Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints*

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

May 9, 2002, revised July 9, 2004

Abstract

Recently, nonlinear programming solvers have been used to 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. SQP is shown to converge superlinearly under reasonable assumptions near a strongly stationary point. A number of examples are presented that 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 [7, 16, 17] for references. (Problem (1.1) is also referred to as a mathematical program with complementarity constraints.)

*Numerical Analysis Report NA/209, Department of Mathematics, University of Dundee.

†Department of Mathematics, University of Dundee, Dundee, DD1 4HN, UK, (fletcher@maths.dundee.ac.uk).

‡Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA, (leyffer@mcs.anl.gov).

§The Judge Institute, University of Cambridge, Cambridge, CB2 1AG, UK, ({d.ralph,s.scholtes}@jims.cam.ac.uk).

Clearly, an MPEC with a more general complementarity condition such as

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

can be written in the form (1.1) by introducing slack variables. One can easily show that the reformulated MPEC has the same properties (such as constraint qualifications or second-order conditions) as the original MPEC. In this sense, nothing is lost by introducing slack variables.

One attractive 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).

The NLP (1.3), obviously has no feasible point that satisfies the inequalities strictly. This fact implies that the Mangasarian-Fromovitz constraint qualification (MFCQ) is violated at every feasible point; see [4, 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) has also been disappointing. Bard [2] reports failure on 50–70% of some bilevel problems for a gradient projection method. Conn et al. [5] and Ferris and Pang [7] attribute certain failures of lancelot to the fact that the problem contains a complementarity constraint. In contrast, Fletcher and Leyffer [10] recently reported encouraging numerical results on a large collection of MPECs [15]. They solved over 150 MPECs with an SQP solver and observed *quadratic* convergence for all but two problems. The two problems that did 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.

Recently, researchers have expressed renewed interest in the global convergence of algorithms for MPECs. Scholtes [20] analyzes a regularization scheme in which a sequence of parametric NLPs is solved. Fukushima and Tseng [11] analyze an algorithm that computes approximate KKT points for a sequence of active sets.

The paper also complements the recently renewed interest in the convergence properties of SQP under weaker assumptions. See for example [8, 13, 21]. These studies suggest modifications to enable SQP solvers to handle NLP problems for which the constraint gradients are linearly dependent at the solution and/or for which strict complementarity fails to hold.

Anitescu [1] extends Wright's analysis [21] 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 therefore applies to MPECs in the given form. However, his assumptions differ from ours, and neither set of assumptions is 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.

In this paper, we argue that the introduction of slack variables is not just a convenience but plays an important role in ensuring convergence. In Section 7.2 we present an example with a nonlinear complementarity constraint for which SQP converges to a nonstationary point. All QP approximations remain consistent during the solve. With the introduction of slack variables, 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 that highlight the key ideas of our approach and illustrate the numerical difficulties associated with MPECs. In Section 3 we review optimality conditions and constraint qualifications for MPECs. Section 4 shows that the optimality conditions of the MPEC and its equivalent NLP are related by a simple formula. In Section 5 we show 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. In Section 7 we present small examples that illustrate the necessity of some of these assumptions. We conclude by briefly emphasizing the importance of degeneracy handling at the QP level and pointing to future research directions.

Notation. 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, for example, $a_i^{(k)} = a_i(z^{(k)}) = \nabla c_i(z^{(k)})$. 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 Examples

The fact that the NLP formulation (1.3) of an MPEC violates MFCQ at any feasible point implies that (1.3) has certain features that 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. Arbitrarily close to a stationary point the linearizations of (1.3) can be *inconsistent.*

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 possess these unpleasant properties, they arise in a well-structured way that allows SQP solvers to tackle MPECs successfully.

In the remainder of this paper, `*.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

In this section we use a small example from Jiang and Ralph [14] (see also `jr*.mod`) to illustrate the key idea of our approach. Consider the two QPECs

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

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.

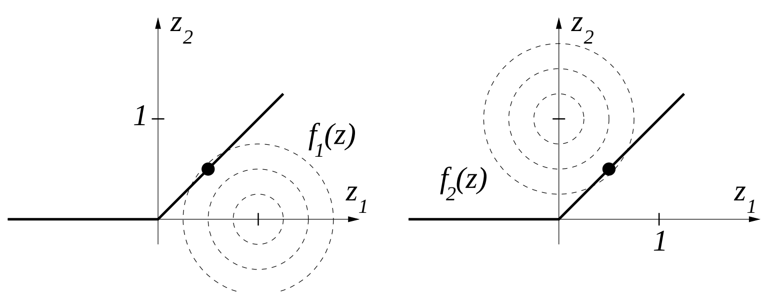

Figure 1: QPEC examples 1 and 2

The equivalent NLP problem to these QPECs is given by

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

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

$$\mathcal{M}_1 = \left\{ (\lambda, \xi) \mid \xi \ge 0,\ \lambda - \tfrac{1}{2}\xi = 1 \right\}$$

$$\mathcal{M}_2 = \left\{ (\lambda, \xi) \mid \lambda \ge 0,\ -\lambda + \tfrac{1}{2}\xi = 1 \right\},$$

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

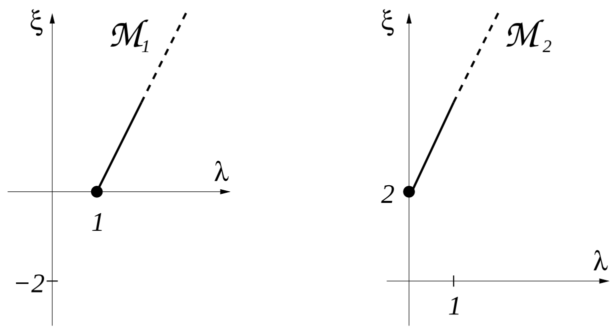

Figure 2: Multiplier sets of QPEC examples 1 and 2

This situation is typical for MPECs that satisfy a strong stationarity condition (see Definition 3.3). 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 that correspond to a *linearly independent set* of constraint normals, then the following reduced sets are obtained:

$$\tilde{\mathcal{M}}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$$

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

These multipliers are bounded and well behaved. We should expect SQP to converge if started near such a stationary point. The KKT multipliers that 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 is essentially an equality constraint. This is exploited in Section 4, where we show that certain MPEC multipliers correspond to multipliers of (1.3). This situation is typical for MPECs under certain assumptions. The key idea is to show that SQP converges to a solution provided the QP solver chooses a linearly independent basis.

2.2 Inconsistent Linearizations

The following example illustrates a possible pitfall for NLP solvers attempting to solve MPECs as NLPs. Consider `sl4.mod`:

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

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). However, linearizing the constraints about a point that satisfies the simple bounds and is *arbitrarily close to the solution*, such as $z^{(0)} = (\epsilon, 1-\delta)^T$ (with $\epsilon, \delta > 0$), gives a QP that is *inconsistent.* The linearizations are

$$\begin{aligned} (1-\delta)^2 + 2(1-\delta)(z_2 - (1-\delta)) \;&\geq\; 1 && (2.4) \\ z_1 \;&\geq\; 0 \\ z_2 \;&\geq\; 0 \\ (1-\delta)\epsilon + (1-\delta)(z_1 - \epsilon) + \epsilon(z_2 - (1-\delta)) \;&\leq\; 0. && (2.5) \end{aligned}$$

One can show that

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

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

Clearly, any NLP solver hoping to tackle MPECs will have to deal with this situation. The solver snopt [12] uses an *elastic mode* that relaxes the linearizations of the QP; filter [9] has a restoration phase. In Section 5 convergence of SQP methods without modifications is analyzed. This analysis is closer in spirit to the results obtained using filter.

3 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], although the presentation is slightly different.

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

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

we 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, we define the *relaxed NLP corresponding to the MPEC (1.1)* as

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

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. In contrast to [19], our definition of the relaxed NLP is independent of a specific point; however, 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 that are both in $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are referred to as the *biactive 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 nondegenerate* if $\mathcal{D}(z^*) = \emptyset$.

First, the linear independence constraint qualification (LICQ) is extended to MPECs.

Definition 3.1 *Let $z_1, z_2 \geq 0$, and define*

$$\mathcal{Z}_| := \{i \;:\; z_{ji} = 0\} \quad \text{for} \;\; j = 1, 2.$$

The MPEC (1.1) is said to satisfy an MPEC-LICQ *at z if the corresponding relaxed NLP (3.2) satisfies an LICQ.*

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

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

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

We note that B-stationarity implies feasibility because if $d = 0$ solves the above LPEC, then $c^*_{\mathcal{E}} = 0$, $c^*_{\mathcal{I}} \geq 0$, and $0 \leq z^*_1 \perp z^*_2 \geq 0$. B-stationarity is difficult to check because it involves the solution of an LPEC that 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 there exist multipliers λ, $\hat{\nu}_1$ and $\hat{\nu}_2$ such that*

$$\begin{aligned} g^* - \left[A^{*^T}_{\mathcal{E}} \; : \; A^{*^T}_{\mathcal{I}} \right] \lambda - \begin{pmatrix} 0 \\ \hat{\nu}_1 \\ \hat{\nu}_2 \end{pmatrix} &= 0 \\ c^*_{\mathcal{E}} &= 0 \\ c^*_{\mathcal{I}} &\geq 0 \\ z^*_1 &\geq 0 \\ z^*_2 &\geq 0 \\ z^*_{1j} = 0 \quad or \quad z^*_{2j} &= 0 \\ \lambda_{\mathcal{I}} &\geq 0 \\ c^*_i \lambda_i &= 0 \\ z^*_{1j} \hat{\nu}_{1j} &= 0 \\ z^*_{2j} \hat{\nu}_{2j} &= 0 \\ if \; z^*_{1j} = z^*_{2j} = 0 \quad then \quad \hat{\nu}_{1j} \geq 0 \quad and \quad \hat{\nu}_{2j} &\geq 0, \end{aligned} \tag{3.3}$$

where $g^* = \nabla f(z^*)$, $A^*_{\mathcal{E}} = \nabla c^T_{\mathcal{E}}(x^*)$, and $A^*_{\mathcal{I}} = \nabla c^T_{\mathcal{I}}(x^*)$.

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, that is,

$$A = \begin{bmatrix} & & 0 & & 0 \\ A^*_{\mathcal{E}} \; : \; A^*_{\mathcal{I}\cap\mathcal{A}^*} & : & I^*_1 & : & 0 \\ & & 0 & & I^*_2 \end{bmatrix} =: \; [a^*_i]_{i\in\mathcal{A}^*} \, ,$$

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 of the relaxed NLP (3.2) as

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

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

Definition 3.4 *A strongly stationary point z^* with multipliers $(\lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$ satisfies the* MPEC-SOSC *if for every direction $s \in S^*$ it follows that*

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

where $\nabla^2 \mathcal{L}^*$ is the Hessian of the Lagrangian of (3.2) evaluated at $(z^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$.

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 and NLP Stationarity

This section shows that there exists a relationship between strong stationarity of the MPEC (1.1) and NLP stationarity conditions for (1.3). In particular, their respective multipliers are shown to be related by a simple formula.

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

$$\begin{aligned} g(z) - \left[A_{\mathcal{E}}^T(z) : A_{\mathcal{I}}^T(z)\right] \lambda - \begin{pmatrix} 0 \\ \nu_1 \\ \nu_2 \end{pmatrix} + \xi \begin{pmatrix} 0 \\ z_2 \\ z_1 \end{pmatrix} &= 0 \\ c_{\mathcal{E}}(z) &\geq 0 \\ c_{\mathcal{I}}(z) &\geq 0 \\ z_1 &\geq 0 \\ z_2 &\geq 0 \\ z_1^T z_2 &\leq 0 \\ \lambda_{\mathcal{I}} &\geq 0 \\ \nu_1 &\geq 0 \\ \nu_2 &\geq 0 \\ \xi &\geq 0 \\ c_i(z)\lambda_i &= 0 \\ z_{1j}\nu_{1j} &= 0 \\ z_{2j}\nu_{2j} &= 0 \quad . \end{aligned} \tag{4.1}$$

The complementarity condition $\xi z_1^T z_2 = 0$ is implied by the feasibility of z_1, z_2. This condition 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

$$g(z) - \left[A_{\mathcal{E}}^T(z) : A_{\mathcal{I}}^T(z)\right] \lambda - \begin{pmatrix} 0 \\ \nu_1 - \xi z_2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \nu_2 - \xi z_1 \end{pmatrix} = 0,$$

which is equivalent to the corresponding first-order condition in (3.3) if

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

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

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 we show 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. From (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) 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), one sees that (4.1) implies (3.3).

Next we show 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 $\nu_{1j} = \xi z_{2j} + \hat{\nu}_{1j} = 0 \geq 0$ for any ξ. To ensure that $\nu_{2j} = \xi z_{1j} + \hat{\nu}_{2j}$ is nonnegative, we need to choose ξ such that $\xi z_{1j} + \hat{\nu}_{2j} \geq 0$, $\forall j$, or equivalently that $\xi \geq -\hat{\nu}_{2j}/z_{1j}$, $\forall j$.

(e) If $z_{2j} > 0$, then transpose above argument.

(f) If $z_{1j} = z_{2j} = 0$, then $\nu_{1j} = \hat{\nu}_{1j} \geq 0$ and $\nu_{2j} = \hat{\nu}_{2j} \geq 0$, for any ξ.

From parts (d) and (e) it follows that choosing ξ to be at least

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

will ensure that $\nu_1, \nu_2 \geq 0$. Examining the expressions on the right-hand side of (4.4), one can see that ξ is bounded. Combining (d) to (f) it follows that (3.3) implies (4.1). $\square$

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 $\nu_{1j} = \hat{\nu}_{1j} + \xi z_{2j}$ nonnegative by choosing ξ sufficiently large.

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 that works with a nonsingular basis will pick the multiplier defined in (4.4).

Definition 4.2 *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 4.3 *If MPEC-LICQ holds at a local minimizer of (1.1), then it is strongly stationary, and the multipliers in (3.3) and the basic multiplier defined by (4.4) are unique. Moreover, the set of constraint normals corresponding to nonzero multipliers is linearly independent.*

Proof. MPEC-LICQ implies the uniqueness of the MPEC multipliers (3.3); see [19]. The uniqueness of the MPEC multiplier implies that all expressions on the right-hand side of (4.4) are unique, hence implying the uniqueness of ξ. Finally, the uniqueness of the corresponding NLP multipliers follows from (4.2) and (4.3) (if the NLP multipliers were not unique, then we could find other MPEC multipliers).

To show that the constraint normals corresponding to nonzero multipliers are linearly independent, we distinguish two cases: $\xi = 0$ and $\xi > 0$.

If $\xi = 0$, then the linear independence of constraint normals corresponding to nonzero 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 5.8 below. □

The conclusions of this section can be readily extended to cover the case where the complementarity condition is of the more 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 counterexamples 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).*

[A3] *z^* is a strongly stationary point of (1.1) with multipliers $\lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*$ (Definition 3.3), and z^* satisfies the MPEC-SOSC (Definition 3.4).*

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

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

The most restrictive assumption is strong stationarity in **[A3]**, which follows if z^* is a local minimizer from **[A2]**. That is **[A3]** (or **[A2]**) removes the combinatorial nature of the problem. It is not clear that **[A2]** can readily be relaxed in the present context, since it allows us 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 that solves only a single LP or QP per iteration can be shown to be convergent to B-stationary points for problems that violate MPEC-LICQ. Note that we do *not* assume that the MPEC (1.1) is second-level nondegenerate, in other words, we do *not* assume that $z_1^* + z_2^* > 0$. Assumption **[A5]** is a reasonable assumption in practice, as most modern SQP solvers are based on active set QP solvers that guarantee this.

This section is divided into two parts. First, we consider the case where complementarity is satisfied at a point sufficiently close to a stationary point. This case corresponds to the situation where all iterates (ultimately) remain on the same face of $0 \leq z_1 \perp z_2 \geq 0$. 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. We make the additional assumption that all QP subproblems remain consistent. This assumption appears to be rather strong, especially in 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 that show that it is not unduly restrictive.

5.1 Local Convergence for Exact Complementarity

In this section we make the following additional assumption:

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

Assumption **[A6]** implies that the correct face has been identified except for degenerate or biactive constraints. Thus, for given index sets $\mathcal{Z}_j = \{i : z_{ji}^{(k)} = 0\},\ j = 1, 2$, the following holds:

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

In particular, it is *not* assumed that the biactive 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 will be shown that the biactive constraints become active after one step of the SQP method as a consequence of [**A4**] (the positivity of biactive multipliers); see Proposition 5.2.

An important consequence of [**A6**] is that $\mathcal{Z}_1$ and $\mathcal{Z}_2$ satisfy

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

in other words, the indices $\mathcal{Z}_1^{*\perp}$ and $\mathcal{Z}_2^{*\perp}$ of the nondegenerate complementarity constraints have been identified correctly.

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 following QP:

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

where

$$\widehat{W^{(k)}} = \nabla^2 f(z^{(k)}) - \sum \lambda_i^{(k)} \nabla^2 c_i(z^{(k)}) = W^{(k)} - \xi^{(k)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & I \\ 0 & I & 0 \end{bmatrix}$$

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 $(QP_R(z^{(k)}))$ 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 $(QP_R(z^{(k)}))$.

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 Assumptions* [**A1**]–[**A6**] *hold, and consider the relaxed NLP for any index sets $\mathcal{Z}_1$, $\mathcal{Z}_2$ (satisfying (5.1) by virtue of* [**A6**]*). Then it follows that*

1. *there exists a neighborhood U of $(z^*, \lambda^*, \nu_1^*, \nu_2^*)$ and a sequence of iterates generated by SQP applied to the relaxed NLP (3.2), $\{(z^{(l)}, \lambda^{(l)}, \nu_1^{(l)}, \nu_2^{(l)})\}_{l>k}$ that lies in U and 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, there exists a neighborhood U of $(z^*, \lambda^*, \nu_1^*, \nu_2^*)$ such that for any $(z^{(l)}, \lambda^{(l)}, \nu_1^{(l)}, \nu_2^{(l)}) \in U$, there exists an SQP iterate $(z^{(l+1)}, \lambda^{(l+1)}, \nu_1^{(l+1)}, \nu_2^{(l+1)})$ that also lies in U; and any sequence of SQP iterates $\{z^{(l)}\}_{l>k} \subset U$ converges at second-order rate when applied to the relaxed NLP. In fact Part 1. is a standard result whose proof can be found, for instance, in [6, Theorem 15.2.2] or in or in [3]. Part 2. is due to [3]. Part 3 follows from the fact that SQP identifies the correct active set in one step by the strict complementarity assumption [**A4**]. □

Next, we show 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 $z_1^{(k)^T} z_2^{(k)} = 0$, that is [**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 $(QP_R(z^{(k)}))$ have different constraints and Hessians. The Hessian of (QP^k) is

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

Despite these differences, however, one can 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 5.3 *Let Assumptions* [**A1**]–[**A6**] *hold. Then, a step* d *is feasible in* (QP^k) *if and only if it is feasible in* $(QP_R(z^{(k)}))$.

Proof. The constraint sets differ only 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* $(QP_R(z^{(k)}))$. Then it follows in particular that d satisfies

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

If these constraints are split into two inequalities, we have 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). 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 **[A6]** that $z_{1j}^{(k)} = 0,\ \forall j \in \mathcal{Z}_2^{\perp}$, and that $z_{2j}^{(k)} = 0,\ \forall j \in \mathcal{Z}_1^{\perp}$. Thus, (QP^k) contains the constraints

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

By **[A6]**, the linearization of the complementarity constraint in (QP^k) simplifies to

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

Since $z_{2j}^{(k)} > 0$, and $z_{1j}^{(k)} > 0$ in this sum, it follows 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 $(QP_R(z^{(k)}))$. □

Next, we show that near z^*, the stationary points of (QP^k) and $(QP_R(z^{(k)}))$ are identical.

Lemma 5.4 *Let Assumptions* **[A1]**–**[A6]** *hold,. Then it follows that the objective function of $(QP_R(z^{(k)}))$ is less than or equal to the objective function of (QP^k) on the feasible set. Moreover, any stationary point of $(QP_R(z^{(k)}))$ near is also a stationary point of (QP^k) and vice versa.*

Proof. For any feasible point of $(QP_R(z^{(k)}))$ (and hence of (QP^k)), it follows from **[A6]** and Lemma 5.3 that $d_1^T d_2 \geq 0$. This gives the first claim. We use a related fact below that if $d_1^T d_2 = 0$ then the corresponding gradients of the objective functions of $(QP_R(z^{(k)}))$ and (QP^k) coincide.

From Part 3 of Proposition 5.2 it can be deduced that any stationary point d of $(QP_R(z^{(k)}))$, near zero, satisfies $d_1^T d_2 = 0$. Thus the gradients of the objective functions of $(QP_R(z^{(k)}))$ and (QP^k) at d coincide. Since these QPs also share the same feasible set, then d must be stationary for the latter. Conversely, if d is stationary for (QP^k), again we have that $d_1^T d_2 = 0$, hence d is stationary for $(QP_R(z^{(k)}))$. □

Lemma 5.5 *Let Assumptions* **[A1]**–**[A6]** *hold. Let $(\lambda, \widehat{\nu}_1, \widehat{\nu}_2)$ be the multipliers of $(QP_R(z^{(k)}))$ (corresponding to a step d near zero). 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 $(QP_R(z^{(k)}))$.

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 nonnegative by construction and satisfy first-order conditions by Lemma 5.4. $\square$

Next, we show that both QPs have the same (unique) solution in a neighborhood of $d = 0$.

Lemma 5.6 *The solution d of $(QP_R(z^{(k)}))$ is the only strict local minimizer in a neighborhood of $d = 0$ that is independent of k, and its corresponding multipliers (λ, ν_1, ν_2) are unique. Moreover, d is also the only strict local minimizer in a neighborhood of $d = 0$ of (QP^k).*

Proof. The result for $(QP_R(z^{(k)}))$ is due to Robinson [18] (see also Conn, Gould and Toint [6]), since the relaxed NLP satisfies **[A1]**–**[A4]**. The statement for (QP^k) follows in two parts. First-order conditions are established in Lemma 5.5. Second-order conditions for (QP^k) follow from second-order conditions of $(QP_R(z^{(k)}))$, as the two problems have the same nullspace, Lemma 5.3. Thus the solution d of $(QP_R(z^{(k)}))$ is also a solution of (QP^k).

If there were another solution of (QP^k) in a neighborhood $\mathcal{N}$ of $d = 0$, then one could construct a corresponding solution to $(QP_R(z^{(k)}))$ in $\mathcal{N}$ (using Lemma 5.5) that would contradict the uniqueness of the solution. $\square$

The following theorem summarizes the results of this section.

Theorem 5.7 *If Assumption* **[A1]**–**[A6]** *hold, then SQP applied to (1.3) generates a sequence $\{(z^{(k)}, \lambda^{(k)}, \nu_1^{(k)}, \nu_2^{(k)}, \xi^{(k)})\}_{l>k}$ that 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]**–**[A4]**, SQP converges quadratically when applied to the relaxed NLP (3.2). Lemmas 5.3–5.6 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)–(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 the basic multiplier (4.4). Finally, (5.9) and (5.10) converge to (4.2) and (4.3) by a similar argument.

Now ${z_1^{(l)}}^T z_2^{(l)} = 0\ ,\ \forall l \geq k$ 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 biactive complementarity constraints. Thus SQP will not move out of the corner but will stay on the same face. □

5.2 Local Convergence for Nonzero Complementarity

This section shows that SQP converges superlinearly even if complementarity does not hold at the starting point, that is 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. To avoid this problem, we make the following assumption, which often holds in practice.

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

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

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

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, we let $B_+^{(k)}$ denote the matrix of strictly active constraint normals at $z = z^{(k)}$, namely,

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

Note that Lemma 4.3 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 special form that can be exploited to prove fast convergence.

Lemma 5.8 *Let Assumptions* **[A1]**–**[A4]** *hold, and let z^* be a solution of the MPEC (1.1). Let $\mathcal{I}^*$ denote the set of* active *inequalities $c_{\mathcal{I}}(x)$, and consider the matrix of active constraint normals at z^*,*

$$B^* \ = \ \left[\begin{array}{ccccc} & & 0 & 0 & 0 \\ A_{\mathcal{E}}^* & A_{\mathcal{I}^*}^* & I & 0 & \begin{pmatrix} 0 \\ z_{22}^* \end{pmatrix} \\ & & 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{array} \right], \tag{5.11}$$

where we have assumed without loss of generality that $\mathcal{Z}_1^{\perp^*} = \emptyset$. 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 idea 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 5.9 *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^k, given by*

$$B^{(k)} = \begin{bmatrix} & & 0 & 0 & 0 \\ A_{\mathcal{E}}^{(k)} & A_{\mathcal{I}^*}^{(k)} & I & 0 & \begin{pmatrix} z_{21}^{(k)} \\ z_{22}^{(k)} \end{pmatrix} \\ & & 0 & \begin{bmatrix} I \\ 0 \end{bmatrix} & \begin{pmatrix} z_{11}^{(k)} \\ z_{12}^{(k)} \end{pmatrix} \end{bmatrix},$$

then it follows that $z_1^{(k)^T} z_2^{(k)} > 0$ and that after the QP step, $z_1^{(k+1)^T} z_2^{(k+1)} = 0$.

Proof. Assume that $z_1^{(k)^T} z_2^{(k)} = 0$, and seek a contradiction. Since $z^{(k)}$ is sufficiently close to z^*, it follows that there exists $\tau > 0$ such that $z_{22}^{(k)} \geq \tau > 0$. Hence, $z_{12}^{(k)} = 0$. Now consider the final three columns of $B^{(k)}$, 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, thus contradicting Assumption [**A5**], and so $z_1^{(k)^T} z_2^{(k)} > 0$.

Now, $z_1^{(k+1)^T} z_2^{(k+1)} = 0$ follows simply by observing that the full basis B implies that $0 = z_1^{(k)} + d_1 = z_1^{(k+1)}$. □

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^*. Second-order convergence then follows from Theorem 5.7.

Corollary 5.10 *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 in which the full basis B is never chosen and $z_1^{(k)^T} z_2^{(k)} > 0$ for all iterates k (otherwise, we have convergence from the results of the previous section).

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

Lemma 5.11 *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 $A_{\mathcal{I}^*}^{(k)}$ of active constraints at the solution.*

Proof. The proof follows by considering the gradient of the Lagrangian of (QP^k),

$$0 \;=\; g^{(k)} + \hat{W}^{(k)} d^{(k)} - \left[A_{\mathcal{E}}^{(k)^T} : A_{\mathcal{I}}^{(k)^T} \right] \lambda^{(k+1)} - \begin{pmatrix} 0 \\ \nu_1^{(k+1)} - \xi^{(k+1)} z_2^{(k)} \\ \nu_2^{(k+1)} - \xi^{(k+1)} z_1^{(k)} \end{pmatrix} + \xi^{(k)} \begin{pmatrix} 0 \\ d_2^{(k)} \\ d_1^{(k)} \end{pmatrix},$$

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

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 5.9. The key idea is now to show that for any such basis, there exists an equality constrained problem for which SQP converges quadratically. Since there exist only a finite number of bases, this implies convergence for SQP.

We now introduce the *reduced NLP*, which is an equality constraint NLP. Its constraints correspond to a linearly independent subset of the basis B^* in (5.11) of Lemma 5.8.

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c_{\mathcal{E}}(z) = 0 \\ & c_{\mathcal{I}^*}(z) = 0 \\ & z_{11} = 0 \\ & z_{21} = 0 \\ & \left.\begin{array}{r} z_{12} = 0 \\ z_1^T z_2 = 0 \end{array}\right\} \text{subset of } B^* \text{ satisfying Lemma 5.8.} \end{array} \tag{5.12}$$

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

Lemma 5.12 *Let Assumptions* **[A1]**–**[A4]** *and* **[A7]** *hold. Then it follows that any reduced NLP satisfies an LICQ and an SOSC.*

Proof. Lemma 5.8 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. $\square$

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

Proposition 5.13 *Let Assumptions* **[A1]**–**[A4]** *and* **[A7]** *hold. Then it follows that SQP applied to any reduced NLP converges locally and quadratically to (z^*, μ^*).*

Proof. Lemma 5.12 shows that the reduced NLP satisfies LICQ and SOSC. 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 the following theorem.

Theorem 5.14 *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 5.13 shows that SQP converges quadratically 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 quadratic convergence follows. Thus, for each basis, a step is computed that satisfies a contraction condition like (5.13) for a constant $c_{\mathcal{B}} > 0$ that depends on the basis. Since there exists a finite number of bases, this condition holds also for $c = \max c_{\mathcal{B}}$ independent of the basis, and SQP converges quadratically independent of the basis. □

5.3 Discussion of the Proofs

An interesting observation about 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. Consequently, one could omit 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, because in general, an SQP method would move off a nonlinear constraint. This is one reason for introducing slacks to deal with complementarity of the form (1.2).

Similar conclusions can easily be derived for other 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, one can now see, that a basis that satisfies the conditions of Lemma 5.9 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. We show in the next section that this assumption holds for several interesting cases. We also show 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]**–**[A5]**. The first such situation arises when there are no general constraints on control and state variables. 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:

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & c(z_0) = 0 \\ & 0 \le G(z) \;\perp\; H(z) \ge 0, \end{array} \tag{6.14}$$

where $G, H : \mathbb{R}^{n+2p} \to \mathbb{R}^p$ are twice continuously differentiable. We note that the general constraints are on the control variables only and that the only complementarity constraint takes the form of a vertical complementarity constraints. This case was brought to our attention by Mihai Anitescu.

In this section, we make the following additional assumption, which is related to the mixed P_0 property (e.g., [16]).

[A8] The matrix $[\nabla c(z_0^*) : \nabla G(z^*) : \nabla H(z^*)]$ has full rank.

The motivation for considering this form of problem (6.14) is that the simple complementarity constraint $0 \le z_1 \;\perp\; z_2 \ge 0$ always produces feasible linearization if there are no other constraints on z_1, z_2.

To see the relationship between Assumption **[A8]** and the mixed P_0 property consider the equivalent MPEC with slacks defined by

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & F(z,s) = 0 \\ & 0 \le s_1 \;\perp\; s_2 \ge 0, \end{array} \tag{6.15}$$

where

$$F(z,s) = \begin{pmatrix} c(z_0) \\ G(z) - s_1 \\ H(z) - s_2. \end{pmatrix}$$

One can see that a sufficient condition for Assumption [**A8**] is that the Jacobian matrix

$$[\nabla_{s_1} F : \nabla_{s_2} F : \nabla_z F]$$

satisfies the mixed P_0 property. This assumption has been used, for instance, in the convergence analysis of MPEC solvers and holds for a range of test problems, such as those arising from obstacle or packaging problems.

Lemma 6.1 *Let Assumptions* [**A1**]–[**A5**] *and* [**A8**] *hold. Then it follows that* (QP^k) *is consistent for all* $z^{(k)}$ *in a neighborhood 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. Let $z^{(k)}$ be sufficiently close to z^* so that the Jacobian matrix

$$\left[\nabla c(z_0^{(k)}) : \nabla G(z^{(k)}) : \nabla H(z^{(k)})\right]$$

has full rank.

The linearizations of the QP approximation to (6.14) has the following constraints:

$$c^{(k)} + \nabla c^{(k)^T} d_0 \;=\; 0 \tag{6.16}$$
$$G^{(k)} + \nabla G^{(k)^T} d \;\geq\; 0 \tag{6.17}$$
$$H^{(k)} + \nabla H^{(k)^T} d \;\geq\; 0 \tag{6.18}$$
$$G^{(k)^T} H^{(k)} + G^{(k)^T} \nabla H^{(k)^T} d + H^{(k)^T} \nabla G^{(k)^T} d \;\leq\; 0. \tag{6.19}$$

We need to show that these constraints are consistent. By [**A8**] it follows that there exists $\hat{d}$ such that constraints (6.16)–(6.18) hold with equality (this corresponds to the origin in the $G-H$ coordinate system).

It can be shown that $\hat{d}$ is also feasible in (6.19). The constraints (6.16) and (6.17) hold with equality, thus implying that $\nabla G^{(k)^T} \hat{d} = -G^{(k)}$ and $\nabla H^{(k)^T} \hat{d} = -H^{(k)}$. Substituting these last two equations into (6.19) simplifies that constraint to

$$G^{(k)^T} H^{(k)} + G^{(k)^T} \nabla H^{(k)^T} \hat{d} + H^{(k)^T} \nabla G^{(k)^T} \hat{d} \;=\; -G^{(k)^T} H^{(k)} \;\leq\; 0,$$

where the last inequality follows from the assumption that $G^{(k)^T} H^{(k)} \geq 0$.

To show that the QP step d^* maintains nonnegative complementarity, we observe that for $z^{(k)}$ sufficiently close to z^*, SQP converges and identifies the correct active set. Thus, there exists a partition

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

and

$$G_i^{(k)} + \nabla G_i^{(k)^T} d^* = 0 \quad , \quad i \in \mathcal{G} \tag{6.20}$$
$$H_i^{(k)} + \nabla H_i^{(k)^T} d^* = 0 \quad , \quad i \in \mathcal{H}. \tag{6.21}$$

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

$$\left(G^{(k)} + \nabla G^{(k)^T} d^*\right)^T \left(H^{(k)} + \nabla H^{(k)^T} d^*\right) = 0. \tag{6.22}$$

Hence, if $G(z)$ and $H(z)$ are convex, it follows that

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

and similarly for $H^{(k+1)}$. Combining this with (6.22) implies that

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

$\square$

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. This assumption is satisfied, for instance, for obstacle and packaging problems.

6.2 Feasibility Restoration for Complementarity

This section examines the properties of (QP^k) where $z_1^{(k)^T} z_2^{(k)} > 0$. In this case, (QP^k) may be inconsistent. This section describes a simple restoration procedure that can be invoked if (QP^k) is inconsistent. The procedure 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 5.7.

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. 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**]–[**A5**] *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 $\mathcal{Z}_1^{\perp} = \emptyset$, namely, that $z_1^* = 0$, and consider the dual feasibility conditions of (LP_F^k) (primal feasibility follows from Assumption [**A2**]),

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} - \left[\begin{array}{c|ccc} & 0 & 0 & 0 \\ A_{\mathcal{A}}^{(k)} & I_1 & & z_2^{(k)} \\ & & I_2 & z_1^{(k)} \\ & & & -1 \end{array}\right] \begin{pmatrix} \lambda_{\mathcal{A}} \\ \nu_1 \\ \nu_2 \\ \xi \end{pmatrix} = 0, \qquad (6.23)$$

where $A_{\mathcal{A}}^{(k)}$ is the matrix of active constraint normals of $c_{\mathcal{E}}(z)$, $c_{\mathcal{I}}(z)$ at z^*, $I_2 = [e_i]_{i \in \mathcal{Z}_2}$ and $I_1 = [e_i]_{i \in \mathcal{Z}_1}$.

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 (6.23) are linearly independent by Assumption **[A2]**. Thus there exists a unique solution to (6.23). **[A2]** implies in particular, that $A_{\mathcal{A}}^{(k)}$ has full column rank, and thus $\lambda_{\mathcal{A}} = 0$ follows from the first line of (6.23). This implies that

$$\nu_1 = z_2^{(k)} \geq 0 \text{ and } \nu_2 = z_1^{(k)} \geq 0.$$

Complementary slackness of (LP_F^k) implies 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 $z_{2i}^{(k)} > 0$. This 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 $z_{1i}^{(k)} + d_{1i} = 0$. The case where $z_{1i}^{(k)} = 0$ but $z_{2i}^{(k)} > 0$ is analogous.

Putting all three cases together and recalling that $\mathcal{Z}_1 = \emptyset$, one has that $z_1^{(k+1)^T} z_2^{(k+1)} = 0$.

It remains to prove that there exist multipliers λ with $\lambda_{\mathcal{I}} \geq 0$ such that (6.23) holds. If $\lambda_{\mathcal{I} \cap \mathcal{A}} \geq 0$, there is nothing to show. Hence assume that there exists a multiplier $\lambda_i < 0$ for $i \in \mathcal{I} \cap \mathcal{A}$. Then one can perform an iteration of an active set method on (LP_F^k) that 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), 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 the elastic mode, some of the constraints are relaxed and an l_∞-QP is solved. The application of snopt to MPECs is explored in [1]. Unlike snopt, however, the present restoration will occur only 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.

We 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 possibility 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 showing 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

The following MPEC shows that if we remove Assumptions [**A2**] and, in particular, Assumption [**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.mod`) 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

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

Next, we show 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. the assertion holds trivially for $k = 0$ (i.e., the 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:

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

We note that all QP approximations are consistent and that the first three constraints are active. Subtracting the second from the first constraint, we have that $d_1 = d_2$. Substituting into the third constraint, we get $d_1 = d_2 = -2^{-(k+1)}$, from which it follows that $d_0 = 4(-2^{-(k+1)})$. We verify the 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. $\square$

Note that (P) satisfies an MPEC-MFCQ [20] but violates an MPEC-LICQ (as can be seen easily by observing that four 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,

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

one can see 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 the set of box constraints). The solvers filter, snopt and lancelot behave in this way.

7.2 Formulations without Slacks

The next example shows that SQP methods can converge to nonstationary points if slacks are not added to replace nonlinear complementarity conditions. Consider the following MPEC (`s12.mod`) 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 solutions satisfy Assumptions **[A1]**–**[A4]**. The feasible set is illustrated by the bold lines in Figure 3.

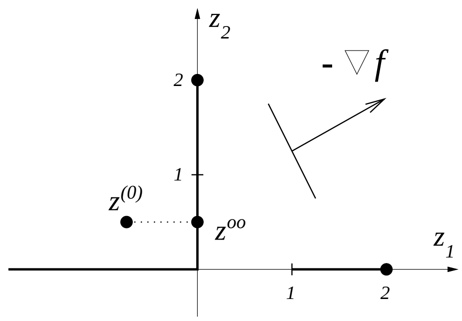

Figure 3: Example `sl2`

Starting at $z^{(0)} = (-\epsilon, t)^T$ gives convergence to the nonstationary point $z^\infty = (0,t)^T$, where $t \geq 0$ is arbitrary. Moreover, one can show 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. Nevertheless, second-order convergence is observed!

It is straightforward to prove quadratic convergence to a nonstationary 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

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

One can see 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 leads to a similar conclusion. This example shows that it is not sufficient to trigger the elastic mode only when QP become inconsistent. Clearly, the elastic mode is also required if the multipliers become too large. The introduction of slacks avoids the need for the elastic mode in this example.

When a slack variable is introduced, SQP converges quadratically. The SQP solver `filter` exhibits this behavior, while `lancelot` and `loqo` converge even for the problem *without* slacks. The reason for this apparently better behavior is 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, one can truncate QP steps such that $z_1^{(k)} \geq 0$, $z_2^{(k)} \geq 0$ for all iterations k. This approach 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 [4] where it was observed that perturbing the right-hand side of the complementarity constraint to $-\epsilon$ renders an inconsistent QP, cannot occur.

7.3 Lack of Second-Order Condition

The following MPEC (`ralph2.mod`) shows that if the second-order sufficient condition [**A3**] is violated, then SQP may converge only linearly:

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

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 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. We illustrate this by the following 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., $\text{conv}(S_M^*)$). 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 that shows that SQP methods converge quadratically when applied to the NLP equivalent of an MPEC. This analysis goes some way toward explaining the extraordinary success of SQP solvers applied to MPECs, as we have observed. The result is remarkable because 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. Further we have given a restoration phase that ensures that this can always be guaranteed sufficiently close to a solution.

We have also experimented with an alternative to the 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:

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

It is easy to see that, for any $z = (\epsilon^4, 1 - \epsilon)$, the (QP^k) relaxed by using (8.1) with $\delta = \kappa = 0.5$ is inconsistent. Note that if we restrict attention to points z that satisfy the linear constraints (e.g. $z = (\epsilon, 1 - \epsilon)$) then (QP^k) using (8.1) is consistent in a neighborhood of $z = (0, 1)$. Thus (8.1) seems to ensure 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 bootstrap a convergence of SQP for MPECs with the 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 that satisfy $z_1^T z_2 = \epsilon > 0$ are near-degenerate, and we would expect this property to play a role in the SQP method. In our numerical experiment we use two SQP solvers, snopt and filter.

The solver `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 approach 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. 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, this is a favorable situation for SQP methods.

Future work will focus on relaxing some assumptions and 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 [21] in this context. Another important question concerns the global convergence of SQP methods. Anitescu [1] 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] M. Anitescu. On solving mathematical programs with complementarity constraints as nonlinear programs. Preprint ANL/MCS-P864-1200, MCS Division, Argonne National Laboratory, Argonne, IL, 2000.

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

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

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

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

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

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

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

[9] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming*, 91:239–270, 2002.

[10] R. Fletcher and S. Leyffer. Solving mathematical program with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004.

[11] M. Fukushima and P. Tseng. An implementable active-set algorithm for computing a B-stationary point of the mathematical program with linear complementarity constraints. *SIAM Journal on Optimization*, 12:724–739, 2002.

[12] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large–scale constrained optimization. *SIAM Journal on Optimization*, 12(4):979–1006, 2002.

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

[14] H. Jiang and D. Ralph. 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] S. Leyffer. MacMPEC: AMPL collection of MPECs. Webpage, `www.mcs.anl.gov/~leyffer/MacMPEC/`, 2000.

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

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

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

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

[20] S. Scholtes. Convergence properties of regularization schemes for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

[21] S.J. Wright. Modifying SQP for degenerate problems. *SIAM Journal on Optimization*, 13(2):470–497, 2002.