \Omega_k,$$

where η is a positive constant, and the search direction d is defined as a certain convex combination of d^0 and d^1. A second order correction $\tilde{d}$ is then obtained as the solution of the quadratic program

$$(\widetilde{QP}(x_k, H_k, d_k, \Omega_k)) \quad \min_{\tilde{d} \in R^n} \quad \frac{1}{2}\langle d_k + \tilde{d}, H_k(d_k + \tilde{d})\rangle + \langle \nabla f(x_k), d_k + \tilde{d}\rangle$$
$$\text{s.t.} \quad \phi(x_k + d_k, \omega) + \langle \nabla_x \phi(x_k,\omega), \tilde{d}\rangle \le$$
$$- \, min(\nu\|d_k\|, \; \|d_k\|^{\tau_2}), \; \forall \omega \in \Omega_k,$$

where ν and τ_2 are positive constant, but is discarded if $\widetilde{QP}(x_k, H_k, d_k, \Omega_k)$ has no solution or if $\|\tilde{d}\|$ is too large. An arc search is performed along $x_k + td_k + t^2\tilde{d}_k$. Finally, $\Omega_{max}(x)$ and $\Omega_\epsilon(x)$ are defined as before except that ψ in the definitions are replaced by zero and Ω_k^b is defined using multipliers associated with constraints in $QP_0(x_k, H_k, \Omega_k)$. Finally, H_k is updated using the BFGS formula with Powell's modification [22]. The complete algorithm is as follows.

Algorithm 6.1.

Parameters. $\alpha \in (0, \frac{1}{2})$, $\beta \in (0,1)$, $\eta = 0.1$, $\nu = 0.01$, $\kappa = 2.1$, $\tau_1 = \tau_2 = 2.5$, $\epsilon = 1$.

Data. $x_0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n\times n}$ and $H_0 = H_0^T > 0$.

Step 0. Initialization. Set $k = 0$ and $\Omega_0 = \Omega_{max}(x_0)$.

Step 1. Computation of search direction and step length.

(i). Compute d_k^0 by solving $QP_0(x_k, H_k, \Omega_k)$. If $\|d_k^0\| = 0$, stop.

(ii). Compute d_k^1 by solving $QP_1(x_k, d_k^0, \Omega_k)$.

(iii). Set $d_k = (1-\rho_k)d_k^0 + \rho_k d_k^1$ with $\rho_k = \|d_k^0\|^\kappa/(\|d_k^0\|^\kappa + v_k)$ and $v_k = max(0.5, \; \|d_k^1\|^{\tau_1})$.

(iv). Compute $\tilde{d}_k$ by solving $\widetilde{QP}(x_k, d_k, H_k, \Omega_k)$. If there is no solution or if $\|\tilde{d}_k\| > \|d_k\|$, set $\tilde{d}_k = 0$.

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

$$\phi(x_k + td_k + t^2\tilde{d}_k, \omega) \le 0, \quad \forall \omega \in \Omega.$$

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

Step 2. Updates. Set $x_{k+1} = x_k + t_k d_k + t_k^2 \tilde{d}_k$. If $t_k < 1$, set

$$\bar{x}_{k+1} = x_k + \frac{t_k}{\beta} d_k + \left(\frac{t_k}{\beta}\right)^2 \tilde{d}_k.$$

Set

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

Compute a new positive definite approximation H_{k+1} to the Hessian of the Lagrangian of (P). Set $k = k+1$. Go back to *Step 1.*

□

7. Appendix 2.

7.1. Proof of Lemma 3.5.

More notations are defined:

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

With this notations, it can be observed that, in view of Lemma 3.1(iii),

$$\lim_{k \to \infty} |\pi_k - \pi_{k+}| = 0 \tag{7.1}$$

and, if $\{H_{k+1}^{-1} H_k\} \to I$ on an infinite index set $K \subset \mathbb{N}$,

$$\lim_{\substack{k \in K \\ k \to \infty}} \|p_k - p_{k+}\| = 0. \tag{7.2}$$

A few more lemmas are first established to facilitate the proof of Lemma 3.5.

LEMMA 7.1. *Let K be an infinite index set such that $\{H_{k+1}^{-1} H_k\} \to I$ on K and $\{d_k\}$ is uniformly bounded away from zero on K, then, given any $\tilde{\alpha} > \alpha$,*

$$\phi(x_{k+1}, \bar{\omega}) + \langle \nabla_x \phi(x_{k+1}, \bar{\omega}), H_{k+1}^{-\frac{1}{2}} H_k^{\frac{1}{2}} d_k \rangle - \psi(x_k) \geq -\tilde{\alpha} \langle d_k, H_k d_k \rangle, \tag{7.3}$$

for $k \in K$, k large enough, and for all $\bar{\omega} \in \Omega_{max}(\bar{x}_{k+1})$.

Proof. Proceeding by contradiction, suppose (7.3) is not true, *i.e.*, there exists $K' \subset K$ such that, for all $k \in K'$ there exists $\bar{\omega}_k \in \Omega_{max}(\bar{x}_{k+1})$, such that

$$\phi(x_{k+1}, \bar{\omega}_k) + \langle \nabla_x \phi(x_{k+1}, \bar{\omega}_k), H_{k+1}^{-\frac{1}{2}} H_k^{\frac{1}{2}} d_k \rangle - \psi(x_k) < -\tilde{\alpha} \langle d_k, H_k d_k \rangle. \tag{7.4}$$

In view of Lemma 3.1(iii), $\{t_k\}$ converges to zero on K and, without loss of generality, we may assume $t_k < 1$ for all $k \in K'$. Clearly, there exists a subset $K'' \subset K'$ 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 3.1(iii), $\{x_{k+1}\}$ also converges to x^* on K''. Furthermore, since $\bar{\omega}_k \in \Omega_{max}(\bar{x}_{k+1})$ for all $k \in K''$, it follows from the continuity of ϕ and ψ (Assumption 1) that $\omega^* \in \Omega_{max}(x^*)$. Thus, since $\{H_{k+1}^{-\frac{1}{2}} H_k^{\frac{1}{2}}\} \to I$ on K, taking limit of (7.4) on K'' yields

$$(7.5) \qquad \langle \nabla_x \phi(x^*, \omega^*), d^* \rangle \le -\tilde{\alpha} \langle d^*, H^* d^* \rangle < -\alpha \langle d^*, H^* d^* \rangle.$$

On the other hand, since $\bar{\omega}_k \in \Omega_{max}(\bar{x}_{k+1})$ and $t_k < 1$, it follows from the line search rule that

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

Thus,

$$(7.6) \qquad \frac{\phi(x_k + \frac{t_k}{\beta} d_k, \bar{\omega}_k) - \phi(x_k, \bar{\omega}_k)}{\frac{t_k}{\beta}} > -\alpha \langle d_k, H_k d_k \rangle.$$

Taking limit of (7.6) as $k \to \infty$ on K'' yields

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

which contradicts (7.5). □

As in [12, Chapter 3], using the dual of $QP(x_k, H_k, \Omega_k)$, denoted by $\overline{QP}(x_k, H_k, \Omega_k)$, facilitates the analysis. Let $\gamma(\omega) = \psi(x) - \phi(x, \omega)$ and $g(\omega) = H^{-\frac{1}{2}} \nabla_x \phi(x, \omega)$.

LEMMA 7.2. *Given any $x \in R^n$, $H = H^T > 0$, and $\hat{\Omega} \subset \Omega$, the dual quadratic program $\overline{QP}(x, H, \hat{\Omega})$ of $QP(x, H, \hat{\Omega})$ is given by*

$$(\overline{QP}(x, H, \hat{\Omega})) \qquad \begin{aligned} &\max_{\mu \in R^{|\hat{\Omega}|}} \quad -\left(\frac{1}{2} \| \sum_{\omega \in \hat{\Omega}} \mu_\omega g(\omega) \|^2 + \sum_{\omega \in \hat{\Omega}} \mu_\omega \gamma(\omega) \right) \\ &\text{s.t.} \qquad \mu_\omega \ge 0 \ \forall \omega \in \hat{\Omega} \ \text{ and } \sum_{\omega \in \hat{\Omega}} \mu_\omega = 1, \end{aligned}$$

where $\{\mu_\omega, \omega \in \hat{\Omega}\}$ are the components of μ.

Proof. The dual is given by

$$max \ \ \phi(\mu) \quad \text{s.t. } \mu \in U,$$

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

$$(7.7) \qquad \phi(\mu) = \min_d \left\{ \frac{1}{2} \langle d, Hd \rangle + \sum_{\omega \in \hat{\Omega}} \mu_\omega (\phi(x, \omega) + \langle \nabla_x \phi(x, \omega), d \rangle) - \psi(x) \right\}$$

with

$$U = \{ \mu \in \mathbb{R}^{|\hat{\Omega}|} : \sum_{\omega \in \hat{\Omega}} \mu_\omega = 1 \ \text{ and } \ \mu_\omega \ge 0 \ \forall \omega \in \hat{\Omega} \}.$$

In view of Assumption 3, the unique minimizer d^* in (7.7) 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

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

Therefore,

$$\begin{aligned}
\phi(\mu) &= -\frac{1}{2}\langle d^*, H d^* \rangle - \sum_{\omega \in \hat{\Omega}} \mu_\omega \{\psi(x) - \phi(x,\omega)\} \\
&= -\left(\frac{1}{2} \| \sum_{\omega \in \hat{\Omega}} \mu_\omega g(\omega) \|^2 + \sum_{\omega \in \hat{\Omega}} \mu_\omega \gamma(\omega) \right)
\end{aligned}$$

and the result follows. □

LEMMA 7.3. *The sequences $\{|v_k|\}$, $\{\|p_{k+}\|\}$ and $\{\|g_{k+1}(\omega)\|\}$ are all bounded.*

Proof. Follows directly from Assumptions 1 — 3 and the boundedness of μ_k and (7.8). □

Proof of Lemma 3.5. We first prove (3.3) for $i = 0$. Let $\bar{\omega}_k \in \Omega_{max}(\bar{x}_{k+1})$ and define $\Omega' = \Omega_k^b \cup \{\bar{\omega}_k\}$. Let v'_{k+1} denote the optimal value of $QP(x_{k+1}, H_{k+1}, \Omega')$. In view of the construction of Ω_{k+1}, $\Omega' \subset \Omega_{k+1}$. Thus, $|v_{k+1}| \leq |v'_{k+1}|$. Therefore, it suffices to prove (3.3) (with $i = 0$) with the left hand side replaced by $|v'_{k+1}|$.

We define the quadratic function in ν

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

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

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

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

$$\begin{aligned}
Q(\nu) &= \frac{1}{2}\nu^2 \|g_{k+1}(\bar{\omega}_k)\|^2 + \frac{1}{2}(1-\nu)^2 \|p_{k+}\|^2 + \nu(1-\nu)\langle g_{k+1}((\bar{\omega})_k, p_{k+}\rangle \\
&\quad + \nu \gamma_{k+1}(\bar{\omega}_k) + (1-\nu)\pi_{k+} \\
&= \frac{1}{2}\|p_{k+}\|^2 + \frac{\nu^2}{2}\|g_{k+1}(\bar{\omega}_k) - p_{k+}\|^2 + \nu\langle g_{k+1}(\bar{\omega}_k), p_{k+}\rangle - \nu\|p_{k+}\|^2 \\
&\quad + \nu \gamma_{k+1}(\bar{\omega}_k) + (1-\nu)\pi_{k+}.
\end{aligned} \tag{7.8}$$

Note that

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

Since $\{x_k\}$ is bounded away from KKT points on K, it follows from Lemma 3.3 that both $\{v_k\}$ and $\{d_k\}$ are bounded away from zero on K and, from Lemma 3.1(iii), $\{t_k\}$ converges to zero on K. Thus $t_k < min\{\delta, \ \|d_k\|\}$ for $k \in K$, k large enough and, from *Step 2* in Algorithm 2.1, $H_{k+1} = H_k$. Thus, assumptions of Lemma 7.1 are all satisfied. Picking $\tilde{\alpha} \in (\alpha, 1/2)$, in view of Lemma 7.1 with $\bar{\omega} = \bar{\omega}_k$ and since $\psi(x_{k+1}) \le \psi(x_k)$, there exists an integer k_1 such that, for all $k \ge k_1$, $k \in K$, it holds

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

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

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

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

$$\begin{aligned}\langle g_{k+1}(\bar{\omega}_k), p_{k+}\rangle &= \langle g_{k+1}(\bar{\omega}_k), p_k\rangle - \langle g_{k+1}(\bar{\omega}_k), p_k - p_{k+}\rangle \\ &\le -\gamma_{k+1}(\bar{\omega}_k) - 2\tilde{\alpha} v_k - \langle g_{k+1}(\bar{\omega}_k), p_k - p_{k+}\rangle.\end{aligned}$$

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, in view of Lemma 7.3, there exists $M > 1$ such that

$$max\{|v_k|, \|p_{k+}\|, \|g_{k+1}(\bar{\omega}_k)\|\} \le M, \quad \forall k,$$

yielding

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

Substituting all these into (7.8) yields

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

In view of Lemma 7.2 and duality,

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

Thus,

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

The minimum of the right hand side is achieved at $\bar{\nu}_k = \frac{(1-2\tilde{\alpha})}{4M^2}|v_k|$. Since $\tilde{\alpha} < 1/2$, $|v_k| \le M$ and $M > 1$, it follows that $\bar{\nu}_k \in [0,1]$ and thus

$$\min_{\nu\in[0,1]} Q(\nu) \le Q(\bar{\nu}_k) \le |v_k| - \frac{(1-2\tilde{\alpha})^2}{8M^2}|v_k|^2 + O(|\pi_k - \pi_{k+}|) + O(||p_k - p_{k+}||)$$

and, in view of (7.1), (7.2) and the fact that v_k is bounded away from zero (Lemma 3.3(i)), there exists an integer N_0 such that, for all $k \ge N_0$, $k \in K$

$$\min_{\nu\in[0,1]} Q(\nu) \le |v_k| - \frac{(1-2\tilde{\alpha})^2}{16M^2}|v_k|^2.$$

Therefore, by letting $c = \frac{(1-2\tilde{\alpha})^2}{16M^2}$, (3.3) with $i = 0$ follows from the inequality

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

Now for any i, in view of Lemma 3.1(iii), $\{x_{k+i}\}$ is also bounded away from KKT points for $k \in K$ and this in turn implies, in view of Lemma 3.3, that $\{v_{k+i}\}$ is bounded away from zero for $k \in K$. Therefore, from above argument, there exits N_i for each given i such that (3.3) holds for all $k \ge N_i$, $k \in K$. Choosing $N = max_{0\le i\le i_0}\{N_i\}$ yields that (3.3) holds for all $i \in [0, i_0]$ and for all $k \ge N$, $k \in K$. □

7.2. Proof of Lemma 3.6. $z_0 - \eta z_0^2$ 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 $\{y_i\}$,

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

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

REFERENCES

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

[2] 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.

[3] I. D. COOPE AND G. A. WATSON, *A Projected Lagrangian Algorithm for Semi-Infinite Programming*, Math. Programming, 32 (1985), pp. 337–356.

[4] H. A. ESCHENAUER, C. MATTHECK AND N. OLHOFF, eds., *Engineering Optimization in Design Processes,*, Springer-Verlag, Berlin, 1991.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

[16] ———, *On Combining Feasibility, Descent and Superlinear Convergence in Inequality Constrained Optimization*, Math. Programming, 1993, to appear.

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

[18] 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.

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

[20] 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.

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

[22] 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.

[23] J. RAKOWSKA, R. T. HAFTKA AND L. T. WATSON, *Multi-Objective Control-Structure Optimization Via Homotopy Methods*, Department of Computer Science, Technical Report TR-90-63 (to appear in SIAM Journal of Optimization), Virginia Polytechnic Institute and State University.

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

[25] 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.

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

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

[28] 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.

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

[30] J. L. ZHOU AND A. L. TITS, *Nonmonotone Line Search for Minimax Problems*, J. Optim. Theory Appl., 76 (March 1993, to appear).