Aachen Institute for Advanced Study in Computational Engineering Science

Preprint: AICES-2012/04-1

30/April/2012

MPCC Solution Approaches for a Class of MINLPs with Applications in Chemical Engineering

M. Herty, S.Steffensen

Financial support from the Deutsche Forschungsgemeinschaft (German Research Foundation) through grant GSC 111 is gratefully acknowledged.

MPCC solution approaches for a class of MINLPs with applications in chemical engineering

M. Herty * S. Steffensen *

April 30, 2012

Abstract

We discuss relaxation and penalization approaches for mixed–integer nonlinear programs arising for example in process engineering applications. We analyse stationarity conditions, feasibility, existence and optimality of limit points of a sequence of stationary points. Some numerical results for relaxation and penalization approaches are given.

Keywords. mixed–integer nonlinear programs, disjunctive optimization problems, mathematical programs with complementarity constraints

AMS Subject Classification. 90C30, 90C11

1 Introduction

We consider the following mixed integer nonlinear programs

$$(MINLP) \qquad \begin{array}{rl} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \geq 0 \\ & x_1 \in \{0,1\}^p \end{array} \tag{1}$$

where $x = (x_0, x_1) \in \mathbb{R}^n \times \mathbb{R}^p$, x_0 denotes the continuous variable and x_1 is the discrete variable and all inequalities are componentwise. Throughout, we assume that $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^{q-1}$ and $g : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ are twice continuously differentiable functions. The MINLP is thus a combination of a nonlinear program and a mixed integer problem. Such optimization problems arise for example in process engineering as a reformulation of a

*RWTH Aachen University, Department of Mathematics, Templergraben 55, D-52064 Aachen, GERMANY. {herty,steffensen}@mathc.rwth-aachen.de

generalized disjunctive optimization problem of the type [24]

$$
(GDP) \qquad \begin{array}{rl}
\min_{x,Y} & \Psi(x) + \sum_{k\in K} b_k \\
\text{s.t.} & H_0(x) = 0 \\
& G_0(x) \geq 0 \\
& \bigvee_{i\in D_k} \begin{bmatrix} Y_{i,k} \\ H_{i,k}(x) = 0 \\ G_{i,k}(x) \geq 0 \\ b_k = \gamma_{i,k} \end{bmatrix}, \qquad k \in K \\
& \Omega(Y) = \text{True} \\
& Y_{i,k} \in \{\text{True, False}\}
\end{array} \tag{2}
$$

where x denotes a continuous variable, $Y_{i,k}$ are the discrete decision variables, b_k is a continuous scalar that equals a fixed charge $\gamma_{i,k}$. Moreover, $K = \{1,..,m\}$, hence the problem contains m disjunctions and each disjunction $k \in K$ consists of n_k terms $i \in D_k = \{1, 2, .., n_k\}$. The operator $\bigvee_{i\in D_k}$ denotes an *exclusive or* and the constraints that are contained in the i-th term become valid only if the decision variable $Y_{i,k}$ is set to *True*. Typically [24, 18, 12] these logical constraints are used to model interrelationships between disjunctive constraints. We refer to [12] for a simple example and to [24] for a process optimization example in chemical engineering. Further examples may also be found in [18, 3].

In order to overcome the difficulty of the binary variables x_1 of (1), we first relax the problem. Instead of the binary constraints $x_1 \in \{0,1\}^p$, we introduce simple bounds on x_1, i.e. $0 \leq x_1 \leq 1$. However, to enforce that for each j ($j = 1,..,p$) one of the bounds is active (such that $x_1 \in \{0,1\}^p$), we assume that the complementarity condition

$$
0 \;\leq\; x_1 \perp 1 - x_1 \geq 0 \tag{3}
$$

holds. This approach was already studied by Biegler et al in [1] and Stein et al. [24]. Replacing the binary constraint with this so–called complementarity constraint we obtain an MPCC (a mathematical program with complementarity constraints) of the form

$$
\begin{array}{rr}
\min & f(x) \\
\text{s.t.} & g(x) \geq 0 \\
& h(x) = 0 \\
& 0 \leq x_1 \perp \mathbf{1} - x_1 \geq 0,
\end{array} \tag{4}
$$

Although we removed the binary variables and obtain a problem with continous variables, the problem is still not well behaved, since the complementarity constraints are the source of some special properties of MPCCs that distinguish them from standard NLPs and may cause serious theoretical and numerical challenges: At all feasible points of an MPCC neither the LICQ nor the MFCQ are satisfied. Thus, these MPCCs also need to be handled with special care. In recent years, significant progress has been made in the numerical solution of MPCCs [7, 21, 8, 17, 14, 2, 4].

We proceed as follows. We reformulate the complementarity condition as the root of an NCP-function, i.e. we use a function $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that satisfies the following condition .

$$
0 \leq a \quad \perp \quad b \geq 0 \qquad \Longleftrightarrow \qquad \varphi(a,b) = 0\,.
$$

Possible choices of NCP-functions are the Fischer-Burmeister, the Minimum and the Natural-Residual function:

$$\begin{aligned} \varphi_{FB}(a,b) &= a + b - \sqrt{a^2+b^2} \\ \varphi_{Min}(a,b) &= \min(a,b) \\ \varphi_{NR}(a,b) &= \frac{1}{2}(a+b-\sqrt{(a-b)^2}) \end{aligned} \tag{5}$$

Moreover, we relax the conditions $\varphi(x_{1j}, 1-x_{1j}) = 0$ $(j = 1,..,p)$ by a positive relaxation parameter μ, which leads to the relaxed problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \geq 0 \\ & h(x) = 0 \\ & 1 \geq x_1 \geq 0, \\ & \Phi(x_1, \mathbf{1} - x_1) \leq \mu, \end{aligned}$$

where $\mathbf{1} = (1, 1, .., 1)^T$ and

$$\Phi(x_1, \mathbf{1} - x_1) = (\varphi(x_{11}, 1 - x_{11}), .., \varphi(x_{1p}, 1 - x_{1p}))^T.$$

Similar relaxation methods for MPCCs have also been discussed e.g. in [21, 17, 8, 23]. Another reformulation that we might consider to solve MINLPs of the form (1) concerns penalty approaches.

$$\begin{aligned} P(\rho) \quad \min \quad & f(x) + \rho\, \Psi(x_1, \mathbf{1} - x_1) \\ \text{s.t.} \quad & g(x) \geq 0 \\ & h(x) = 0 \\ & 1 \geq x_1 \geq 0 \end{aligned} \tag{6}$$

where ρ denotes a penalty parameter and $\Psi(x)$ might be one of the following functions

$$\Psi(a,b) = \sum_{j=1}^{p} \Phi_j(a,b) \qquad \text{or} \qquad \Psi(a,b) = \sum_{j=1}^{p} a_j b_j \tag{7}$$

where $\Phi(a, b)$ is defined as before. A similar approach for MINLPs of the form (1) has been considered by Biegler et al. in [1]. Moreover, in the literature, there exist a variety of such penalty approaches for MPCCs, see e.g. in [22, 16] and in particular, Hu and Ralph consider problems similar to this type [10].

We contribute to the existing results by discussing two approaches to solve (1). We propose a sequential solution of relaxed problems $NLP(\mu)$ given by (9) using NCP functions and analyse convergence properties for a general class of NCP–functions. Furthermore, we analyse a penalty formulation of (1) given by (6). In both cases we discuss stationarity conditions, feasibility, existence and optimality of limit points of a sequence of stationary points (x^k).

The following notation is used: e_{ij} denotes the unit vector that corresponds to x_{ij} and $B_\varepsilon(\hat{x})$ denotes the ball around $\hat{x}$ with radius ε. Furthermore, we define the support of $\lambda \in \mathbb{R}^m$ as

$$\text{supp}(\lambda) := \{j \in \{1, \dots, m\} \,:\, \lambda_j \neq 0\}.$$

2 A relaxation approach to MINLPs

In this section we introduce a slack variable $x_2 \in \mathbb{R}^p$ and rewrite (3) as

$$h_q(x) := x_1 + x_2 - 1 = 0, \qquad 0 \;\le\; x_1 \perp x_2 \ge 0$$

to simplify the notation. Then, our MPCC has the form

$$\begin{array}{lr} \min & f(x) \\ \text{s.t.} & g(x) \ge 0 \\ & h(x) = 0 \\ & 0 \le x_1 \perp x_2 \ge 0, \end{array} \tag{8}$$

where $x = (x_0, x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^p$ and the relaxed problem is

$$\begin{array}{llr} NLP(\mu) & \min & f(x) \\ & \text{s.t.} & g(x) \ge 0 \\ & & h(x) = 0 \\ & & x_1, x_2 \ge 0 \\ & & \Phi(x_1, x_2) \le \mu, \end{array} \tag{9}$$

where $\Phi(x_1, x_2) = (\varphi(x_{11}, x_{21}), .., \varphi(x_{1p}, x_{2p}))^T$.

Let us define the index sets of active constraints for (9)

$$\begin{aligned} I_g(x) &:= \{i \in \{1, \ldots, m\} \,:\, g_i(x) = 0\}, \\ I_1(x) &:= \{j \in \{1, \ldots, p\} \,:\, x_{1j} = 0\}, \\ I_2(x) &:= \{j \in \{1, \ldots, p\} \,:\, x_{2j} = 0\} \\ I_\Phi(x, \mu) &:= \{j \in \{1, \ldots, p\} \,:\, \Phi_j(x_1, x_2) = \mu\} \end{aligned}$$

The derivatives of the NCP-functions of (5) satisfy the following limits.

Lemma 2.1. *Let $\bar{x}$ be feasible for (8) and consider a sequence (x^k) that converges to $\bar{x}$. Define*

$$\alpha_j^k := \frac{\partial \Phi}{\partial x_{1j}}(x_1^k, x_2^k) \qquad \text{and} \qquad \beta_j^k := \frac{\partial \Phi}{\partial x_{2j}}(x_1^k, x_2^k),$$

then for $\varphi_{FB}, \varphi_{Min}$ and φ_{NR} it holds that

$$\lim_{k\to\infty} \alpha_j^k \;=\; \begin{cases} 1 & j \in (I_1 \setminus I_2)(\bar{x}) \\ 0 & j \in (I_2 \setminus I_1)(\bar{x}) \end{cases}$$

$$\lim_{k\to\infty} \beta_j^k \;=\; \begin{cases} 0 & j \in (I_1 \setminus I_2)(\bar{x}) \\ 1 & j \in (I_2 \setminus I_1)(\bar{x}) \,. \end{cases}$$

We use the notation $(I_1 \backslash I_2)(\bar{x})$ for $I_1(\bar{x}) \backslash I_2(\bar{x})$ and $(I_2 \backslash I_1)(\bar{x})$ for $I_2(\bar{x}) \backslash I_1(\bar{x})$, respectively.

Proof. Note that the gradients of $\varphi_{FB}(a, b)$ and $\varphi_{NR}(a, b)$ for all $a, b \in \mathbb{R}$ with $a \ne b$ are given by

$$\nabla \varphi_{FB}(a, b) = \begin{pmatrix} 1 - \frac{a}{\sqrt{a^2+b^2}} \\ 1 - \frac{b}{\sqrt{a^2+b^2}} \end{pmatrix}, \quad \nabla \varphi_{NR}(a, b) = \frac{1}{2} \begin{pmatrix} 1 - \frac{a-b}{\sqrt{(a-b)^2}} \\ 1 + \frac{a-b}{\sqrt{(a-b)^2}} \end{pmatrix}$$

and the gradient of $\varphi_{Min}(a, b)$ is either given by the unit vector $e_1 \in \mathbb{R}^2$ or by $e_2 \in \mathbb{R}^2$. Hence, the conclusion directly follows from the convergence of the sequence (x^k). □

2.1 Properties of $NLP(\mu)$

Let the feasible set of $NLP(\mu)$ be defined by

$$Z(\mu) := \{\ x \in \mathbb{R}^{n+p+p}\ :\ g(x) \geq 0,\ h(x) = 0,\ x_1 \geq,\ x_2 \geq 0,\ \Phi(x_1, x_2) \leq \mu\ \}.$$

Then we have the following result on the relation of the feasible sets of $NLP(\mu)$ for different values of $\mu > 0$.

Lemma 2.2. *Let Z_{MPCC} be the feasible set of the MPCC (8) and assume that the following implication holds*

$$\varphi(a, b) < 0 \qquad \Rightarrow \qquad a < 0 \ \ or \ \ b < 0. \qquad \forall a, b \in \mathbb{R}$$

Then the following relations hold

1.

$$\bigcap_{\mu>0} Z(\mu) = Z(0) = Z_{MPCC}$$

2.

$$Z(\mu_1) \subseteq Z(\mu_2) \qquad \forall \mu_1 \leq \mu_2\,.$$

Proof. Note, that the feasible sets only differ in the complementarity constraint and all other constraints are directly satisfied.

First, we assume that $x \in Z(0)$ then $\Phi(x_1, x_2) \leq 0$ and $x_1, x_2 \geq 0$ and hence $\Phi(x_1, x_2) = 0$, such that (3) is satisfied since φ is supposed to be an NCP-function. On the other hand, if $x \in Z_{MPCC}$, then (3) is satisfied and therefore $\Phi(x_1, x_2) = 0$ such that $x \in Z(0)$. The relations $\bigcap_{\mu>0} Z(\mu) = Z(0)$ and $Z(\mu_1) \subseteq Z(\mu_2)$ for all $\mu_1 \leq \mu_2$ follow directly from the definition of $Z(\mu)$. □

Next, this result can be used to relate the solutions of the MPCC and of $NLP(\mu)$.

Theorem 2.1.

1. *Let $\hat{\mu} > 0$ and assume that $\bar{x}$ is a (strict) local solution for $NLP(\hat{\mu})$ in $B_\epsilon(\bar{x})$ for some $\epsilon > 0$, which is feasible for (8), i.e. $\bar{x} \in Z_{MPCC}$. Then $\bar{x}$ is a (strict) local solution for $NLP(\mu)$ in $B_\epsilon(\bar{x})$ for all $\mu \in [0, \hat{\mu}]$.*

2. *Let $\hat{\mu} > 0$ and assume that $\bar{x}$ is a (strict) global solution for $NLP(\hat{\mu})$, which is feasible for (8). Then $\bar{x}$ is a (strict) global solution for $NLP(\mu)$ for all $\mu \in [0, \hat{\mu}]$.*

3. *Let $(\mu_k) \subseteq \mathbb{R}^+$ be a sequence that satisfies $\mu_k \to 0$ and let (x^k) be a sequence of global solutions of the corresponding problems $NLP(\mu_k)$. Furthermore, let $\bar{x}$ be an accumulation point of (x^k). Then, $\bar{x}$ is a global solution of (8). Moreover, $\bar{x}$ is a global solution to the MINLP.*

Proof. For the first part, suppose that $\bar{x}$ is a local solution for $NLP(\hat{\mu})$ in a neighborhood $B_\epsilon(\bar{x})$, then by Lemma 2.2 it follows that

$$\forall x \in Z(\mu) \cap B_\epsilon(\bar{x}) \subset Z(\hat{\mu}) \cap B_\epsilon(\bar{x})\ :\quad f(x) \geq f(\bar{x}) \tag{10}$$

for all $\mu \in [0, \hat{\mu}]$ with a strict inequality if $\bar{x}$ is a strict local minimum.

The proof of the second part is analog to the first one, since the (10) holds without the restriction to $B_\epsilon(\bar{x})$.

Finally,consider an accumulation point $\bar{x}$ of (x^k). Then there exist a subsequence $(x^k)_{k\in\mathcal{K}}$, such that

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} x^k = \bar{x}\,.$$

It follows by Lemma 2.2, that $\bar{x}$ is feasible for (8). Now, suppose $\bar{x}$ is not a global minimum of (8), then there exists an $\hat{x} \in Z_{MPCC}$ with $f(\bar{x}) > f(\hat{x})$. This, however, implies, by the continuity of f and the convergence of the subsequence $(x^k)_{k\in\mathcal{K}}$, that for all $k \in \mathcal{K}$ sufficiently large $f(x^k) > f(\hat{x})$ which contradicts the assumption that x^k is a global minimum of $NLP(\mu_k)$.

Clearly the feasible set of the MINLP and the feasible set of its reformulation as an MPCC coincide. Hence, $\bar{x}$ is also a global solution of the MINLP. □

However, most NLP solver are designed to find stationary points. Hence in the following, we are concerned with the set of stationary points of both problems, the MPCC and the relaxed problem $NLP(\mu)$. The fact that the MPCC does not satisfy the general conditions under which the KKT-conditions hold, led to various specially designed constraint qualifications and stationarity concepts for MPCCs, see for example [9, 20, 25, 15, 11]. The most important of them using the existence of multipliers are *C-, M-* and *strong stationarity*. All of these differ only in a sign condition for the multipliers corresponding to the constraints $x_1 \geq 0$ and $x_2 \geq 0$ for indices where both inequalities are active (i.e. $x_{1j} = x_{2j} = 0$). As we noticed that in our case, any feasible point x of the MPCC (8) cannot contain such degenerate indices, all these stationarity concepts are equivalent for our particular MPCC. Therefore, we will only use the terminology *"stationary for (8)"*.

The Lagrangian of (8) is given by

$$\mathcal{L}_{\text{MPCC}}(x, \lambda, \mu, \hat{\nu}_1, \hat{\nu}_2) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x) - \mu h(x) - \hat{\nu}_1^T x_1 - \hat{\nu}_2^T x_2. \tag{11}$$

Definition 2.1. A point x^* is called *stationary for (8)*, if x^* is feasible for (8) and there exist multipliers $\lambda^* \in \mathbb{R}^m_+$, $\lambda^*_h \in \mathbb{R}$, $\hat{\nu}^*_1 \in \mathbb{R}^p$ and $\hat{\nu}^*_2 \in \mathbb{R}^p$, such that the conditions

$$\begin{aligned} \nabla_x \mathcal{L}_{\text{MPCC}}(x^*, \lambda^*, \lambda^*_h, \hat{\nu}^*_1, \hat{\nu}^*_2) &= 0 \\ g_i(x^*)\lambda^*_i &= 0, \quad (1 \leq i \leq m) \\ x^*_{1j}\hat{\nu}^*_{1j} &= 0, \quad (1 \leq j \leq p) \\ x^*_{2j}\hat{\nu}^*_{2j} &= 0, \quad (1 \leq j \leq p) \end{aligned} \tag{12}$$

are satisfied.

Remark 2.1. Note that the stationarity condition corresponds to the general KKT-conditions for NLPs applied to the so-called *Relaxed Nonlinear Problem* of (8) [6, 9, 20].

$$\begin{array}{llrcll} \text{RNLP} & \min & f(x) & & & \\ & \text{s.t.} & g(x) & \geq & 0 & \\ & & h(x) & = & 0 & \\ & x_{1j} = 0, & x_{2j} & \geq & 0 & j \in (I_1 \backslash I_2)(x^*) \\ & x_{2j} = 0, & x_{1j} & \geq & 0 & j \in (I_2 \backslash I_1)(x^*). \end{array}$$

Note again, that for x being feasible, the vectors x_1 and x_2 satisfy strict complementarity, thus the index set $\{1, \dots, p\}$ partitions into $(I_1 \setminus I_2)(x^*)$ and $(I_2 \setminus I_1)(x^*)$.

Remark 2.2. Moreover, if f is convex, g concave and h linear and the MPEC multipliers $\hat{\nu}_1^*$ and $\hat{\nu}_2^*$ of a stationary point x^* of (8) are nonnegative, then x^* is a global minimum of the convex nonlinear problem

$$\begin{array}{lrcl} \min & f(x) & & \\ \text{s.t.} & g(x) & \geq & 0 \\ & h(x) & = & 0 \\ & x_1, x_2 & \geq & 0 \end{array}$$

and hence x^* is also global minimum of (1).

The auxiliary problem $RNLP$ can also be used to define a specific constraint qualification for MPCCs.

Definition 2.2. The MPEC-LICQ is said to hold in x^* for (8) if the LICQ holds in x^* for the relaxed problem $RNLP$ (determined in x^*).

The following lemma states that the MPEC-LICQ implies the general LICQ for NLPs for the relaxed problem $NLP(\mu)$ if $\mu > 0$ is sufficiently small.

Lemma 2.3. *Suppose the MPEC-LICQ holds at a feasible point $\bar{x}$ of (8). Then there exists a neighbourhood $\mathcal{B}(\bar{x})$ and a positive scalar $\bar{\mu} > 0$ such that for every $\mu \in (0, \bar{\mu})$ the general LICQ holds at every feasible point $x \in \mathcal{B}(\bar{x})$ for $NLP(\mu)$.*

Proof. The linear independent family of active constraint gradients in $\bar{x}$ for (8) is given by

$$\begin{array}{rl} \nabla h(\bar{x})^T & i \in \{1, \dots, q\} \\ \nabla g_j(\bar{x})^T & j \in I_g(\bar{x}) \\ e_{1j}^T & j \in (I_1 \setminus I_2)(\bar{x}) \\ e_{2j}^T & j \in (I_2 \setminus I_1)(\bar{x}). \end{array} \tag{13}$$

whereas the set of active constraint gradients in x for $NLP(\mu)$ is

$$\begin{array}{rl} \nabla h(x^k)^T & i \in \{1, \dots, q\} \\ \nabla g_j(x^k)^T & j \in I_g(x), \\ e_{1j}^T & j \in I_1(x) \\ e_{2j}^T & j \in I_2(x) \\ \alpha_j^k e_{1j}^T + \beta_j^k e_{2j}^T & j \in I_\Phi(x, \mu), \end{array} \tag{14}$$

Now note that the continuous differentiability of g implies that for x close enough to $\bar{x}$ we have $I_g(x) \subseteq I_g(\bar{x})$ as well as $I_1(x) \subseteq I_1(\bar{x})$ and $I_2(x) \subseteq I_2(\bar{x})$. Moreover, by the definition of Φ, we have that $(I_1 \cup I_2)(x) \cap I_\Phi(x) = \emptyset$. Therefore, we consider the family

$$\begin{array}{rl} \nabla h(x)^T & i \in \{1, \dots, q\} \\ \nabla g_j(x)^T & j \in I_g(\bar{x}), \\ e_{1j}^T & j \in (I_1 \setminus I_2)(\bar{x}) \setminus I_\Phi(x, \mu), \\ \alpha_j e_{1j}^T + \beta_j e_{2j}^T & j \in I_\Phi(x, \mu), \\ e_{2j}^T & j \in (I_2 \setminus I_1)(\bar{x}) \setminus I_\Phi(x, \mu) \end{array} \tag{15}$$

that can be larger than (14).

Let A be the matrix that consists of the row vectors of (13). Then by the MPEC-LICQ it holds that AA^T is regular. Furthermore, let $\tilde{A}$ be the matrix that consists of the row vectors of (15). Then by the continuous differentiability of h and g and by Lemma 2.1 we can choose a neighbourhood $\mathcal{B}(\bar{x})$ such that $\tilde{A}\tilde{A}^T$ is close enough to AA^T which in turn by the perturbation lemma then implies that $\tilde{A}\tilde{A}^T$ is also regular. Therefore $\tilde{A}$ must have full row rank which proves that the LICQ condition holds for all $x \in \mathcal{B}(\bar{x})$ for $NLP(\mu)$. □

In the following, we will relate the set of stationary points of the MPCC (8) and its reformulation $NLP(\mu)$. The first one considers the limiting problem $NLP(0)$ and corresponds to Theorem 3.7 of [13].

Theorem 2.2. *A point $(x^*, \lambda^*, \lambda_h^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$ is a stationary point for (8) if and only if there exist multipliers λ^*, λ_h^*, ν_1^*, ν_2^*, ξ^* such that $(x^*, \lambda^*, \lambda_h^*, \nu_1^*, \nu_2^*, \xi^*)$ satisfies the KKT-conditions of $NLP(0)$, where φ is supposed to be one of the NCP-functions of (5). Moreover, the multipliers satisfy the equations*

$$\begin{array}{lll} \hat{\nu}_{1j}^* = \nu_{1j}^* - \xi_j^* & \text{and} \quad \hat{\nu}_{2j}^* = \nu_{2j}^* = 0 & \text{for } j \in (I_1 \setminus I_2)(x^*) \\ \hat{\nu}_{2j}^* = \nu_{2j}^* - \xi_j^* & \text{and} \quad \hat{\nu}_{1j}^* = \nu_{1j}^* = 0 & \text{for } j \in (I_2 \setminus I_1)(x^*). \end{array} \tag{16}$$

However, since the considered solution method is to solve a sequence of relaxed problems $NLP(\mu_k)$, in the next theorem we prove that the limit of a convergent sequence of stationary (KKT-) points (x^k) of a sequence $NLP(\mu_k)$ (where $\mu_k \searrow 0$) is a stationary point of the MPCC (8).

2.2 Convergence of the relaxation approximation

Theorem 2.3. *Let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence with $\mu_k > 0$ for all $k \in \mathbb{N}$ that satisfies $\mu_k \to 0$, further let $(x^k)_{k\in\mathbb{N}}$ be a sequence of stationary points of $NLP(\mu_k)$ that satisfies $x^k \to \bar{x}$ and suppose the MPEC-LICQ holds at $\bar{x}$. Then $\bar{x}$ is a stationary point of the MPCC (8) with unique multipliers $\bar{\lambda}, \bar{\lambda}_h, \bar{\nu}_1$ and $\bar{\nu}_2$ that satisfy the limit conditions*

$$\begin{array}{rcll} \bar{\lambda}_j &=& \lim_{k\to\infty} \lambda_j^k \geq 0 & j \in I_g(\bar{x}) \\ \bar{\lambda}_j &=& \lim_{k\to\infty} \lambda_j^k = 0 & j \notin I_g(\bar{x}) \\ \bar{\lambda}_h &=& \lim_{k\to\infty} \lambda_h^k & \\ \bar{\nu}_{1j} &=& \lim_{k\to\infty} \nu_{1j}^k & j \notin I_1^\infty \\ \bar{\nu}_{1j} &=& -\lim_{k\to\infty} \xi_j^k & j \in I_1^\infty \\ \bar{\nu}_{2j} &=& \lim_{k\to\infty} \nu_{2j}^k & j \notin I_2^\infty \\ \bar{\nu}_{2j} &=& -\lim_{k\to\infty} \xi_j^k & j \in I_2^\infty \end{array}$$

where

$$\begin{array}{rcll} I_1^\infty &=& \{j \in (I_1 \setminus I_2)(\bar{x}) & : \quad j \in I_\Phi(x^k, \mu_k) \text{ for infinitely many } k\ \} \\ I_2^\infty &=& \{j \in (I_2 \setminus I_1)(\bar{x}) & : \quad j \in I_\Phi(x^k, \mu_k) \text{ for infinitely many } k\ \}. \end{array}$$

Proof. Since x^k is supposed to be a stationary point of $NLP(\mu_k)$ there exist multipliers λ^k, λ_h^k, ν_1^k, ν_2^k and ξ^k such that

$$\begin{aligned}\nabla f(x^k) &= \nabla h(x^k)\lambda_h^k + \sum_{j\in I_g(x^k)} \nabla g_j(x^k)\lambda_j^k \\ &\quad + \sum_{j\in I_1(x^k)} \nu_{1j}^k e_{1j} + \sum_{j\in I_2(x^k)} \nu_{2j}^k e_{2j} - \sum_{j\in I_\Phi(x^k,\mu_k)} \nabla\Phi_j(x_1^k, x_2^k)\xi_j^k\end{aligned} \tag{17}$$

As g is continuous and (x^k) converges to $\bar{x}$, we have $I_g(x^k) \subseteq I_g(\bar{x})$ for sufficiently large $k \in \mathbb{N}$. Moreover, it holds that $I_1(x^k) \subset (I_1 \setminus I_2)(\bar{x})$ as well as $I_2(x^k) \subset (I_2 \setminus I_1)(\bar{x})$ for k large enough. By the definition of $\Phi(x_1, x_2)$ this equation can therefore be rewritten as $f(x^k) = A_k^T \omega^k$, where the matrix A_k consists of the row vectors of (15) evaluated in (x^k, μ_k) and $\omega^k = (\lambda_h^k, \lambda^k, \gamma^k)$ with

$$\begin{aligned}\gamma_j^k &= \nu_{1j}^k \qquad \forall j \in (I_1 \setminus I_2)(\bar{x}) \setminus I_\Phi(x^k, \mu_k) \\ \gamma_j^k &= -\xi^k \qquad \forall j \in I_\Phi(x^k, \mu_k) \\ \gamma_j^k &= \nu_{2j}^k \qquad \forall j \in (I_2 \setminus I_1)(\bar{x}) \setminus I_\Phi(x^k, \mu_k)\end{aligned}$$

As we have seen in the proof of Lemma 2.3, A_k converges to the matrix A that consists of the row vectors of (13). Since the MPEC-LICQ is assumed to hold at $\bar{x}$, A has full row rank. Hence, there exists a unique solution vector $\bar{\omega}$ solving $A^T\bar{\omega} = \nabla f(\bar{x})$. Moreover, the full row rank of A implies that AA^T is invertible, such that by the convergence of (A_k) and the perturbation lemma $A_k A_k^T$ is invertible for sufficiently large $k \in \mathbb{N}$. Hence there exists a unique solution vector $\omega_k = (A_k A_k^T)^{-1}(A_k \nabla f(x^k))$. Since $\nabla f(x^k)$ converges to $\nabla f(\bar{x})$ it follows that $\omega_k = (A_k A_k^T)^{-1}(A_k \nabla f(x^k)) \longrightarrow (AA^T)^{-1}(A\nabla f(\bar{x})) = \bar{\omega}$. Finally, since $\nu_{1j}^k = 0$ for $j \notin I_1(\bar{x})$ and $\nu_{2j}^k = 0$ for $j \notin I_2(\bar{x})$ and by the definition of ω^k, the limiting expression holds.

□

Finally we state a result, where our starting point is opposite. Consider a stationary point $\bar{x}$ of the MPCC (8). If $\bar{x}$ satisfies suitable assumptions, then for sufficiently small relaxation parameter $\mu > 0$ we obtain a "path" $\sigma(\mu)$ that converges to $\bar{x}$ for $\mu \to 0$ and for every μ the point $x_\mu = \sigma(\mu)$ denotes the unique stationary point of $NLP(\mu)$ in some neighbourhood of $\bar{x}$. However, we first need another definition, namely the strong second order sufficient condition as defined in [21].

Definition 2.3. Let x^* be a stationary point of (8) and suppose the MPEC-LICQ holds and $(\lambda_h^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)$ is the unique MPEC multiplier. Then the strong second order sufficient condition (SSOSC) is said to hold, if

$$d^T\, \nabla_{xx}^2 \mathcal{L}_{\text{MPCC}}(x^*, \lambda_h^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)\, d > 0$$

for all

$$\begin{aligned} d \in \mathcal{S}(x^*, \lambda_h^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*) \;:=\; \{\, d \in \mathbb{R}n+2p \setminus \{0\} :& \\ \nabla h_i(x^*)^T d = 0, &\quad i \in \{1, \ldots, q\} \\ \nabla g_j(x^*)^T d = 0, &\quad j \in I_g(x^*) \quad \text{and} \quad \lambda_i^* > 0 \\ d_{1j} = 0, &\quad j \in I_1(x^*) \quad \text{and} \quad \hat{\nu}_{1j}^* \neq 0 \\ d_{2j} = 0, &\quad j \in I_2(x^*) \quad \text{and} \quad \hat{\nu}_{2j}^* \neq 0 \,\}. \end{aligned}$$

Theorem 2.4. *Let x^* be a stationary point of (8) and assume that the MPEC-LICQ and the SSOSC hold at x^*. Moreover, suppose that $\hat{\nu}_{1j} \neq 0$ for all $j \in I_1(x^*)$ and $\hat{\nu}_{2j} \neq 0$ for all $j \in I_2(x^*)$. Then there exists some open neighborhood $\mathcal{B}(x^*)$, a scalar $\hat{\mu}$ and a piecewise smooth function $\sigma : (0, \hat{\mu}) \longrightarrow \mathcal{B}(x^*)$ such that $x_\mu = \sigma(\mu)$ is the unique stationary point of $NLP(\mu)$ for all $\mu \in (0, \hat{\mu})$ in $\mathcal{B}(x^*)$. Moreover, x_μ satisfies the second order sufficient condition for $NLP(\mu)$.*

Proof. First define the two sets

$$I_i^+(x) = \{\, j \in I_i(x) \,:\, \hat{\nu}_{ij}^* > 0 \,\} \text{ and } I_i^-(x) = \{\, j \in I_i(x) \,:\, \hat{\nu}_{ij}^* < 0 \,\}$$

for $i = 1, 2$. Since the strict complementarity is supposed to hold, we have $I_i(x^*) = (I_i^+ \cup I_i^-)(x^*)$ for $i = 1, 2$. Next, consider the relaxed problem

$$\begin{array}{llrcllcl}
R(\mu) & \min & f(x) & & & & & \\
 & \text{s.t.} & h(x) & = & 0 & & : & \eta \\
 & & g(x) & \geq & 0 & & : & \gamma \\
 & & x_{1j} & \geq & 0 & j \in I_1^+(x^*) & : & \omega_1 \\
 & & x_{2j} & \geq & 0 & j \in I_2^+(x^*) & : & \omega_2 \\
 & & \Phi_j(x_1, x_2) & \leq & \mu & j \in (I_1^- \cup I_2^-)(x^*) & : & \vartheta
\end{array}$$

It follows from Theorem 2.2 that x^* is a stationary point of $R(0)$ with unique multipliers $(\eta^*, \gamma^*, \omega_1^*, \omega_2^*, \vartheta^*)$. Moreover, by the MPEC-LICQ, the definition of Φ and a choice of $\varphi(a, b)$ according to (5) it satisfies the general LICQ for $R(0)$. Furthermore, the SSOSC implies that

$$d^T \nabla_{xx}^2 \mathcal{L}_{R(0)}(x^*, \eta^*, \gamma^*, \omega_1^*, \omega_2^*, \vartheta^*)\, d = d^T \nabla_{xx}^2 \mathcal{L}_{\text{MPCC}}(x^*, \lambda_h^*, \lambda^*, \hat{\nu}_1^*, \hat{\nu}_2^*)\, d > 0$$

(note that $\eta^* = \lambda_h^*$ and $\gamma^* = \lambda^*$) for all $d \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^p \setminus \{0\}$ with

$$\begin{array}{rcl}
\nabla h_i(x^*)^T d & = & 0, \quad i \in \{1, \ldots, q\} \\
\nabla g_j(x^*)^T d & = & 0, \quad j \in I_g(x^*) \quad \text{and} \quad \lambda_i^* > 0 \\
d_{1j} & = & 0, \quad j \in I_1(x^*) \\
d_{2j} & = & 0, \quad j \in I_2(x^*)
\end{array}$$

and hence for all critical directions for $R(0)$ such that the SOSC is satisfied in x^* for $R(0)$. We can therefore conclude that there exist a $\hat{\mu} > 0$, a neighborhood $\mathcal{N}$ of $(x^*, \eta^*, \gamma^*, \omega_1^*, \omega_2^*, \vartheta^*)$ and a piecewise smooth function $\Sigma : (0, \hat{\mu}) \to \mathcal{N}$ with $\Sigma(\mu) = (x^\mu, \eta^\mu, \gamma^\mu, \omega_1^\mu, \omega_2^\mu, \vartheta^\mu)$ and x^μ is a unique stationary (KKT-) point of $R(\mu)$ in some neighbourhood $\mathcal{B}(x^*)$ (according to $\mathcal{N}$) with unique multipliers $(\eta^\mu, \gamma^\mu, \omega_1^\mu, \omega_2^\mu, \vartheta^\mu)$ [19]. Moreover, each x^μ satisfies the SOSC for $R(\mu)$. Since the feasible region of $NLP(\mu)$ is contained in the feasible region of $R(\mu)$ it remains to show that x^μ is feasible and unique for $NLP(\mu)$. However, because $\Sigma(\mu)$ is piecewise smooth and the multipliers are unique and converge to the corresponding components of $(\lambda_h^*, \lambda^*, \pm\hat{\nu}_1^*, \pm\hat{\nu}_2^*)$, it holds

$$\begin{array}{rcll}
x_{1j}^\mu & = & 0, & \forall\, j \in I_1^+(x^*) \\
x_{2j}^\mu & = & 0, & \forall\, j \in I_2^+(x^*) \\
\Phi_j(x_1^\mu, x_2^\mu) & = & \mu, & \forall\, j \in (I_1^- \cup I_2^-)(x^*)
\end{array}$$

and therefore, by the definition of Φ and for all choices of (5) it follows that

$$\begin{aligned} x_{1j}^{\mu} &> 0 \quad \text{and} \quad x_{2j}^{\mu} > 0, & \forall\, j \in (I_1^- \cup I_2^-)(x^*) \\ \Phi_j(x_1^{\mu}, x_2^{\mu}) &= 0 < \mu, & \forall\, j \in (I_1^+ \cup I_2^+)(x^*) \end{aligned}$$

The remaining constraints on x are directly satisfied by the definition of $R(\mu)$ and moreover by the fact that x^{μ} converges to x^*. The uniqueness and the SOSC for $NLP(\mu)$ follow from the fact that they are satisfied for $R(\mu)$ and the fact that both problems differ only in constraints that are inactive in x^* and hence remain inactive in a sufficiently small neighbourhood of x^*. □

Remark 2.3. Note that, assuming that strict complementarity also holds for the inequality constraints $g(x) \geq 0$, the mapping $\Sigma : (0, \hat{\mu}) \to \mathcal{N}$ becomes continuously differentiable [5].

Remark 2.4. The results of the previous section can easily be transferred to the special case of MPCC (4), which justifies the introduction of the slack variable earlier on. Moreover, since the analysis is independent of the special relationship of x_1 and x_2, the results hold for the general type of (8).

3 A penalty approach to MINLPs

In this section, we consider the the reformulation of (1) as a penalty problem $P(\rho)$ as in (6). Let the feasible set of $P(\rho)$ be defined by

$$\mathcal{Z}_P := \{\ x \in \mathbb{R}^{n+p}\ :\ g(x) \geq 0,\ h(x) = 0,\ 1 \geq x_1 \geq 0\}.$$

We make the following assumptions on the function $\Psi(x)$ appearing in the penalty term of $P(\rho)$.

Assumptions 3.1.

(1) The function values satisfy

(a) $\Psi(x_1) \geq 0$ *for all* $x_1 \in [0, 1]$.

(b) $\Psi(x_1) = 0$ *if and only if* $0 \leq x_1 \perp 1 - x_1 \geq 0$.

(2) The values of the derivatives fulfill:

(a) $\frac{\partial \Psi}{\partial a_j}(x_1, \mathbf{1} - x_1) = 0$ *if* $x_{1j} = 1$.

(b) $\frac{\partial \Psi}{\partial b_j}(x_1, \mathbf{1} - x_1) = 0$ *if* $x_{1j} = 0$.

(c) $\frac{\partial \Psi}{\partial a_j}(x_1, \mathbf{1} - x_1) > 0$ *and* $\frac{\partial \Psi}{\partial b_j}(x_1, \mathbf{1} - x_1) > 0$ *if* $1 \geq x_1 \geq 0$ *and* $x_{1j} \in (0, 1)$.

(3) The limits of the derivative values satisfy for any sequence $x^k \longrightarrow \bar{x}$

(a) $\frac{\partial \Psi}{\partial b_j} / \frac{\partial \Psi}{\partial a_j}\,(x_1^k, \mathbf{1} - x_1^k) \longrightarrow 0$ *if* $\bar{x}_{1j} = 0$.

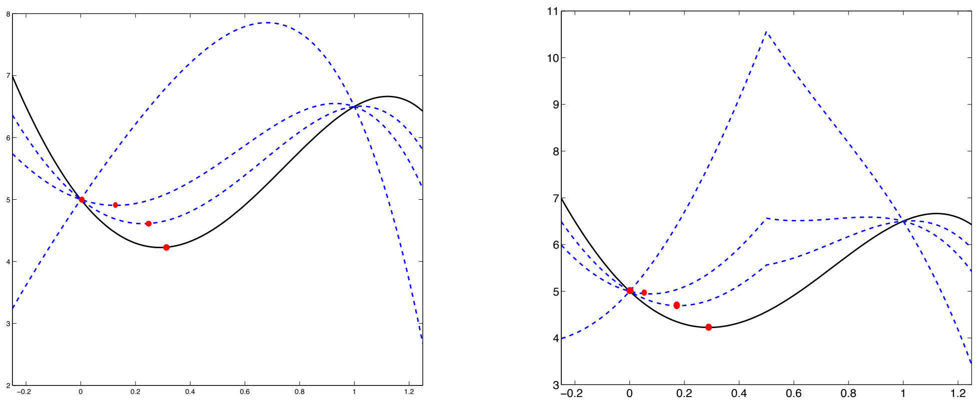
(b) $\frac{\partial \Psi}{\partial a_j} / \frac{\partial \Psi}{\partial b_j}\,(x_1^k, \mathbf{1} - x_1^k) \longrightarrow 0$ *if* $\bar{x}_{1j} = 1$.

Remark 3.1. Note, that both alternatives given previously are continuously differentiable and satisfy these conditions at least locally near a solution of (4). Furthermore if we use φ^2 instead of φ, then all conditions are satisfied globally.

Example 3.1. *Consider the MINLP*

$$\min \qquad f(x) := -3x^4 + 10x^2 - 5.5x + 5 \qquad \text{subject to} \quad x \in \{0,1\}.$$

The minimum of f on $[0,1]$ is $x^* \approx 0.29$, whereas the minimum of the MINLP is obviously $x^* = 0$. In Figure 1 we display the minimization of $P(\rho^k)$ for increasing parameter values of ρ^k and two different penalty terms $\Psi(x, 1-x) = x\,(1-x)$ and $\Psi(x, 1-x) = \min(x,\, 1-x)$. It can be observed that in both cases the minimum of $P(\rho^k)$ converges to the minimum x^*.



(a) $\Psi(x, 1-x) := x\,(1-x)$ (b) $\Psi(x, 1-x) := \min(x,\, 1-x)$

Figure 1: Minimization of $P(\rho^k)$ on $[0,1]$ for $\rho^k = 0, 2, 8, 12$

3.1 Stationarity of a limit point

In this section we will derive a convergence result for the penalty method. We consider a sequence of stationary points (x^k) of $P(\rho^k)$ with $\rho_k \longrightarrow +\infty$ and assume that $x^k \longrightarrow \bar{x}$ and moreover that $\bar{x}$ is feasible for (4). We prove that $\bar{x}$ is stationary for (4) if the so-called MPEC-CPLD holds in $\bar{x}$. The question of how to achieve feasibility of the limit point $\bar{x}$ will be addressed later. We define a constraint qualification for (4) that we call *reduced MPEC-CPLD*, which is similar to the *constant positive linear dependence constraint qualification* for MPECs.

Definition 3.1. Let $\hat{x} \in \mathbb{R}^n \times \mathbb{R}^p$ be feasible for (4). Then the *reduced MPEC-CPLD* is said to hold in $\hat{x}$, if for every $\mathcal{K}_g \subseteq I_g(\hat{x})$,$\mathcal{K}_h \subseteq I_h(\hat{x})$ and $\mathcal{K}_1 \subseteq \{1, .., p\}$ such that the family of gradient vectors

$$\{\nabla g_j(\hat{x}) \,:\, j \in \mathcal{K}_g\} \cup \{\nabla h_j(\hat{x}) \,:\, j \in \mathcal{K}_h\} \cup \{e_{1j} \,:\, j \in \mathcal{K}_1\}$$

is positive linearly dependent, i.e. there exist vectors $\mu \in \mathbb{R}^{|\mathcal{K}_h|}$, $\lambda \in \mathbb{R}_+^{|\mathcal{K}_g|}$ and $\xi \in \mathbb{R}^{|\mathcal{K}_1|}$ with $(\mu, \lambda, \xi) \neq (0, 0, 0)$ and

$$\sum_{i \in \mathcal{K}_h} \mu_i \nabla h_i(\hat{x}) + \sum_{j \in \mathcal{K}_g} \lambda_j \nabla g_i(\hat{x}) + \sum_{j \in \mathcal{K}_1} \xi_j e_{1j} = 0,$$

there exists a neighborhood $\mathcal{U}(\hat{x})$ such that for every $y \in \mathcal{U}(\hat{x})$ the family

$$\{\nabla g_j(\hat{x}) \,:\, j \in \mathcal{K}_g\} \cup \{\nabla h_j(\hat{x}) \,:\, j \in \mathcal{K}_h\} \cup \{e_{1j} \,:\, j \in \mathcal{K}_1\}$$

is linearly dependent.

Theorem 3.1. *Let $\Psi(a,b)$ satisfy the Assumptions 3.1. Suppose that (x^k) is a sequence, where each x^k is a stationary point of $P(\rho^k)$ and $\rho^k \longrightarrow +\infty$. Moreover, assume that (x^k) has a limit point $\bar{x}$, which is feasible for (4) and satisfies the reduced MPEC-CPLD for (4). Then $\bar{x}$ is stationary for (4), i.e. there exist multipliers $\bar{\lambda} \geq 0$, $\bar{\lambda}_h$, $\bar{\nu}_0$ and $\bar{\nu}_1$ such that the conditions*

$$\nabla f(\bar{x}) - \sum_{i \in \mathcal{I}_h} \bar{\lambda}_{h,i} \nabla h_i(\bar{x}) - \sum_{j \in \mathcal{I}_g(\bar{x})} \bar{\lambda}_j \nabla g_i(\bar{x}) - \sum_{j \in \mathcal{I}_1(\bar{x})} \nu_{0j} e_{1j} + \sum_{j \in \mathcal{I}_1^{\perp}(\bar{x})} \nu_{1j} e_{1j} = 0$$

$$\begin{aligned} g_i(\bar{x})\bar{\lambda}_i &= 0, \quad (1 \leq i \leq m) \\ \bar{x}_{1j}\,\bar{\nu}_{0j} &= 0, \quad (1 \leq j \leq p) \\ (1 - \bar{x}_{1j})\,\bar{\nu}_{1j} &= 0, \quad (1 \leq j \leq p) \end{aligned} \tag{18}$$

are satisfied.

Proof. Since each x^k is supposed to be a stationary point of $P(\rho^k)$ it holds $x^k \in \mathcal{Z}_P$ and there exist multipliers λ_h^k, $\lambda^k \geq 0$, $\xi_0^k \geq 0$ and $\xi_1^k \geq 0$ such that

$$\begin{aligned} \nabla f(x^k) - \sum_{i \in \mathcal{I}_h} \lambda_{h,i}^k \nabla h_i(x^k) - \sum_{j \in I_g(x^k)} \lambda_j^k \nabla g_i(x^k) & \\ - \sum_{j=1}^{p} \left(\xi_{0j}^k - \rho^k \frac{\partial \Psi}{\partial a_j}(x_1^k, \mathbf{1} - x_1^k) \right) e_{1j} & \\ + \sum_{j=1}^{p} \left(\xi_{1j}^k - \rho_k \frac{\partial \Psi}{\partial b_j}(x_1^k, \mathbf{1} - x_1^k) \right) e_{1j} &= 0. \end{aligned} \tag{19}$$

and

$$\begin{aligned} g_i(x^k)\lambda_i^k &= 0, \quad (1 \leq i \leq m) \\ x_{1j}^k\,\xi_{0j}^k &= 0, \quad (1 \leq j \leq p) \\ (1 - x_{1j}^k)\,\xi_{1j}^k &= 0, \quad (1 \leq j \leq p). \end{aligned} \tag{20}$$

We define

$$\nu_0^k := \xi_{0j}^k - \rho^k \frac{\partial \Psi}{\partial a_j}(x_1^k, \mathbf{1} - x_1^k) \quad \text{and} \quad \nu_1^k := \xi_{1j}^k - \rho^k \frac{\partial \Psi}{\partial b_j}(x_1^k, \mathbf{1} - x_1^k)$$

and construct sequence $(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)$ which has a limit point $(\bar{\lambda}_h, \bar{\lambda}, \bar{\nu}_0, \bar{\nu}_1)$ that is a suitable multiplier vector for $\bar{x}$. We therefore prove that our sequence $(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)$ is bounded.

Applying Lemma A.1, for every x^k we can find a multiplier vector $(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)$ such that $\tilde{\lambda}^k \geq 0$ and the family

$$\begin{aligned} &\{\nabla g_j(x^k) \,:\, j \in \operatorname{supp}(\tilde{\lambda}^k)\} \cup \{\nabla h_j(x^k) \,:\, j \in \operatorname{supp}(\tilde{\lambda}_h^k)\} \\ &\cup \{e_{1j} \,:\, j \in \operatorname{supp}(\tilde{\nu}_0^k) \cup \operatorname{supp}(\tilde{\nu}_1^k)\} \end{aligned} \tag{21}$$

is linearly independent and

$$\begin{array}{rcccl} \operatorname{supp}(\tilde{\lambda}^k) & \subseteq & \operatorname{supp}(\lambda^k) & \subseteq & I_g(\bar{x}) \\ \operatorname{supp}(\tilde{\lambda}_h^k) & \subseteq & \operatorname{supp}(\lambda_h^k) & \subseteq & I_h \\ \operatorname{supp}(\tilde{\nu}_0^k) \cup \operatorname{supp}(\tilde{\nu}_1^k) & \subseteq & \operatorname{supp}(\nu_0^k) \cup \operatorname{supp}(\nu_1^k) & \subseteq & \{1, .., p\} \end{array} \tag{22}$$

for sufficiently large k. There exist only finitely many configurations of different supports hence there exists at least one configuration $\mathcal{C}_\lambda$, ..., $\mathcal{C}_{\nu_1}$ of supports such that

$$\operatorname{supp}(\tilde{\lambda}^k) = \mathcal{C}_\lambda, \quad \ldots, \quad \operatorname{supp}(\tilde{\nu}_1^k) = \mathcal{C}_{\nu_1} \tag{23}$$

for infinitely many k. Let $\mathcal{M} = \{k \in \mathbb{N} : (23) \text{ holds for } k\}$ be the associated index set. We next prove that the corresponding subsequence of $(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)$ is bounded. Assume that this is not the case. Then there exist an unbounded subsequence $(k \in \mathcal{M}_u \subseteq \mathcal{M})$ such that $\gamma^k := \|(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)\| \neq 0$ for all $k \in \mathcal{M}_u$ and the sequence

$$(\hat{\lambda}_h^k, \hat{\lambda}^k, \hat{\nu}_0^k, \hat{\nu}_1^k) := \frac{(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)}{\|(\tilde{\lambda}_h^k, \tilde{\lambda}^k, \tilde{\nu}_0^k, \tilde{\nu}_1^k)\|}$$

is well-defined. Moreover, since it is bounded, $(\hat{\lambda}_h^k, \hat{\lambda}^k, \hat{\nu}_0^k, \hat{\nu}_1^k)$ has a limit point. Denote $\mathcal{K} \subseteq \mathcal{M}_u$ the index subset such that

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} (\hat{\lambda}_h^k, \hat{\lambda}^k, \hat{\nu}_0^k, \hat{\nu}_1^k) = (\check{\lambda}_h^k, \check{\lambda}^k, \check{\nu}_0^k, \check{\nu}_1^k).$$

Deviding (19) by γ^k and taking the limit $\mathcal{K} \ni k \to \infty$ yields

$$\sum_{i\in\mathcal{C}_{\lambda_h}} \check{\lambda}_{h,i} \nabla h_i(\bar{x}) + \sum_{j\in\mathcal{C}_\lambda} \check{\lambda}_j \nabla g_i(\bar{x}) + \sum_{j\in\mathcal{C}_{\nu_0}\cup\mathcal{C}_{\nu_1}} (\check{\nu}_{0j} - \check{\nu}_{0j})\, e_{1j} = 0. \tag{24}$$

and hence the positive linear dependence of the family

$$\{\nabla g_j(\bar{x}) \,:\, j \in \mathcal{C}_\lambda\} \cup \{\nabla h_j(\bar{x}) \,:\, j \in \mathcal{C}_{\lambda_h}\} \cup \{e_{1j} \,:\, j \in \mathcal{C}_{\nu_0} \cup \mathcal{C}_{\nu_1}\}.$$

This however, contradicts the reduced MPEC-CPLD to hold in $\bar{x}$, as $x^k \longrightarrow \bar{x}$. Hence $(\tilde{\lambda}_h^k$, $\tilde{\lambda}^k$, $\tilde{\nu}_0^k$, $\tilde{\nu}_1^k)$ is bounded and has a limit point $(\bar{\lambda}_h, \bar{\lambda}, \bar{\nu}_0, \bar{\nu}_1)$ satisfying the first equation of (18).

It remains to prove that $\bar{\nu}_0$ and $\bar{\nu}_1$ satisfy the complementarity conditions of (18). Assume there exists $\delta > 0$ with $|\bar{\nu}_{1j}| > \delta$ for some $j \in I_1(\bar{x})$. Then, since $\xi_{1j}^k = 0$, it follows that $\frac{\partial \Psi}{\partial b_j}^k > 0$ (evaluated at $(x_1^k \mathbf{1} - x_1)$) for all $k \in \mathcal{K}$ sufficiently large. By Assumption 3.1 (2) b this implies $x_{1j}^k > 0$ and thus $\xi_0^k = 0$ for all $k \in \mathcal{K}$ sufficiently large. Hence, by Assumption 3.1 (3) we have

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \frac{|\nu_{1j}^k|}{|\nu_{0j}^k|} = \lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \frac{|\rho^k \frac{\partial \Psi}{\partial b_j}^k|}{|\rho^k \frac{\partial \Psi}{\partial a_j}^k|} = 0.$$

Because $|\nu_{1j}^k| \longrightarrow |\bar{\nu}_{1j}| > \delta$ this contradicts $\lim_{k\in\mathcal{K}, k\to\infty} \nu_{0j}^k = \bar{\nu}_{0j}$. The complementarity of $\bar{\nu}_0$ can be proved similarly. □

3.2 Feasibility of a limit point

Lemma 3.1. *Let $\rho^k \to +\infty$ and let (x^k) be a sequence of stationary points of $P(\rho^k)$. Assume that it has a limit point $\bar{x}$. Let $\hat{x}$ be a feasible point of (4) at which the MPEC-LICQ holds. Then there exists $\varepsilon > 0$ such that if $\bar{x} \in B_\varepsilon(\hat{x})$, then $\bar{x}$ is feasible for (4).*

Proof. If the MPEC-LICQ holds at $\tilde{x}$ for (4), then there exists an $\varepsilon > 0$ such that for any $x \in B_\varepsilon(\tilde{x})$ the vectors

$$\{\nabla g_j(x) : j \in I_g(\tilde{x})\} \cup \{\nabla h_j(x) : j \in I_h(\tilde{x})\} \cup \{e_{1j} : j \in \{1,..,p\}\} \tag{25}$$

remain linear independent.

Since each x^k is a stationary point of $P(\rho^k)$ we have that $x^k \in \mathcal{Z}_P$ for all $k \in \mathbb{N}$ and there exist multipliers λ_h^k, $\lambda^k \geq 0$, $\xi_0^k \geq 0$ and $\xi_1^k \geq 0$ such that

$$\begin{aligned} 0 = &- \nabla f(x^k) + \sum_{i \in \mathcal{I}_h} \lambda_{h,i}^k \nabla h_i(x^k) + \sum_{j \in I_g(x^k)} \lambda_j^k \nabla g_j(x^k) \\ &+ \sum_{j=1}^{p} \left(\xi_{0j}^k - \rho^k \frac{\partial \Psi}{\partial a_j} (x_1^k, \mathbf{1} - x_1^k) \right) e_{1j} \\ &- \sum_{j=1}^{p} \left(\xi_{1j}^k - \rho_k \frac{\partial \Psi}{\partial b_j} (x_1^k, \mathbf{1} - x_1^k) \right) e_{1j}. \end{aligned} \tag{26}$$

Furthermore, if $\varepsilon > 0$ is small enough, then for all k sufficiently large

$$I_g(x^k) \subseteq I_g(\bar{x}) \subseteq I_g(\tilde{x}). \tag{27}$$

Define

$$\nu^k := (\xi_{0j}^k - \xi_{1j}^k) - \rho^k \left(\frac{\partial \Psi}{\partial a_j} - \frac{\partial \Psi}{\partial b_j} \right) (x_1^k, \mathbf{1} - x_1^k)$$

Assume that there exists an index $j \in \{1,..,p\}$ with $0 < \bar{x}_{1j} < 1$. Then, for sufficiently large k it holds $0 < x_{1j}^k < 1$. More precisely if ε is small enough then either $0 < \bar{x}_{1j} << 1$ for $j \in I_1(\tilde{x})$ or $0 << \bar{x}_{1j} < 1$ for $j \in I_1(\tilde{x})^\perp$ and therefore either $0 < x_{1j}^k << 1$ or $0 << x_{1j}^k < 1$, respectively, for all k sufficiently large. Hence, by Assumptions 3.1 there exist $1 >> \delta_1 > 0$ and $\delta_2 > 0$ such that

$$\begin{aligned} |\nu^k| &= \rho^k \left| \left(\frac{\partial \Psi}{\partial a_j} - \frac{\partial \Psi}{\partial b_j} \right) (x_1^k, \mathbf{1} - x_1^k) \right| \\ &\geq \rho^k (1 - \delta_1) \begin{cases} \frac{\partial \Psi}{\partial a_j} (x_1^k, \mathbf{1} - x_1^k) & j \in I_1(\tilde{x}) \\ \frac{\partial \Psi}{\partial b_j} (x_1^k, \mathbf{1} - x_1^k) & j \in I_1(\tilde{x})^\perp \end{cases} \\ &\geq \rho^k (1 - \delta_1)\, \delta_2 \longrightarrow +\infty \end{aligned}$$

Hence, without restriction we assume that $\beta^k := \|(\lambda_h^k,\ \lambda^k,\ \nu^k)\| > 0$. Dividing (26) by β^k and taking the limit $k \to \infty$ we obtain a nontrivial linear combination of the vectors of (25) which contradicts the MPEC-LICQ in $\tilde{x}$. Hence, $\bar{x}$ must be feasible for (4). □

Lemma 3.2. *Let x^* be a strict local minimum of (4) in $B_\delta(x^*)$. Define*

$$\mathcal{S}(\rho) := \{x \in \mathbb{R}^{n+p} : \ x \text{ is a local minimum of } P(\rho) \ \} \tag{28}$$

Then for each $\varepsilon \in (0, \delta)$ there exists a $\rho(\varepsilon)$ such that for each $\rho \geq \rho(\varepsilon)$ the set $\mathcal{S}(\rho) \cap B_\varepsilon(x^*)$ is not empty.

Proof. We prove that the global minimum of $P_\varepsilon(\rho)$, which is $P(\rho)$ with the additional constraint $\|x - x^*\| \leq \varepsilon$, satisfies $\|x - x^*\| < \varepsilon$, i.e. it occurs in the interior of $B_\varepsilon(x^*)$. If this is the case, then the same point is a local minimum of $P(\rho)$ in $B_\varepsilon(x^*)$, hence $\mathcal{S}(\rho) \cap B_\varepsilon(x^*) \neq \emptyset$.

Since x^* is a strict local minimum of (4), there exists a $\delta > 0$ such that $f(x^*) < f(x)$ for all $x \in B_\delta(x^*) \cap \mathcal{Z}_{MPCC}$. In particular, this inequality holds for all x on the sphere $S_\delta(x^*) \cap \mathcal{Z}_{MPCC}$. Let $\tilde{x}$ be the minimum of f on $S_\delta(x^*) \cap \mathcal{Z}_{MPCC}$ (which exists, since f is continuous and $S_\delta(x^*) \cap \mathcal{Z}_{MPCC}$ is compact). Define $r := f(\tilde{x}) - f(x^*)$. Then,

$$f(x) \geq f(x^*) + r \qquad \text{for all} \qquad x \in S_\delta(x^*) \cap \mathcal{Z}_{MPCC}. \tag{29}$$

Moreover, (29) implies that for any $\sigma \in (0, r)$ there exists a $\bar{\rho} > 0$ such that

$$f(x) + \rho\, \Psi(x_1, \mathbf{1} - x_1) \geq f(x^*) + \sigma \qquad \text{for all} \qquad x \in S_\delta(x^*) \cap \mathcal{Z}_P. \tag{30}$$

for all $\rho \geq \bar{\rho}$. Assume that this is not the case. Then, for any $\rho^k \to +\infty$ we can find $x^k \in S_\delta(x^*) \cap \mathcal{Z}_P$ such that

$$f(x^k) + \rho\, \Psi(x_1^k, \mathbf{1} - x_1^k) \leq f(x^*) + \sigma.$$

where $\sigma \in (0, r)$. Since (x^k) is a bounded sequence, there exists a limit point $\bar{x} \in S_\delta(x^*) \cap \mathcal{Z}_P$ and

$$0 \leq \Psi(\bar{x}_1, \mathbf{1} - \bar{x}_1) = \lim_{\substack{k\to\infty \\ k \in \mathcal{K}}} \Psi(x_1^k, \mathbf{1} - x_1^k) \leq \lim_{\substack{k\to\infty \\ k \in \mathcal{K}}} \frac{1}{\rho^k}(f(x^*) - f(x^k) + \sigma) \to 0$$

Hence, $\bar{x} \in S_\delta(x^*) \cap \mathcal{Z}_{MPCC}$ and thus by (29)

$$f(x^*) + r \leq f(\bar{x}) = \lim_{\substack{k\to\infty \\ k \in \mathcal{K}}} f(x^k) + \rho\, \Psi(x_1^k, \mathbf{1} - x_1^k) \leq f(x^*) + \sigma$$

which contradicts our choice $\sigma \in (0, r)$.

Since (30) implies that for all ρ sufficiently large it holds

$$f(x^*) + \rho\, \Psi(x_1^*, \mathbf{1} - x_1^*) = f(x^*) < f(x) + \rho\, \Psi(x_1, \mathbf{1} - x_1) \quad \text{for all} \quad x \in S_\delta(x^*) \cap \mathcal{Z}_P.$$

and $x^* \in \operatorname{int} B_\varepsilon(x^*) \cap \mathcal{Z}_P$ for all $\varepsilon > 0$, the global minimum of $P_\varepsilon(\rho)$ occurs in the interior of $B_\varepsilon(x^*)$.

□

Theorem 3.2. *Let x^* be a local minimum of (4) at which the MPEC-LICQ holds. Let $V(\rho)$ be the set of stationary points of $P(\rho)$. Then there exists an $\varepsilon > 0$ and $\rho(\varepsilon)$ such that $V(\rho) \cap B_\varepsilon(x^*) \neq \emptyset$ for all $\rho \geq \rho(\varepsilon)$ and moreover, if (x^k) is a sequence of stationary points of $P(\rho^k)$ with $\rho^k \to +\infty$ and $x^k \in V(\rho^k) \cap B_\varepsilon(x^*)$ for all k, then the limit point of (x^k) is x^*.*

Proof. Since x^* is a local minimum of (4), there exists a neighbourhood $B_\delta(x^*)$ such that x^* is the unique minimum and stationary point of (4) in $B_\delta(x^*)$ (see e.g. Theorem .. [20]). Hence by Lemma 3.2 for any $\varepsilon_1 \in (0, \delta)$ there exists $\rho(\varepsilon_1)$ such that $V(\rho) \cap B_{\varepsilon_1}(x^*) \neq \emptyset$ for all $\rho \geq \rho(\varepsilon_1)$.

Furthermore, since the MPEC-LICQ holds in x^* there exists a $\delta_1 > 0$ such that the MPEC-LICQ holds at any point $x \in B_{\delta_1}(x^*) \cap \mathcal{Z}_{MPCC}$.

Now, consider a sequence (x^k) such that $x^k \in V(\rho^k) \cap B_\varepsilon(x^*)$ of stationary points of $P(\rho^k)$ with $\rho^k \to +\infty$, where $\varepsilon = \min(\varepsilon_1, \varepsilon_2, \delta_1)$ with ε_2 as in Lemma 3.1. Then (x^k) has a limit point $\bar{x} \in B_\varepsilon(x^*)$, that is feasible for (4) by Lemma 3.1. Moreover, it satisfies the MPEC-LICQ. Hence by Theorem 3.1 $\bar{x}$ is stationary for (4). As x^* is the unique stationary point of (4) in $x \in B_\varepsilon(x^*) \cap \mathcal{Z}_{MPCC}$, it follows that $\bar{x} = x^*$. □

4 Numerical Results

The numerical results for the relaxation approximation (9) are computed using the Fischer–Burmeister NCP function. The initial relaxation value was $\mu = 0.1$ and in each iteration the value decreases by a factor of 1/10. Similarly, we use the Fischer–Burmeister function as penalty function within Ψ in (7). The initial penalty parameter is one and within each iteration its value increases by a factor of ten. The termination tolerance is in all cases 10^{-6}.

We consider a prototype of a generalized disjunctive program [12, 24] given by

$$\begin{array}{rl} \min_{x,Y} & (x_1-3)^2 + (x_2-2)^2 + c \\ \text{subject to} & 0 \leq x_1, x_2 \leq 8, \qquad Y_j \in \{\text{true,false}\}\ j = 1,2,3 \quad , \end{array}$$

$$\begin{bmatrix} Y_1 \\ x_1^2 + x_2^2 - 1 \leq 0 \\ c = 2 \end{bmatrix} \vee \begin{bmatrix} Y_2 \\ (x_1-4)^2 + (x_2-1)^2 - 1 \leq 0 \\ c = 1 \end{bmatrix} \vee \begin{bmatrix} Y_3 \\ (x_1-2)^2 + (x_2-4)^2 - 1 \leq 0 \\ c = 3 \end{bmatrix}$$

The problem is reformulated using big-M constraints as

$$\begin{array}{rl} \min_{x,y} & (x_1-3)^2 + (x_2-2)^2 + 2y_1 + y_2 + 3y_3 \\ \text{s.t.} & x_1^2 + x_2^2 - 1 \leq M(1-y_1), \\ & (x_1-4)^2 + (x_2-1)^2 - 1 \leq M(1-y_2), \\ & (x_1-2)^2 + (x_2-4)^2 - 1 \leq M(1-y_3), \\ & 0 \leq x_i \leq 8, \qquad y_1, y_2, y_3 \in \{0,1\}. \end{array}$$

For this problem the big–M parameter is assigned as $M = 30$. The later MINLP formulation is now solved using the penalty approach. The optimal solution is

$$(x^*, y^*) = (3.293, 1.707, 0, 1, 0)$$

with the corresponding cost of $f^* = 1.172$. Using a penalty approach with the parameter $\rho_0 = 1$ we obtain the solution

$$(x, y) = (3.1, 1.9, 0.02, 0.98, 0),$$

and the corresponding cost of $f = 1.04$. Now, increasing the penalty parameter to $\rho_1 = 10$ we obtain the correct solution $(x, y) = (x^*, y^*)$.

Further numerical tests have been performed on a selection of MINLP problems from the MINLPLib [3]. We have chosen 29 MINLP problems with each having less than 200 variables and equations. The selected problems originate from process engineering applications. For the inner iterations for relaxation and penalization approaches we used CONOPT as a black–box NLP solver. The results are reported in Table 1 and Figure 2. We report on the number of solved problems where we consider a problem solved when the residual in the KKT–system is less than the given tolerance (problems solved). Further, we report on the problems where the value of the cost function is equal to the best known solution (global solutions found). The relative error between the computed optimal functional value and the best known functional value is given in column three for both approaches. The number of NLP solved during the sequential optimization is also stated as well as the average number of CONOPT steps per inner optimization step. In Figure 2 we show the computational effort of both approaches. We observe that the penalty approach is in all cases faster than the relaxation approximation. The measure used is the total number of CONOPT iterations. A similar result has been observed in [1]. Therein, also the penalty approach to a model of distillation columns has shown superior convergence properties compared with a relaxation approach.

	Penalty Approach (6)	Relaxation Approach (9)
% problems solved	76	68
% global solutions found	11	3
⊘ relative error	0.70	0.58
avarage nr. of solver iter.	13.4	62.8
avarage nr. of outer iter.	1	4

Table 1: Comparison of the penalty and relaxation approach for a selection of MINLP problems taken from MINLPLib[3].

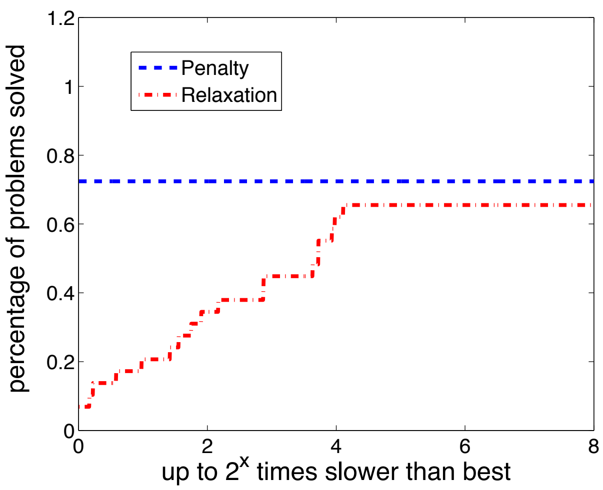


Figure 2: Comparison of penalty and relaxation approximation by comparing the total number of CONOPT iterations.

5 Summary

Using techniques to analyze MPCCs two approaches to MINLPs have been theoretically and numerically discussed. For a relaxation reformulation of the MINLP as well as for a reformulation using penalty functions rigorous results on feasibility, stationarity and convergence of the sequence of stationary points has been shown. Numerical results comparing both approaches show a slight superior behavior of the penalty approach.

Acknowledgments

The work has been supported by DFG HE5386/8-1 and STE2063/1-1. We thank W. Marquardt for pointing out this interesting problem.

References

[1] B. T. Baumruckner, J. G. Renfro, and L. T. Biegler, *Mpec problem formulations in chemical engineering applications*, (2007).

[2] L. T. Biegler and A. Raghunathan, *An interior point method for mathematical programs with complementarity constraints (mpccs)*, SIAM Journal on Optimization, 15 (2005), pp. 720–750.

[3] M. R. Bussieck, A. S. Drud, and A. Meeraus, *MINLPLib - A Collection of Test Models for Mixed-Integer Nonlinear Programming*, Informs J. Comput., 15 (2003).

[4] V. DeMiguel, M. P. Friedlander, F. J. Nogales, and S. Scholtes, *A two-sided relaxation scheme for mathematical programs with equilibrium constraints*, SIAM Journal on Optimization, 16 (2005), pp. 587–609.

[5] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, 1983.

[6] R. Fletcher and S. Leyffer, *Solving mathematical programs with complementary constraints as nonlinear programs.*, Optimization Methods and Software, 19 (2004), pp. 15–40.

[7] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local convergence of SQP methods for mathematical programs with equilibrium constraints.*, SIAM Journal on Optimization, 17 (2006), pp. 259–286.

[8] M. Fukushima and G. H. Lin, *A modified relaxation scheme for mathematical programs with complementarity constraints.*, Annals of Operations Research, 133 (2005), pp. 63–84.

[9] M. Fukushima and J. S. Pang, *Complementarity constraint qualifications and simplified B-stationary conditions for mathematical programs with equilibrium constraints.*, Computational Optimization and Applications, 13 (1999), pp. 111–136.

[10] X. M. Hu and D. Ralph, *Convergence of a penalty method for mathematical programming with complementarity constraints*, Journal of Optimization Theory and Applications, 123 (2004), pp. 365–390.

[11] M. Kočvara, J. Outrata, and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints. Theory, applications and numerical results.*, Nonconvex Optimization and Its Applications. 28. Dordrecht: Kluwer Academic Publishers. , 1998.

[12] S. Lee and I. E. Grossman, *New algorithms for nonlinear generalized disjunctive programming*, Computers and Chemical Engineering, 24 (2000), pp. 2125–2141.

[13] S. Leyffer, *Complementarity constraints as nonlinear equations: Theory and numerical experience*, in S. Dempe and V. Kalashnikov, editors, Optimization and Multivalued Mappings, (2006), pp. 169–208.

[14] S. Leyffer, G. López-Calva, and J. Nocedal, *Interior methods for mathematical programs with complementarity constraints*, SIAM Journal on Optimization, 17 (2006), pp. 52–77.

[15] Z. Q. Luo, J. S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints.*, Cambridge: Cambridge University Press., 1997.

[16] Z. Q. Luo, J. S. Pang, D. Ralph, and S. Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints.*, Mathematical Programming, Series A, 75 (1996), pp. 19–76.

[17] D. Ralph and S. Wright, *Some properties of regularization and penalization schemes for MPECs.*, Optimization Methods and Software, 19 (2004), pp. 527–556.

[18] R. Raman and I. E. Grossman, *Modelling and computational techniques for logic based integer programming*, Computers and Chemical Engineering, 27 (1994), pp. 563–578.

[19] S. M. Robinson, *Strongly Regular Generalized Equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.

[20] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity.*, Mathematical Operations Research, 25 (2000), pp. 1–22.

[21] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM Journal on Optimization, 11 (2001), pp. 918–936.

[22] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM Journal on Control and Optimization, 37 (1999), pp. 617–652.

[23] S. Steffensen and M. Ulbrich, *A new relaxation scheme for mathematical programs with equilibrium constraints*, SIAM Journal on Optimization, (2010).

[24] O. Stein, J. Oldenburg, and W. Marquardt, *Continuous reformulations of discrete-continuous optimization problems*, Computers & Chemical Engineering, 28 (2004), pp. 1951–1966.

[25] J. J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints.*, Journal of Mathematical Analysis and Applications, 307 (2005), pp. 350–369.

Appendix

The following Lemma is an auxiliary result that we use in the proof of Theorem 3.1.

Lemma .1. *Let (x, λ, μ) be a KKT-triple of the NLP*

$$\begin{array}{rl} \mathit{min} & f(x) \\ \mathit{subject\ to} & g(x) \geq 0 \\ & h(x) = 0. \end{array}$$

Then there exist feasible multipliers $(\bar{\lambda}, \bar{\mu})$, such that

$$\bar{\lambda}_j = 0 \quad \forall\ j : \lambda_j = 0, \qquad\qquad \bar{\mu}_j = 0 \quad \forall\ j : \mu_j = 0,$$

and the system of vectors

$$\begin{array}{ll} \nabla g_j(x) & j \in I_g(x),\ \bar{\lambda}_j > 0 \\ \nabla h_j(x) & j \in \{i \in I_h \,|\, \bar{\mu} \neq 0\} \end{array}$$

is linearly independent.

Proof. Since (x, λ, μ) is a KKT-triple, there holds

$$\nabla f(x) - \sum_{j \in I_g(x)} \lambda_j \nabla g_j(x) - \sum \mu_j \nabla h_j(x) = 0, \qquad \lambda_j \geq 0. \tag{31}$$

If we substitute $\mu_j := \mu_j^+ - \mu_j^-$ in (31) with $\mu_j^+ = \max(0, \mu_j)$, $\mu_j^- = \min(0, -\mu_j)$, then $\mu_j^+, \mu_j^- \geq 0$, and finding multipliers satisfying (31) corresponds to finding a solution to

$$Az = b, \qquad z \geq 0, \tag{32}$$

where $b = \nabla f(x)$ and the columns of A are composed by the gradient vectors $\nabla g_j(x)$ such that $\lambda_j > 0$, $\nabla h_j(x)$ such that $\mu_j^+ > 0$, and $-\nabla h_j(x)$ such that $\mu_j^- > 0$. Applying a result of linear programming it is possible to find a $\bar{z}$ solving (32) such that the columns of A corresponding to $\{j \ :\ \bar{z}_j \neq 0\}$ are linearly independent. Now $\bar{\lambda}$ and $\bar{\mu}$ can be easily obtained from $\bar{z}$.

□