Local convergence of SQP methods for Mathematical Programs with Equilibrium Constraints*

ROGER FLETCHER, SVEN LEYFFER†, DANNY RALPH AND STEFAN SCHOLTES‡

9 May 2002

Abstract

Recently, it has been shown that Nonlinear Programming solvers can successfully solve a range of Mathematical Programs with Equilibrium Constraints (MPECs). In particular, Sequential Quadratic Programming (SQP) methods have been very successful.

This paper examines the local convergence properties of SQP methods applied to MPECs. It is shown that SQP converges superlinearly under reasonable assumptions near a strongly stationary point. A number of illustrative examples are presented which show that some of the assumptions are difficult to relax.

Keywords: Nonlinear programming, SQP, MPEC, MPCC, equilibrium constraints.
AMS-MSC2000: 90C30, 90C33, 90C55, 49M37, 65K10.

1 Introduction

We consider Mathematical Programs with Equilibrium Constraints (MPECs) of the form

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c_{\mathcal{E}}(z) = 0 \\ & c_{\mathcal{I}}(z) \geq 0 \\ & 0 \leq z_1 \ \perp \ z_2 \geq 0, \end{array} \tag{1.1}$$

where $z = (z_0, z_1, z_2)$ is a decomposition of the problem variables into controls $z_0 \in \mathbb{R}^n$ and states $(z_1, z_2) \in \mathbb{R}^{2p}$. The equality constraints $c_i(z) = 0,\ i \in \mathcal{E}$ are abbreviated as $c_{\mathcal{E}}(z) = 0$ and similarly, $c_{\mathcal{I}}(z) \geq 0$ represents the inequality constraints. Problems of this type arise frequently in applications, see [8, 16, 17] for references. Problem (1.1) is also referred to as a Mathematical Program with Complementarity Constraints (MPCC).

Clearly, an MPEC with a more general complementarity condition like

$$0 \leq G(z) \ \perp \ H(z) \geq 0 \tag{1.2}$$

can be written in the form (1.1) by introducing slack variables. It is easy to show that the reformulated MPEC has the same properties (such as constraint qualifications or second

*Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee.
†Department of Mathematics, University of Dundee, ({fletcher,sleyffer}@maths.dundee.ac.uk).
‡The Judge Institute, University of Cambridge ({d.ralph,s.scholtes}@jims.cam.ac.uk).

order conditions) as the original MPEC. In this sense, nothing is lost by introducing slacks.

One possible way of solving (1.1) is to consider its equivalent nonlinear programming NLP formulation,

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c_{\mathcal{E}}(z) = 0 \\ & c_{\mathcal{I}}(z) \geq 0 \\ & z_1 \geq 0 \\ & z_2 \geq 0 \\ & z_1^T z_2 \leq 0, \end{array} \tag{1.3}$$

and solve (1.3) with existing NLP solvers. This paper examines the local convergence properties of Sequential Quadratic Programming (SQP) methods applied to (1.3).

It is obvious that the NLP (1.3) has no feasible point which satisfies the inequalities strictly. This implies that the Mangasarian Fromovitz Constraint Qualification (MFCQ) is violated at every feasible point, see [5, 19]. Since MFCQ is a sufficient condition for stability of an NLP, the lack of MFCQ in (1.3) has been advanced as a theoretical argument against the use of standard NLP solvers.

Numerical experience with (1.3) in the past has also been rather disappointing. Bard [3] reports failure on 50 - 70 % of some bilevel problems for a gradient projection method. Conn et al. [6] and Ferris and Pang [8] attribute certain failures of lancelot to the fact that the problem contains a complementarity constraint. In contrast to this, Fletcher and Leyffer [10] reported recently very encouraging numerical results on a large collection of MPECs [15]. They solved over 100 MPECs with an SQP solver, and observe *quadratic* convergence for all but two problems. The two problems which do not give quadratic convergence violate certain MPEC regularity conditions and are rather pathological. The present work complements these numerical observations by giving a theoretical explanation for the good performance of the SQP method on apparently ill-posed problems of the type (1.3). We show that SQP is guaranteed to converge quadratically near a stationary point under relatively mild assumptions.

The paper also complements the recently renewed interest in the convergence properties of SQP under weaker assumptions, see e.g. [9, 13, 20]. These studies suggest modifications for SQP solvers to enable them to handle NLP problems for which the constraint gradients are linearly dependent at the solution and/or strict complementarity fails to hold.

Anitescu [2] extends Wright's analysis [20] to NLPs with unbounded multiplier sets. The fact that (1.3) violates MFCQ is equivalent to the unboundedness of its multiplier set. Anitescu's work does therefore apply to MPECs in the given form. However, his assumptions are different from ours and neither assumptions are implied by the others. Most notably, Anitescu assumes that the QP solver employs an elastic mode, relaxing constraint linearizations if they are inconsistent. We do not require such a modification and provide a local analysis of the SQP method in its pure form.

An alternative approach to MPECs via NLP is considered by Andreani and Martínez [1]. They argue that an MPEC which satisfies strict complementarity and strong stationarity can be solved successfully by algorithms which satisfy a certain AGP property. A new formulation is also introduced, namely $z_{1i} z_{2i} = 0, z_{1i} + z_{2i} \geq 0$.

In this paper, we argue that the introduction of slacks is not just a convenience, but plays an important role in ensuring convergence. In Section 7.2 an example with a nonlinear complementarity constraint is presented for which SQP converges to a nonstationary point. All QP approximations remain consistent during the solve. By introducing slacks, on the other hand, SQP converges to a stationary point. Of course, this does not mean that the use of slacks makes an elastic mode or a feasibility restoration unnecessary. The example in Section 2.2 clearly shows that NLP solvers must be able to handle inconsistent QPs.

This paper is organized as follows. The next section gives a few simple motivating examples which highlight the key ideas of our approach and illustrate the numerical difficulties associated with MPECs. In Section 3 optimality conditions and constraint qualifications for MPECs are reviewed. Section 4 shows that the optimality conditions of (1.1) and (1.3) are related by a simple formula. In Section 5 it is shown that SQP converges quadratically in two distinct situations. The first arises when SQP is started close to a complementary stationary point. If the starting point is not complementary, then we show convergence under the assumption that all QP subproblems remain consistent. Sufficient conditions for this assumption are introduced in Section 6. Finally, in Section 7 more small examples are presented which illustrate the necessity of some of the assumptions. The paper concludes by briefly emphasizing the importance of degeneracy-handling at the QP level before pointing to future research directions.

Throughout the paper, $g(z) = \nabla f(z)$ is the objective gradient and the constraint gradients are denoted by $a_i(z) = \nabla c_i(z)$. Superscripts refer to the point at which functions or gradients are evaluated, e.g. $a_i^{(k)} = a_i(z^{(k)}) = \nabla c_i(z^{(k)})$. Finally, the Jacobian matrices are denoted by $A_{\mathcal{E}} := [a_i]_{i \in \mathcal{E}}$ and $A_{\mathcal{I}} := [a_i]_{i \in \mathcal{I}}$, respectively.

2 Some illustrative example

The fact that the NLP formulation (1.3) of an MPEC violates MFCQ at any feasible point implies that (1.3) has certain features which pose numerical challenges to NLP solvers.

1. The active constraint normals are *linearly dependent* at any feasible point.

2. The set of multipliers is *unbounded.*

3. The linearizations of (1.3) can be *inconsistent, arbitrarily close* to a stationary point.

These features are illustrated by the following examples. The examples also motivate the analysis in subsequent sections. The main conclusion of this section is that while MPECs posses these unpleasant properties, these arise in a well-structured way which allows SQP solvers to tackle MPECs successfully.

Below and in the remainder, `*.mod` refers to the AMPL model of the problem in MacMPEC, an AMPL collection of MPECs [15].

2.1 Dependent constraint normals and unbounded multipliers (`jr*.mod`)

In this section we use a small example due to Jiang and Ralph [14] to illustrate the key idea of our approach. Consider the two QPECs

$$\left\{ \begin{array}{ll} \underset{z}{\text{minimize}} & f_i(z) \\ \text{subject to} & 0 \leq z_2 \perp z_2 - z_1 \geq 0 \end{array} \right. \tag{2.1}$$

with $f_1(z) = (z_1 - 1)^2 + z_2^2$ and $f_2(z) = z_1^2 + (z_2 - 1)^2$. The problems differ only in their objectives. The solution to both problems is $z^* = (1/2, 1/2)^T$, see Figure 1.

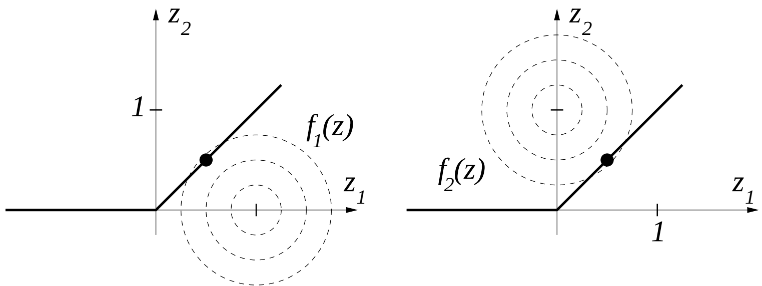

Figure 1: QPEC examples 1 and 2

The equivalent NLP problem to these QPECs is given by

$$\left\{ \begin{array}{lll} \underset{z}{\text{minimize}} & f_i(z) & \text{multiplier} \\ \text{subject to} & z_2 \geq 0 & \nu \geq 0 \\ & z_2 - z_1 \geq 0 & \lambda \geq 0 \\ & z_2\,(z_2 - z_1) \leq 0 & \xi \geq 0. \end{array} \right. \tag{2.2}$$

The first order conditions for these NLPs differ only in the objective gradient and are

$$\begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda^* \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \xi^* \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Clearly, the two active constraint normals are linearly dependent. Since $z_2^* = \frac{1}{2} > 0$ it follows that $\nu^* = 0$. The multiplier sets, given by

$$\begin{aligned} \mathcal{M}_1 &= \left\{ (\lambda, \xi) \mid \lambda \geq 0,\ \lambda - \tfrac{1}{2}\xi = 1 \right\} \\ \mathcal{M}_2 &= \left\{ (\lambda, \xi) \mid \lambda \geq 0,\ -\lambda + \tfrac{1}{2}\xi = 1 \right\}. \end{aligned}$$

are unbounded, as expected. The sets are shown in Figure 2.

This situation is typical for MPECs which satisfy a strong stationarity condition. The multiplier set is a ray and there is exactly one degree of freedom in the choice of multipliers.

Note however, that if we restrict attention to multipliers which correspond to a *linearly independent set* of constraint normals, then the following reduced sets are obtained

$$\begin{aligned} \tilde{\mathcal{M}}_1 &= \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \\ \tilde{\mathcal{M}}_2 &= \left\{ \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\}. \end{aligned}$$

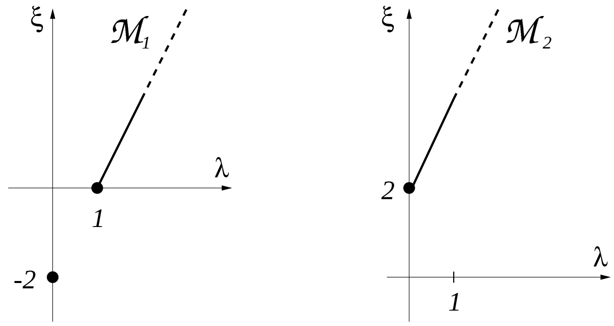

Figure 2: Multiplier sets of QPEC examples 1 and 2

These multipliers are bounded and well behaved. We should expect SQP to converge if started near such a stationary point. The KKT multipliers which correspond to a solution with linearly independent strictly active constraints are illustrated by the black circles in Figure 2. The half-line shows the unbounded multiplier set.

Observe that in the first example, $\lambda \geq 0$ at the solution which implies that this is also the solution for the NLP with the complementarity condition removed. In the second example, no $\lambda \geq 0$ can on its own satisfy the stationarity conditions and $\xi > 0$ is required. If we had interpreted $z_2 - z_1 \geq 0$ as an equality constraint, then we could have chosen $\lambda = -1$ in the stationarity conditions. However, an NLP solver would never return $\lambda < 0$ for an inequality constraint and hence $\xi = 2$ ensures that the stationarity conditions are satisfied.

The effect of the multiplier of the complementarity constraint is to relax the condition that $\lambda, \nu \geq 0$ for what in effect amounts to an equality constraint. This is exploited in Section 4, where it is shown that certain MPEC multipliers correspond to multipliers of (1.3). This situation is typical for MPECs under certain assumptions. The key idea is to then show that SQP converges to a solution provided the QP solver chooses a linearly independent basis.

2.2 Inconsistent linearizations (`sl4.mod`)

The following example illustrates a possible pitfall for NLP solvers attempting to solve MPECs as NLPs. Consider

$$\left\{ \begin{array}{ll} \underset{z}{\text{minimize}} & z_1 + z_2 \\ \text{subject to} & z_2^2 \geq 1 \\ & 0 \leq z_1 \;\perp\; z_2 \geq 0. \end{array} \right. \tag{2.3}$$

Its solution is $z^* = (0,1)^T$ with NLP multipliers $\lambda^* = 0.5$ of $z_2^2 \geq 1$, $\nu_1^* = 1$ of $z_1 \geq 0$ and $\xi^* = 0$ of $z_1 z_2 \leq 0$. In particular, this solution is a strongly stationary point (see Definition 3.3 below). However, linearizing the constraints about a *linear feasible point* $z^{(0)} = (\epsilon, 1-\delta)^T$ (with $\epsilon, \delta > 0$), gives a QP which is *inconsistent*. Note that $z^{(0)}$ can be *arbitrarily close* to the solution. The linearizations are

$$\begin{array}{rcl} (1-\delta)^2 + 2(1-\delta)(z_2 - (1-\delta)) & \geq & 1 \\ z_1 & \geq & 0 \end{array} \tag{2.4}$$

$$\begin{aligned} z_2 &\geq 0 \\ (1-\delta)\epsilon + (1-\delta)(z_1 - \epsilon) + \epsilon(z_2 - (1-\delta)) &\leq 0. \end{aligned} \tag{2.5}$$

It is easily shown that

$$(2.4) \Rightarrow z_2 \geq \frac{1 + (1-\delta)^2}{2(1-\delta)} > 1$$
$$(2.5) \Rightarrow z_2 \leq 1 - \delta < 1,$$

which shows that the QP approximation is inconsistent. This is also observed during our filter solves (where we enter restoration at this point).

Clearly, any NLP solver hoping to tackle MPECs will have to deal with this situation. snopt [12] uses an *elastic mode* which relaxes the linearizations of the QP. filter [11] has a restoration phase. In Section 5 convergence of SQP methods without modifications is analysed. This analysis is closer in spirit to the results obtained using filter.

3 Review of optimality conditions for MPECs

This section reviews stationarity concepts for MPECs in the form (1.1) and introduces a second order condition. It follows loosely the development of Scheel and Scholtes [19], though the presentation is slightly different.

Given two index sets $\mathcal{Z}_1, \mathcal{Z}_2 \subset \{1, \ldots, p\}$ with

$$\mathcal{Z}_1 \cup \mathcal{Z}_2 = \{1, \ldots, p\} \tag{3.1}$$

denote their respective complements in $\{1, \ldots, p\}$ by $\mathcal{Z}_1^\perp$ and $\mathcal{Z}_2^\perp$. For any such pair of index sets, define the *relaxed NLP corresponding to the MPEC (1.1)* as

$$\begin{array}{lll} \underset{z}{\text{minimize}} & f(z) & \\ \text{subject to} & h(z) = 0 & \\ & c(z) \geq 0 & \\ & z_{1j} = 0 & \forall j \in \mathcal{Z}_2^\perp \\ & z_{2j} = 0 & \forall j \in \mathcal{Z}_1^\perp \\ & z_{1j} \geq 0 & \forall j \in \mathcal{Z}_1 \\ & z_{2j} \geq 0 & \forall j \in \mathcal{Z}_2. \end{array} \tag{3.2}$$

Concepts such as constraint qualifications, stationarity and a second order condition for MPECs will be defined in terms of the relaxed NLPs. The term relaxed NLP stems from the observation, that if z^* is a local solution of a relaxed NLP (3.2) *and* satisfies complementarity ${z_1^*}^T z_2^* = 0$, then z^* is also a local solution of the original MPEC (1.1). One can naturally associate with every feasible point $\hat{z} = (\hat{z}_0, \hat{z}_1, \hat{z}_2)$ of the MPEC a relaxed NLP (3.2) by choosing $\mathcal{Z}_1$ and $\mathcal{Z}_2$ to contain the indices of the vanishing components of $\hat{z}_1$ and $\hat{z}_2$, respectively. Whilst, in contrast to [19], our definition of the relaxed NLP is independent of a specific point, it will occasionally be convenient to identify the above sets of vanishing components associated with a specific point $\hat{z}$, in which case we denote them by $\mathcal{Z}_1(\hat{z}), \mathcal{Z}_2(\hat{z})$ or use suitable superscripts. Note that for these sets the condition (3.1) is equivalent to $\hat{z}_1^T \hat{z}_2 = 0$.

The indices which are both in $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are referred to as the *bi-active components* (or second level degenerate indices) and are denoted by

$$\mathcal{D} := \mathcal{Z}_1 \cap \mathcal{Z}_2.$$

Obviously, in view of (3.1), $(\mathcal{Z}_1^\perp, \mathcal{Z}_2^\perp, \mathcal{D})$ is a partition of $\{1, \ldots, p\}$. A solution z^* to the problem (1.1) is said to be *second level non-degenerate* if and only if $\mathcal{D}(z^*) = \emptyset$.

First two well known constraint qualifications are extended to MPECs, namely the Mangasarian-Fromowitz constraint qualification (MFCQ) and the linear independence constraint qualification (LICQ).

Definition 3.1 *The MPEC (1.1) is said to satisfy an* MPEC-MFCQ (MPEC-LICQ)*, if and only if the corresponding relaxed NLP (3.2) satisfies an MFCQ (LICQ).*

In [19], four stationarity concepts are introduced for MPEC (1.1). The stationarity definition which allows the strongest conclusions is Bouligand or B-stationarity.

Definition 3.2 *A point z^* is called* Bouligand or B-stationary*, if and only if $d = 0$ solves the LPEC obtained by linearizing f and c about z^*,*

$$\begin{array}{ll} \underset{d}{\text{minimize}} & {g^*}^T d \\ \textit{subject to} & c_\mathcal{E}^* + {A_\mathcal{E}^*}^T d = 0 \\ & c_\mathcal{I}^* + {A_\mathcal{E}^*}^T d \geq 0 \\ & 0 \leq z_1^* + d_1 \ \perp \ z_2^* + d_2 \geq 0. \end{array}$$

B-stationarity is very difficult to check, since it involves the solution of an LPEC which is a combinatorial problem and may require the solution of an exponential number of LPs, unless *all* these LPs share a common multiplier vector. Such a common multiplier vectors exists, if an MPEC-LICQ holds.

The results of this paper relate to the following notion of strong stationarity.

Definition 3.3 *A point z^* is called* strongly stationary*, if and only if there exist multipliers λ, $\hat{\nu}_1$ and $\hat{\nu}_2$ such that*

$$\begin{array}{r} \nabla f^* - \nabla {c^*}^T \lambda - \begin{pmatrix} 0 \\ \hat{\nu}_1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \hat{\nu}_2 \end{pmatrix} = 0 \\ c_\mathcal{E}^* = 0 \\ c_\mathcal{I}^* \geq 0 \\ z_1^* \geq 0 \\ z_2^* \geq 0 \\ z_{1j}^* = 0 \ \ \textit{or} \ \ z_{2j}^* = 0 \\ \lambda_\mathcal{I} \geq 0 \\ c_i^* \lambda_i = 0 \\ z_{1j}^* \hat{\nu}_{1j} = 0 \\ z_{2j}^* \hat{\nu}_{2j} = 0 \\ \textit{if } z_{1j}^* = z_{2j}^* = 0 \ \textit{ then } \ \hat{\nu}_{1j} \geq 0 \quad \textit{and} \quad \hat{\nu}_{2j} \geq 0. \end{array} \tag{3.3}$$

Note that (3.3) are the stationarity conditions of the relaxed NLP (3.2) at z^*. B-stationarity is equivalent to strong stationarity, if the MPEC-LICQ holds (e.g. [19]).

Next, a second order sufficient condition (SOSC) for MPECs is given. Since strong stationarity is related to the relaxed NLP (3.2) it seems plausible to use the same NLP to define a second order condition. For this purpose, let $\mathcal{A}^*$ denote the set of active constraints of (3.2) and $\mathcal{A}_+^* \subset \mathcal{A}^*$ the set of active constraints with nonzero multipliers (some could be negative). Let A denote the matrix of active constraint normals, i.e.

$$A = \left[\begin{array}{ccccccc} A_{\mathcal{E}}^* & : & A_{\mathcal{I}\cap\mathcal{A}^*}^* & : & \begin{array}{c} 0 \\ I_1^* \\ 0 \end{array} & : & \begin{array}{c} 0 \\ 0 \\ I_2^* \end{array} \end{array} \right] =: [a_i^*]_{i\in\mathcal{A}^*} ,$$

where $A_{\mathcal{I}\cap\mathcal{A}^*}^*$ are the active inequality constraint normals and

$$I_1^* := [e_i]_{i\in\mathcal{Z}_1^*} \quad \text{and} \quad I_2^* := [e_i]_{i\in\mathcal{Z}_2^*}$$

are parts of the $p \times p$ identity matrices corresponding to active bounds. Define the set of feasible directions of zero slope as

$$S^* = \left\{ s \mid s \neq 0 \ , \ g^{*^T} s = 0 \ , \ a_i^{*^T} s = 0 \ , \ i \in \mathcal{A}_+^* \ , \ a_i^{*^T} s \geq 0 \ , \ i \in \mathcal{A}^* \backslash \mathcal{A}_+^* \right\} .$$

We can now give an MPEC-SOSC. This condition is also sometimes referred to as the strong-SOSC.

Definition 3.4 *A point z^* satisfies an* MPEC-SOSC*, if and only if strong stationarity (3.3) holds and*

$$s^T \nabla^2 \mathcal{L}^* s > 0 \ , \ \forall s \in S^*,$$

where $\nabla^2 \mathcal{L}^*$ is the Hessian of the Lagrangian evaluated at $(z^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$ and S^* is the set of feasible directions for the relaxed NLP (3.2).

The definitions of this section are readily extended to the case where a more general complementarity condition such as (1.2) is used. Moreover, any reformulation using slacks preserves all of these definitions. In that sense, there is no loss of generality in assuming that slacks are being used.

4 Strong stationarity versus NLP stationarity

This section shows that there is a relationship between strong stationarity of the MPEC (1.1) and NLP stationarity conditions for (1.3). It is shown that their respective multipliers are related by a simple formula.

The NLP stationarity conditions of (1.3) are that there exists multipliers $\mu := (\lambda, \nu_1, \nu_2, \xi)$

such that

$$\begin{aligned}
\nabla f(z) - \nabla c(z)^T\lambda - \begin{pmatrix} 0 \\ \nu_1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \nu_2 \end{pmatrix} + \xi \begin{pmatrix} 0 \\ z_2 \\ z_1 \end{pmatrix} &= 0 \\
c_{\mathcal{E}}(z) &\geq 0 \\
c_{\mathcal{I}}(z) &\geq 0 \\
z_1 &\geq 0 \\
z_2 &\geq 0 \\
z_1^T z_2 &\leq 0 \\
\lambda_{\mathcal{I}} &\geq 0 \\
\nu_1 &\geq 0 \\
\nu_2 &\geq 0 \\
\xi &\geq 0 \\
c_i(z)\lambda_i &= 0 \\
z_{1j}\nu_{1j} &= 0 \\
z_{2j}\nu_{2j} &= 0
\end{aligned} \tag{4.1}$$

There is also a complementarity condition $\xi z_1^T z_2 = 0$ which is implied by feasibility of z_1, z_2 and has been omitted.

We examine the difference between (4.1) and the strong-stationarity condition (3.3). In (3.3), the multipliers $\hat{\nu}_1$ and $\hat{\nu}_2$ may be negative for components that satisfy second level nondegeneracy, while in (4.1), $\nu_1 \geq 0, \quad \nu_2 \geq 0$ is required. We will relate the multipliers of (3.3) and (4.1) to show that stationarity in both senses is equivalent.

The main observation in proving the following result is that the first order condition of (4.1) can be written as

$$\nabla f(z) - \nabla c(z)^T\lambda - \begin{pmatrix} 0 \\ \nu_1 - \xi z_2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \nu_2 - \xi z_1 \end{pmatrix} = 0,$$

which is equivalent to the corresponding first order condition in (3.3), if and only if

$$\hat{\nu}_1 = \nu_1 - \xi z_2 \tag{4.2}$$

$$\hat{\nu}_2 = \nu_2 - \xi z_1. \tag{4.3}$$

Proposition 4.1 *A point z is strongly stationary in the MPEC (1.1) if and only if it is a stationary point of the NLP (1.3).*

Proof. First it is shown that (4.1) $\Rightarrow$ (3.3) by distinguishing three cases:
(a) If $z_{1j} > 0$, then $z_{2j} = 0 = \nu_{1j}$ from complementarity and slackness. Using (4.2) it follows that $\hat{\nu}_{1j} = 0$ and $\hat{\nu}_{2j} = \nu_{2j} - \xi z_{1j}$ satisfies (3.3).
(b) If $z_{2j} > 0$, then transpose above argument.
(c) Finally, if $z_{1j} = z_{2j} = 0$, then (4.2) and (4.3) imply that $\hat{\nu}_{1j} = \nu_{1j} \geq 0$ and $\hat{\nu}_{2j} = \nu_{2j} \geq 0$. Combining (a) - (c) it follows that (4.1) implies (3.3).

Next it is shown that (3.3) $\Rightarrow$ (4.1) by distinguishing three cases:
(d) If $z_{1j} > 0$, then $\hat{\nu}_{1j} = 0$ and $z_{2j} = 0$. This implies that $\nu_{1j} = \xi z_{2j} + \hat{\nu}_{1j} = 0 \geq 0$. To ensure that $\nu_{2j} = \xi z_{1j} + \hat{\nu}_{2j}$ is non-negative, it may be necessary to increase ξ such that $\xi z_{1j} + \hat{\nu}_{2j} \geq 0, \ \forall j$.

(e) If $z_{2j} > 0$, then transpose above argument.
(f) Finally, if $z_{1j} = z_{2j} = 0$, then $\nu_{1j} = \hat{\nu}_{1j} \geq 0$ and $\nu_{2j} = \hat{\nu}_{2j} \geq 0$. Combining (d) - (f) it follows that (3.3) implies (4.1). □

The interesting point about the proof is that it relates the multiplier ξ to the fact that the NLP conditions (4.1) are more restrictive in the sense that they enforce $\nu_1, \nu_2 \geq 0$, while $\hat{\nu}_1, \hat{\nu}_2$ may be negative. In a way, ξ compensates for this: if, for instance, $\hat{\nu}_{1j} < 0$, then $z_{2j} > 0$ and we can get the corresponding NLP multiplier $\nu_{1j} = \hat{\nu}_{1j} + \xi z_{2j}$ non-negative by choosing ξ sufficiently large.

It is easy to derive an expression for the minimum value which ξ must take in order to ensure that all multipliers ν_1, ν_2 are non-negative. If $z^*_{2i} = 0$, then $\hat{\nu}_{1i} \geq 0$ (either because $z^*_{1i} > 0$ or because of the last equation in (3.3)) and thus $\nu_{1i} = \hat{\nu}_{1i} \geq 0$. If on the other hand $z^*_{2i} > 0$, then $\hat{\nu}_{1i}$ may be negative. Hence, in order to ensure that $\nu_1 \geq 0$, it is sufficient to choose ξ such that

$$\xi \geq \max_{i \in \mathcal{Z}_2^{\perp}} \frac{-\hat{\nu}_{1i}}{z^*_{2i}}.$$

A similar expression is obtained for $z^*_{1i} > 0$, $\hat{\nu}_{2i} < 0$, so that choosing

$$\xi = \max \left\{ 0 \,,\, \max_{i \in \mathcal{Z}_2^{\perp}} \frac{-\hat{\nu}_{1i}}{z^*_{2i}} \,,\, \max_{i \in \mathcal{Z}_1^{\perp}} \frac{-\hat{\nu}_{2i}}{z^*_{1i}} \right\} \tag{4.4}$$

will ensure that $\nu_1, \nu_2 \geq 0$. Examining the expressions on the right hand side of (4.4) it can be seen that ξ is bounded.

Clearly, any value $\hat{\xi} > \xi$ in (4.4) would also satisfy the stationarity conditions (4.1) and this is how the unboundedness of the multiplier set arises. However, any such $\hat{\xi} > \xi$ would not correspond to a *basic solution*, in the sense that the constraint normals corresponding to nonzero multipliers are linearly dependent. The main argument in our convergence analysis is to show that an SQP solver which works with a non-singular basis will pick the multiplier defined in (4.4).

Definition 4.1 *The multiplier defined by (4.4) is referred to as the* basic multiplier.

The terminology of this definition is justified by the following Lemma, which shows that if MPEC-LICQ holds, then the MPEC multipliers and the multiplier in (4.4) are unique and correspond to a linearly independent set of constraint normals.

Lemma 4.1 *If MPEC-LICQ holds at a local minimizer, then it is strongly stationary and there the multipliers in (3.3) as well as the basic multiplier defined by (4.4) are unique. Moreover, the set of constraint normals corresponding to non-zero multipliers is linearly independent.*

Proof. For a proof of the existence and uniqueness of MPEC multipliers (3.3) see [19]. Because of the uniqueness of the MPEC multiplier, all expressions on the right-hand side of (4.4) are unique which implies the uniqueness of ξ. The uniqueness of the corresponding NLP multipliers follows from (4.2) and (4.3).
Now distinguish two cases: $\xi = 0$ and $\xi > 0$ in (4.4).
If $\xi = 0$ then the linear independence of constraint normals corresponding to non-zero multipliers follows from MPEC-LICQ.

If $\xi > 0$, then there exists at least one component $i \in \mathcal{Z}_1^\perp$ or $i \in \mathcal{Z}_2^\perp$ such that $\nu_{2i} = 0$ or $\nu_{1i} = 0$. It remains to show that the set of constraint normals corresponding to non-zero multipliers is linearly independent. By MPEC-LICQ, this is true for all but the complementarity constraint. Then we can exchange the normal of the complementarity constraint for any normal whose multiplier is driven to zero by (4.4) and (4.2) or (4.3) in the basis as explained in Lemma 5.1 below. $\square$

The conclusions of this section can be readily extended to cover the case where the complementarity condition is of more the general form (1.2).

5 Local convergence of SQP methods

This section shows that SQP methods converge quadratically near a strongly stationary point under mild conditions. Section 7 discusses the assumptions and provides counter-examples for situations where (some of) these assumptions are not satisfied. In particular, we are interested in the situation, where $z^{(k)}$ is close to a strongly stationary point, z^*, but $z_1^{(k)^T} z_2^{(k)}$ is *not* necessarily zero. SQP then solves a sequence of quadratic programming approximations, given by

$$(QP^k) \left\{ \begin{array}{ll} \underset{d}{\text{minimize}} & g^{(k)^T} d + \frac{1}{2} d^T W^{(k)} d \\ \text{subject to} & c_{\mathcal{E}}^{(k)} + A_{\mathcal{E}}^{(k)^T} d = 0 \\ & c_{\mathcal{I}}^{(k)} + A_{\mathcal{I}}^{(k)^T} d \geq 0 \\ & z_1^{(k)} + d_1 \geq 0 \\ & z_2^{(k)} + d_2 \geq 0 \\ & z_1^{(k)^T} z_2^{(k)} + z_2^{(k)^T} d_1 + z_1^{(k)^T} d_2 \leq 0, \end{array} \right.$$

where $W^{(k)} = \nabla^2 \mathcal{L}(z^{(k)}, \mu^{(k)})$ is the Hessian of the Lagrangian of (1.3) and $\mu^{(k)} = (\lambda^{(k)}, \nu_1^{(k)}, \nu_2^{(k)}, \xi^{(k)})$. The last constraint of (QP^k) is the linearization of the complementarity condition $z_1^T z_2 \leq 0$.

Assumptions 5.1 *The following assumptions are made*

[A1] *f and c are twice Lipschitz continuously differentiable.*

[A2] *(1.1) satisfies an MPEC-LICQ (Definition 3.1), i.e. the relaxed NLP satisfies an LICQ.*

[A3] *At z^* an MPEC-SOSC (Definition 3.4) holds. Note that this implies that z^* is a strongly stationary point of (1.1), (Definition 3.3).*

[A4] $\lambda_i \neq 0$, $\forall i \in \mathcal{E}^*$, $\lambda_i^* > 0$, $\forall i \in \mathcal{A}^* \cap \mathcal{I}$, *and either* $\nu_{1j}^* > 0$ *or* $\nu_{2j}^* > 0$, $\forall j \in \mathcal{D}^*$.

[A5] *The QP solver always chooses a linearly independent basis.*

The most restrictive assumption is **[A3]**, which assumes the existence of a common multiplier. In this sense, **[A3]** removes the combinatorial nature of the problem. It is not clear that **[A2]** can readily be relaxed in the present context, since **[A2]** and **[A3]** imply that z^* is B-stationary. Thus it is possible to check B-stationarity by solving exactly one

LP or QP. Without assumption **[A2]** it would not be possible to verify B-stationarity without solving several LPs (one for every possible combination of second level degenerate indices $i \in \mathcal{D}^*$). It seems unlikely therefore, that any method which solves only a single LP or QP per iteration can be shown to be convergent to B-stationary points for problems which violate MPEC-LICQ. However, it may be possible to replace **[A2]** by an MPEC-MFCQ and use similar ideas to [20] to show convergence (at the cost of possibly modifying the SQP algorithm) to strongly stationary points. Note that we do *not* assume that the MPEC (1.1) is second level nondegenerate. Assumption **[A5]** is a reasonable assumption in practice, as most modern SQP solvers are based on QP solvers which guarantee this.

This section is divided into two parts. First, the case where complementarity is satisfied at a point sufficiently close to a stationary point is considered. This case corresponds to the situation where all iterates (ultimately) remain on the same face of $0 \le z_1 \perp z_2 \ge 0$. The key idea is to show that SQP applied to (1.3) behaves identical to SQP applied to (3.2).

The second case considered arises, when $z_1^{(k)^T} z_2^{(k)} > 0$ for all iterates k. In this case, the previous ideas cannot be applied and a separate proof is required and we make the additional assumption that all QP subproblems remain consistent. This last assumption **[A6]** appears to be rather strong, especially in the light of example (2.3), which shows that the QP approximation may be inconsistent *arbitrarily close to a solution*. However, we will give several sufficient conditions for it later which show that is it not unduly restrictive.

Without loss of generality, we assume that $\mathcal{Z}_1^{*^\perp} = \emptyset$, i.e. we will assume that the solution has the form $z_1^* = 0$ and $z_2^* = (0, z_{22}^*)$ and that $z_{22}^* > 0$. This greatly simplifies the notation.

5.1 Local convergence for exact complementarity

In this section we make the additional assumption that

[A6] $z_1^{(k)^T} z_2^{(k)} = 0$ and $(z^{(k)}, \mu^{(k)})$ is sufficiently close to a strongly stationary point.

Assumption **[A6]** implies that for given index sets $\mathcal{Z}_1 = \mathcal{Z}_1^{(k)}$ and $\mathcal{Z}_2 = \mathcal{Z}_2^{(k)}$, the following implications hold

$$\begin{array}{ll} z_{1j}^{(k)} = 0 & \forall j \in \mathcal{Z}_2^\perp \\ z_{2j}^{(k)} = 0 & \forall j \in \mathcal{Z}_1^\perp \\ z_{1j}^{(k)} = 0 \text{ or } z_{2j}^{(k)} = 0 & \forall j \in \mathcal{D}. \end{array}$$

In particular, it is *not* assumed that the bi-active complementarity constraints $\mathcal{D}^*$ are active at $z^{(k)}$. Thus it may be possible that $\mathcal{Z}_1 \neq \mathcal{Z}_1^*$ (and similarly for $\mathcal{Z}_2$). However, it is shown that the bi-active constraints become active after one step of the SQP method as a consequence of **[A4]** (the positivity of bi-active multipliers).

The following well-known result shows that if strong stationarity and **[A6]** hold, then *all* relaxed NLPs have the same multiplier.

Proposition 5.1 *Let* **[A1]** *to* **[A5]** *hold at* z^* *and let* $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{D}$ *be any index sets satisfying*

$$\begin{array}{l} \mathcal{Z}_1,\ \mathcal{Z}_2 \subset \{1, \ldots, p\} \\ \mathcal{Z}_1 \subset \mathcal{Z}_1^* \qquad \mathcal{Z}_2 \subset \mathcal{Z}_2^* \\ \mathcal{D} = \mathcal{Z}_1 \cap \mathcal{Z}_2 \end{array}$$

Then it follows that all corresponding relaxed NLPs (3.2) have the same (unique) multipliers.

Proof. Elsewhere. □

The conditions on $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{D}$ in Proposition 5.1 are equivalent to requiring that $\mathcal{Z}_1^\perp, \mathcal{Z}_2^\perp, \mathcal{D}$ are a partition of $\{1, \ldots, p\}$ for which the following conditions hold

$$\begin{array}{ccccc} \mathcal{Z}_1^{*\perp} & \subset & \mathcal{Z}_1^\perp & \subset & \mathcal{Z}_1^{*\perp} \cup \mathcal{D}^* \\ \mathcal{Z}_2^{*\perp} & \subset & \mathcal{Z}_2^\perp & \subset & \mathcal{Z}_2^{*\perp} \cup \mathcal{D}^* \\ & & \mathcal{D} & \subset & \mathcal{D}^*. \end{array}$$

The key idea of the proof is to show that SQP applied to (1.3) is equivalent to SQP applied to the relaxed NLP (3.2) on a face. For a given partition $(\mathcal{Z}_1^\perp, \mathcal{Z}_2^\perp, \mathcal{D})$, an SQP step for (3.2) is obtained by solving the QP

$$(QP_R(z^{(k)}, \mathcal{Z})) \left\{ \begin{array}{lll} \underset{d}{\text{minimize}} & g^{(k)^T} d + \frac{1}{2} d^T \widehat{W^{(k)}} d & \\ \text{subject to} & c_\mathcal{E}^{(k)} + A_\mathcal{E}^{(k)^T} d = 0 & \\ & c_\mathcal{I}^{(k)} + A_\mathcal{I}^{(k)^T} d \geq 0 & \\ & d_{1j} = 0 & \forall j \in \mathcal{Z}_2^\perp \\ & d_{2j} = 0 & \forall j \in \mathcal{Z}_1^\perp \\ & z_{1j}^{(k)} + d_{1j} \geq 0 & \forall j \in \mathcal{Z}_1 \\ & z_{2j}^{(k)} + d_{2j} \geq 0 & \forall j \in \mathcal{Z}_2, \end{array} \right.$$

where

$$\widehat{W^{(k)}} = \nabla^2 f(z^{(k)}) - \sum \lambda_i^{(k)} \nabla^2 c_i(z^{(k)})$$

is the Hessian of the Lagrangian of the relaxed NLP (3.2). Note that the relaxed NLP (3.2) is never set up, nor is $(QP_R(z^{(k)}, \mathcal{Z}))$ ever solved. These two problems are merely used in the convergence proof. The key idea is to show that SQP applied to the ill-conditioned NLP (1.3) is equivalent to SQP applied to the well-behaved relaxed NLP (3.2), given by the sequence defined by $(QP_R(z^{(k)}, \mathcal{Z}))$.

Throughout this section, $\mathcal{Z}_1 = \mathcal{Z}_1^{(k)}$ and $\mathcal{Z}_2 = \mathcal{Z}_2^{(k)}$ and it is *not* assumed that the correct sets have been identified. However, an important consequence of **[A5]** is that $\mathcal{Z}_1^{(k)}$, $\mathcal{Z}_2^{(k)}$ satisfy

$$\begin{array}{ccccc} \mathcal{Z}_1^{*\perp} & \subset & \mathcal{Z}_1^{(k)^\perp} & \subset & \mathcal{Z}_1^{*\perp} \cup \mathcal{D}^* \\ \mathcal{Z}_2^{*\perp} & \subset & \mathcal{Z}_2^{(k)^\perp} & \subset & \mathcal{Z}_2^{*\perp} \cup \mathcal{D}^* \\ & & \mathcal{D}^{(k)} & \subset & \mathcal{D}^*, \end{array} \tag{5.1}$$

i.e. the indices $\mathcal{Z}_1^{*\perp}$ and $\mathcal{Z}_2^{*\perp}$ of the non-degenerate complementarity constraints have been identified correctly (see Proposition 5.1).

The following proposition states the fact that SQP applied to the relaxed NLP converges quadratically and identifies the correct index sets $\mathcal{Z}_1^*$ and $\mathcal{Z}_2^*$ in one step.

Proposition 5.2 *Let* **[A1]** *to* **[A5]** *hold and consider the relaxed NLP for any index sets* $\mathcal{Z}_1^k$, $\mathcal{Z}_2^k$ *(satisfying (5.1) by virtue of* **[A5]**). *Then it follows that*

1. *the sequence of iterates* $\{(z^{(l)}, \lambda^{(l)}, \nu_1^{(l)}, \nu_2^{(l)})\}_{l>k}$ *converges Q-quadratically to* $(z^*, \lambda^*, \nu_1^*, \nu_2^*)$;

2. *the sequence* $\{z^{(l)}\}_{l>k}$ *converges Q-superlinearly to* z^*; *and*

3. $\mathcal{Z}_1^{(l)} = \mathcal{Z}_1^*$ *and* $\mathcal{Z}_2^{(l)} = \mathcal{Z}_2^*$ *for* $l > k$.

Proof. The relaxed NLP satisfies LICQ and a second order sufficient condition. Therefore, SQP converges at second order rate when applied to the relaxed NLP. Part 1. is a standard result whose proof can be found for instance in [7, Theorem 15.2.2]. Part 2. is due to [4] and part 3 follows from the fact that SQP identifies the correct active set in one step. □

Next, it is shown that the sequence of steps generated by SQP applied to the relaxed NLP (3.2) is identical to the sequence of steps generated by applying SQP to the equivalent NLP (1.3), provided that $z_1^{(k)^T} z_2^{(k)} = 0$, i.e. **[A6]** holds for some k. If $z_1^{(k)^T} z_2^{(k)} = 0$, then an SQP step for (1.3) is obtained by solving the following (QP^k) with $z_1^{(k)^T} z_2^{(k)} = 0$ in the last constraint.

The two QPs (QP^k) and $(QP_R(z^{(k)}, \mathcal{Z}))$ have different constraints and Hessians. The Hessian of (QP^k) is

$$W^{(k)} = \widehat{W^{(k)}} + \xi^{(k)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{bmatrix}.$$

However, despite these differences, it is possible to show that the two QPs have the same solution (from which second order convergence follows). The following Lemma shows that the constraint sets are the same.

Lemma 5.1 *Let Assumptions* **[A1]** *tp* **[A6]** *hold. Then, a step* d *is feasible in* (QP^k) *if and only if it is feasible in* $(QP_R(z^{(k)}, \mathcal{Z}))$.

Proof. The constraint sets only differ in the way in which indices $j \in \mathcal{Z}_2^\perp$ and $j \in \mathcal{Z}_1^\perp$ are handled. Thus it suffices to consider those constraints.
(a) Let d *be feasible in* $(QP_R(z^{(k)}, \mathcal{Z}))$. Then it follows in particular that d satisfies

$$\begin{aligned} d_{1j} &= 0 \quad \forall j \in \mathcal{Z}_2^\perp \\ d_{2j} &= 0 \quad \forall j \in \mathcal{Z}_1^\perp. \end{aligned}$$

Splitting these constraints into two inequalities, it follows that d satisfies

$$d_{1j} \geq 0 \quad \forall j \in \mathcal{Z}_2^\perp \tag{5.2}$$
$$d_{1j} \leq 0 \quad \forall j \in \mathcal{Z}_2^\perp \tag{5.3}$$
$$d_{2j} \geq 0 \quad \forall j \in \mathcal{Z}_1^\perp \tag{5.4}$$
$$d_{2j} \leq 0 \quad \forall j \in \mathcal{Z}_1^\perp. \tag{5.5}$$

Summing (5.3) over all $j \in \mathcal{Z}_2^\perp$ weighted with $z_{2j}^{(k)} > 0$ and (5.5) over all $j \in \mathcal{Z}_1^\perp$ weighted with $z_{1j}^{(k)} > 0$, it follows that d satisfies the last constraint of (QP^k) (the simple bounds follow from (5.2) and (5.4)).
(b) Let d *be feasible in* (QP^k). Then it follows that

$$d_{1j} \geq 0 \; \forall j \in \mathcal{Z}_2^\perp \;\text{ and }\; d_{2j} \geq 0 \; \forall j \in \mathcal{Z}_1^\perp.$$

Since $z_{2j}^{(k)} > 0,\ \forall j \in \mathcal{Z}_2^\perp$ and $z_{1j}^{(k)} > 0,\ \forall j \in \mathcal{Z}_1^\perp$, it follows from

$$\sum_{j\in\mathcal{Z}_2^\perp} z_{2j}^{(k)} d_{1j} + \sum_{j\in\mathcal{Z}_1^\perp} z_{1j}^{(k)} d_{2j} \leq 0$$

that

$$d_{1j} \leq 0\ \forall j \in \mathcal{Z}_2^\perp \ \text{ and } \ d_{2j} \leq 0\ \forall j \in \mathcal{Z}_1^\perp.$$

Thus, d is feasible in $(QP_R(z^{(k)}, \mathcal{Z}))$. □

Next, it is shown that the first order stationary points of (QP^k) and $(QP_R(z^{(k)}, \mathcal{Z}))$ are identical.

Lemma 5.2 *Let assumptions* **[A1]** - **[A6]** *hold, then it follows that the objective functions of* $(QP_R(z^{(k)}, \mathcal{Z}))$ *and* (QP^k) *are identical on the feasible set. Moreover, any first order stationary point of* $(QP_R(z^{(k)}, \mathcal{Z}))$ *is also a stationary point of* (QP^k) *and vise-versa.*

Proof. For any feasible point of $(QP_R(z^{(k)}, \mathcal{Z}))$ (and hence of (QP^k)), it follows from **[A6]** and Lemma 5.1 that $d_1^T d_2 = 0$. Hence, the objective functions of the two QPs are identical on the feasible set. Thus it follows, that the reduced gradients are the same and hence that any first order stationary point of $(QP_R(z^{(k)}, \mathcal{Z}))$ is also a stationary point of (QP^k) and vise-versa. □

Lemma 5.3 *Let Assumptions* **[A1]** *to* **A6]** *hold. Let* $(\lambda, \widehat{\nu_1}, \widehat{\nu_2})$ *be the multipliers of* $(QP_R(z^{(k)}, \mathcal{Z}))$ *(corresponding to a step d). Then it follows that the multipliers of* (QP^k)*, corresponding to the same step d are* $\mu = (\lambda, \nu_1, \nu_2, \xi)$*, where*

$$\begin{aligned}
\xi &= \max\left(0\ ,\ \max_{j\in\mathcal{Z}_1\setminus\mathcal{D}} \frac{-\widehat{\nu_{1j}} - \xi^{(k)} d_{2j}}{z_{2j}^{(k)}}\ ,\ \max_{j\in\mathcal{Z}_2\setminus\mathcal{D}} \frac{-\widehat{\nu_{2j}} - \xi^{(k)} d_{1j}}{z_{1j}^{(k)}}\right) && (5.6)\\
\nu_{1j} &= \widehat{\nu_{1j}} > 0\ ,\ \forall j \in \mathcal{D} && (5.7)\\
\nu_{2j} &= \widehat{\nu_{2j}} > 0\ ,\ \forall j \in \mathcal{D} && (5.8)\\
\nu_{1j} &= \widehat{\nu_{1j}} + \xi^{(k)} d_{2j} + \xi z_{2j}^{(k)}\ ,\ \forall j \in \mathcal{Z}_1\setminus\mathcal{D} && (5.9)\\
\nu_{2j} &= \widehat{\nu_{2j}} + \xi^{(k)} d_{1j} + \xi z_{1j}^{(k)}\ ,\ \forall j \in \mathcal{Z}_2\setminus\mathcal{D}. && (5.10)
\end{aligned}$$

Conversely, given a solution d and multipliers μ *of* (QP^k)*, (5.7) to (5.10) show how to construct multipliers, so that* $(d, \lambda, \widehat{\nu_1}, \widehat{\nu_2})$ *solves* (QP_R^k)*.*

Proof. If $z^{(k)}$ is sufficiently close to z^*, then the sign of the multipliers in (5.7) and (5.8) follows from **[A4]** and the value for the multipliers of (QP^k) follows similar to Proposition 4.1. Similarly, the multipliers of (QP^k) in (5.9) and (5.10) are non-negative by construction and satisfy first order conditions by Lemma 5.2. □

Next, it is shown that both QPs have the same (unique) solution in a neighbourhood of $d = 0$.

Lemma 5.4 *The solution d of* $(QP_R(z^{(k)}, \mathcal{Z}))$ *is the only strict local minimizer in a neighbourhood of* $d = 0$ *and its corresponding multipliers* (λ, ν_1, ν_2) *are unique. Moreover, d is also the only strict local minimizer in a neighbourhood of* $d = 0$ *of* (QP^k)*.*

Proof. The result for $(QP_R(z^{(k)}, \mathcal{Z}))$ is due to Robinson [18], see also Conn, Gould and Toint [7], since the relaxed NLP satisfies **[A1]** to **[A4]**. The statement for (QP^k) follows in two parts. First order conditions are established in Lemma 5.3. Second order conditions for (QP^k) follow from second order conditions of $(QP_R(z^{(k)}, \mathcal{Z}))$, as the two problems have the same null-space (Lemma 5.1) and their objectives are identical on the feasible set (Lemma 5.2). Thus the solution d of $(QP_R(z^{(k)}, \mathcal{Z}))$ is also a solution of (QP^k).
If there were another solution of (QP^k) in a neighbourhood $\mathcal{N}$ of $d = 0$, then it would be possible to construct a corresponding solution to $(QP_R(z^{(k)}, \mathcal{Z}))$ in $\mathcal{N}$ (using Lemma 5.3) which would contradict the uniqueness of that solution. □

The following theorem summarizes the results of this section.

Theorem 5.1 *If assumption* **[A1]** *to* **[A6]** *hold, then SQP applied to (1.3) generates a sequence* $\{(z^{(k)}, \lambda^{(k)}, \nu_1^{(k)}, \nu_2^{(k)}, \xi^{(k)})\}_{l>k}$ *which converges Q-quadratically to* $\{(z^*, \lambda^*, \nu_1^*, \nu_2^*, \xi^*)\}$ *of (4.1), satisfying strong stationarity. Moreover, the sequence* $\{z^{(k)}\}_{l>k}$ *converges Q-superlinearly to* z^* *and* $z_1^{(l)^T} z_2^{(l)} = 0$ *for all* $l \geq k$.

Proof. Under assumption **[A1]** to **[A4]**, SQP converges quadratically when applied to the relaxed NLP (3.2). Lemmas 5.1 to 5.4 show that the sequence of iterates generated by this SQP method is equivalent to the sequence of steps generated by SQP applied to (1.3). This implies Q-superlinear convergence of $\{z^{(k)}\}_{l>k}$. Convergence of the multipliers follows by considering (5.6) to (5.10). Clearly, the multipliers in (5.7) and (5.8) converge, as they are just the multipliers of the relaxed NLP, which converge by virtue of Proposition 5.2. Now observe, that (5.6) becomes

$$\widehat{\xi}^{(k+1)} = \max\left(0\ ,\ \max_{j \in \mathcal{Z}_1^{*\perp}} \frac{-\widehat{\nu_{1j}^{(k+1)}} - \xi^{(k)} d_{2j}^{(k)}}{z_{2j}^{(k)}}\ ,\ \max_{j \in \mathcal{Z}_2^{*\perp}} \frac{-\widehat{\nu_{2j}^{(k+1)}} - \xi^{(k)} d_{1j}^{(k)}}{z_{1j}^{(k)}}\right).$$

The right-hand-side of this expression converges, since $\widehat{\nu_{1j}^{(k+1)}}, \widehat{\nu_{2j}^{(k+1)}}$ and $z_{1j}^{(k)}, z_{2j}^{(k)}$ converge and $d_{1j}^{(k)}, d_{2j}^{(k)} \to 0$. Note that the limit of (5.6) is (4.4). Finally, (5.9) and (5.10) converge to (4.2) and (4.3) by a similar argument.

$z_1^{(l)^T} z_2^{(l)} = 0\ ,\ \forall l \geq k$ follows from the convergence of SQP for the relaxed NLP (3.2) and the fact that SQP retains feasibility with respect to linear constraints. Assumption **[A4]** ensures that $d_{1j}^{(k)} = d_{2j}^{(k)} = 0, \forall j \in \mathcal{D}^*$, since $\nu_{1j}^{(k)}, \nu_{2j}^{(k)} > 0$ for bi-active complementarity constraints. Thus SQP will not move out of the corner and stay on the same face.
□

5.2 Local convergence for non-zero complementarity

This section shows that SQP converges superlinearly, even if complementarity does not hold at the starting point, i.e. if $z_1^{(k)^T} z_2^{(k)} > 0$. Example (2.3) shows that the QP approximations can be inconsistent arbitrarily close to a stationary point. In order to avoid this, we make the following assumption which often appears to hold in practice.

[A7] All QP approximations (QP^k) are consistent.

This is clearly an undesirable assumption, as it makes an assumption on the progress of the method. However, we show in the next section, that this assumption is satisfied for some important practical applications.

Our convergence analysis is concerned with showing that for any "basic" active set, SQP converges. To this end, we introduce the set of basic constraints

$$\mathcal{B}(z) := \mathcal{E} \cup \mathcal{I} \cap \mathcal{A}^* \cup \mathcal{Z}_1(z) \cup \mathcal{Z}_2(z) \cup \{z_1^T z_2 = 0\}$$

and the set of strictly active constraints (defined in terms of the basic multiplier, μ),

$$\mathcal{B}_+(z) \ := \ \{i \in \mathcal{B}(z) \mid \mu_i \neq 0\} .$$

Moreover, let $B_+^{(k)}$ denote the matrix of strictly active constraint normals at $z = z^{(k)}$, i.e.

$$B_+^{(k)} \ := \ \left[a_i^{(k)} \right]_{i \in \mathcal{B}_+(z^{(k)})} .$$

Note that Lemma 4.1 shows that the optimal multiplier is unique. However, it may be possible that for some iterates $\mathcal{B}_+^{(k)} \neq \mathcal{B}_+(z^*)$ and our analysis will have to allow for this.

The failure of any constraint qualification at a solution z^* of the equivalent NLP (1.3) implies that the active constraint normals at z^* are linearly dependent. However, the linear dependence occurs in a very special form which can be exploited to prove fast convergence.

Lemma 5.1 *Let assumptions* **[A1]** - **[A4]** *hold and let z^* be a solution of the MPEC (1.1). Consider the matrix of active constraint normals at z^*,*

$$B \ = \ \left[\begin{array}{ccccc} & & 0 & 0 & 0 \\ A_{\mathcal{E}}^* & A_{\mathcal{I}}^* & I & 0 & \begin{pmatrix} 0 \\ z_{22}^* \end{pmatrix} \\ & & 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} & 0 \end{array} \right] ,$$

where we have assumed without loss of generality that ${\mathcal{Z}_\infty}^{\perp^*} = \emptyset$. Note, that the last column is the gradient of the complementarity constraint.

Then it follows that B is linearly dependent and any submatrix of columns of B has full rank, provided that it contains $[A_{\mathcal{E}}^* \ A_{\mathcal{I}}^*]$ and either the last column of B is missing or any column corresponding to $z_{12} = 0$ is missing.

Proof. The fact that the columns of B are linearly dependent is clear by looking at the last three columns of B. Assumption **[A2]**, MPEC-LICQ, implies that B without the last column has full rank. The final statement follows by exchanging any column corresponding to $z_{12}^* = 0$ with the final column of B and observing that $z_{22}^* > 0$. □

The proof shows, that in order to obtain a linearly independent basis, any column of $z_{12} = 0$ can be exchanged with the normal of the complementarity constraint. This is precisely what lies behind (4.2) and (4.3). The corresponding basic multipliers are shown as dots in Figure 2.

Next, we show that if we are close to z^* and the QP solver chooses the full basis B, then exact complementarity holds for all subsequent iterations. Thus in this case, the development of the previous section shows second order convergence.

Lemma 5.2 *Let $z^{(k)}$ be sufficiently close to z^* and let Assumptions* **A1**] - [**A5**] *and* [**A7**] *hold. If the QP solver chooses the full basis B, then it follows that ${z_1^{(k)}}^T z_2^{(k)} > 0$ and that after the QP step, ${z_1^{(k+1)}}^T z_2^{(k+1)} = 0$ and moreover there exists $c > 0$ such that*

$$\| \left(z^{(k+1)}, \mu^{(k+1)}\right) - (z^*, \mu^*) \| \ \leq \ c \,\| \left(z^{(k)}, \mu^{(k)}\right) - (z^*, \mu^*) \|. \tag{5.11}$$

Proof. Assume that ${z_1^{(k)}}^T z_2^{(k)} = 0$ and seek a contradiction. Since $z^{(k)}$ be sufficiently close to z^*, it follows that $z_{22}^{(k)} = \mathcal{O}(1)$. Hence, $z_{12}^{(k)} = 0$. Now consider the final 3 columns of B and observe that if $z_{12}^{(k)} = 0$, then the last column lies in the range of the other two. Hence the basis would be singular, which contradicts Assumption [**A5**] and so ${z_1^{(k)}}^T z_2^{(k)} > 0$.
${z_1^{(k+1)}}^T z_2^{(k+1)} = 0$ follows simply by observing that the full basis B implies that $0 = z_1^{(k)} + d_1 = z_1^{(k+1)}$.
The third part, follows by observing that Assumption [**A2**] and [**A3**] imply that the relaxed NLP satisfies an LICQ and a SOSC. Hence, the basis B without the final column gives a feasible point close to $z^{(k)}$. Denote this solution by $(\hat{z}, \hat{\mu})$ and let the corresponding step be denoted by $\hat{d}$. Clearly, if this step also satisfies the linearization of the complementarity constraint, i.e. if

$${z_1^{(k)}}^T z_2^{(k)} + {z_2^{(k)}}^T \hat{d}_1 + {z_1^{(k)}}^T \hat{d}_2 = 0,$$

then (5.11) follows by second order convergence of SQP for relaxed NLP.
If on the other hand,

$${z_1^{(k)}}^T z_2^{(k)} + {z_2^{(k)}}^T \hat{d}_1 + {z_1^{(k)}}^T \hat{d}_2 > 0,$$

then the SQP step of the relaxed NLP is not feasible in (QP^k). In this case consider the following decomposition of the SQP step. Let

$$\hat{d}^n = \left(\begin{array}{c} 0 \\ \hat{d}_1 \\ \left(\begin{array}{c} \hat{d}_{21} \\ 0 \end{array} \right) \end{array} \right) = \left(\begin{array}{c} 0 \\ -z_1^{(k)} \\ \left(\begin{array}{c} -z_{21}^{(k)} \\ 0 \end{array} \right) \end{array} \right)$$

be the normal component and let $\hat{d}^t := \hat{d} - \hat{d}^n$ be the tangential component. Then it follows, that the step of (QP^k) satisfies $d^{(k)} = \hat{d}^n + \sigma \hat{d}^t$ for some $\sigma \in [0, 1]$. The desired bound on the distance follows now from the convergence of $\hat{d}$. □

Thus, once a full basis is chosen, the corresponding step will give ${z_1^{(k+1)}}^T z_2^{(k+1)} = 0$ for a point close to z^* and second order convergence then follows using Theorem 5.1.

Corollary 5.1 *Let $z^{(k)}$ be sufficiently close to z^* and let Assumptions* [**A1**] - [**A5**] *and* [**A7**] *hold. If the QP solver chooses the full basis B, then it follows that SQP converges quadratically from iteration $k+1$.*

In the remainder we can therefore concentrate on the case, that the full basis B is never chosen and that ${z_1^{(k)}}^T z_2^{(k)} > 0$ for all iterates k (otherwise, we have convergence from the results of the previous section).

Next, we show that for $z^{(k)}$ be sufficiently close to z^*, the basis at $z^{(k)}$ contains both $\mathcal{E}$ and $\mathcal{I}^*$.

Lemma 5.3 *Let $z^{(k)}$ be sufficiently close to z^* and let Assumptions* **[A1]** - **[A5]** *and* **[A7]** *hold. Then it follows, that the optimal basis B of (QP^k) contains the normals $A_{\mathcal{E}}^{(k)}$ and $A_{\mathcal{I}^*}^{(k)}$.*

Proof. This follows by considering the gradient of (QP^k),

$$0 \;=\; \nabla f^{(k)} + \hat{W}^{(k)} d^{(k)} - \nabla c^{(k)^T} \lambda^{(k+1)} - \begin{pmatrix} 0 \\ \nu_1^{(k+1)} - \xi^{(k+1)} z_2^{(k)} \\ \nu_2(k+1) - \xi^{(k+1)} z_1(k) \end{pmatrix} + \xi^{(k)} \begin{pmatrix} 0 \\ d_2^{(k)} \\ d_1^{(k)} \end{pmatrix},$$

where $\hat{W}^{(k)}$ is the Hessian of the Lagrangian without the term corresponding to the complementarity constraint (which is the last term above). Now, for $z^{(k)}$ be sufficiently close to z^*, it follows from **[A4]** that $\lambda_i^{(k+1)} \neq 0$ for all $i \in \mathcal{E} \cup \mathcal{I}^*$. □

Thus, as long as the QP approximations remain consistent, the optimal basis of (QP^k) will be a subset of B satisfying the conditions in Lemma 5.2. The key idea is now to show that for any such basis, there exists an equality constraint problem for which SQP converges quadratically. Since there is only a finite number of basis, this implies convergence for SQP.

Now introduce the *reduced NLP*, which is an equality constraint NLP corresponding to a basis with properties as in Lemma 5.2 and given by

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c_{\mathcal{E}}(z) = 0 \\ & c_{\mathcal{I}^*}(z) = 0 \\ & z_{11} = 0 \\ & z_{21} = 0 \\ & \left.\begin{array}{r} z_{12} = 0 \\ z_1^T z_2 = 0 \end{array}\right\} \text{subset satisfying Lemma 5.2.} \end{array} \tag{5.12}$$

The next Lemma shows, that any reduced NLP satisfies an LICQ and an SOCS.

Lemma 5.4 *Let Assumptions* **[A1]** *to* **[A4]** *and* **[A7]** *hold. Then it follows, that any reduced NLP satisfies an LICQ and an SOSC.*

Proof. Lemma 5.2 shows that the normals of the equality constraints of each reduced NLP are linearly independent. The SOSC follows from the MPEC-SOSC and the observation, that the MPEC and the reduced NLP have the same null space. □

Thus applying SQP to the reduced NLP results in second order convergence.

Proposition 5.1 *Let Assumptions* **[A1]** *to* **[A4]** *and* **[A7]** *hold. Then it follows, that SQP applied to any reduced NLP converges quadratically to z^*.*

Proof. Lemma 5.4 shows that the reduced NLP satisfy LICQ and SOSC and therefore, convergence of SQP follows. In particular, it follows that for a given reduced NLP corresponding to a basis $\mathcal{B}$, there exists a constant $c_{\mathcal{B}} > 0$ such that

$$\| \left(z^{(k+1)}, \mu^{(k+1)}\right) - (z^*, \mu^*) \| \;\leq\; c_{\mathcal{B}} \, \| \left(z^{(k)}, \mu^{(k)}\right) - (z^*, \mu^*) \|^2. \tag{5.13}$$

□

Summarizing the results of this section, we obtain

Theorem 5.1 *Let assumptions* **[A1]** *-* **[A5]** *and* **[A7]** *hold. Then it follows that SQP applied to the NLP formulation (1.3) of the MPEC (1.1) converges quadratically near a solution* (z^*, μ^*).

Proof. Proposition 5.1 shows that SQP converges superlinearly for any possible choice of basis $\mathcal{B}$ and assumption **[A7]** shows that (QP^k) is consistent and remains consistent. Therefore, there exists a basis for which superlinear convergence follows. Thus for each basis, a step is computed which satisfies a contraction condition like (5.13) for a constant $c_{\mathcal{B}} > 0$ which depends on the basis. Since there is a finite number of basis, this condition holds also for $c = \max c_{\mathcal{B}}$ independent of the basis and SQP converges superlinearly independent of the basis. $\square$

5.3 Discussion of proofs

An interesting observation of the convergence proofs of this section is that if $z_1^{(k)^T} z_2^{(k)} = 0$, then the actual value of $\xi^{(k)}$ has no effect on the step computed by SQP. This shows, that the curvature information contained in the complementarity constraint $z_1^T z_2 \leq 0$ is not important. As a consequence, one could simply leave out this contribution to the Hessian of the Lagrangian. This can be easily implemented and convergence results follow along similar lines to the observation above.

The conclusions and proofs presented in this section also carry through for linear complementarity constraints, but *not* for general nonlinear complementarity constraints. The reason is that the implication

$$z_1^{(k)^T} z_2^{(k)} = 0 \;\Rightarrow\; z_1^{(k+1)^T} z_2^{(k+1)} = 0$$

holds for linear complementarity problems, but *not* for nonlinear complementarity problems, since in general, an SQP method would move off a nonlinear constraint. This is one reason for the introduction of slacks to deal with complementarity of the form (1.2).

Similar conclusions can easily be derived for alternative NLP formulations of the MPEC (1.1). For instance, the complementarity constraint in (1.3) can be replaced by

$$z_{1j} z_{2j} \leq 0 \;,\; \forall j = 1, \ldots, p.$$

In this case, a similar construction to (5.6) is possible, where $\widehat{\xi}$ is replaced by a vector of complementarity multipliers, one for each constraint. Equations (4.2) and (4.3) then become componentwise conditions and similarly, (5.9) and (5.10). In addition, it can now be seen, that a basis that satisfies the conditions of Lemma 5.2 satisfies a complementarity condition between the multipliers ξ_i and ν_{1i} (and ν_{2i}).

The strongest assumption in the present convergence analysis is Assumption **[A7]**, namely that all (QP^k) remain consistent. It is shown in the next section that this assumption holds for several interesting cases. Finally, it is shown that a simple restoration procedure always ensures consistency after one step.

6 Sufficient conditions for consistency of (QP^k)

Example (2.3) shows that the QP approximation to an MPEC can be inconsistent arbitrarily close to a stationary point. This section gives two situations in which consistency of (QP^k) can be guaranteed under assumptions **[A1]** to **[A5]**. The first such situation arises, when there are no general constraints. Next, we show that one step of a simple restoration procedure is guaranteed to find an iterate with $z_1^{(k)^T} z_2^{(k)} = 0$, thus ensuring consistency.

6.1 Vertical complementarity constraints

This section shows that the QP approximations (QP^k) are consistent arbitrarily close to a strongly stationary point, provided that the MPEC has the following form

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & 0 \leq G(z) \ \perp \ H(z) \geq 0, \end{array}$$

i.e. there are no general constraints apart from the vertical complementarity constraints. This case has been brought to our attention by Anitescu. This argument can also be extended to the case, where the complementarity condition is written in terms of slacks.

In this section, we make the following additional assumption.

[A8] $G(z)$ and $H(z)$ are twice continuously differentiable and $\nabla G(z^*)$ and $\nabla H(z^*)$ have full rank.

Define index sets

$$\mathcal{G} := \{i \ : \ G_i(z^*) = 0\} \quad \text{and} \quad \mathcal{H} := \{i \ : \ H_i(z^*) = 0\},$$

which take the place of $\mathcal{Z}_1$ and $\mathcal{Z}_2$. Note that the full rank assumption **[A8]** is in addition to the MPEC-LICQ which in the present case implies that the columns of the matrix

$$\left[\nabla G^*_{\mathcal{G}} \ : \ \nabla H^*_{\mathcal{H}}\right],$$

are linearly independent.

Lemma 6.1 *Let assumptions* **[A1]** - **[A5]** *and* **[A8]** *hold, then it follows that* (QP^k) *is consistent for all* $z^{(k)}$ *in a neighbourhood of* z^* *where* $G^{(k)^T} H^{(k)} \geq 0$. *If in addition, the functions* $G(z)$ *and* $H(z)$ *are convex, then* $G^{(k+1)^T} H^{(k+1)} \geq 0$.

Proof. From assumption **[A2]** it follows that there exists a solution $\hat{d}$ to the set of equations

$$G_i^{(k)} + \nabla G_i^{(k)^T} d = 0 \quad , \quad i \in \mathcal{G} \tag{6.14}$$
$$H_i^{(k)} + \nabla H_i^{(k)^T} d = 0 \quad , \quad i \in \mathcal{H}, \tag{6.15}$$

which is strictly feasible with respect to the remaining inequalities, i.e.

$$G_i^{(k)} + \nabla G_i^{(k)^T} \hat{d} > 0 \ , \ \forall i \in \mathcal{G}^{\perp} \ \text{ and } \ H_i^{(k)} + \nabla H_i^{(k)^T} \hat{d} > 0 \ , \ \forall i \in \mathcal{H}^{\perp}.$$

Note, that $\hat{d}$ is feasible for an LPEC approximation, since $\mathcal{G} \cup \mathcal{H} \supset \{1, \ldots, p\}$ implies that

$$\left(G_i^{(k)} + \nabla G_i^{(k)^T} \hat{d}\right)^T \left(H_i^{(k)} + \nabla H_i^{(k)^T} \hat{d}\right) = 0. \tag{6.16}$$

It remains to show that $\hat{d}$ is also feasible in the linearization of the complementarity constraint, which is given by

$$\begin{aligned} 0 \;\geq\; G^{(k)^T} H^{(k)} & + G^{(k)^T} \nabla H^{(k)^T} \hat{d} + H^{(k)^T} \nabla G^{(k)^T} \hat{d} \\ = G^{(k)^T} H^{(k)} & + \sum_{i \in \mathcal{H}} G_i^{(k)^T} (-H_i^{(k)}) + \sum_{i \in \mathcal{G}} H_i^{(k)^T} (-G_i^{(k)}) \\ & + \sum_{i \in \mathcal{H}^\perp} G_i^{(k)^T} \nabla H_i^{(k)^T} \hat{d} + \sum_{i \in \mathcal{G}^\perp} H_i^{(k)^T} \nabla G_i^{(k)^T} \hat{d} \end{aligned}$$

where we have split the summation into $\mathcal{G}$ and $\mathcal{G}^\perp$ and used (6.14) and (6.15). Now note that the first three term on the right hand side simplify to give

$$= - \sum_{i \in \mathcal{G} \cap \mathcal{H}} G_i^{(k)} H_i^{(k)} + \sum_{i \in \mathcal{H}^\perp} G_i^{(k)^T} \nabla H_i^{(k)^T} \hat{d} + \sum_{i \in \mathcal{G}^\perp} H_i^{(k)^T} \nabla G_i^{(k)^T} \hat{d}.$$

Since $\hat{d} \in \text{range}(\nabla H_i^{(k)},\ i \in \mathcal{H})$ and by **[A8]** it follows that

$$\nabla H_i^{(k)^T} \hat{d} = 0, \ \forall i \in \mathcal{H}^\perp,$$

and similarly for $i \in \mathcal{G}^\perp$. Thus the right hand side now becomes

$$= - \sum_{i \in \mathcal{G} \cap \mathcal{H}} G_i^{(k)} H_i^{(k)} \leq 0$$

which follows from $G^{(k)^T} H^{(k)} \geq 0$ and thus, we have shown that $\hat{d}$ satisfies the linearization of the complementarity constraint.
Finally, if $G(z)$ and $H(z)$ are convex, then it follows that

$$G^{(k+1)} \;=\; G(z^{(k)} + d) \;\geq\; G^{(k)} + \nabla G^{(k)^T} d$$

and similarly for $H^{(k+1)}$. Using this, it follows that

$$G^{(k+1)^T} H^{(k+1)} \;\geq\; \left(G^{(k)} + \nabla G^{(k)^T} d \right)^T \left(H^{(k)} + \nabla H^{(k)^T} d \right) \;\geq\; 0,$$

by (6.16). □

Lemma 6.1 can be generalized to the situation where the MPEC has the form

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c(z_0) \geq 0 \\ & 0 \leq G(z) \;\perp\; H(z) \geq 0, \end{array}$$

provided that $0 \leq G(z) \;\perp\; H(z) \geq 0$ is consistent for any z_0 in a neighbourhood of z^*. Finally, note that the conclusions of this section remain valid, if slacks are added to the vertical complementarity problem.

The main conclusion of this section is that Assumption **[A8]** turns out to be satisfied for a range of practical problems as long as the vertical complementarity problem has certain properties. These assumption are satisfied for instance, for obstacle problems.

6.2 Feasibility restoration for complementarity

Finally, this section examines the properties of (QP^k), in the case, where $z_1^{(k)^T} z_2^{(k)} > 0$. In this case, it is possible that (QP^k) is inconsistent. This section describes a simple restoration procedure which can be invoked, if (QP^k) is inconsistent and which finds a new iterate $z^{(k+1)}$ with $z_1^{(k+1)^T} z_2^{(k+1)} = 0$. Thus after one step, all subsequent iterates retain feasibility of the QP approximations by virtue of Theorem 5.1.

If (QP^k) is inconsistent, then we consider solving the following LP

$$(LP_F^k) \left\{ \begin{array}{ll} \underset{d,\theta}{\text{minimize}} & \theta \\ \text{subject to} & c_{\mathcal{E}}^{(k)} + A_{\mathcal{E}}^{(k)^T} d = 0 \\ & c_{\mathcal{I}}^{(k)} + A_{\mathcal{I}}^{(k)^T} d \geq 0 \\ & z_1^{(k)} + d_1 \geq 0 \\ & z_2^{(k)} + d_2 \geq 0 \\ & z_1^{(k)^T} z_2^{(k)} + z_2^{(k)^T} d_1 + z_1^{(k)^T} d_2 \leq \theta. \end{array} \right.$$

It follows from Assumption [**A2**] that any QP approximation to the relaxed NLP (3.2) is consistent for $z^{(k)}$ sufficiently close to z^* and thus that (LP_F^k) is consistent (since it is a relaxation of the relaxed QP). If $z^{(k)}$ is far away from z^*, then clearly, (LP_F^k) need not be consistent and in that case we enter a restoration phase.

The following Lemma shows that the solution d of (LP_F^k) satisfies $(z_1^{(k)} + d_1)^T(z_2^{(k)} + d_2) = 0$. The key idea of the proof is to show that the optimal active set includes $\mathcal{Z}_1$ and $\mathcal{Z}_2$.

Lemma 6.2 *Let assumptions* [**A1**] *to* [**A4**] *hold and assume that* $z^{(k)}$ *is sufficiently close to* z^*, *so that the linearizations of* $c_{\mathcal{E}}(z)$, $c_{\mathcal{I}}(z)$ *are consistent and* $z_1^{(k)}, z_2^{(k)} \geq 0$. *Then it follows that* (LP_F^k) *has a solution* d *such that* $z^{(k+1)} = z^{(k)} + d$ *satisfies* $z_1^{(k+1)^T} z_2^{(k+1)} = 0$.

Proof. Assume without loss of generality that $\mathcal{Z}_1^{\perp} = \emptyset$, i.e. that $z_1^* = 0$ and consider the dual feasibility conditions of (LP_F^k) (primal feasibility follows from Assumption [**A2**]).

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} - \begin{bmatrix} A_{\mathcal{A}}^{(k)} & I_1 & & z_2^{(k)} \\ & & I_2 & z_1^{(k)} \\ & & & -1 \end{bmatrix} \begin{pmatrix} \lambda_{\mathcal{A}} \\ \nu_1 \\ \nu_2 \\ \xi \end{pmatrix} = 0 \qquad (6.17)$$

where $A_{\mathcal{A}}^{(k)}$ is the matrix of active constraint normals of $c_{\mathcal{E}}(z)$, $c_{\mathcal{I}}(z)$ at z^* and $I_2 = [e_i]_{i \in \mathcal{Z}_2}$. It follows immediately, that $\xi = -1$ and that this active set gives rise to a primal feasible solution. Moreover, the columns of the basis matrix in (6.17) are linearly independent by Assumption [**A2**]. Thus there exists a unique solution to (6.17) which satisfies

$$\nu_1 = z_2^{(k)} \geq 0 \text{ and } \nu_2 = z_1^{(k)} \geq 0 \; \forall i \in \mathcal{Z}_2.$$

This choice implies by complementary slackness of (LP_F^k), that $z_1^{(k+1)^T} z_2^{(k+1)} = 0$. To see, how this follows, consider three cases:

Case 1: $i \in \mathcal{Z}_2^{\perp}$ implies that $z_{2i}^{(k)} > 0$ which implies that $\nu_{1i} > 0$ and thus $z_{1i}^{(k)} + d_{1i} = 0$.

Case 2: $i \in \mathcal{Z}_2$ and $z_{1i}^{(k)}, z_{2i}^{(k)} > 0$. This implies that $\nu_{1i}, \nu_{2i} > 0$ and thus $z_{1i}^{(k)} + d_{1i} = 0$ and $z_{2i}^{(k)} + d_{2i} = 0$.
Case 3: $i \in \mathcal{Z}_2$ and $z_{1i}^{(k)} > 0$ but $z_{2i}^{(k)} = 0$. This implies that $\nu_{2i} > 0$ and thus $z_{1i}^{(k)} + d_{1i} = 0$. Finally, the case where $z_{1i}^{(k)} = 0$ but $z_{2i}^{(k)} > 0$ is analogous.
Putting all 3 cases together and recalling that $\mathcal{Z}_1 = \emptyset$ it follows that $z_1^{(k+1)^T} z_2^{(k+1)} = 0$.
It remains to be proven that there exists multipliers λ with $\lambda_{\mathcal{I}} \geq 0$ such that (6.17) holds. If $\lambda_{\mathcal{I} \cap \mathcal{A}} \geq 0$, then there is nothing to show. Hence assume that there exists a multiplier $\lambda_i < 0$ for $i \in \mathcal{I} \cap \mathcal{A}$. Then it is possible to perform an iteration of an active set method on (LP_F^k), which will not remove any columns of I_1 or I_2 from the basis. Since (LP_F^k) is bounded ($\theta > 0$, since (QP^k) is inconsistent) it follows that after a finite number of such pivots, a basis is found with ν_1, ν_2 as above and the conclusion follows. □

Solving (LP_F^k), if (QP^k) is inconsistent is related to the elastic mode of snopt. In elastic mode, some of the constraints are relaxed and an l_∞-QP is solved. The application of snopt to MPECs s explored in [2]. However, unlike snopt, the present restoration will only occur at one iteration.

An alternative to solving (LP_F^k) would be to move $z^{(k)}$ onto the "nearest" axis. This is the effect of (LP_F^k) as can be seen from Lemma 6.2. However, solving (LP_F^k) avoids the need to choose tolerances to break ties between "close" values.

Finally, note that this restoration does not address the wider issue of *global* convergence. It may be possible, that the solution to (LP_F^k) is not acceptable to the global convergence criterion of the SQP method. Clearly, this has to be taken into account in designing a globally convergent SQP method. It is beyond the scope of the present paper, which deals exclusively with local convergence issues.

7 Discussion of assumptions

This section discusses some of the assumptions made in the proof above. In particular, examples are presented which show that SQP will fail to converge at second order rate, if some or all of the assumptions are removed. The following table shows which assumptions seem difficult to remove. Below, each example is presented in turn.

Example	**[A2]**	MFCQ	**[A3]**	slacks	SOSC	comments
`scholtes4`	no	yes	no	yes	yes	$\xi \to \infty$, linear convergence
`sl2`	yes	yes	yes	no	yes	$\xi \to \infty$, nonstationary limit
`ralph2`	yes	yes	yes	yes	no	$\xi < \infty$, linear convergence

7.1 Unbounded multipliers & slow convergence (`scholtes4.mod`)

The following MPEC shows that if we remove the assumptions **[A2]** and in particular **[A3]**, then the NLP multipliers are not bounded (and may not even exist). Despite this, SQP converges linearly to the solution in the example presented here, although quadratic convergence is lost.

Consider the following MPEC (`scholtes4`) from MacMPEC, see also [19]

$$(P)\left\{\begin{array}{ll} \underset{z}{\text{minimize}} & z_1+z_2-z_0 \\ \text{subject to} & -4z_1+z_0\le 0 \\ & -4z_2+z_0\le 0 \\ & 0\le z_1 \;\perp\; z_2\ge 0\ , \end{array}\right.$$

whose optimal solution is $z^*=(0,0,0)^T$. Writing (P) as an NLP gives

$$(P')\left\{\begin{array}{llll} \underset{z}{\text{minimize}} & z_1+z_2-z_0 & & \text{multiplier} \\ \text{subject to} & -4z_1+z_0 & \le 0 & \lambda_1\ge 0 \\ & -4z_2+z_0 & \le 0 & \lambda_2\ge 0 \\ & z_1z_2 & \le 0 & \xi\ge 0 \\ & z_1 & \ge 0 & \nu_1\ge 0 \\ & z_2 & \ge 0 & \nu_2\ge 0. \end{array}\right.$$

Next, it is shown that SQP converges linearly for this problem.

Proposition 7.1 *SQP applied to (P') generates the following sequence of iterates*

$$z^{(k)}=\begin{pmatrix} 2^{2-k} \\ 2^{-k} \\ 2^{-k}\end{pmatrix}\ ,\ \lambda^{(k)}=\begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix}\ ,\ \xi^{(k)}=2^{k-1}+\xi^{(k-1)}/2=\sum_{j=0}^{k-1}2^{(k-1)-2j}$$

for suitable starting values (e.g. $z=(4,1,1)^T$). Moreover, SQP converges linearly.

Proof. By induction. Holds trivially for $k=0$, i.e. starting point. Now assume the assertion holds for k and show it also holds for $k+1$. At iteration k, SQP solves the following QP for a step d

$$(QP^{(k)})\left\{\begin{array}{lll} \underset{z}{\text{minimize}} & d_1\xi^{(k)}d_2+d_1+d_2-d_0 & \\ \text{subject to} & -4d_1+d_0 & \le 0 \\ & -4d_2+d_0 & \le 0 \\ & z_1^{(k)}z_2^{(k)}+z_2^{(k)}d_1+z_1^{(k)}d_2 & \le 0 \\ & z_1^{(k)}+d_1 & \ge 0 \\ & z_2^{(k)}+d_2 & \ge 0. \end{array}\right.$$

Since all QP approximations are consistent, it is possible to choose $\delta=0$ and not to relax the linearization of the complementarity constraint for the sake of simplicity. It is easy to see that the first three constraints are active. Subtracting the second from the first constraint it follows that $d_1=d_2$. Substituting into the third constraint, it follows that $d_1=d_2=-2^{-(k+1)}$ from which it follows that $d_0=4(-2^{-(k+1)})$. Now verify KKT conditions of $(QP^{(k)})$:

$$0=\begin{pmatrix}-1\\ 1\\ 1\end{pmatrix}+\begin{pmatrix}0\\ -2^{-(k+1)}\xi^{(k)}\\ -2^{-(k+1)}\xi^{(k)}\end{pmatrix}+\lambda_1\begin{pmatrix}1\\ -4\\ 0\end{pmatrix}+\lambda_2\begin{pmatrix}1\\ 0\\ -4\end{pmatrix}+\xi\begin{pmatrix}0\\ 2^{-k}\\ 2^{-k}\end{pmatrix}$$

Subtracting the second from the first equation shows that $\lambda_1=\lambda_2$. Substituting into the third equation then verifies that $\lambda_1^{(k+1)}=\lambda_2^{(k+1)}=\frac{1}{2}$. Finally, the second equation shows

$\xi^{(k)} = 2^{k-1} + \xi^{(k-1)}/2$, the recurrence relation for ξ. The explicit formula for ξ follows easily. The iterates clearly converge linearly to the solution. □

Note that (P) satisfies an MPEC-MFCQ, but violates an MPEC-LICQ (which can be seen easily by observing that 4 constraints are active at the solution). In addition, (P) fails to satisfy strong complementarity. For strong complementarity, it would be necessary that $\lambda_i \geq 0$ and $\nu_i \geq 0$, since $z_1 = z_2 = 0$. Checking the first order condition,

$$0 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ -4 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \\ -4 \end{pmatrix} - \hat{\nu}_1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \hat{\nu}_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

it can be seen that the system is underdetermined. Setting $\lambda_1 = t$ we obtain $\lambda_2 = 1 - t$, $\nu_1 = 1 - 4t$ and $\nu_2 = -3 + 4t$. The condition $\nu_i \geq 0$ now implies that $t \leq \frac{1}{4}$ *and* $t \geq \frac{3}{4}$ which cannot hold simultaneously. Thus the solution of (P) is not strongly stationary.

The linear inequalities always ensure that $z_1^{(0)} = z_2^{(0)} \geq 0$ and the above analysis goes through for alternative starting points. It is not clear, what would happen, if we allowed $z_1 < 0$, but sensible NLP solvers will always project the starting point into the set of linear constraints (or at least of the box constraints). filter, snopt and lancelot behave in this way.

7.2 No slacks & convergence to nonstationary point (`sl2.mod`)

The next example shows that SQP methods could converge to nonstationary points, if slacks are not added to replace nonlinear complementarity conditions. Consider the following MPEC (`sl2`) from MacMPEC which involves a nonlinear expression in the complementarity condition

$$(P) \begin{cases} \underset{z}{\text{minimize}} & -z_1 - \frac{1}{2} z_2 \\ \text{subject to} & z_1 + z_2 \leq 2 \\ & 0 \leq z_1^2 - z_1 \perp z_2 \geq 0 \,. \end{cases}$$

The problem has a global solution at $z^* = (2,0)^T$ with $f^* = -2$ and a local solution at $z^* = (0,2)^T$ with $f^* = -1$. Both solution satisfy assumptions **[A1]** to **[A4]**. The feasible set is illustrated by the bold lines in Figure 3.

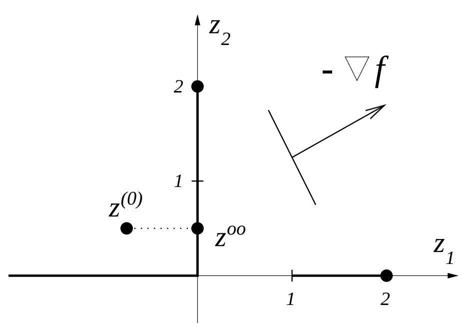

Figure 3: Example `sl2`

Starting at $z^{(0)} = (-\epsilon, t)^T$ gives convergence to the non-stationary point $z^\infty = (0,t)^T$, where $t \geq 0$ is arbitrary. Moreover, it can be shown that $\xi \to \infty$ and that *both* the complementarity constraint *and* $0 \leq z_1^2 + z_1$ remain in the active set. Thus, the active set is singular in the limit. Despite this, second order convergence is observed!

It is straightforward to prove quadratic convergence to a non-stationary limit. Let $z^{(k)} = (-\epsilon, t)^T$ with $t \leq 1$. Then the following problem is solved for a step of the SQP method

$$(P) \left\{ \begin{array}{lrcl} \underset{d}{\text{minimize}} & -d_1 - \frac{1}{2}d_2 & & \\ \text{subject to} & d_1 + d_2 & \leq & 2 + \epsilon - t \\ & (\epsilon^2+\epsilon) - (2\epsilon+1)d_1 & \geq & 0 \\ & t + d_2 & \geq & 0 \\ & t(\epsilon^2+\epsilon) - t(2\epsilon+1)d_1 + (\epsilon^2+\epsilon) & \leq & 0 \end{array} \right.$$

whose solution is

$$d = \begin{pmatrix} \frac{\epsilon^2+\epsilon}{2\epsilon+1} \\ 0 \end{pmatrix}, \ \xi = \frac{1}{2(\epsilon^2+\epsilon)}, \ \nu_1 = \frac{1}{2\epsilon+1} + \xi t.$$

It can be seen that $z^{(k+1)} = (-\mathcal{O}(\epsilon^2), t)^T$ and quadratic convergence occurs to $z^\infty = (0, t)^T$. On the other hand, the multiplier ξ clearly diverges to infinity. Note that including the Hessian of the Lagrangian leeds to a similar conclusion. This example shows that it is not sufficient to trigger the elastic mode only when QP become inconsistent. Clearly, elastic mode is also required, if the multipliers become too large. The introduction of slacks avoids the need for elastic mode in this example.

Introducing a slack, SQP converges quadratically. The SQP solver `filter` exhibits this behaviour, while `lancelot` and `loqo` converge even for the problem *without* slacks. The reason for this apparently better behaviour is of course the fact that both introduce slacks internally before solving the problem!

Another reason for using slacks (rather than linear or even nonlinear complementarity) is that SQP solvers maintain linear feasibility throughout the iteration. Thus they *guarantee* that $z_1^{(k)} \geq 0,\ z_2^{(k)} \geq 0$ for all iterations k in *exact* arithmetic. In *inexact* arithmetic, it is possible to truncate QP steps such that $z_1^{(k)} \geq 0,\ z_2^{(k)} \geq 0$ for all iterations k. This is *not* possible for general *linear* complementarity conditions even if iterative refinement were used.

Thus the use of slacks ensures that $z_1^{(k)^T} z_2^{(k)} \geq 0$ for all iterations k and the trivial pitfall of [5] where it was observed that perturbing the right-hand-side of the complementarity constraint to $-\epsilon$ renders an inconsistent QP cannot occur.

7.3 No second order condition (`ralph2.mod`)

The following MPEC shows that if the second order sufficient condition **[A3]** is violated, then SQP may converge only linearly.

$$(P) \left\{ \begin{array}{ll} \underset{z}{\text{minimize}} & z_1^2 + z_2^2 - 4z_1z_2 \\ \text{subject to} & 0 \leq z_1 \ \perp \ z_2 \geq 0 \ . \end{array} \right.$$

The problem has a global solution at $(0,0)$. Starting at $z = (1,1)$ causes SQP to converge linearly to the solution. Note that (P) also violates any upper level strict complementarity condition.

The MPEC-SOSC is rather stronger than needed for MPECs in the sense that the set of directions over which positive curvature is required for SQP is larger than the set of MPEC-feasible-directions. This is illustrated by this example. The set of MPEC-feasible-directions at $(0,0)$ is

$$S_M^* = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

while the set of directions over which curvature is required to be positive for SQP to converge is the whole positive orthant (i.e. $\mathrm{conv}(S_M^*)$). The linear rate of convergence is due to the fact that the curvature in the direction $(1,1)$ is negative.

8 Conclusions and future work

We have presented a convergence analysis which shows that SQP methods converge quadratically when applied to the NLP equivalent of an MPEC. This analysis goes some way towards explaining the extraordinary success of SQP solvers applied to MPECs which we have observed. This is a remarkable result since MPEC violate the Mangasarian Fromowitz constraint qualification.

Conditions are identified under which local second order convergence occurs. These conditions include the assumption that all QP approximations remain consistent. It can be shown, that this assumption always holds, if $z_1^{(k)^T} z_2^{(k)} = 0$, i.e. for iterates which satisfy complementarity and this is often observed in practice. We have also shown that MPECs whose lower level problem is a certain vertical complementarity problem generate consistent QP approximations. Finally, we have given a restoration phase which ensure that this can always be guaranteed sufficiently close to a solution.

We have also experimented with an alternative to a restoration problem. In this approach, the linearization of the complementarity condition is relaxed as

$$z_1^{(k)^T} z_2^{(k)} + z_2^{(k)^T} d_1 + z_1^{(k)^T} d_2 \leq \delta \left(z_1^{(k)^T} z_2^{(k)} \right)^{1+\kappa} , \tag{8.1}$$

where $0 < \delta, \kappa < 1$ are constants. Note that the perturbation to the right hand side of the complementarity condition is $o(\|d_{NR}\|)$ where d_{NR} is the Newton step. This form of perturbation allows the superlinear convergence proof to be extended by virtue of the Dennis-Moré characterization theorem.

However, the perturbation alone is not sufficient to guarantee consistency of (QP^k). The following example illustrates the need for further assumptions. Consider the following feasible set

$$z_1 + z_2 - 1 \geq 0 \ , \ 0 \leq z_1 \perp z_2 \geq 0.$$

Then it is easy to see, that for any $z = (\epsilon^4, 1-\epsilon)$, the (QP^k) relaxed using (8.1) with $\delta = \kappa = 0.5$ is inconsistent. Note that if we restrict attention to points z which satisfy the linear constraints, e.g. $z = (\epsilon, 1-\epsilon)$, then (QP^k) using (8.1) is consistent in a neighbourhood of $z = (0,1)$. Thus it seems as though (8.1) ensures consistency of (QP^k), as long as $z^{(k)}$ satisfies the linearizations of $c_{\mathcal{E}}(z)$, $c_{\mathcal{I}}(z)$ about $z^{(k-1)}$. Unfortunately, we have been unable to prove any general results along those lines. Such a proof would clearly allow us to boot-strap a convergence of SQP for MPECs with a relaxed equation (8.1).

We finish this paper with some observations on the role of degeneracy and point to some future work. It has been observed that any QP approximation about a feasible point of (1.3) is degenerate. Moreover, approximations about points which satisfy $z_1^T z_2 = \epsilon > 0$ are near-degenerate and we would expect this to play a role in the SQP method. In our numerical experiment we use two SQP solvers, snopt and filter.

snopt uses EXPAND to handle degeneracy. This procedure, perturbs the bounds of (QP^k) to *remove* degeneracy. Some numerical experiments suggest that this is not the best way to handle degeneracy in the case of MPECs. The QP solver in filter, bqpd applies a different methodology to handle degeneracy. It *creates* degeneracy whenever *near degeneracy* is detected and then handles the degenerate situation. This has two implications.

1. If *exact degeneracy* exists (i.e. if $z_1^{(k)^T} z_2^{(k)} = 0$), then bqpd will deal with it.

2. Secondly, if *near degeneracy* exists (i.e. if $z_1^{(k)^T} z_2^{(k)} = \epsilon > 0$), then bqpd *creates degeneracy* by perturbing the bound ϵ to zero. This has the effect of *pushing* the solution onto the axis. As we have shown above, this is a very favourable situation for SQP methods.

Future work will focus on the questions of relaxing some assumptions and in providing a global convergence analysis. Some numerical results suggest that SQP converges under even weaker assumptions than those made above and it may be possible to pursue the ideas of [20] in this context. Another important question concerns the global convergence of SQP methods. [2] provides a framework for convergence (possibly under additional assumptions) of Sl_∞QP methods. However, the numerical results suggest, that a similar proof may be possible for filter methods.

References

[1] Andreani, R. and Martínez, J.M. On the solution of mathematical programming problems with equilibrium constraints using nonlinear programming algorithms. Technical Report 42/00, Department of Applied Mathematics, IMECC-UNICAMP, www.ime.unicamp.br/rel_pesq/relatorio.html, November 2000.

[2] Anitescu, M. On solving Mathematical Programs with Complementarity Constraints as Nonlinear Programs. Preprint ANL/MCS-P864-1200, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, USA, 2000.

[3] Bard, J.F. Convex two-level optimization. *Mathematical Programming*, 40(1):15–27, 1988.

[4] Bonnans, J.F. Local convergence analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[5] Chen, Y. and Florian, M. The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. *Optimization*, 32:193–209, 1995.

[6] Conn, A.R. Gould, N.I.M. and Toint, Ph.L. Numerical experiments with the lancelot package (Release A) for large–scale nonlinear optimization. *Mathematical Programming*, 73(1):73–110, 1996.

[7] Conn, A.R. Gould, N.I.M. and Toint, Ph.L. *Trust-region Methods.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[8] Ferris, M.C. and Pang, J.S. Engineering and economic applications of complementarity problems. *SIAM Review*, 39(4):669–713, 1997.

[9] Fischer, A. Modied Wilson method for nonlinear programs with nonunique multipliers. *Mathematics of Operations Research*, 24:699–727, 1999.

[10] Fletcher, R. and Leyffer, S. Numerical experience with solving MPECs by nonlinear programming methods. Numerical Analysis Report NA/YYY, Department of Mathematics, University of Dundee, Dundee, DD1 4HN, UK, 2001. In preparation.

[11] Fletcher, R. and Leyffer, S. Nonlinear programming without a penalty function. *Mathematical Programming*, 91(2):239–269, January 2002.

[12] Gill, P.E., Murray, W. and Saunders, M.A. SNOPT: An SQP algorithm for large–scale constrained optimization. Report NA 97-2, Dept. of Mathematics, University of California, San Diego, November 1997.

[13] Hager, W.W. Stabilized sequential quadratic programming. *Computational Optimization and Application*, 12:253–273, 1999.

[14] Jiang, H. and Ralph, D. QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints. Technical report, University of Melbourne, Department of Mathematics, December 1997.

[15] Leyffer, S. MacMPEC: AMPL collection of MPECs. Technical report, `www.maths.dundee.ac.uk/~sleyffer/MacMPEC/`, 2000.

[16] Luo, Z.-Q., Pang, J.-S. and Ralph, D. *Mathematical Programs with Equilibrium Constraints.* Cambridge University Press, 1996.

[17] Outrata, J., Kocvara, M. and Zowe, J. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publishers, Dordrecht, 1998.

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

[19] Scheel, H. and Scholtes, S. Mathematical program with complementarity constraints: Stationarity, optimality and sensitivity. *Mathematics of Operations Research*, 25:1–22, 2000.

[20] Wright, S.J. Modifying SQP for degenerate problems. Preprint ANL/MCS-P699-1097, Mathematics and Computer Science Division, Argonne National Laboratory, November 1997. (Revised June, 2000).