1

FEASIBLE SEQUENTIAL QUADRATIC PROGRAMMING FOR FINELY DISCRETIZED PROBLEMS FROM SIP

Craig T. Lawrence and André L. Tits

Department of Electrical Engineering
and Institute for Systems Research
University of Maryland, College Park
College Park, Maryland 20742
USA

ABSTRACT

A Sequential Quadratic Programming algorithm designed to efficiently solve nonlinear optimization problems with many inequality constraints, e.g. problems arising from finely discretized Semi-Infinite Programming, is described and analyzed. The key features of the algorithm are *(i)* that only a few of the constraints are used in the QP sub-problems at each iteration, and *(ii)* that every iterate satisfies all constraints.

1 INTRODUCTION

Consider the Semi-Infinite Programming (SIP) problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \Phi(x) \leq 0, \end{array} \qquad (SI)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and $\Phi : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\Phi(x) \stackrel{\Delta}{=} \sup_{\xi \in [0,1]} \phi(x, \xi),$$

with $\phi : \mathbb{R}^n \times [0,1] \to \mathbb{R}$ continuously differentiable in the first argument. For an excellent survey of the theory behind the problem (SI), in addition to some algorithms and applications, see [9] as well as the other papers in the present volume. Many globally convergent algorithms designed to solve (SI)

rely on approximating $\Phi(x)$ by using progressively finer discretizations of $[0,1]$ (see, e.g. [5, 7, 8, 16, 18, 19, 20, 23]). Specifically, such algorithms generate a sequence of problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \phi(x,\xi) \leq 0, \quad \forall \xi \in \Xi, \end{array} \tag{DSI}$$

where $\Xi \subset [0,1]$ is a (presumably large) finite set. For example, given $q \in \mathbb{N}$, one could use the *uniform discretization*

$$\Xi \triangleq \left\{ 0, \frac{1}{q}, \ldots, \frac{q-1}{q}, 1 \right\}.$$

Clearly these algorithms are crucially dependent upon being able to efficiently solve problem (DSI).

Of course, (DSI) involves only a finite number of smooth constraints, thus could be solved in principle via classical constrained optimization techniques. Note however that when $|\Xi|$ is large compared to the number of variables n, it is likely that only a small subset of the constraints are active at a solution. A scheme which exploits this fact by cleverly using an appropriate small subset of the constraints at each step should, in most cases, enjoy substantial savings in computational effort without sacrificing global and local convergence properties.

Early efforts at employing such a scheme appear in [19, 16] in the context of first order methods of feasible directions. In [19], at iteration k, a search direction is computed based on the method of Zoutendijk [28] using only the gradients of those constraints satisfying $\phi(x_k,\xi) \geq -\epsilon$, where $\epsilon > 0$ is small. Clearly, close to a solution, such "ϵ-active" constraints are sufficient to ensure convergence. However, if the discretization is very fine, such an approach may still produce sub-problems with an unduly large number of constraints. It was shown in [16] that, by means of a scheme inspired by the bundle-type methods of nondifferentiable optimization (see, e.g. [11, 13]), the number of constraints used in the sub-problems can be further reduced without jeopardizing global convergence. Specifically, in [16], the constraints to be used in the computation of the search direction d_{k+1} at iteration $k+1$ are chosen as follows. Let $\Xi_k \subseteq \Xi$ be the set of constraints used to compute the search direction d_k, and let x_{k+1} be the next iterate. Then Ξ_{k+1} includes:

- All $\xi \in \Xi$ such that $\phi(x_{k+1},\xi) = 0$ (i.e. the "active" constraints),
- All $\xi \in \Xi_k$ which affected the computation of the search direction d_k, and

- A $\xi \in \Xi$, if it exists, which caused a step reduction in the line search at iteration k.

While the former is obviously needed to ensure that d_k is a feasible direction, it is argued in [16] that the latter two are necessary to avoid zig-zagging or other jamming phenomena.

The number of constraints required to compute the search direction is thus typically small compared to $|\Xi|$, hence each iteration of such a method is computationally less costly. Unfortunately, for a fixed level of discretization, the algorithms in [19, 16] converge at best at a linear rate.

Sequential Quadratic Programming (SQP)-type algorithms exhibit fast local convergence and are well-suited for problems in which the number of variables is not too large but the evaluation of objective/constraint functions and their gradients is costly. In such algorithms, quadratic programs (QPs) are used as models to construct the search direction. For an excellent recent survey of SQP algorithms, see [2]. A number of attempts at applying the SQP scheme to problems with a large number of constraints, e.g. our discretized problem from SIP, have been documented in the literature. In [1], Biggs treats all active inequality constraints as equality constraints in the QP sub-problem, while ignoring all constraints which are not active. Polak and Tits [20], and Mine et al. [14], apply to the SQP framework an ϵ-active scheme similar to that used in [19]. Similar to the ϵ-active idea, Powell proposes a "tolerant" algorithm for linearly constrained problems in [22]. Finally, in [26], Schittkowski proposes another modification of the SQP scheme for problems with many constraints, but does not prove convergence. In practice, the algorithm in [26] may or may not converge, dependent upon the heuristics applied to choose the constraints for the QP sub-problem.

In this paper, the scheme introduced in [16] in the context of first-order feasible direction methods is extended to the SQP framework, specifically, the Feasible SQP (FSQP) framework introduced in [17] (the qualifier "feasible" signifies that all iterates x_k satisfy the constraints, i.e. $\phi(x_k, \xi) \leq 0$, for all $\xi \in \Xi$). Our presentation and analysis significantly borrow from [27], where an important special case of (DSI) is considered, the unconstrained minimax problem.

Let the feasible set be denoted

$$X \stackrel{\Delta}{=} \{x \in \mathbb{R}^n \mid \phi(x, \xi) \leq 0, \ \forall \xi \in \Xi \ \}.$$

For $x \in X$, $\hat{\Xi} \subseteq \Xi$, and $H \in \mathbb{R}^{n\times n}$ with $H = H^T > 0$, let $d^0(x, H, \hat{\Xi})$ be the corresponding SQP direction, i.e. the unique solution of the QP

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle d^0, Hd^0\rangle + \langle \nabla f(x), d^0\rangle \\ \text{subject to} & \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^0\rangle \leq 0, \quad \forall \xi \in \hat{\Xi}. \end{array} \qquad QP^0(x, H, \hat{\Xi})$$

At iteration k, given an estimate $x_k \in X$ of the solution, a constraint index set $\Xi_k \subseteq \Xi$, and a symmetric positive definite estimate H_k of the Hessian of the Lagrangian, first compute $d_k^0 = d^0(x_k, H_k, \Xi_k)$. Note that d_k^0 may not be a feasible search direction, as required in the FSQP context, but that at worst it is tangent to the feasible set. Since all iterates are to remain in the feasible set, following [17], an essentially arbitrary feasible descent direction d_k^1 is computed and the search direction is taken to be the convex combination $d_k = (1-\rho_k)d_k^0 + \rho_k d_k^1$. The coefficient $\rho_k = \rho(d_k^0) \in [0,1]$ goes to zero fast enough, as x_k approaches a solution, to ensure the fast convergence rate of the standard SQP scheme is preserved. An Armijo-type line search is then performed along the direction d_k, yielding a step-size $t_k \in (0,1]$. The next iterate is taken to be $x_{k+1} = x_k + t_k d_k$. Finally, H_k is updated yielding H_{k+1}, and a new constraint index set Ξ_{k+1} is constructed following the ideas of [16].

As is pointed out in [27], the construction of [16] cannot be used meaningfully in the SQP framework without modifying the update rule for the new metric H_{k+1}. The reason is as follows. As noted above, following [16], if $t_k < 1$, Ξ_{k+1} is to include, among others, the index $\bar{\xi} \in \Xi$ of a constraint which was infeasible for the last trial point in the line search.[1] The rationale for including $\bar{\xi}$ in Ξ_{k+1} is that if $\bar{\xi}$ had been in Ξ_k, then it is likely that the computed search direction would have allowed a longer step. Such reasoning is clearly justified in the context of first-order search directions as is used in [16], but it is not clear that $\bar{\xi}$ is the right constraint to include under the new metric H_{k+1}. To overcome this difficulty, it is proposed in [27] that H_k not be updated whenever $t_k < \delta$, δ a prescribed small positive number, and $\bar{\xi} \notin \Xi_k$. We will show in Section 3 that, as is the case for the minimax algorithm of [27], for k large enough, $\bar{\xi}$ will always be in Ξ_k, thus normal updating will take place eventually, preserving the local convergence rate properties of the SQP scheme.

There is an important additional complication, with the update of Ξ_k, which was not present in the minimax case considered in [27]. As just pointed out, any $\xi \in \Xi_k$ which affected the search direction is to be included in Ξ_{k+1}. In [27] (unconstrained minimax problem) this is accomplished by including those objectives whose multipliers are non-zero in the QP used to compute the search

[1] Assuming that it was a constraint, and not the objective function, which caused a failure in the line search.

direction (analogous to $QP^0(x_k, H_k, \Xi_k)$ above), i.e. the "binding" objectives. In our case, in addition to the binding constraints from $QP^0(x_k, H_k, \Xi_k)$, we must also include those constraints which affect the computation of the feasible descent direction d_k^1. If this is not done, convergence is not ensured and a "zig-zagging" phenomenon as discussed in [16] could result.

As a final matter on the update rule for Ξ_k, following [27], we allow for additional constraint indices to be added to the set Ξ_k. While not necessary for global convergence, cleverly choosing additional constraints can significantly improve performance, especially in early iterations. In the context of discretized SIP, exploiting the possible regularity properties of the SIP constraints with respect to the independent parameter can give useful heuristics for choosing additional constraints.

In order to guarantee fast (superlinear) local convergence, it is necessary that, for k large enough, the line search always accept the step-size $t_k = 1$. It is well-known in the SQP context that the line search could truncate the step size arbitrarily close to a solution (the so-called Maratos effect), thus preventing superlinear convergence. Various schemes have been devised to overcome such a situation. We will argue that a second-order correction, as used in [17], will overcome the Maratos effect without sacrificing global convergence.

The balance of the paper is organized as follows. In Section 2 we introduce the algorithm and present some preliminary material. Next, in Section 3, we give a complete convergence analysis of the algorithm proposed in Section 2. The local convergence analysis assumes the just mentioned second-order correction is used. To improve the continuity of the development, a few of the proofs are deferred to an appendix. In Section 4, the algorithm is extended to handle the constrained minimax case. Some implementation details, in addition to numerical results, are provided in Section 5. Finally, in Section 6, we offer some concluding remarks.

2 ALGORITHM

We begin by making a few assumptions that will be in force throughout. The first is a standard regularity assumption, while the second ensures that the set of active constraint gradients is always linearly independent.

Assumption 1: The functions $f : \mathbb{R}^n \to \mathbb{R}$ and $\phi(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}$, $\xi \in \Xi$ are continuously differentiable.

Define the set of active constraints for a point $x \in X$ as

$$\Xi_{\text{act}}(x) \stackrel{\Delta}{=} \{\xi \in \Xi \mid \phi(x,\xi) = 0\}.$$

Assumption 2: For all $x \in X$ with $\Xi_{\text{act}}(x) \neq \emptyset$, the set $\{\nabla_x \phi(x,\xi) \mid \xi \in \Xi_{\text{act}}(x)\}$ is linearly independent.

A point $x^* \in \mathbb{R}^n$ is called a *Karush-Kuhn-Tucker (KKT)* point for the problem (DSI) if there exist KKT multipliers λ^*_ξ, $\xi \in \Xi$, satisfying

$$\begin{cases} \nabla f(x^*) + \sum_{\xi\in\Xi} \lambda^*_\xi \nabla_x \phi(x^*,\xi) = 0, \\ \phi(x^*,\xi) \leq 0, \quad \forall \xi \in \Xi, \\ \lambda^*_\xi \phi(x^*,\xi) = 0 \text{ and } \lambda^*_\xi \geq 0, \quad \forall \xi \in \Xi. \end{cases} \tag{1.1}$$

Under our assumptions, any local minimizer x^* for (DSI) is a KKT point. Thus, (1.1) provides a set of first-order necessary conditions of optimality.

Throughout our analysis, we will often make use of the KKT conditions for $QP^0(x,H,\hat{\Xi})$. Specifically, given $x \in X$, $H = H^T > 0$, and $\hat{\Xi} \subseteq \Xi$, d^0 is a KKT point for $QP^0(x,H,\hat{\Xi})$ if there exist multipliers λ^0_ξ, $\xi \in \hat{\Xi}$, satisfying

$$\begin{cases} Hd^0 + \nabla f(x) + \sum_{\xi\in\hat{\Xi}} \lambda^0_\xi \nabla_x \phi(x,\xi) = 0, \\ \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^0 \rangle \leq 0, \quad \forall \xi \in \hat{\Xi}, \\ \lambda^0_\xi \left(\phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^0 \rangle \right) = 0 \text{ and } \lambda^0_\xi \geq 0, \quad \forall \xi \in \hat{\Xi}. \end{cases} \tag{1.2}$$

In fact, since the objective for $QP^0(x,H,\hat{\Xi})$ is strictly convex, such a d^0 is the unique KKT point, as well as the unique global minimizer (stated formally in Lemma 1 below).

As noted above, d^0 need not be a feasible direction. The search direction d will be taken as a convex combination of d^0 and a first-order feasible descent direction d^1. For $x \in X$ and $\hat{\Xi} \subseteq \Xi$, we compute $d^1 = d^1(x,\hat{\Xi})$, and $\gamma = \gamma(x,\hat{\Xi})$, as the solution of the QP

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|d^1\|^2 + \gamma \\ \text{subject to} & \langle \nabla f(x), d^1 \rangle \leq \gamma, \\ & \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^1 \rangle \leq \gamma, \quad \forall \xi \in \hat{\Xi}. \end{array} \qquad QP^1(x,\hat{\Xi})$$

The notation $\|\cdot\|$ will be used throughout to denote the standard Euclidean norm. The pair (d^1,γ) is a KKT point for $QP^1(x,\hat{\Xi})$ if there exist multipliers μ^1 and λ^1_ξ, $\xi \in \hat{\Xi}$, satisfying

$$\begin{cases} \begin{bmatrix} d^1 \\ 1 \end{bmatrix} + \mu^1 \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix} + \sum_{\xi \in \hat{\Xi}} \lambda^1_\xi \begin{bmatrix} \nabla_x \phi(x,\xi) \\ -1 \end{bmatrix} = 0, \\ \langle \nabla f(x), d^1 \rangle \leq \gamma, \\ \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^1 \rangle \leq \gamma, \quad \forall \xi \in \hat{\Xi}, \\ \mu^1 \left(\langle \nabla f(x), d^1 \rangle - \gamma \right) = 0 \text{ and } \mu^1 \geq 0, \\ \lambda^1_\xi \left(\phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^1 \rangle - \gamma \right) = 0 \text{ and } \lambda^1_\xi \geq 0, \quad \forall \xi \in \hat{\Xi}. \end{cases} \tag{1.3}$$

In Section 1 we stated that the feasible descent direction d^1 was essentially arbitrary. In the subsequent analysis we assume that d^1 is chosen specifically as the solution of $QP^1(x,\hat{\Xi})$, though it can be shown that the results still hold if some minor variation is used. To be precise, following [17], we require that $d^1 = d^1(x,\hat{\Xi})$ satisfy:

- $d^1(x,\hat{\Xi}) = 0$ if x is a KKT point,
- $\langle \nabla f(x), d^1(x,\hat{\Xi}) \rangle < 0$ if x is not a KKT point,
- $\langle \nabla_x \phi(x,\xi), d^1(x,\hat{\Xi}) \rangle < 0$, for all $\xi \in \Xi_{\text{act}}$ if x is not a KKT point, and
- for $\hat{\Xi}$ fixed, $d^1(x,\hat{\Xi})$ is bounded over bounded subsets of X.

It will be shown in Lemma 2 that the solution of $QP^1(x,\hat{\Xi})$ satisfies these requirements. In our context, d^1 must fulfill one additional property, which is essentially captured by Lemma 7 in the appendix.

Thus, at iteration k, the search direction d_k is taken as a convex combination of d^0_k and d^1_k, i.e. $d_k = (1-\rho_k)d^0_k + \rho_k d^1_k$, $\rho_k \in [0,1]$. In order to guarantee a fast local rate of convergence while providing a suitably feasible search direction, we require the coefficient of the convex combination $\rho_k = \rho(d^0_k)$ to satisfy certain properties. Namely, $\rho(\cdot) : \mathbb{R}^n \to [0,1]$ must satisfy

- $\rho(d^0)$ is bounded away from zero outside every neighborhood of zero, and
- $\rho(d^0) = O(\|d^0\|^2)$.

For example, we could take $\rho(d^0) = \min\{1, \|d^0\|^\nu\}$, where $\nu \geq 2$.

It remains to explicitly specify the key feature of the proposed algorithm: the update rule for Ξ_k. As discussed above, following [16], Ξ_{k+1} will include (in addition to possible heuristics) three crucial components. The first one is the set $\Xi_{\text{act}}(x_{k+1})$ of indices of active constraints at the new iterate. The second component of Ξ_{k+1} is the set $\Xi_k^b \subseteq \Xi_k$ of indices of constraints that affected d_k. In particular, Ξ_k^b will include all indices of constraints in $QP^0(x_k, H_k, \Xi_k)$ and $QP^1(x_k, \Xi_k)$ which have positive multipliers, i.e. the *binding* constraints for these QPs. Specifically, let $\lambda_{k,\xi}^0$, and $\lambda_{k,\xi}^1$, for $\xi \in \Xi_k$, be the QP multipliers from $QP^0(x_k, H_k, \Xi_k)$ and $QP^1(x_k, \Xi_k)$, respectively. Defining

$$\Xi_k^{b,0} \triangleq \{\ \xi \in \Xi_k \mid \lambda_{k,\xi}^0 > 0\ \}, \quad \Xi_k^{b,1} \triangleq \{\ \xi \in \Xi_k \mid \lambda_{k,\xi}^1 > 0\ \},$$

we let

$$\Xi_k^b \triangleq \Xi_k^{b,0} \cup \Xi_k^{b,1}.$$

Finally, the third component of Ξ_{k+1} is the index $\bar{\xi}$ of one constraint, if any exists, that forced a reduction of the step in the previous line search. While the exact type of line search we employ is not critical to our analysis, we assume from this point onward that it is an Armijo-type search. That is, given constants $\alpha \in (0, 1/2)$ and $\beta \in (0, 1)$, the step-size t_k is taken as the first number t in the set $\{1, \beta, \beta^2, \ldots\}$ such that

$$f(x_k + td_k) \leq f(x_k) + \alpha t \langle \nabla f(x_k), d_k \rangle, \tag{1.4}$$

and

$$\phi(x_k + td_k, \xi) \leq 0, \quad \forall \xi \in \Xi. \tag{1.5}$$

Thus, $t_k < 1$ implies that either (1.4) or (1.5) is violated at $x_k + \frac{t_k}{\beta} d_k$. In the event that (1.5) is violated, there exists $\bar{\xi} \in \Xi$ such that

$$\phi\left(x_k + \frac{t_k}{\beta} d_k, \bar{\xi}\right) > 0, \tag{1.6}$$

and in such a case we will include $\bar{\xi}$ in Ξ_{k+1}.

Algorithm FSQP-MC

Parameters. $\alpha \in (0, \frac{1}{2})$, $\beta \in (0,1)$, and $0 < \delta \ll 1$.

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

Step 0 - Initialization. Set $k \leftarrow 0$ and choose $\Xi_0 \supseteq \Xi_{\text{act}}(x_0)$.

Step 1 - Computation of search direction.

(i). Compute $d_k^0 = d^0(x_k, H_k, \Xi_k)$. If $d_k^0 = 0$, stop.
(ii). Compute $d_k^1 = d^1(x_k, \Xi_k)$.
(iii). Compute $\rho_k = \rho(d_k^0)$ and set $d_k \leftarrow (1 - \rho_k)d_k^0 + \rho_k d_k^1$.

Step 2 - Line search. Compute t_k, the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying (1.4) and (1.5).

Step 3 - Updates.

(i). Set $x_{k+1} \leftarrow x_k + t_k d_k$.
(ii). If $t_k < 1$ and (1.5) was violated at $x_k + \frac{t_k}{\beta} d_k$, then let $\bar{\xi}$ be such that (1.6) holds.
(iii). Pick

$$\Xi_{k+1} \supseteq \Xi_{\text{act}}(x_{k+1}) \cup \Xi_k^b.$$

If $t_k < 1$ and (1.6) holds for some $\bar{\xi} \in \Xi$, then set

$$\Xi_{k+1} \leftarrow \Xi_{k+1} \cup \{\bar{\xi}\}.$$

(iv). If $t_k \leq \delta$ and $\bar{\xi} \notin \Xi_k$, set $H_{k+1} \leftarrow H_k$. Otherwise, obtain a new symmetric positive definite estimate H_{k+1} to the Hessian of the Lagrangian.
(v). Set $k \leftarrow k + 1$ and go back to *Step 1.*

3 CONVERGENCE ANALYSIS

While there are some critical differences, the analysis in this section closely parallels that of [27]. We begin by establishing that, under a few additional assumptions, algorithm **FSQP-MC** generates a sequence which converges to a KKT point for (DSI). Then, upon strengthening our assumptions slightly, we show that the rate of convergence is two-step superlinear.

3.1 Global Convergence

The following will be assumed to hold throughout our analysis.

Assumption 3: The level set $\{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \cap X$ is compact.

Assumption 4: There exist scalars $0 < \sigma_1 \leq \sigma_2$ such that for all k,

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

Given the scalars $0 < \sigma_1 \leq \sigma_2$ from Assumption 4, define

$$\mathcal{H} \stackrel{\Delta}{=} \{H = H^T \mid \sigma_1 \|d\|^2 \leq \langle d, Hd \rangle \leq \sigma_2 \|d\|^2, \quad \forall d \in \mathbb{R}^n\}.$$

First, we derive some properties of $d^0(x, H, \hat{\Xi})$ and $d^1(x, \hat{\Xi})$.

Lemma 1 *For all $x \in X$, $H \in \mathcal{H}$, and $\hat{\Xi} \subseteq \Xi$ such that $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$, the search direction $d^0 = d^0(x, H, \hat{\Xi})$ is well-defined, is continuous in x and H for $\hat{\Xi}$ fixed, and is the unique KKT point for $QP^0(x, H, \hat{\Xi})$. Furthermore, $d^0 = 0$ if, and only if, x is a KKT point for the problem (DSI). If x is not a KKT point for this problem, then d^0 satisfies*

$$\langle \nabla f(x), d^0 \rangle < 0, \tag{1.7}$$

$$\langle \nabla_x \phi(x, \xi), d^0 \rangle \leq 0, \quad \forall \xi \in \Xi_{\text{act}}(x). \tag{1.8}$$

Proof: $H > 0$ implies that $QP^0(x, H, \hat{\Xi})$ is strictly convex. Further $d^0 = 0$ is always feasible for the QP constraints. It follows that the solution d^0 is well-defined and the unique KKT point for the QP. As the set $\mathcal{H}$ is uniformly positive definite, continuity in x and H for fixed $\hat{\Xi}$ is a direct consequence of Theorem 4.4 in [4]. Now suppose d^0 is 0 and let $\{ \lambda_\xi^0 \mid \xi \in \hat{\Xi} \}$ be the QP multipliers. In view of the KKT conditions (1.2) for $QP^0(x, H, \hat{\Xi})$, since $d^0 = 0$ and $x \in X$, we see that x satisfies the KKT conditions (1.1) for (DSI) with

$$\lambda_\xi = \begin{cases} \lambda_\xi^0, & \xi \in \hat{\Xi}, \\ 0, & \xi \notin \hat{\Xi}. \end{cases}$$

The converse is proved similarly, appealing to the uniqueness of the KKT point for $QP^0(x, H, \hat{\Xi})$ and the fact that $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$. Finally, since $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$, (1.7) and (1.8) follow from Proposition 3.1 in [17]. □

Lemma 2 *For all $x \in X$, and $\hat{\Xi} \subseteq \Xi$ such that $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$, the direction $d^1 = d^1(x, \hat{\Xi})$ is well-defined and the pair (d^1, γ), where $\gamma = \gamma(x, \hat{\Xi})$, is the unique KKT point of $QP^1(x, \hat{\Xi})$. Furthermore, for given $\hat{\Xi}$, $d^1(x, \hat{\Xi})$ is bounded over bounded subsets of X. In addition, $d^1 = 0$ if, and only if, x is a KKT point for the problem (DSI). If x is not a KKT point for this problem, then d^1 satisfies*

$$\langle \nabla f(x), d^1 \rangle < 0, \tag{1.9}$$

$$\langle \nabla_x \phi(x, \xi), d^1 \rangle < 0, \quad \forall \xi \in \Xi_{\text{act}}(x), \tag{1.10}$$

and γ satisfies $\gamma < 0$.

Proof: We begin by noting that (d^1, γ) solves $QP^1(x, \hat{\Xi})$ if, and only if, d^1 solves

$$\text{minimize} \quad \frac{1}{2}\|d^1\|^2 + \max\left\{\langle \nabla f(x), d^1 \rangle,\ \max_{\xi \in \hat{\Xi}}\{\phi(x, \xi) + \langle \nabla_x \phi(x, \xi), d^1 \rangle\}\right\}. \tag{1.11}$$

and

$$\gamma = \max\left\{\langle \nabla f(x), d^1 \rangle,\ \max_{\xi \in \hat{\Xi}}\{\phi(x, \xi) + \langle \nabla_x \phi(x, \xi), d^1 \rangle\}\right\}.$$

Since the objective function in (1.11) is strictly convex and radially unbounded, it follows that $QP^1(x, \hat{\Xi})$ has $(d^1, \gamma) = (d^1(x, \hat{\Xi}), \gamma(x, \hat{\Xi}))$ as its unique global minimizer. Since $QP^1(x, \hat{\Xi})$ is convex, (d^1, γ) is also its unique KKT point. Boundedness of $d^1(x, \hat{\Xi})$ over bounded subsets of X follows from the first equation of the optimality conditions (1.3), noting that the QP multipliers must all lie in $[0, 1]$. Now suppose $d^1 = 0$. Since $x \in X$, it is clear that $\gamma = 0$. Substitute $d^1 = 0$ and $\gamma = 0$ into (1.3) and let $\mu^1 \in \mathbb{R}$ and $\lambda^1 \in \mathbb{R}^{|\hat{\Xi}|}$ be the corresponding multipliers. Note that, in view of Assumption 2, $\mu^1 > 0$. Since $x \in X$, it follows that x satisfies (1.1) with multipliers

$$\lambda_\xi = \begin{cases} \lambda^1_\xi / \mu^1, & \xi \in \hat{\Xi}, \\ 0, & \xi \notin \hat{\Xi}. \end{cases}$$

Therefore, x is a KKT point for (DSI). The converse is proved similarly, appealing to uniqueness of the KKT point for $QP^1(x, \hat{\Xi})$, and the fact that $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$. To prove (1.9) and (1.10) note that if x is not a KKT point for (DSI), then as just shown $d^1 \neq 0$, hence $\gamma < 0$ (since $d^1 = 0$ and $\gamma = 0$ form a feasible pair, the optimal value of $QP^1(x_k, \Xi_k)$ must be non-positive). The result then follows directly from the form of the QP constraints and the fact that $\Xi_{\text{act}}(x) \subseteq \hat{\Xi}$. □

Next, we establish that the line search is well-defined.

Lemma 3 *For each k, the line search in Step 2 yields a step $t_k = \beta^j$ for some finite $j = j(k)$.*

Proof: If x_k is a KKT point, then *Step 2* is not reached, hence assume x_k is not KKT. In view of Lemma 1 and the properties of $\rho(\cdot)$, $d_k^0 \neq 0$ and $\rho(d_k^0) > 0$. Lemmas 1 and 2 imply, since $\Xi_{\text{act}}(x_k) \subseteq \Xi_k$, that

$$\langle \nabla f(x_k), d_k \rangle < 0,$$

$$\langle \nabla_x \phi(x_k, \xi), d_k \rangle < 0, \quad \forall \xi \in \Xi_{\text{act}}(x_k).$$

Further, feasibility of x_k requires $\phi(x_k, \xi) < 0$ for all $\xi \in \Xi \setminus \Xi_{\text{act}}(x_k)$. The result then follows by considering first order expansions of $f(x_k + t_k d_k)$ and $\phi(x_k + t_k d_k, \xi)$, $\xi \in \Xi$, about x_k and by appealing to our regularity assumptions. □

The previous three lemmas imply that the algorithm is well-defined. In addition, Lemma 1 shows that if Algorithm **FSQP-MC** generates a finite sequence terminating at the point x_N, then x_N is a KKT point for the problem (DSI). We now concentrate on the case in which the algorithm never satisfies the termination condition in *Step 1(i)* and generates an infinite sequence $\{x_k\}$. Given an infinite index set $\mathcal{K}$, we use the notation

$$x_k \stackrel{k\in\mathcal{K}}{\longrightarrow} x^*$$

to mean

$$x_k \to x^* \quad \text{as} \quad k \to \infty, \ k \in \mathcal{K}.$$

Lemma 4 *Let $\mathcal{K}$ be an infinite index set such that $\Xi_k = \Xi^*$ for all $k \in \mathcal{K}$, $x_k \stackrel{k\in\mathcal{K}}{\longrightarrow} x^*$, $H_k \stackrel{k\in\mathcal{K}}{\longrightarrow} H^*$, $d_k^0 \stackrel{k\in\mathcal{K}}{\longrightarrow} d^{0,*}$, $d_k^1 \stackrel{k\in\mathcal{K}}{\longrightarrow} d^{1,*}$, and $\gamma_k \stackrel{k\in\mathcal{K}}{\longrightarrow} \gamma^*$. Then (i) $d^{0,*}$ is the unique solution of $QP^0(x^*, H^*, \Xi^*)$, and (ii) $(d^{1,*}, \gamma^*)$ is the unique solution of $QP^1(x^*, \Xi^*)$.*

Proof: Part (i) follows from continuity of $d^0(x, H, \hat{\Xi})$ for fixed $\hat{\Xi}$ (Lemma 1). To prove part *(ii)*, recall that in view of Lemma 2, (d_k^1, γ_k) is the unique KKT point for $QP^1(x_k, \Xi^*)$, i.e. is the unique solution of (1.3) with corresponding multipliers $\mu_k^1 \geq 0$ and $\lambda_{k,\xi}^1 \geq 0$, $\xi \in \Xi^*$. Note that the multipliers satisfy

$$\mu_k^1 + \sum_{\xi \in \Xi^*} \lambda_{k,\xi}^1 = 1,$$

for all k, hence are bounded. Let $\mathcal{K}' \subseteq \mathcal{K}$ be an infinite index set such that $\mu_k^1 \stackrel{k\in\mathcal{K}'}{\longrightarrow} \mu^{1,*}$ and $\lambda_{k,\xi}^1 \stackrel{k\in\mathcal{K}'}{\longrightarrow} \lambda_\xi^{1,*}$, $\xi \in \Xi^*$. Taking limits in the optimality conditions (1.3) shows that $(d^{1,*}, \gamma^*)$ is a KKT point for $QP^1(x^*, \Xi^*)$ with multipliers $\mu^{1,*}$ and $\lambda_\xi^{1,*}$, $\xi \in \Xi^*$. Uniqueness of such points proves the result. □

The following lemma establishes a few basic properties of some of the sequences generated by the algorithm.

Lemma 5 *(i) The sequences* $\{x_k\}$, $\{d_k^0\}$, *and* $\{d_k\}$ *are bounded; (ii)* $\{f(x_k)\}$ *converges; (iii)* $t_k d_k \longrightarrow 0$.

Proof: In view of Assumption 3 and the fact that the line search guarantees $\{f(x_k)\}$ is a monotonically decreasing sequence, it follows that $\{x_k\}$ is bounded, and since $f(\cdot)$ is continuous, that $\{f(x_k)\}$ converges. Boundedness of $\{d_k^0\}$ follows from boundedness of $\{x_k\}$, Assumption 4, continuity of $d^0(x, H, \hat{\Xi})$ for fixed $\hat{\Xi}$, and the fact that there are only finitely many subsets $\hat{\Xi}$ of Ξ. Boundedness of $\{d_k^1\}$ follows from Lemma 2 and boundedness of $\{x_k\}$. Since $\rho_k \in [0,1]$, $\{d_k\}$ is bounded as well. Finally, suppose $\{t_k d_k\} \not\to 0$. Then there exists an infinite index set $\mathcal{K} \subseteq \mathbb{N}$ such that $t_k d_k$ is bounded away from zero on $\mathcal{K}$. Since all sequences of interest are bounded and Ξ is finite, we may suppose without loss of generality that $x_k \overset{k\in\mathcal{K}}{\longrightarrow} x^*$, $H_k \overset{k\in\mathcal{K}}{\longrightarrow} H^*$, $d_k^0 \overset{k\in\mathcal{K}}{\longrightarrow} d^{0,*}$, $d_k^1 \overset{k\in\mathcal{K}}{\longrightarrow} d^{1,*}$, $\gamma_k \overset{k\in\mathcal{K}}{\longrightarrow} \gamma^*$, $\rho_k \overset{k\in\mathcal{K}}{\longrightarrow} \rho^*$, and $\Xi_k = \Xi^*$ for all $k \in \mathcal{K}$. Lemma 4 tells us that $d^{0,*}$ is the unique solution of $QP^0(x^*, H^*, \Xi^*)$ and $(d^{1,*}, \gamma^*)$ is the unique solution pair for $QP^1(x^*, \Xi^*)$. Furthermore, since $t_k d_k$ is bounded away from zero on $\mathcal{K}$, there exists $\underline{t} > 0$ such that $t_k \geq \underline{t}$ for all $k \in \mathcal{K}$, and since $\{t_k\}$ is bounded ($t_k \in [0,1]$), it follows that either $d^{0,*} \neq 0$ or $d^{1,*} \neq 0$. Applying Lemmas 1 and 2 in both directions shows that x^* is not a KKT point for (DSI) and *both* $d^{0,*} \neq 0$ and $d^{1,*} \neq 0$. In addition, $\gamma^* < 0$ and $\gamma_k < 0$, for all $k \in \mathcal{K}$ (from Lemma 2) and ρ_k is bounded away from zero on $\mathcal{K}$ (from the assumptions on $\rho(\cdot)$ as given in Section 2 and the fact that d_k^0 is bounded away from zero on $\mathcal{K}$). As a consequence, there exists $\underline{\rho} > 0$ such that $\rho_k = \rho(d_k^0) \geq \underline{\rho}$ for all $k \in \mathcal{K}$ and there exists $\bar{\gamma} < 0$ such that $\gamma_k \leq \bar{\gamma}$ for all $k \in \mathcal{K}$. Now, since $\langle \nabla f(x_k), d_k^1 \rangle \leq \gamma_k$, for all k,

$$\begin{aligned} t_k \langle \nabla f(x_k), d_k \rangle &= t_k(1-\rho_k)\langle \nabla f(x_k), d_k^0 \rangle + t_k \rho_k \langle \nabla f(x_k), d_k^1 \rangle \\ &\leq t_k \rho_k \gamma_k \leq \underline{t}\underline{\rho}\bar{\gamma} < 0, \end{aligned}$$

for all $k \in \mathcal{K}$, where we have used (1.7). Thus, by the line search criterion of *Step 2*,

$$f(x_{k+1}) \leq f(x_k) + \underline{t}\underline{\rho}\bar{\gamma}$$

for all $k \in \mathcal{K}$. Since $f(x_k)$ is monotone non-increasing, it follows that $\{f(x_k)\}$ diverges, which contradicts (ii). □

In order to establish convergence to a KKT point, it will be convenient to consider the value functions for the search direction QPs, $QP^0(x, H, \hat{\Xi})$ and

$QP^1(x,\hat{\Xi})$. In particular, given the solutions $d^0 = d^0(x,H,\hat{\Xi})$ and $(d^1,\gamma) = (d^1(x,\hat{\Xi}),\gamma(x,\hat{\Xi}))$, define

$$v^0(x,H,\hat{\Xi}) \stackrel{\Delta}{=} -\left(\frac{1}{2}\langle d^0, Hd^0\rangle + \langle\nabla f(x), d^0\rangle\right),$$

$$v^1(x,\hat{\Xi}) \stackrel{\Delta}{=} -\left(\frac{1}{2}\|d^1\|^2 + \hat{\gamma}\right).$$

Further, let $v(x,H,\hat{\Xi}) \stackrel{\Delta}{=} v^0(x,H,\hat{\Xi}) + v^1(x,\hat{\Xi})$, and, at iteration k, define $v_k^0 = v^0(x_k,H_k,\Xi_k)$, $v_k^1 = v^1(x_k,\Xi_k)$, and $v_k = v(x_k,H_k,\Xi_k)$. Note that, since 0 is always feasible for both QPs, $v_k^0 \geq 0$ and $v_k^1 \geq 0$, for all k.

Lemma 6 *Let $\mathcal{K}$ be an infinite index set. Then (i) $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$ if and only if $v_k^0 \xrightarrow{k\in\mathcal{K}} 0$, (ii) $(d_k^1,\gamma_k) \xrightarrow{k\in\mathcal{K}} (0,0)$ if and only if $v_k^1 \xrightarrow{k\in\mathcal{K}} 0$, and (iii) if $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$, then all accumulation points of $\{x_k\}_{k\in\mathcal{K}}$ are KKT points for (DSI).*

Proof: First, if $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$ then it is clear from the definition of v_k^0 that $v_k^0 \xrightarrow{k\in\mathcal{K}} 0$. Next, from (1.2) and the last statement in Lemma 1, it follows that $\langle\nabla f(x_k), d_k^0\rangle \leq -\langle d_k^0, H_k d_k^0\rangle$. Thus, using again the definition of v_k^0, we get

$$\begin{aligned} v_k^0 &= -\left(\tfrac{1}{2}\langle d_k^0, H_k d_k^0\rangle + \langle\nabla f(x_k), d_k^0\rangle\right) \\ &\geq -\tfrac{1}{2}\langle d_k^0, H_k d_k^0\rangle + \langle d_k^0, H_k d_k^0\rangle \;\geq\; \tfrac{\sigma_1}{2}\|d_k^0\|^2 \;>\; 0, \end{aligned}$$

where we have used Assumption 4. Thus, if $v_k^0 \xrightarrow{k\in\mathcal{K}} 0$, then $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$. If $(d_k^1,\gamma_k) \xrightarrow{k\in\mathcal{K}} (0,0)$, then from the definition of v_k^1 we see that $v_k^1 \xrightarrow{k\in\mathcal{K}} 0$. Now suppose $v_k^1 \xrightarrow{k\in\mathcal{K}} 0$. To show $d_k^1 \xrightarrow{k\in\mathcal{K}} 0$, note that from the optimality conditions (1.3),

$$\begin{aligned} \|d_k^1\|^2 &= -\mu_k^1\langle\nabla f(x_k), d_k^1\rangle - \sum_{\xi\in\Xi_k}\lambda_{k,\xi}^1\langle\nabla_x\phi(x_k,\xi), d_k^1\rangle \\ &= -\mu_k^1\gamma_k - \sum_{\xi\in\Xi_k}\lambda_{k,\xi}^1(\gamma_k - \phi(x_k,\xi)) \\ &= -\gamma_k + \sum_{\xi\in\Xi_k}\lambda_{k,\xi}^1\phi(x_k,\xi) \;\leq\; -\gamma_k. \end{aligned}$$

Thus, again using the definition of v_k^1,

$$v_k^1 = -\frac{1}{2}\|d_k^1\|^2 - \gamma_k \geq \frac{1}{2}\|d_k^1\|^2,$$

and it immediately follows that $d_k^1 \xrightarrow{k\in\mathcal{K}} 0$ and $\gamma_k \xrightarrow{k\in\mathcal{K}} 0$. To prove *(iii)*, suppose $\mathcal{K}$ is such that $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$. Let x^* be an accumulation point of $\{x_k\}_{k\in\mathcal{K}}$ and let $\mathcal{K}' \subseteq \mathcal{K}$ be an infinite index set such that $x_k \xrightarrow{k\in\mathcal{K}'} x^*$ and, for some $\hat{\Xi}$, $\Xi_k = \hat{\Xi} \subseteq \Xi$ for all $k \in \mathcal{K}'$. Let $\lambda_k^0 \in \mathbb{R}^{|\hat{\Xi}|}$ be the multiplier vector from $QP^0(x_k, H_k, \hat{\Xi})$ and define

$$\Xi_k^0 \triangleq \{\ \xi \in \hat{\Xi} \mid \phi(x_k,\xi) + \langle \nabla_x \phi(x_k,\xi), d_k^0\rangle = 0\ \}.$$

Suppose, without loss of generality, $\Xi_k^0 = \hat{\Xi}^0$, for all $k \in \mathcal{K}'$. As $d_k^0 \xrightarrow{k\in\mathcal{K}'} 0$, it is clear that $\hat{\Xi}^0 \subseteq \Xi_{\text{act}}(x^*)$ and, in view of Assumption 2, the set $\{\ \nabla_x\phi(x_k,\xi) \mid \xi \in \hat{\Xi}^0\ \}$ is linearly independent, for k large enough. Thus, from (1.2), a unique expression for the QP multipliers (for k large enough) is given by

$$\lambda_k^0 = -\left(\hat{R}(x_k)^T \hat{R}(x_k)\right)^{-1} \hat{R}(x_k)^T \left(H_k d_k^0 + \nabla f(x_k)\right),$$

where $\hat{R}(x_k) \triangleq [\ \nabla_x\phi(x_k,\xi) \mid \xi \in \hat{\Xi}^0\ ] \in \mathbb{R}^{n\times|\hat{\Xi}^0|}$. In view of Assumption 4, boundedness of $\{x_k\}$, the regularity assumptions, and the fact that $d_k^0 \xrightarrow{k\in\mathcal{K}'} 0$, we see that

$$\lambda_k^0 \xrightarrow{k\in\mathcal{K}'} \lambda^{0,*} = -\left(\hat{R}(x^*)^T \hat{R}(x^*)\right)^{-1} \hat{R}(x^*)^T \nabla f(x^*).$$

Taking limits in the optimality conditions (1.2) for $QP^0(x_k, H_k, \hat{\Xi})$ shows that x^* is a KKT point for (DSI) with multipliers

$$\lambda_\xi^* = \begin{cases} \lambda_\xi^{0,*} & \xi \in \hat{\Xi}, \\ 0 & \xi \notin \hat{\Xi}. \end{cases}$$

□

We are now in a position to show that there exists an accumulation point of $\{x_k\}$ which is a KKT point for (DSI). This result is, in fact, weaker than that obtained in [27] for the unconstrained minimax case, where under similar assumptions, but with a more involved argument, it is shown that *all* accumulation points are KKT. The price to be paid is the introduction of Assumption 5 below for proving Theorem 1.

The proof of the following result is inspired by that of Theorem T in [16].

Proposition 1 $\liminf_k v_k = 0$.

Corollary 1 *There exists an accumulation point x^* of $\{x_k\}$ which is a KKT point for (DSI).*

Proof: Since $v_k^0 \geq 0$ and $v_k^1 \geq 0$ for all k, Proposition 1 implies $\liminf_k v_k^0 = 0$, i.e. there exists an infinite index set $\mathcal{K}$ such that $v_k^0 \xrightarrow{k \in \mathcal{K}} 0$. In view of Lemma 6, all accumulation points of $\{x_k\}_{k \in \mathcal{K}}$ are KKT points. Finally, boundedness of $\{x_k\}$ implies at least one such point exists. □

Define the Lagrangian function for (DSI) as

$$L(x, \lambda) \stackrel{\Delta}{=} f(x) + \sum_{\xi \in \Xi} \lambda_\xi \phi(x, \xi).$$

In order to show that the entire sequence converges to a KKT point x^*, we strengthen our assumptions as follows.

Assumption 1′: The functions $f : \mathbb{R}^n \to \mathbb{R}$ and $\phi(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}$, $\xi \in \Xi$ are twice continuously differentiable.

Assumption 5: Some accumulation point x^* of $\{x_k\}$ which is a KKT point for (DSI) also satisfies the second order sufficiency conditions with strict complementary slackness, i.e. there exists $\lambda^* \in \mathbb{R}^{|\Xi|}$ satisfying (1.1) as well as

- $\nabla^2_{xx} L(x^*, \lambda^*)$ is positive definite on the subspace

$$\{h \mid \langle \nabla_x \phi(x^*, \xi), h \rangle = 0, \ \forall \xi \in \Xi_{\text{act}}(x^*)\},$$

- and $\lambda^*_\xi > 0$ for all $\xi \in \Xi_{\text{act}}(x^*)$.

It is well-known that such an assumption implies that x^* is an isolated KKT point for (DSI) as well as an isolated local minimizer. The following theorem is the main result of this section.

Theorem 1 *The sequence $\{x_k\}$ generated by algorithm* **FSQP-MC** *converges to a strict local minimizer x^* of (DSI).*

Proof: First we show that there exists a neighborhood of x^* in which no other accumulation points of $\{x_k\}$ can exist, KKT points or not. As x^* is a strict local minimizer, there exists $\epsilon > 0$ such that $f(x) > f(x^*)$ for all $x \neq x^*$, $x \in \mathcal{S} \stackrel{\Delta}{=} B(x^*, \epsilon) \cap X$, where $B(x^*, \epsilon)$ is the open ball of radius ϵ centered at

x^*. Proceeding by contradiction, suppose $x' \in B(x^*, \epsilon)$, $x' \neq x^*$, is another accumulation point of $\{x_k\}$. Feasibility of the iterates implies that $x' \in \mathcal{S}$. Thus $f(x') > f(x^*)$, which is in contradiction with Lemma 5(ii). Next, in view of Lemma 5(iii), $(x_{k+1} - x_k) \longrightarrow 0$. Suppose $\mathcal{K}$ is an infinite index set such that $x_k \stackrel{k\in\mathcal{K}}{\longrightarrow} x^*$. Then there exists k_1 such that $\|x_k - x^*\| < \epsilon/4$, for all $k \in \mathcal{K}$, $k \geq k_1$. Further, there exists k_2 such that $\|x_{k+1} - x_k\| < \epsilon/4$, for all $k > k_2$. Therefore, if there were another accumulation point outside of $B(x^*, \epsilon)$, then the sequence would have to pass through the compact set $\overline{B(x^*, \epsilon)} \setminus B(x^*, \epsilon/4)$ infinitely many times. This contradicts the established fact that there are no accumulation points of $\{x_k\}$, other than x^*, in $B(x^*, \epsilon)$. □

3.2 Local Convergence

We have thus shown that, with a likely dramatically reduced amount of work per iteration, global convergence can be preserved. This would be of little interest, though, if the speed of convergence were to suffer significantly. In this section we establish that, under a few additional assumptions, the sequence $\{x_k\}$ generated by a slightly modified version of algorithm **FSQP-MC** (to avoid the Maratos effect) exhibits 2-step superlinear convergence. To do this, the bulk of our effort is focussed on showing that for k large the set of constraints Ξ_k^b which affect the search direction is precisely the set of active constraints at the solution, i.e. $\Xi_{\text{act}}(x^*)$. In addition, we show that, eventually, no constraints outside of $\Xi_{\text{act}}(x^*)$ affect the line search, and that H_k is updated normally at every iteration. Thus, for k large enough, the algorithm behaves as if it were solving the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \phi(x,\xi) \leq 0, \quad \forall \xi \in \Xi_{\text{act}}(x^*), \end{array} \tag{P^*}$$

using *all* constraints at every iteration. Establishing this allows us to apply known results concerning local convergence rates.

The following is proved in the appendix.

Proposition 2 *For k large enough,*

(i) $\Xi_k^{b,0} = \Xi_k^{b,1} = \Xi_{\text{act}}(x^*)$, *and*

(ii) $\phi(x_k + td_k, \xi) \leq 0$ *for all* $t \in [0,1]$, $\xi \in \Xi \setminus \Xi_{\text{act}}(x^*)$.

In order to achieve superlinear convergence, it is crucial that a unit step, i.e. $t_k = 1$, always be accepted for all k sufficiently large. Several techniques have been introduced to avoid the so-called Maratos effect. We chose to include a second order correction such as that used in [17]. Specifically, at iteration k, let $\tilde{d}(x,d,H,\hat{\Xi})$ be the unique solution of the QP $\widetilde{QP}(x,d,H,\hat{\Xi})$, defined for $\tau \in (2,3)$ as follows

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle d+\tilde{d}, H(d+\tilde{d})\rangle + \langle \nabla f(x), d+\tilde{d}\rangle \\ \text{subject to} & \phi(x+d,\xi) + \langle \nabla_x \phi(x,\xi), d+\tilde{d}\rangle \leq \|d\|^\tau, \quad \forall \xi \in \hat{\Xi}, \end{array} \qquad \widetilde{QP}(x,d,H,\hat{\Xi})$$

if it exists and has norm less that $\min\{\|d\|, C\}$, where C is a large number. Otherwise, set $\tilde{d}(x,d,H,\hat{\Xi}) = 0$. The following step is added to algorithm **FSQP-MC**:

Step 1(iv). Compute $\tilde{d}_k = \tilde{d}(x_k, d_k, H_k, \Xi_k)$.

In addition, the line search criterion (1.4) and (1.5) are replaced with

$$f(x_k + td_k + t^2\tilde{d}_k) \leq f(x_k) + \alpha t \langle \nabla f(x_k), d_k\rangle, \tag{1.12}$$

and

$$\phi(x_k + td_k + t^2\tilde{d}_k) \leq 0, \quad \forall \xi \in \Xi. \tag{1.13}$$

Finally, the condition (1.6) is replaced with

$$\phi\left(x_k + \frac{t_k}{\beta}d_k + \left(\frac{t_k}{\beta}\right)^2 \tilde{d}_k, \bar{\xi}\right) > 0. \tag{1.14}$$

With some effort, it can be shown that these modifications do not affect any of the results obtained to this point. Further, for k sufficiently large, the set of binding constraints in $\widetilde{QP}(x_k, d_k, H_k, \Xi_k)$ is again $\Xi_{\text{act}}(x^*)$. Hence, it is established that for k large enough, the modified algorithm **FSQP-MC** behaves identically to that given in [17], applied to (P^*).

Assumption 1 is now further strengthened and a new assumption concerning the Hessian approximations H_k is given. These assumptions allow us to use the local convergence rate result from [17].

Assumption 1″: The functions $f : \mathbb{R}^n \to \mathbb{R}$, and $\phi(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}$, $\xi \in \Xi$, are three times continuously differentiable.

Assumption 6: As a result of the update rule chosen for *Step 3(iv)*, H_k approaches the Hessian of the Lagrangian in the sense that

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

where λ^* is the KKT multiplier vector associated with x^* and

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

with $R_k = [\nabla_x \phi(x_k, \xi) \mid \xi \in \Xi_{\text{act}}(x^*)]$.

Theorem 2 *For all k sufficiently large, the unit step $t_k = 1$ is accepted in Step 2. Further, the sequence $\{x_k\}$ converges to x^* 2-step superlinearly, i.e.*

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

4 EXTENSION TO CONSTRAINED MINIMAX

The algorithm we have discussed may be extended following the scheme of [27] to handle problems with many objective functions, i.e. large-scale constrained minimax. Specifically, consider the problem

$$\begin{array}{ll} \text{minimize} & \max\limits_{\omega\in\Omega} f(x,\omega) \\ \text{subject to} & \phi(x,\xi) \le 0, \quad \forall \xi \in \Xi, \end{array}$$

where Ω and Ξ are finite (again, presumably large) sets, and $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ and $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ are both three times continuously differentiable with respect to their first argument. Given $\hat{\Omega} \subseteq \Omega$, define

$$F_{\hat{\Omega}}(x) \stackrel{\Delta}{=} \max_{\omega\in\hat{\Omega}} f(x,\omega).$$

Given a direction $d \in \mathbb{R}^n$, define a first-order approximation of $F_{\hat{\Omega}}(x+d) - F_{\hat{\Omega}}(x)$ by

$$F'_{\hat{\Omega}}(x,d) \stackrel{\Delta}{=} \max_{\omega\in\hat{\Omega}} \{f(x+d,\omega) + \langle \nabla_x f(x,\omega), d\rangle\} - F_{\hat{\Omega}}(x),$$

and, finally, given a direction $\tilde{d} \in \mathbb{R}^n$, let

$$\tilde{F}'_{\hat{\Omega}}(x,d,\tilde{d}) \stackrel{\Delta}{=} \max_{\omega\in\hat{\Omega}} \{f(x+d,\omega) + \langle \nabla_x f(x,\omega), \tilde{d}\rangle\} - F_{\hat{\Omega}}(x+d).$$

Let Ω_k be the set of objective functions used to compute the search direction at iteration k. The modified QPs follow. To compute $d^0(x, H, \hat{\Omega}, \hat{\Xi})$, we solve

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle d^0, Hd^0\rangle + F'_{\hat{\Omega}}(x, d^0) \\ \text{subject to} & \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^0\rangle \leq 0, \quad \forall \xi \in \hat{\Xi}, \end{array} \qquad QP^0(x, H, \hat{\Omega}, \hat{\Xi})$$

and to compute $d^1(x, \hat{\Omega}, \hat{\Xi})$, we solve

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|d^1\|^2 + \gamma \\ \text{subject to} & F'_{\hat{\Omega}}(x, d^1) \leq \gamma, \\ & \phi(x,\xi) + \langle \nabla_x \phi(x,\xi), d^1\rangle \leq \gamma, \quad \forall \xi \in \hat{\Xi}. \end{array} \qquad QP^1(x, \hat{\Omega}, \hat{\Xi})$$

Finally, to compute $\tilde{d}(x, d, H, \hat{\Omega}, \hat{\Xi})$, we solve

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\langle d + \tilde{d}, H(d + \tilde{d})\rangle + \tilde{F}'_{\hat{\Omega}}(x, d, \tilde{d}) \\ \text{subject to} & \phi(x + d,\xi) + \langle \nabla_x \phi(x,\xi), d + \tilde{d}\rangle \leq \|d\|^\tau, \quad \forall \xi \in \hat{\Xi}, \end{array}$$
$$\widetilde{QP}(x, d, H, \hat{\Omega}, \hat{\Xi})$$

where again, if the QP has no solution or if the solution has norm greater than $\min\{\|d\|, C\}$, we set $\tilde{d}(x, d, H, \hat{\Omega}, \hat{\Xi}) = 0$.

In order to describe the update rules for Ω_k, following [27], we define a few index sets for the objectives (in direct analogy with the index sets for the constraints as introduced in Section 2). The set of indices of "maximizing" objectives is defined in the obvious manner as

$$\Omega_{\max}(x) \stackrel{\Delta}{=} \{\omega \in \Omega \mid f(x,\omega) = F_\Omega(x)\}.$$

At iteration k, let $\mu^0_{k,\omega}$, $\omega \in \Omega_k$, be the multipliers from $QP^0(x_k, H_k, \Omega_k, \Xi_k)$ associated with the objective functions. Likewise, let $\mu^1_{k,\omega}$, $\omega \in \Omega_k$, be the multipliers from $QP^1(x_k, \Omega_k, \Xi_k)$ associated with the objective functions. The set of indices of objective functions which affected the computation of the search direction d_k is given by[2]

$$\Omega^b_k \stackrel{\Delta}{=} \{\omega \in \Omega_k \mid \mu^0_{k,\omega} > 0 \text{ or } \mu^1_{k,\omega} > 0\}.$$

The line search criterion (1.12) is replaced with

$$F_\Omega(x_k + td_k + t^2\tilde{d}_k) \leq F_\Omega(x_k) + \alpha t F'_\Omega(x_k, d_k). \tag{1.15}$$

[2] $QP^1(x, \hat{\Omega}, \hat{\Xi})$ is not needed in the unconstrained case. Accordingly, in [27], Ω^b_k is defined based on a single set of multipliers.

If $t_k < 1$ and the truncation is due to an objective function, then define $\bar{\omega} \in \Omega$ such that

$$f\left(x_k + \frac{t_k}{\beta} d_k + \left(\frac{t_k}{\beta}\right)^2 \tilde{d}_k, \bar{\omega}\right) > F_\Omega(x_k) + \alpha \frac{t_k}{\beta} F'_\Omega(x_k, d_k). \tag{1.16}$$

We are now in a position to state the extended algorithm.

Algorithm FSQP-MOC

Parameters. $\alpha \in (0, \frac{1}{2})$, $\beta \in (0,1)$, and $0 < \delta \ll 1$.

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

Step 0 - Initialization. Set $k \leftarrow 0$. Choose $\Omega_0 \supseteq \Omega_{\max}(x_0)$, $\Xi_0 \supseteq \Xi_{\text{act}}(x_0)$.

Step 1 - Computation of search direction.

(i). Compute $d_k^0 = d^0(x_k, H_k, \Omega_k, \Xi_k)$. If $d_k^0 = 0$, stop.
(ii). Compute $d_k^1 = d^1(x_k, \Omega_k, \Xi_k)$.
(iii). Compute $\rho_k = \rho(d_k^0)$ and set $d_k \leftarrow (1-\rho_k)d_k^0 + \rho_k d_k^1$.
(iv). Compute $\tilde{d}_k = \tilde{d}(x_k, d_k, H_k, \Omega_k, \Xi_k)$.

Step 2 - Line search. Compute t_k, the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying (1.15) and (1.13).

Step 3 - Updates.

(i). Set $x_{k+1} \leftarrow x_k + t_k d_k + t_k^2 \tilde{d}_k$.

(ii). If $t_k < 1$ and (1.15) was violated at $\bar{x}_{k+1} = x_k + \frac{t_k}{\beta} d_k + \left(\frac{t_k}{\beta}\right)^2 \tilde{d}_k$, then let $\bar{\omega}$ be such that (1.16) holds. If (1.13) was violated at $\bar{x}_{k+1}$, then let $\bar{\xi}$ be such that (1.14) holds.

(iii). Pick

$$\Omega_{k+1} \supseteq \Omega_{\max}(x_{k+1}) \cup \Omega_k^b, \quad \text{and}$$
$$\Xi_{k+1} \supseteq \Xi_{\text{act}}(x_{k+1}) \cup \Xi_k^b.$$

If $t_k < 1$ and (1.16) holds for some $\bar{\omega} \in \Omega$, then set $\Omega_{k+1} \leftarrow \Omega_{k+1} \cup \{\bar{\omega}\}$. If $t_k < 1$ and (1.14) holds for some $\bar{\xi} \in \Xi$, then set $\Xi_{k+1} \leftarrow \Xi_{k+1} \cup \{\bar{\xi}\}$.

(iv). If $t_k \leq \delta$ and $\bar{\omega} \notin \Omega_k$ or $\bar{\xi} \notin \Xi_k$ set $H_{k+1} \leftarrow H_k$. Otherwise, obtain a new symmetric positive definite estimate H_{k+1} to the Hessian of the Lagrangian.

(v). Set $k \leftarrow k+1$ and go back to *Step 1*.

5 IMPLEMENTATION AND NUMERICAL RESULTS

Algorithm **FSQP-MOC** has been implemented as part of the code CFSQP[3] [12]. The numerical test results reported in this section were obtained with a modified copy of CFSQP Version 2.4 (the relevant changes will be included in subsequent releases, beginning with Version 2.5). All test problems we consider here are instances of (DSI). Thus in this section we only discuss implementation details relevant to solving such problems, i.e. implementation details of algorithm **FSQP-MC** modified to include the second order correction $\tilde{d}_k$.[4]

The implementation allows for multiple discretized SIP constraints and contains special provisions for those which are affine in x. Specifically, problem (DSI) is generalized to

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \phi_j(x,\xi) \stackrel{\Delta}{=} \langle c_j(\xi), x\rangle - d_j(\xi) \le 0, \quad \forall \xi \in \Xi^{(j)},\ j=1,\ldots,m_\ell, \\ & \phi_j(x,\xi) \le 0, \qquad\qquad\qquad\qquad\quad \forall \xi \in \Xi^{(j)},\ j=m_\ell+1,\ldots,m, \end{array}$$

where $c_j : \Xi_\ell^{(j)} \longrightarrow \mathbb{R}^n$, $j = 1,\ldots,m_\ell$, $d_j : \Xi_\ell^{(j)} \longrightarrow \mathbb{R}$, $j=1,\ldots,m_\ell$, and $\Xi^{(j)}$ is finite, $j=1,\ldots,m$. The assumptions and algorithm statement are generalized in the obvious manner. As far as the analysis of Section 3 is concerned, such a formulation could readily be adapted to the format of (DSI) by grouping all constraints together, i.e. letting $\Xi = \cup_{j=1}^m \Xi^{(j)}$, and ignoring the fact that some may be affine in x.

In the case that the initial point x_0 provided is not feasible for the affine constraints, CFSQP first computes $v \in \mathbb{R}^n$ as the solution of the strictly convex QP

$$\begin{array}{ll} \text{minimize} & \langle v, v\rangle \\ \text{subject to} & \langle c_j(\xi), x_0 + v\rangle - d_j(\xi) \le 0, \quad \forall \xi \in \Xi^{(j)}, \quad j=1,\ldots,m_\ell, \end{array}$$

for $\hat{v}$ such that $x_0 + \hat{v}$ is feasible for linear constraints. If the new initial point is not feasible for all nonlinear constraints, then CFSQP iterates, using the algorithm **FSQP-MOC**, on the constrained minimax problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\max_{j=m_\ell+1,\ldots,m}\ \max_{\xi\in\Xi^{(j)}} \{\phi_j(x,\xi)\} \\ \text{subject to} & \langle c_j(\xi), x\rangle - d_j(\xi) \le 0, \qquad \forall \xi \in \Xi_\ell^{(j)}, \quad j=1,\ldots,m_\ell, \end{array}$$

[3] Available from the authors. See `http://www.isr.umd.edu/Labs/CACSE/FSQP/fsqp.html`

[4] That is, we will not discuss the implementation details specific to the minimax algorithm, even though in the case that the initial guess is infeasible for nonlinear constraints, the minimax algorithm *is* used to generate a feasible point.

until $\max\limits_{j=m_\ell+1,\ldots,m}\ \max\limits_{\xi\in\Xi^{(j)}}\{\phi_j(x,\xi)\} \le 0$ is achieved. The final iterate will be feasible for all constraints, allowing the algorithm to be applied to the original problem.

Recall that it is only required that Ξ_k *contain* certain subsets of Ξ. The algorithm allows for additional elements of Ξ to be included in order to speed up initial convergence. Of course, there is a trade-off between speeding up initial convergence and increasing (i) the number of gradient evaluations and (ii) the size of the QPs. In the implementation, heuristics are applied to add potentially useful elements to Ξ_k (see, e.g. [26] for a discussion of such heuristics). In the case of discretized SIP, one may wish to exploit the knowledge that adjacent discretization points are likely to be closely related. Following [27, 16, 6], for some $\epsilon > 0$, the CFSQP implementation includes in Ξ_k the set $\Xi_\epsilon^{\ell\ell m}(x_k)$ of ϵ-active "left local maximizers" at x_k. At a point $x \in X$, for $j = 1,\ldots,m$, define the ϵ-active discretization points as

$$\Xi_\epsilon^{(j)}(x) \stackrel{\Delta}{=} \{\xi \in \Xi^{(j)} \mid \phi_j(x,\xi) \ge -\epsilon\}.$$

Such a discretization point $\xi_i^{(j)} \in \Xi^{(j)} = \{\xi_1^{(j)},\ldots,\xi_{|\Xi^{(j)}|}^{(j)}\}$ is a left local maximizer if it satisfies *one* of the following three conditions: (i) $i \in \{2,\ldots,|\Xi^{(j)}|-1\}$ and

$$\phi_j(x,\xi_i^{(j)}) > \phi_j(x,\xi_{i-1}^{(j)}) \tag{1.17}$$

and

$$\phi_j(x,\xi_i^{(j)}) \ge \phi_j(x,\xi_{i+1}^{(j)}); \tag{1.18}$$

(ii) $i = 1$ and (1.18); (iii) $i = |\Xi^{(j)}|$ and (1.17). The set $\Xi_\epsilon^{\ell\ell m}(x)$ is the set of all left local maximizers in $\Xi_\epsilon(x) = \cup_{j=1}^m \Xi_\epsilon^{(j)}(x)$. The first part of the update (i.e. before updates due to line search violations) in *Step 3*(iii) of the algorithm becomes

$$\Xi_{k+1} = \Xi_{\text{act}}(x_k) \cup \Xi_k^b \cup \Xi_\epsilon^{\ell\ell m}(x_k).$$

Finally, we have found that in practice, including the end-points (whether or not they are close to being active) during the first iteration often leads to a better initial search direction. Thus we set

$$\Xi_0 = \Xi_{\text{act}}(x_0) \cup \Xi_\epsilon^{\ell\ell m}(x_0) \cup \left(\bigcup_{j=1}^m (\{\xi_1^{(j)}\} \cup \{\xi_{|\Xi^{(j)}|}^{(j)}\})\right).$$

A few other specifics of the CFSQP implementation are as follows. First, as was discussed in Section 2, it is not required to use $QP^1(x,\hat{\Xi})$ to compute our feasible descent direction d^1. In fact, at iteration k, CFSQP uses the following QP, which is a function of the SQP direction d_k^0,

$$\begin{array}{ll} \text{minimize} & \frac{\eta}{2}\|d_k^0 - d^1\|^2 + \gamma \\ \text{subject to} & \langle \nabla f(x_k), d^1 \rangle \le \gamma, \\ & \phi(x_k,\xi) + \langle \nabla_x \phi(x_k,\xi), d^1 \rangle \le \gamma, \quad \forall \xi \in \Xi_k, \end{array}$$

where η was set to 0.1 for our numerical tests. Using such an objective function encourages d_k^1 to be "close" to d_k^0, a condition we have found to be beneficial in practice. It can be verified that the arguments given in Section 3 go through with little modification if we disable the inclusion of d_k^0 in the QP objective function when the step size t_{k-1} from the previous iteration is less than a given small threshold.[5] The expression used for ρ_k is given by

$$\rho_k \triangleq \frac{\|d_k^0\|^\kappa}{\|d_k^0\|^\kappa + \nu_k},$$

where $\nu_k = \max\{0.5, \|d_k^1\|^\tau\}$, with $\kappa = 2.1$ and $\tau = 2.5$ for our numerical experiments. The matrices H_k are updated using the BFGS formula with Powell's modification [21]. The multiplier estimates used for the updates are those obtained from $QP^0(x_k, H_k, \Xi_k)$, with all multipliers corresponding to discretization points outside of Ξ_k set to zero. The QP sub-problems were solved using the routine QLD due to Powell and Schittkowski [25]. Finally, the following parameter values were used for all numerical testing: $\alpha = 0.1$, $\beta = 0.5$, $\epsilon = 1$, and δ was set to the square root of the machine precision.

In order to judge the efficiency of algorithm **FSQP-MOC**, we ran the same numerical tests with two other algorithms differing only in the manner in which Ξ_k is updated. The results are given in Tables 1 and 2. All test algorithms were implemented by making the appropriate modifications to CFSQP Version 2.4. In the tables, the implementation of **FSQP-MOC** just discussed is denoted NEW. A simple ϵ-active strategy was employed in the algorithm we call ϵ-ACT, i.e. we set $\Xi_k = \Xi_\epsilon(x_k)$ for all k, where $\epsilon = 0.1$. The standard FSQP scheme of [17] was applied in algorithm FULL by simply setting $\Xi_k = \Xi$, for all k. All three algorithms were set to stop when $\|d_k^0\| \le 1 \times 10^{-4}$.

We report the numerical results for 13 discretized SIP test problems with discretization levels of 101 and 501. A uniform discretization was used in all cases. Problems `cw_3`, `cw_5`, and `cw_6` are borrowed from [3]. Problems `oet_1` through `oet_7` are from [15]. Finally, `hz_1` is from [10] and `sch_3` is from [26]. In all cases except for `oet_7`, the initial guess x_0 is the same as that given

[5] In the numerical experiments reported here, t_k remained bounded away from 0.

in the reference from which the problem was taken. For `oet_7`, ϵ-ACT and FULL had difficulty generating a feasible point, thus we used the feasible initial guess $x_0 = (0,0,0,-7,-3,-1,3)$. The first two columns of the tables are self-explanatory. A description of the remaining columns is as follows. The third column, n, indicates the number of variables, while m_ℓ and m_n in the next two columns indicate the number of linear SIP constraints and nonlinear SIP constraints ($m_n = m - m_\ell$), respectively. Next, `NF` is the number of objective function evaluations, `NG` is the number of "scalar" constraint function evaluations (i.e. evaluation of some $\phi_j(x,\xi)$ for a given x and ξ), and `IT` indicates the number of iterations required before the stopping criterion was satisfied. Finally, $\sum |\Xi_k|$ is the sum over all iterations of the size of Ξ_k (it is equal to the number of gradient evaluations in the case of NEW and FULL), $|\Xi^*|$ is the size of Ξ_k at the final iterate, and `TIME` is the time of execution in seconds on a Sun Sparc 4 workstation. For all test problems and for all three algorithms, the value of the objective function at the final iterate agreed (to within four significant figures) with the optimal value as given in the original references.

A few conclusions may be drawn from the results. In general, NEW requires the most iterations to "converge" to a solution, whereas FULL requires the least. Typically, though, the difference is not large. Of course, such behavior is expected since NEW uses a simpler QP model at each iteration. It is clear from comparing the results for $\sum |\Xi_k|$ that NEW provides significant savings in the number of gradient evaluations and the size of the QP sub-problems. The savings for ϵ-ACT are not as dramatic. In almost all cases, comparing `TIME` of execution confirms that, indeed, NEW requires far less computational effort than either of the other two approaches. Further, note that $|\Xi^*|$ remains, in general, unchanged for NEW when the discretization level is increased and is typically equal to, or less than, n. This is not the case for ϵ-ACT and FULL, as would be expected. Such behavior suggests that computational effort does not increase with respect to discretization level for NEW at the same rate as it does for ϵ-ACT and FULL. This conclusion is supported by the increase in execution `TIME` when discretization level increases.

PROB	ALGO	n	m_ℓ	m_n	NF	NG	IT	$\sum \lvert\Xi_k\rvert$	$\lvert\Xi^*\rvert$	TIME
cw_3	NEW	3	0	1	10	1625	14	22	2	0.43
	ϵ-ACT				14	2358	18	25	3	0.43
	FULL				20	1866	16	1616	101	0.86
cw_5	NEW	3	1	0	13		13	38	3	0.20
	ϵ-ACT				5		5	347	101	0.19
	FULL				4		4	404	101	0.53
cw_6	NEW	2	0	1	14	1858	15	15	1	0.33
	ϵ-ACT				15	1534	12	32	8	0.26
	FULL				14	1752	15	1515	101	0.71
oet_1	NEW	3	2	0	12		12	57	4	0.32
	ϵ-ACT				7		7	292	46	0.29
	FULL				6		6	1212	202	0.62
oet_2	NEW	3	0	2	6	1283	6	26	3	0.27
	ϵ-ACT				9	3132	9	412	111	0.49
	FULL				4	817	4	808	202	1.23
oet_3	NEW	4	2	0	12		12	62	4	0.34
	ϵ-ACT				8		8	759	202	0.40
	FULL				6		6	1212	202	0.60
oet_4	NEW	4	0	2	19	5711	21	91	4	0.88
	ϵ-ACT				14	5895	16	920	202	1.19
	FULL				12	3147	14	2828	202	2.19
oet_5	NEW	5	0	2	23	8067	24	106	4	1.19
	ϵ-ACT				31	10930	27	3839	202	4.27
	FULL				31	10777	29	5858	202	6.22
oet_6	NEW	5	0	2	23	7099	21	111	6	1.18
	ϵ-ACT				17	6236	18	2359	202	3.15
	FULL				15	4479	15	3030	202	3.46
oet_7	NEW	7	0	2	48	11744	29	188	7	2.36
	ϵ-ACT				647	61831	113	21552	202	50.44
	FULL				372	39721	77	15554	202	37.82
hz_1	NEW	2	0	2	4	1206	6	8	1	0.23
	ϵ-ACT				4	1206	6	68	31	0.24
	FULL				8	1887	10	2020	202	1.03
sch_3	NEW	3	1	0	14		14	41	3	0.19
	ϵ-ACT				5		5	347	101	0.20
	FULL				4		4	404	101	0.31

Table 1 Numerical results for problems with $\lvert\Xi^{(j)}\rvert = 101$.

PROB	ALGO	n	m_ℓ	m_n	NF	NG	IT	$\sum \lvert\Xi_k\rvert$	$\lvert\Xi^*\rvert$	TIME
cw_3	NEW	3	0	1	10	7972	14	22	2	1.08
	ϵ-ACT				15	13599	19	86	12	1.37
	FULL				20	9216	16	8016	501	3.75
cw_5	NEW	3	1	0	47		47	142	2	1.28
	ϵ-ACT				6		6	2213	501	0.71
	FULL				5		5	2505	501	1.17
cw_6	NEW	2	0	1	14	9012	15	15	1	0.88
	ϵ-ACT				15	7509	12	144	40	0.79
	FULL				14	8585	15	7515	501	2.98
oet_1	NEW	3	2	0	18		18	89	4	1.34
	ϵ-ACT				8		8	1582	224	1.02
	FULL				6		6	6012	1002	2.58
oet_2	NEW	3	0	2	6	6270	6	26	3	0.76
	ϵ-ACT				9	15415	9	2010	557	1.83
	FULL				4	4017	4	4008	1002	2.67
oet_3	NEW	4	2	0	15		15	86	4	1.21
	ϵ-ACT				8		8	3734	1002	1.55
	FULL				6		6	12024	1002	2.62
oet_4	NEW	4	0	2	19	26511	21	95	4	2.71
	ϵ-ACT				14	29145	16	4508	1002	5.07
	FULL				12	15531	14	14028	1002	9.80
oet_5	NEW	5	0	2	24	39314	23	102	4	3.91
	ϵ-ACT				22	43177	24	15987	1002	16.9
	FULL				31	52255	29	29058	1002	28.4
oet_6	NEW	5	0	2	23	35073	21	118	7	3.70
	ϵ-ACT				19	33602	19	12688	1002	16.3
	FULL				15	22114	15	15030	1002	15.7
oet_7	NEW	7	0	2	109	149623	73	483	9	17.8
	ϵ-ACT				647	305237	113	106860	1002	250
	FULL				376	196081	77	77154	1002	172
hz_1	NEW	2	0	2	4	5968	6	8	1	0.69
	ϵ-ACT				4	5968	6	320	159	0.88
	FULL				10	11386	12	12024	1002	5.60
sch_3	NEW	3	1	0	48		48	144	2	1.28
	ϵ-ACT				7		7	2714	501	0.79
	FULL				6		6	3006	501	1.54

Table 2 Numerical results for problems with $\lvert\Xi^{(j)}\rvert = 501$.

6 CONCLUSIONS

We have presented and analyzed a feasible SQP algorithm for tackling smooth nonlinear programming problems with a large number of constraints, e.g. those arising from discretization of SIP problems. At each iteration, only a small subset of the constraints are used in the QP sub-problems. Thus, fewer gradient evaluations are required and the computational effort to solve the QP sub-problems is decreased. We showed that the scheme for choosing which constraints are to be included in the QP sub-problems preserves global and fast local convergence. Numerical results obtained from the CFSQP implementation show that, indeed, the algorithm performs favorably.

Acknowledgments. The authors wish to thank Stephan Görner for his many valuable comments on a preliminary draft of the paper. This research was supported in part by NSF grant No. DMI-93-13286 and by the NSF Engineering Research Centers Program NSFD-CDR-88-03012.

REFERENCES

[1] M. C. Biggs. Constrained minimization using recursive equality quadratic programming. In F. A. Lootsma, editor, *Numerical Methods for Non-Linear Optimization*, pages 411–428. Academic Press, New York, 1972.

[2] P. T. Boggs and J. W. Tolle. Sequential quadratic programming. *Acta Numerica*, pages 1–51, 1995.

[3] I. D. Coope and G. A. Watson. A projected lagrangian algorithm for semi-infinite programming. *Math. Programming*, 32:337–356, 1985.

[4] J. W. Daniel. Stability of the solution of definite quadratic programs. *Math. Programming*, 5:41–53, 1973.

[5] C. Gonzaga and E. Polak. On constraint dropping schemes and optimality functions for a class of outer approximation algorithms. *SIAM J. Control Optim.*, 17:477–493, 1979.

[6] C. Gonzaga, E. Polak, and R. Trahan. An improved algorithm for optimization problems with functional inequality constraints. *IEEE Transactions on Automatic Control*, AC-25:49–54, 1980.

[7] S. A. Gustafson. A three-phase algorithm for semi-infinite programs. In A. V. Fiacco and K. O. Kortanek, editors, *Semi-Infinite Programming*

and Applications, Lecture Notes in Control and Information Sciences 215, pages 138–157. Springer Verlag, 1983.

[8] R. Hettich. An implementation of a discretization method for semi-infinite programming. *Math. Programming*, 34:354–361, 1986.

[9] R. Hettich and K. O. Kortanek. Semi-infinite programming: theory, methods, and applications. *SIAM Rev.*, 35:380–429, 1993.

[10] R. Hettich and P. Zencke. *Numerische Methoden der Approximation und Semi-Infiniten Optimierung.* Teubner Studienbücher Mathematik, Stuttgart, Germany, 1982.

[11] K. C. Kiwiel. *Methods of Descent in Nondifferentiable Optimization, Lecture Notes in Mathematics No. 1183.* Springer-Verlag, Berlin, 1985.

[12] C. T. Lawrence, J. L. Zhou, and A. L. Tits. *User's Guide for CFSQP Version 2.4: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints*, 1996. ISR TR-94-16r1, Institute for Systems Research, University of Maryland (College Park, MD).

[13] C. Lemaréchal. Nondifferentiable optimization. In G. Nemhauser, A. Rinooy-Kan, and M. Todd, editors, *Optimization, Handbooks in Operations Research and Management Science.* Elsevier Science, North Holland, 1989.

[14] 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 Transactions on Automatic Control*, AC-29:1040–1042, 1984.

[15] K. Oettershagen. *Ein Superlinear Konvergenter Algorithmus zur Lösung Semi-Infiniter Optimierungsprobleme.* PhD thesis, Bonn University, 1982.

[16] 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 Transactions on Automatic Control*, AC-34(8):903–908, 1989.

[17] E. R. Panier and A. L. Tits. On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Math. Programming*, 59:261–276, 1993.

[18] E. Polak and L. He. Rate preserving discretization strategies for semi-infinite programming and optimal control. In *Proceedings of the 29th Conference on Decision and Control*, December 1990.

[19] E. Polak and D. Q. Mayne. An algorithm for optimization problems with functional inequality constraints. *IEEE Transactions on Automatic Control*, AC-21:184–193, 1976.

[20] E. Polak and A. L. Tits. A recursive quadratic programming algorithm for semi-infinite optimization problems. *Appl. Math. Optim.*, 8:325–349, 1982.

[21] M. J. D. Powell. A fast algorithm for nonlinearly constrained optimization calculations. In G. A. Watson, editor, *Numerical Analysis, Dundee, 1977, Lecture Notes in Mathematics 630*, pages 144–157. Springer Verlag, 1978.

[22] M. J. D. Powell. A tolerant algorithm for linearly constrained optimization calculations. *Math. Programming*, 45:547–566, 1989.

[23] R. Reemtsen. Discretization methods for the solution of semi-infinite programming problems. *J. Optim. Theory Appl.*, 71:85–103, 1991.

[24] S. M. Robinson. Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms. *Math. Programming*, 7:1–16, 1974.

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

[26] K. Schittkowski. Solving nonlinear programming problems with very many constraints. Technical Report No. 294, Mathematisches Institut, Universität Bayreuth, Germany, 1991.

[27] J. L. Zhou and A. L. Tits. An SQP algorithm for finely discretized continuous minimax problems and other minimax problems with many objective functions. *SIAM J. on Optimization*, pages 461–487, May 1996.

[28] G. Zoutendijk. *Methods of Feasible Directions*. Elsevier Science, Amsterdam, The Netherlands, 1960.

APPENDIX A

PROOFS

The following lemmas will be used in the proof of Proposition 1.

Lemma 7 *Given $x \in \mathbb{R}^n$ and $H > 0$, suppose $\Xi' \subset \Xi'' \subseteq \Xi$. (i) If $d^0(x, H, \Xi')$ is not feasible for $QP^0(x, H, \Xi'')$, then $v^0(x, H, \Xi'') < v^0(x, H, \Xi')$, and (ii) if $d^1(x, \Xi')$ is not feasible for $QP^1(x, \Xi'')$, then $v^1(x, \Xi'') < v^1(x, \Xi')$.*

Proof: First $d^0(x, H, \Xi'') \neq d^0(x, H, \Xi')$, since by assumption $d^0(x, H, \Xi')$ is not feasible for $QP^0(x, H, \Xi'')$. On the other hand, since $\Xi' \subset \Xi''$, $d^0(x, H, \Xi'')$ is feasible for $QP^0(x, H, \Xi')$. Uniqueness of the solution of $QP^0(x, H, \Xi')$ then implies the claim. Part *(ii)* is proved similarly. □

Lemma 8 *Suppose $\mathcal{K}$ is an infinite index set such that*

$$x_k \xrightarrow{k \in \mathcal{K}} x^*, \quad H_k \xrightarrow{k \in \mathcal{K}} H^*, \quad d_k^0 \xrightarrow{k \in \mathcal{K}} d^{0,*}, \quad d_k^1 \xrightarrow{k \in \mathcal{K}} d^{1,*}, \quad \gamma_k \xrightarrow{k \in \mathcal{K}} \gamma^*,$$

where x^* is not a KKT point for (DSI), and suppose $\Xi_k = \hat{\Xi}$ for all $k \in \mathcal{K}$. Then there exists $\underline{t} > 0$ such that for all $t \in [0, \underline{t}]$, $\phi(x_k + t d_k, \xi) \leq 0$, for all $\xi \in \hat{\Xi}$, and for all $k \in \mathcal{K}$ sufficiently large.

Proof: By definition of d_k^1 and γ_k, for all $k \in \mathcal{K}$, $\phi(x_k, \xi) + \langle \nabla_x \phi(x_k, \xi), d_k^1 \rangle \leq \gamma_k$, for all $\xi \in \hat{\Xi}$. Since x_k is not a KKT point, $d_k^1 \neq 0$ and $\gamma_k < 0$, for all $k \in \mathcal{K}$ (Lemma 2). Further, in view of Lemma 4 $(d^{1,*}, \gamma^*)$ solves $QP^1(x^*, \hat{\Xi})$, and since x^* is not a KKT point, $d^{1,*} \neq 0$ and $\gamma^* < 0$ (Lemma 2). Thus, there exists $\bar{\gamma} < 0$ (e.g. $\bar{\gamma} = \gamma^*/2$) such that for all $k \in \mathcal{K}$, $\phi(x_k, \xi) + \langle \nabla_x \phi(x_k, \xi), d_k^1 \rangle \leq \bar{\gamma}$, for all $\xi \in \hat{\Xi}$. It follows that there exists $\delta > 0$ and $\underline{k}$ such that for all $k \in \mathcal{K}$, $k \geq \underline{k}$,

$$\langle \nabla_x \phi(x_k, \xi), d_k^1 \rangle \leq -\delta, \quad \forall \xi \in \hat{\Xi} \cap \Xi_{\text{act}}(x^*)$$

$$\phi(x_k, \xi) \leq -\delta, \quad \forall \xi \in \hat{\Xi} \setminus (\hat{\Xi} \cap \Xi_{\text{act}}(x^*)).$$

Next, in view of Lemma 1, $d_k^0 \neq 0$ and $\langle \nabla_x \phi(x_k, \xi), d_k^0 \rangle \leq 0$, for all $\xi \in \Xi_{\text{act}}(x_k)$, for all $k \in \mathcal{K}$. On the other hand, applying Lemma 4 allows us to conclude $d^{0,*}$ solves $QP(x^*, H^*, \hat{\Xi})$. Hence, from Lemma 6, since x^* is not a KKT point for

(DSI), $d^{0,*} \neq 0$. Since $\rho(\cdot)$ is assumed to be bounded away from zero outside every neighborhood of zero, there exists $\underline{\rho} > 0$ such that $\rho_k = \rho(d_k^0) \geq \underline{\rho}$, for all $k \in \mathcal{K}$. It follows that

$$\begin{aligned}\langle \nabla_x \phi(x_k, \xi), d_k \rangle &= (1-\rho_k)\langle \nabla_x \phi(x_k, \xi), d_k^0 \rangle + \rho_k \langle \nabla_x \phi(x_k, \xi), d_k^1 \rangle \\ &\leq -\underline{\rho}\delta, \quad \forall \xi \in \hat{\Xi} \cap \Xi_{\text{act}}(x^*),\end{aligned}$$

for all $k \in \mathcal{K}$, $k \geq \underline{k}$. Now let $\mathcal{Q} \triangleq \{x_k |\ k \in \mathcal{K}\} \cup \{x^*\}$, $\mathcal{D} \triangleq \{d_k |\ k \in \mathcal{K}\} \cup \{d^*\}$ and define

$$M(t, \xi) \triangleq \max_{x \in \mathcal{Q}} \max_{d \in \mathcal{D}} \max_{\eta \in [0,1]} \|\nabla_x \phi(x + t\eta d, \xi) - \nabla_x \phi(x, \xi)\| \cdot \|d\|,$$

which is well-defined and continuous in t for all $\xi \in \hat{\Xi}$, since $\mathcal{Q}$ and $\mathcal{D}$ are compact. Now for all $k \in \mathcal{K}$, $\xi \in \hat{\Xi}$ we have

$$\begin{aligned}&\phi(x_k + td_k, \xi) - \phi(x_k, \xi) \\ &= \int_0^1 \langle \nabla_x \phi(x_k + t\eta d_k, \xi), d_k \rangle d\eta \\ &= t\left\{\int_0^1 \langle \nabla_x \phi(x_k + t\eta d_k, \xi) - \nabla_x \phi(x_k, \xi), d_k \rangle d\eta + \langle \nabla_x \phi(x_k, \xi), d_k \rangle\right\} \\ &\leq t\left\{\sup_{\eta \in [0,1]} \|\nabla_x \phi(x_k + t\eta d_k, \xi) - \nabla_x \phi(x_k, \xi)\| \cdot \|d_k\| + \langle \nabla_x \phi(x_k, \xi), d_k \rangle\right\} \\ &\leq t\left\{M(t, \xi) + \langle \nabla_x \phi(x_k, \xi), d_k \rangle\right\}. \qquad \text{(A.1)}\end{aligned}$$

Further note that $M(0, \xi) = 0$, for all $\xi \in \hat{\Xi}$. For $\xi \in \hat{\Xi} \cap \Xi_{\text{act}}(x^*)$, define $\underline{t}_\xi$ such that $M(t, \xi) < \underline{\rho}\delta$ for all $t \in [0, \underline{t}_\xi]$. For all $\xi \in \hat{\Xi} \setminus (\hat{\Xi} \cap \Xi_{\text{act}}(x^*))$, our regularity assumptions and boundedness of $\{x_k\}$ and $\{d_k\}$ imply there exist $M_{1,\xi} > 0$ and $M_{2,\xi} > 0$ such that

$$|\langle \nabla_x \phi(x_k, \xi), d_k \rangle| \leq M_{1,\xi},\ \forall k, \quad \text{and} \quad \max_{t \in [0,1]} |M(t, \xi)| \leq M_{2,\xi}.$$

For such ξ, define $\underline{t}_\xi = \delta/(M_{1,\xi} + M_{2,\xi})$. Then $t\{M(t, \xi) + \langle \nabla_x \phi(x_k, \xi), d_k \rangle\} \leq \delta$, for all $t \in [0, \underline{t}_\xi]$, $\xi \in \hat{\Xi} \setminus (\hat{\Xi} \cap \Xi_{\text{act}}(x^*))$. Finally, set $\underline{t} = \max_{\xi \in \hat{\Xi}} \underline{t}_\xi$. From (A.1) it is easily verified that $\underline{t}$ is as claimed. □

Proof of Proposition 1. We argue by contradiction. Suppose that

$$v^* \triangleq \liminf_k v_k > 0. \qquad \text{(A.2)}$$

As all sequences of interest are bounded, there exists an infinite index set $\mathcal{K}$ such that

$$\begin{array}{llll} v_k \xrightarrow{k\in\mathcal{K}} v^*, & x_k \xrightarrow{k\in\mathcal{K}} x^*, & H_k \xrightarrow{k\in\mathcal{K}} H^*, & \rho_k \xrightarrow{k\in\mathcal{K}} \rho^*, \\ d_k^0 \xrightarrow{k\in\mathcal{K}} d^{0,*}, & d_{k+1}^0 \xrightarrow{k\in\mathcal{K}} d_+^{0,*}, & d_k^1 \xrightarrow{k\in\mathcal{K}} d^{1,*}, & d_{k+1}^1 \xrightarrow{k\in\mathcal{K}} d_+^{1,*}, \\ v_k^0 \xrightarrow{k\in\mathcal{K}} v^{0,*}, & v_{k+1}^0 \xrightarrow{k\in\mathcal{K}} v_+^{0,*}, & v_k^1 \xrightarrow{k\in\mathcal{K}} v^{1,*}, & v_{k+1}^1 \xrightarrow{k\in\mathcal{K}} v_+^{1,*}, \\ \gamma_k \xrightarrow{k\in\mathcal{K}} \gamma^*, & \gamma_{k+1} \xrightarrow{k\in\mathcal{K}} \gamma_+^*. & & \end{array}$$

Further, since the number of possible subsets of Ξ is finite, we may assume that on $\mathcal{K}$, the sets $\Xi_k^{b,0}$ and $\Xi_k^{b,1}$ are constant and equal to $\hat{\Xi}^{b,0}$ and $\hat{\Xi}^{b,1}$, respectively. Thus, for all k, d_k^0 solves $QP^0(x_k, H_k, \hat{\Xi}^{b,0})$ and d_k^1 solves $QP^1(x_k, \hat{\Xi}^{b,1})$. Note that in view of the definition of $\Xi_k^{b,0}$ and $\Xi_k^{b,1}$, the sequences constructed by the algorithm are identical to those that would have been constructed with $\Xi' \triangleq \hat{\Xi}^{b,0} \cup \hat{\Xi}^{b,1}$ substituted for Ξ_k for all k. Without loss of generality we thus assume that $\Xi_k = \Xi'$, for all k. Finally, define $d^* = (1-\rho^*)d^{0,*} + \rho^* d^{1,*}$.

In view of Lemma 4, $d^{0,*}$ and $d^{1,*}$ are the unique solutions of $QP^0(x^*, H^*, \Xi')$ and $QP^1(x^*, \Xi')$. Now, of course, x^* is not a KKT point for (DSI), otherwise, in view of Lemmas 1 and 2, $d^{0,*} = d^{1,*} = 0$, which would imply $v^* = v^{0,*} + v^{1,*} = 0$, contradicting (A.2). Hence, $d^{0,*} \neq 0$ and $d^{1,*} \neq 0$, and both are directions of descent for $f(\cdot)$ at x^*. This further implies $d^* \neq 0$ and $\langle \nabla f(x^*), d^* \rangle < 0$. Therefore, applying Lemma 5*(iii)*, we conclude that $t_k \xrightarrow{k\in\mathcal{K}} 0$. Without loss of generality, assume that $t_k < \min\{\delta, \underline{t}\}$, for all $k \in \mathcal{K}$, where $\underline{t}$ is as given by Lemma 8 and $\delta > 0$ is as in the algorithm. The fact that $t_k < \delta < 1$ implies that for all $k \in \mathcal{K}$ the line search criterion of *Step 2* is not satisfied at $\bar{x}_{k+1} = x_k + \frac{t_k}{\beta} d_k$. Since $\alpha < 1$ (indeed, $\alpha < 1/2$) using a standard argument it follows that (1.4) is violated at $\bar{x}_{k+1}$ only finitely many times. Thus, without loss of generality, assume (1.6) holds for all $k \in \mathcal{K}$, i.e.

$$\phi(\bar{x}_{k+1}, \bar{\xi}) > 0, \quad \forall k \in \mathcal{K}.$$

Further, we have assumed (since there are only a finite number of constraints) that the violation is caused by the same constraint, with index $\bar{\xi}$, for all $k \in \mathcal{K}$. In view of Lemma 8, we may conclude that $\bar{\xi} \notin \Xi'$. Thus, according to *Step 3(iv)*, $H_{k+1} = H_k$, for all $k \in \mathcal{K}$.

Since, for all $k \in \mathcal{K}$, $H_{k+1} = H_k$ and $\Xi_k = \Xi' = \Xi_k^{b,0} \cup \Xi_k^{b,1}$, the directions d_{k+1}^0 and d_{k+1}^1 solve $QP^0(x_{k+1}, H_k, \Xi_{k+1})$ and $QP^1(x_{k+1}, \Xi_{k+1})$, for all $k \in \mathcal{K}$, where, for some Ξ'',

$$\Xi_{k+1} = \Xi'' \supseteq \Xi' \cup \{\bar{\xi}\}.$$

Without loss of generality, since the number of constraints is finite, we may assume that the set of indices in Ξ_{k+1} and not in $\Xi' \cup \{\bar{\xi}\}$ is constant for all $k \in \mathcal{K}$. Further, in view of Lemma 5(iii), $x_{k+1} \xrightarrow{k\in\mathcal{K}} x^*$. It follows that, in view of Lemma 4, the limits $d_+^{0,*}$ and $d_+^{1,*}$ are the unique solutions of $QP^0(x^*, H^*, \Xi'')$ and $QP^1(x^*, \Xi'')$. Since, $\phi(\bar{x}_{k+1}, \bar{\xi}) > 0$ and $\phi(x_{k+1}, \bar{\xi}) \leq 0$, for all $k \in \mathcal{K}$, and since Lemma 5(iii) also implies $\bar{x}_{k+1} \xrightarrow{k\in\mathcal{K}} x^*$, we see that $\phi(x^*, \bar{\xi}) = 0$. By considering a first-order expansion of $\phi(\bar{x}_{k+1}, \bar{\xi}) - \phi(x_{k+1}, \bar{\xi})$, and taking limits, we see that $\langle \nabla_x \phi(x^*, \bar{\xi}), d^* \rangle \geq 0$. Note that since $d^{0,*} \neq 0$ and $d_k^0 \neq 0$, for all $k \in \mathcal{K}$, d_k^0 is bounded away from zero. By our assumptions on $\rho(\cdot)$, ρ_k is thus bounded away from zero and $\rho^* > 0$. This implies that either $\langle \nabla_x \phi(x^*, \bar{\xi}), d^{1,*} \rangle \geq 0$, or $\langle \nabla_x \phi(x^*, \bar{\xi}), d^{0,*} \rangle > 0$. If the first inequality holds, then $(d^{1,*}, \gamma^*)$ is infeasible for $QP^1(x^*, \Xi'')$ (recall that $\phi(x^*, \bar{\xi}) = 0$ and, from Lemmas 4 and 2, $\gamma^* < 0$) and, in view of Lemma 7, $v_+^{1,*} < v^{1,*}$. Similarly, if the second inequality holds, then $d^{0,*}$ is infeasible for $QP^0(x^*, H^*, \Xi'')$, and $v_+^{0,*} < v^{0,*}$. In view of (A.2), in both cases we have a contradiction. □

The following sequence of Lemmas will be used in the proof of Proposition 2.

Lemma 9 *There exists an infinite index set $\mathcal{K}$ such that, for all $k \in \mathcal{K}$, (i) $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,0}$, and (ii) $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,1}$.*

Proof: In view of Proposition 1, since for all k $v_k^0 \geq 0$ and $v_k^1 \geq 0$, there exists an infinite index set $\mathcal{K}$ such that both $v_k^0 \xrightarrow{k\in\mathcal{K}} 0$ and $v_k^1 \xrightarrow{k\in\mathcal{K}} 0$. By Lemma 6, $d_k^0 \xrightarrow{k\in\mathcal{K}} 0$ and $d_k^1 \xrightarrow{k\in\mathcal{K}} 0$. To prove (i), let $\lambda_{k,\xi}^0$, $\xi \in \Xi_k$, be the multipliers from $QP^0(x_k, H_k, \Xi_k)$ and let $\lambda_{k,\xi}^0 = 0$, for all $\xi \notin \Xi_k$. Assume, without loss of generality, that $\Xi_k^{b,0} = \hat{\Xi}^0$ for all $k \in \mathcal{K}$ and $H_k \xrightarrow{k\in\mathcal{K}} H^*$. Since $\mathcal{H}$ is compact, and in view of Assumptions 2 and 5, we may apply Theorem 2.1 of [24] to show that $\lambda_{k,\xi}^0 \xrightarrow{k\in\mathcal{K}} \lambda_\xi^{0,*}$, $\xi \in \Xi$, the KKT multipliers for $QP^0(x^*, H^*, \hat{\Xi}^0)$. Note that the KKT conditions (1.2) for $QP^0(x^*, H^*, \hat{\Xi}^0)$ are equivalent to the KKT conditions (1.1) for (DSI) at x^* with multipliers $\lambda_\xi^{0,*}$, $\xi \in \Xi$. Uniqueness of the multipliers at x^* (Assumption 2) and strict complementarity imply $\lambda_\xi^{0,*} > 0$ if $\xi \in \Xi_{\text{act}}(x^*)$. Therefore, $\Xi_{\text{act}}(x^*) \subseteq \hat{\Xi}^{b,0}$, which means $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,0}$, for all $k \in \mathcal{K}$. Part (ii) is proved similarly. □

Lemma 10 *Given $\epsilon > 0$, there exists $\delta > 0$ such that for every $x \in X$ satisfying $\|x - x^*\| < \delta$, every $H \in \mathcal{H}$, and every $\hat{\Xi} \subseteq \Xi$ with $\Xi_{\text{act}}(x^*) \subseteq \hat{\Xi}$,*

(i) all $\xi \in \Xi_{\text{act}}(x^*)$ are binding for $QP^0(x, H, \hat{\Xi})$ and $\|d^0(x, H, \hat{\Xi})\| < \epsilon$, and

(ii) all $\xi \in \Xi_{\text{act}}(x^*)$ are binding for $QP^1(x, \hat{\Xi})$ and $\|d^1(x, \hat{\Xi})\| < \epsilon$.

Proof: Given $H \in \mathcal{H}$ and $\hat{\Xi} \subseteq \Xi$ such that $\Xi_{\text{act}}(x^*) \subseteq \hat{\Xi}$, Lemmas 1 and 2 imply that $d^0(x^*, H, \hat{\Xi}) = d^1(x^*, \hat{\Xi}) = 0$. Since $\mathcal{H}$ is compact, Assumptions 2 and 5 allow us to apply Theorem 2.1 of [24] to conclude that, given $\epsilon > 0$, there exists $\delta_{\hat{\Xi}} > 0$ such that for all x satisfying $\|x - x^*\| < \delta_{\hat{\Xi}}$ and all $H \in \mathcal{H}$, the QP multipliers from $QP^0(x, H, \hat{\Xi})$ and $QP^1(x, \hat{\Xi})$ are positive for all $\xi \in \Xi_{\text{act}}(x^*)$, $\|d^0(x, H, \hat{\Xi})\| < \epsilon$, and $\|d^1(x, \hat{\Xi})\| < \epsilon$. As Ξ is a finite set, δ may be chosen independently of $\hat{\Xi}$. □

Lemma 11 *For k sufficiently large $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,0}$ and $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,1}$.*

Proof: For an arbitrary $\epsilon > 0$, let $\delta > 0$ be as given by Lemma 10. In view of Theorem 1, there exists $\underline{k}$ such that $\|x - x^*\| < \delta$ for all $k \geq \underline{k}$. By Lemma 9, there exists an infinite index set $\mathcal{K}$ such that $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,0}$, and $\Xi_{\text{act}}(x^*) \subseteq \Xi_k^{b,1}$, for all $k \in \mathcal{K}$. Choose $\underline{k}' \geq \underline{k}$, $\underline{k}' \in \mathcal{K}$. It follows that $\Xi_{\text{act}}(x^*) \subseteq \Xi_{\underline{k}'+1}$. The result follows by induction and Lemma 10. □

Lemma 12 *$d_k^0 \longrightarrow 0$ and $d_k^1 \longrightarrow 0$.*

Proof: Follows immediately from Lemma 11, *Step 3(iii)* of algorithm **FSQP-MC**, Assumption 4, and Lemma 10. □

Proof of Proposition 2. For (i), in view of Lemma 11, it suffices to show that, for k sufficiently large, $\Xi_k^{b,0} \subseteq \Xi_{\text{act}}(x^*)$ and $\Xi_k^{b,1} \subseteq \Xi_{\text{act}}(x^*)$. Suppose $\hat{\xi} \in \Xi \setminus \Xi_{\text{act}}(x^*)$, i.e. $\phi(x^*, \hat{\xi}) < 0$. Since $x_k \longrightarrow x^*$, by continuity we have $\phi(x_k, \hat{\xi}) < 0$ for all k sufficiently large. In view of Lemma 12, for k sufficiently large we have

$$\phi(x_k, \hat{\xi}) + \langle \nabla_x \phi(x_k, \hat{\xi}), d_k^0 \rangle < 0.$$

Therefore, $\lambda_{k,\hat{\xi}}^0 = 0$ (hence $\hat{\xi} \notin \Xi_k^{b,0}$) for all k sufficiently large. The argument is identical for $\Xi_k^{b,1}$. Part (ii) follows from Theorem 1, Lemma 12, and our regularity assumptions. □