DFG Deutsche Forschungsgemeinschaft
Priority Programme 1962

Reformulation of the M-stationarity Conditions as a System of Discontinuous Equations and its Solution by a Semismooth Newton Method

Felix Harder, Patrick Mehlitz, Gerd Wachsmuth

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-135

received on February 24, 2020

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Preprint, Faculty 1
Institute of Mathematics

Brandenburg
University of Technology
Cottbus - Senftenberg

Reformulation of the M-stationarity conditions as a system of discontinuous equations and its solution by a semismooth Newton method

Felix Harder, Patrick Mehlitz, Gerd Wachsmuth

February 24, 2020

Chair of Optimal Control, BTU Cottbus-Senftenberg

We show that the Mordukhovich-stationarity system associated with a mathematical program with complementarity constraints (MPCC) can be equivalently written as a system of discontinuous equations which can be tackled with a semismooth Newton method. We show that the resulting algorithm can be interpreted as an active set strategy for MPCCs. Local fast convergence of the method is guaranteed under validity of an MPCC-tailored version of LICQ and a suitable second-order condition. In case of linear-quadratic MPCCs, the LICQ-type constraint qualification can be replaced by a weaker condition which depends on the underlying multipliers. We discuss a suitable globalization strategy for our method. Some numerical results are presented in order to illustrate our theoretical findings.

Keywords: Active set method, Mathematical program with complementarity constraints, M-stationarity, Nonlinear M-stationarity function, Semismooth Newton method

MSC: 49M05, 49M15, 90C30, 90C33

1 Introduction

We aim for the numerical solution of so-called *mathematical programs with complementarity constraints* (MPCCs for short) which are nonlinear optimization problems of the form

$$\begin{aligned}\min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) \le 0, \quad h(x) = 0, \\ & G(x) \ge 0, \quad H(x) \ge 0, \quad G(x)^\top H(x) = 0.\end{aligned} \tag{MPCC}$$

Throughout the article, we assume that the data functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^\ell$, $h\colon \mathbb{R}^n \to \mathbb{R}^m$, and $G, H\colon \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable. Observing that most of the standard constraint qualifications fail to hold at the feasible points of (MPCC) while the feasible set of it is likely to be (almost) disconnected, complementarity-constrained programs form a challenging class of optimization problems. On the other hand, several real-world optimization scenarios from mechanics, finance, or natural sciences naturally comprise equilibrium conditions which is why they can be modeled in the form (MPCC). For an introduction to the topic of complementarity-constrained programming, the interested reader is referred to the monographs Luo, Pang, Ralph, 1996; Outrata, Kočvara, Zowe, 1998. In the past, huge effort has been put into the development of problem-tailored constraint qualifications and stationarity notions which apply to (MPCC), see e.g. Scheel, Scholtes, 2000; Ye, 2005 for an overview. Second-order necessary and sufficient optimality conditions for (MPCC) are discussed in Gfrerer, 2014; Guo, Lin, Ye, 2013; Scheel, Scholtes, 2000. There exist several different strategies in order to handle the inherent difficulties of (MPCC) in the context of its numerical solution. A common idea is to relax the complementarity constraints and to solve the resulting standard nonlinear surrogate problems with a standard method, see e.g. Hoheisel, Kanzow, Schwartz, 2013 for an overview. Problem-tailored penalization approaches are discussed e.g. in Hu, Ralph, 2004; Huang, Yang, Zhu, 2006; Leyffer, López-Calva, Nocedal, 2006; Ralph, Wright, 2004. Possible approaches for adapting the well-known SQP method of nonlinear programming to (MPCC) are investigated in Benko, Gfrerer, 2016; Fletcher et al., 2006; Luo, Pang, Ralph, 1998. Active set strategies for the numerical solution of (MPCC) with affine complementarity constraints are under consideration in Fukushima, Tseng, 2002; Júdice et al., 2007. In Izmailov, Pogosyan, Solodov, 2012, the authors combine a lifting approach as well as a globalized semismooth Newton-type method in order to solve (MPCC). Furthermore, we would like to mention the paper Guo, Lin, Ye, 2015 where the authors reformulate different stationarity systems of (MPCC) as (over-determined) nonlinear systems of equations subject to a polyhedron, and the latter systems are solved via a Levenberg–Marquardt method.

Using so-called NCP-functions, where NCP abbreviates *nonlinear complementarity problem*, complementarity restrictions can be transferred into systems of equations which are possibly nonsmooth. Recall that a function $\pi\colon \mathbb{R}^2 \to \mathbb{R}$ is called NCP-function whenever it satisfies

$$\forall (a,b) \in \mathbb{R}^2\colon \quad \pi(a,b) = 0 \iff a, b \ge 0 \wedge ab = 0.$$

Two popular examples of such NCP-functions are the minimum-function $\pi_{\min}\colon \mathbb{R}^2 \to \mathbb{R}$ as well as the Fischer–Burmeister-function $\pi_{\text{FB}}\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$\pi_{\min}(a,b) := \min(a,b), \quad \pi_{\text{FB}}(a,b) := \sqrt{a^2 + b^2} - a - b.$$

A convincing overview of existing NCP-functions and their properties can be found in Galántai, 2012; Kanzow, Yamashita, Fukushima, 1997; Sun, Qi, 1999. We note that most of the established NCP-functions like $\pi_{\min}$ or π_{FB} are nonsmooth. Classically, NCP-function have been used to transfer Karush–Kuhn–Tucker (KKT) systems of standard nonlinear problems with inequality constraints into systems of equations which then are tackled with the aid of a Newton-type method which is capable of handling the potentially arising nonsmoothness, see De Luca, Facchinei, Kanzow, 1996; 2000; Facchinei, Soares, 1997 for an overview. Furthermore, these papers report on the differentiability of the function π_{FB}^2 which can be exploited in order to globalize the resulting Newton method. In Izmailov, Solodov, 2008, the authors extended this idea to (MPCC) by interpreting it as a nonlinear problem. Under reasonable assumptions, local quadratic convergence to so-called strongly stationary points has been obtained and suitable globalization strategies have been presented.

In this paper, we aim to reformulate Mordukhovich's system of stationarity (the so-called system of M-stationarity) associated with (MPCC) as a system of nonsmooth equations which can be solved by a semismooth Newton method. Our study is motivated by several different aspects. First, we would like to mention that the set

$$M := \left\{ (a, b, \mu, \nu) \in \mathbb{R}^4 \,\middle|\, \begin{array}{l} 0 \le a \perp b \ge 0,\ a\mu = 0,\ b\nu = 0, \\ (\mu\nu = 0 \ \vee\ \mu < 0, \nu < 0) \end{array} \right\}, \tag{1.1}$$

which is closely related to the M-stationarity system of (MPCC), see Definition 2.4, is closed. In contrast, the set

$$\tilde{M} := \left\{ (a, b, \mu, \nu) \in \mathbb{R}^4 \,\middle|\, \begin{array}{l} 0 \le a \perp b \ge 0,\ a\,\mu = 0,\ b\,\nu = 0, \\ a = b = 0 \ \Rightarrow\ \mu, \nu \le 0 \end{array} \right\},$$

which characterizes the system of strongly stationary points associated with (MPCC), is not closed. In fact, M is the closure of $\tilde{M}$. Based on this topological observation, it is clear that searching for M-stationary points is far more promising than searching for strongly stationary points as long as both stationarity systems are transferred into systems of nonsmooth equations which can be solved by suitable methods. In Guo, Lin, Ye, 2015, the authors transferred the M-stationarity system of (MPCC) into a smooth (over-determined) system of equations subject to a polyhedron, and they solved it with the aid of a modified Levenberg–Marquardt method. It has been shown that the resulting algorithm converges superlinearly to an M-stationary point whenever a suitable error bound condition holds. However, one cannot expect local quadratic convergence of the method. Our aim in this paper is, thus, to use a nonsmooth reformulation of the M-stationarity system which can be tackled with a semismooth Newton method in order to ensure local fast convergence of the resulting algorithm under suitable assumptions, namely MPCC-LICQ, an MPCC-tailored variant of the prominent Linear Independence Constraint Qualification (LICQ), and MPCC-SSOC, an MPCC-tailored strong second-order condition, have to hold at the limit point, see Definitions 2.5 and 2.6 as well as Theorem 4.2. Using the squared residual, we are in position to globalize our method, see

Section 5. Observing that the strongly stationary points of (MPCC) can be found among its M-stationary points, the resulting method may converge to strongly stationary points, too. Let us mention that even in the absence of MPCC-LICQ, local fast convergence of the method is possible if the linearly dependent gradients do not appear in the Newton system, and, anyway, local slow convergence will be always guaranteed via our globalization strategy. It will turn out that whenever the objective f of (MPCC) is quadratic while the constraint functions g, h, G, and H are affine, then we actually can replace MPCC-LICQ by a slightly weaker condition depending on the multipliers at the limit point, see Section 6.

The manuscript is organized as follows: In Section 2, we summarize the essential preliminaries. Particularly, we recall some terminology from complementarity-constrained programming and review the foundations of semismooth Newton methods. Section 3 is dedicated to the reformulation of the M-stationarity system associated with (MPCC) as a system of nonsmooth equations. In order to guarantee that a Newton-type method can be applied in order to solve the resulting system, we first motivate the general structure of this system. Afterwards, we introduce a so-called *nonlinear M-stationarity function* whose roots are precisely the elements of the set M from (1.1). Although the nonlinear M-stationarity function of our interest is nonsmooth and even discontinuous, we prove that it is Newton differentiable on the set M. Based on this function, we construct a semismooth Newton method which solves the M-stationarity system of (MPCC) in Section 4. Furthermore, we provide a local convergence analysis which shows that our method ensures local quadratic convergence under validity of MPCC-LICQ and MPCC-SSOC at the limit point. Moreover, we illustrate that our Newton-type method can be interpreted as an active set strategy for (MPCC). In Section 5, the globalization of the algorithm is discussed. We exploit the standard idea to minimize the squared residual. In Section 6, we show that it is possible to relax MPCC-LICQ in the setting of a linear-quadratic model problem (MPCC) while keeping all the desired convergence properties. Numerical experiments addressing a discretized version of the optimal control of the obstacle problem whose local minimizers are M- but not S-stationary are presented in Section 7. Some remarks close the paper in Section 8.

2 Preliminaries

2.1 Notation

We introduce the index sets $I^\ell := \{1, \ldots, \ell\}$, $I^m := \{1, \ldots, m\}$, and $I^p := \{1, \ldots, p\}$. The component mappings of g, h, G, and H are denoted by g_i ($i \in I^\ell$), h_i ($i \in I^m$), G_i ($i \in I^p$), and H_i ($i \in I^p$), respectively.

We use 0 in order to denote the scalar zero as well as the zero matrix of appropriate dimensions. For $i \in \{1, \ldots, n\}$, we use $e_i \in \mathbb{R}^n$ to represent the i-th unit vector. For a vector $v \in \mathbb{R}^n$ and a set $I \subset \{1, \ldots, n\}$, $v_I \in \mathbb{R}^{|I|}$ denotes the vector which results from v

by deleting all entries corresponding to indices from $\{1, \ldots, n\} \setminus I$. Similarly, for a matrix $V \in \mathbb{R}^{n\times m}$, $V_I \in \mathbb{R}^{|I|\times m}$ denotes the matrix which is obtained by deleting all those rows from V whose indices correspond to the elements of $\{1, \ldots, n\} \setminus I$. If (row) vectors v_i, $i \in I$, are given, then $[v_i]_I$ denotes the matrix whose rows are precisely the vectors v_i, $i \in I$. Finally, let us mention that for $x \in \mathbb{R}^n$ and $\varepsilon > 0$, $B_\varepsilon(x)$ represents the closed ε-ball around x.

Lemma 2.1. Let A and B be matrices of suitable dimensions satisfying

$$A^\top \lambda + B^\top \eta = 0,\ \lambda \geq 0 \quad \Longrightarrow \quad \lambda = 0,$$

i.e., the rows of A are positive linearly independent from the rows of B. Then it holds

$$\operatorname{span}\{d \in \mathbb{R}^n \mid Ad \leq 0,\ Bd = 0\} = \{d \in \mathbb{R}^n \mid Bd = 0\}.$$

The proof follows from standard arguments. The next two lemmas are classical.

Lemma 2.2. Consider the saddle-point matrix

$$C := \begin{pmatrix} A & B^\top \\ B & 0 \end{pmatrix},$$

where A and B are matrices of compatible sizes. If the constraint block B is surjective, i.e., the rows of B are linearly independent, and if A is positive definite on the kernel of B, i.e., $x^\top A\, x > 0$ for all $x \in \ker(B) \setminus \{0\}$, then C is invertible.

Lemma 2.3. Let the matrix A be invertible. Then there exist constants $\varepsilon > 0$ and $C > 0$ such that

$$\|(A + \delta A)^{-1}\| \leq C$$

holds for all δA with $\|\delta A\| \leq \varepsilon$.

2.2 MPCCs

Here, we briefly summarize the well-known necessary essentials on stationarity conditions, constraint qualifications, and second-order conditions for complementarity constrained optimization problems. As mentioned earlier, most of the standard constraint qualifications do not hold at the feasible points of (MPCC) which is why stationarity notions, weaker than the KKT conditions, have been introduced. Let us recall some of them. For that purpose, we first introduce the MPCC-tailored Lagrangian $\mathcal{L}\colon \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ associated with (MPCC) via

$$\mathcal{L}(x, \lambda, \eta, \mu, \nu) := f(x) + \lambda^\top g(x) + \eta^\top h(x) + \mu^\top G(x) + \nu^\top H(x).$$

Furthermore, for a feasible point $\bar{x} \in \mathbb{R}^n$ of (MPCC), we will make use of the index sets

$$\begin{aligned} I^g(\bar{x}) &:= \{i \in I^\ell \,|\, g_i(\bar{x}) = 0\}, \\ I^{+0}(\bar{x}) &:= \{i \in I^p \,|\, G_i(\bar{x}) > 0 \;\wedge\; H_i(\bar{x}) = 0\}, \\ I^{0+}(\bar{x}) &:= \{i \in I^p \,|\, G_i(\bar{x}) = 0 \;\wedge\; H_i(\bar{x}) > 0\}, \\ I^{00}(\bar{x}) &:= \{i \in I^p \,|\, G_i(\bar{x}) = 0 \;\wedge\; H_i(\bar{x}) = 0\}. \end{aligned}$$

Clearly, $\{I^{+0}(\bar{x}), I^{0+}(\bar{x}), I^{00}(\bar{x})\}$ is a disjoint partition of I^p.

Definition 2.4. Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (MPCC). Then $\bar{x}$ is said to be

(a) Mordukhovich-stationary (M-stationary) if there exist multipliers $\lambda \in \mathbb{R}^\ell$, $\eta \in \mathbb{R}^m$, and $\mu, \nu \in \mathbb{R}^p$ which solve the system

$$\nabla_x \mathcal{L}(\bar{x}, \lambda, \eta, \mu, \nu) = 0, \tag{2.1a}$$
$$\lambda_{I^g(\bar{x})} \geq 0, \quad \lambda_{I^\ell \setminus I^g(\bar{x})} = 0, \tag{2.1b}$$
$$\mu_{I^{+0}(\bar{x})} = 0, \tag{2.1c}$$
$$\nu_{I^{0+}(\bar{x})} = 0, \tag{2.1d}$$
$$\forall i \in I^{00}(\bar{x})\colon\ \mu_i \nu_i = 0 \;\vee\; (\mu_i < 0 \;\wedge\; \nu_i < 0), \tag{2.1e}$$

(b) strongly stationary (S-stationary) if there exist multipliers $\lambda \in \mathbb{R}^\ell$, $\eta \in \mathbb{R}^m$, and $\mu, \nu \in \mathbb{R}^p$ which satisfy (2.1a)-(2.1d) and

$$\mu_{I^{00}(\bar{x})} \leq 0, \quad \nu_{I^{00}(\bar{x})} \leq 0. \tag{2.2}$$

Let us briefly note that there exist several more stationarity notions which apply to (MPCC), see e.g. Ye, 2005 for an overview. For later use, let $\Lambda^{\mathrm{M}}(\bar{x})$ and $\Lambda^{\mathrm{S}}(\bar{x})$ be the sets of all multipliers which solve the system of M- and S-stationarity w.r.t. a feasible point $\bar{x} \in \mathbb{R}^n$ of (MPCC), respectively.

In this paper, we will make use of a popular MPCC-tailored version of the Linear Independence Constraint Qualification.

Definition 2.5. Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of (MPCC). Then the MPCC-tailored Linear Independence Constraint Qualification (MPCC-LICQ) is said to hold at $\bar{x}$ whenever the matrix

$$\begin{bmatrix} g'(\bar{x})_{I^g(\bar{x})} \\ h'(\bar{x}) \\ G'(\bar{x})_{I^{0+}(\bar{x}) \cup I^{00}(\bar{x})} \\ H'(\bar{x})_{I^{+0}(\bar{x}) \cup I^{00}(\bar{x})} \end{bmatrix}$$

possesses full row rank.

It is a classical result that a local minimizer of (MPCC) where MPCC-LICQ holds is S-stationary. Furthermore, the associated multipliers $(\lambda, \eta, \mu, \nu)$, which solve the system (2.1a)-(2.1d), (2.2) are uniquely determined. It has been reported in Flegel, Kanzow, 2005a; Ye, 2005 that even under validity of mild MPCC-tailored constraint qualifications, local minimizers of (MPCC) are M-stationary. Therefore, it is a reasonable strategy to identify the M-stationary points of a given complementarity-constrained optimization problem in order to tackle the problem of interest.

We review existing second-order optimality conditions addressing (MPCC) which are based on S-stationary points. We adapt the considerations from Scheel, Scholtes, 2000. For some point $\bar{x} \in \mathbb{R}^n$, we first introduce the so-called MPCC-critical cone

$$\mathcal{C}(\bar{x}) := \left\{ \delta x \in \mathbb{R}^n \;\middle|\; \begin{array}{rl} \nabla f(\bar{x})^\top \delta x \leq 0 & \\ \nabla g_i(\bar{x})^\top \delta x \leq 0 & i \in I^g(\bar{x}) \\ h'(\bar{x}) \delta x = 0 & \\ \nabla G_i(\bar{x})^\top \delta x = 0 & i \in I^{0+}(\bar{x}) \\ \nabla H_i(\bar{x})^\top \delta x = 0 & i \in I^{+0}(\bar{x}) \\ \nabla G_i(\bar{x})^\top \delta x \geq 0 & i \in I^{00}(\bar{x}) \\ \nabla H_i(\bar{x})^\top \delta x \geq 0 & i \in I^{00}(\bar{x}) \\ (\nabla G_i(\bar{x})^\top \delta x)(\nabla H_i(\bar{x})^\top \delta x) = 0 & i \in I^{00}(\bar{x}) \end{array} \right\}.$$

We note that this cone is likely to be not convex if the index set $I^{00}(\bar{x})$ of biactive complementarity constraints is nonempty. In case where $\bar{x}$ is an S-stationary point of (MPCC) and $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{S}}(\bar{x})$ is arbitrarily chosen, we obtain the representation

$$\mathcal{C}(\bar{x}) = \left\{ \delta x \in \mathbb{R}^n \;\middle|\; \begin{array}{rl} \nabla g_i(\bar{x})^\top \delta x = 0 & i \in I^g(\bar{x}), \lambda_i > 0 \\ \nabla g_i(\bar{x})^\top \delta x \leq 0 & i \in I^g(\bar{x}), \lambda_i = 0 \\ h'(\bar{x}) \delta x = 0 & \\ \nabla G_i(\bar{x})^\top \delta x = 0 & i \in I^{0+}(\bar{x}) \cup I^{00}_{\pm\mathbb{R}}(\bar{x}, \mu, \nu) \\ \nabla H_i(\bar{x})^\top \delta x = 0 & i \in I^{+0}(\bar{x}) \cup I^{00}_{\mathbb{R}\pm}(\bar{x}, \mu, \nu) \\ \nabla G_i(\bar{x})^\top \delta x \geq 0 & i \in I^{00}_{00}(\bar{x}, \mu, \nu) \\ \nabla H_i(\bar{x})^\top \delta x \geq 0 & i \in I^{00}_{00}(\bar{x}, \mu, \nu) \\ (\nabla G_i(\bar{x})^\top \delta x)(\nabla H_i(\bar{x})^\top \delta x) = 0 & i \in I^{00}_{00}(\bar{x}, \mu, \nu) \end{array} \right\}$$

by elementary calculations, see Mehlitz, 2019, Lemma 4.1 as well, where we used

$$I^{00}_{\pm\mathbb{R}}(\bar{x}, \mu, \nu) := \{ i \in I^{00}(\bar{x}) \,|\, \mu_j \neq 0 \}, \tag{2.3a}$$

$$I^{00}_{\mathbb{R}\pm}(\bar{x}, \mu, \nu) := \{ i \in I^{00}(\bar{x}) \,|\, \nu_j \neq 0 \}, \tag{2.3b}$$

$$I^{00}_{00}(\bar{x}, \mu, \nu) := \{ i \in I^{00}(\bar{x}) \,|\, \mu_j = \nu_j = 0 \}. \tag{2.3c}$$

If $\bar{x} \in \mathbb{R}^n$ is a local minimizer of (MPCC) where MPCC-LICQ holds, then the unique multiplier $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{S}}(\bar{x})$ satisfies

$$\forall \delta x \in \mathcal{C}(\bar{x})\colon \quad \delta x^\top \nabla^2_{xx} \mathcal{L}(\bar{x}, \lambda, \eta, \mu, \nu) \delta x \geq 0.$$

Let us note that necessary second-order conditions for (MPCC) which are based on M-stationary points can be found in Guo, Lin, Ye, 2013. On the other hand, if $\bar{x}$ is an arbitrary S-stationary point of (MPCC) where the so-called MPCC-tailored Second-Order Sufficient Condition (MPCC-SOSC) given by

$$\forall \delta x \in \mathcal{C}(\bar{x}) \setminus \{0\}\, \exists (\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{S}}(\bar{x})\colon \qquad \delta x^\top \nabla^2_{xx} \mathcal{L}(\bar{x}, \lambda, \eta, \mu, \nu) \delta x > 0$$

holds, then $\bar{x}$ is a strict local minimizer of (MPCC). More precisely, the second-order growth condition holds for (MPCC) at $\bar{x}$.

Finally, we are going to state a second-order condition which we are going to exploit for our convergence analysis. Observe that this condition is based on M-stationary points.

Definition 2.6. Let $\bar{x} \in \mathbb{R}^n$ be an M-stationary point of (MPCC). Furthermore, let $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{M}}(\bar{x})$ be fixed. Then the MPCC-tailored Strong Second-Order Condition (MPCC-SSOC) is said to hold at $\bar{x}$ w.r.t. $(\lambda, \eta, \mu, \nu)$ whenever

$$\forall \delta x \in S(\bar{x}, \lambda, \mu, \nu) \setminus \{0\}\colon \quad \delta x^\top \nabla^2_{xx} \mathcal{L}(\bar{x}, \lambda, \eta, \mu, \nu) \delta x > 0$$

holds true. Here, the set $S(\bar{x}, \lambda, \mu, \nu)$ is given by

$$S(\bar{x}, \lambda, \mu, \nu) := \left\{ \delta x \in \mathbb{R}^n \;\middle|\; \begin{array}{rl} \nabla g_i(\bar{x})^\top \delta x = 0 & i \in I^g(\bar{x}),\ \lambda_i > 0 \\ h'(\bar{x}) \delta x = 0 & \\ \nabla G_i(\bar{x})^\top \delta x = 0 & i \in I^{0+}(\bar{x}) \cup I^{00}_{\pm\mathbb{R}}(\bar{x}, \mu, \nu) \\ \nabla H_i(\bar{x})^\top \delta x = 0 & i \in I^{+0}(\bar{x}) \cup I^{00}_{\mathbb{R}\pm}(\bar{x}, \mu, \nu) \\ (\nabla G_i(\bar{x})^\top \delta x)(\nabla H_i(\bar{x})^\top \delta x) = 0 & i \in I^{00}_{00}(\bar{x}, \mu, \nu) \end{array} \right\}.$$

Fix a feasible point $\bar{x} \in \mathbb{R}^n$ of (MPCC) which is M-stationary and let $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{M}}(\bar{x})$ be an associated multiplier. For any set $\beta \subset I^{00}_{00}(\bar{x}, \mu, \nu)$, we define the complement $\bar{\beta} := I^{00}_{00}(\bar{x}, \mu, \nu) \setminus \beta$ as well as

$$S_\beta(\bar{x}, \lambda, \mu, \nu) := \left\{ \delta x \in \mathbb{R}^n \;\middle|\; \begin{array}{rl} \nabla g_i(\bar{x})^\top \delta x = 0 & i \in I^g(\bar{x}),\ \lambda_i > 0 \\ h'(\bar{x}) \delta x = 0 & \\ \nabla G_i(\bar{x})^\top \delta x = 0 & i \in I^{0+}(\bar{x}) \cup I^{00}_{\pm\mathbb{R}}(\bar{x}, \mu, \nu) \cup \beta \\ \nabla H_i(\bar{x})^\top \delta x = 0 & i \in I^{+0}(\bar{x}) \cup I^{00}_{\mathbb{R}\pm}(\bar{x}, \mu, \nu) \cup \bar{\beta} \end{array} \right\}.$$

Then it holds

$$S(\bar{x}, \lambda, \mu, \nu) = \bigcup_{\beta \subset I^{00}_{00}(\bar{x}, \mu, \nu)} S_\beta(\bar{x}, \lambda, \mu, \nu).$$

Due to Lemma 2.1, under validity of MPCC-LICQ at $\bar{x}$, we have $S_\beta(\bar{x}, \lambda, \mu, \nu) = \operatorname{span} \mathcal{C}_\beta(\bar{x}, \lambda, \mu, \nu)$ where we used

$$\mathcal{C}_\beta(\bar{x}, \lambda, \mu, \nu) := \left\{ \delta x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^\top \delta x = 0 & i \in I^g(\bar{x}), \lambda_i > 0 \\ \nabla g_i(\bar{x})^\top \delta x \leq 0 & i \in I^g(\bar{x}), \lambda_i = 0 \\ h'(\bar{x}) \delta x = 0 & \\ \nabla G_i(\bar{x})^\top \delta x = 0 & i \in I^{0+}(\bar{x}) \cup I^{00}_{\pm\mathbb{R}}(\bar{x}, \mu, \nu) \cup \beta \\ \nabla H_i(\bar{x})^\top \delta x = 0 & i \in I^{+0}(\bar{x}) \cup I^{00}_{\mathbb{R}\pm}(\bar{x}, \mu, \nu) \cup \bar{\beta} \\ \nabla G_i(\bar{x})^\top \delta x \geq 0 & i \in \bar{\beta} \\ \nabla H_i(\bar{x})^\top \delta x \geq 0 & i \in \beta \end{array} \right\},$$

i.e., $S(\bar{x}, \lambda, \mu, \nu)$ is the finite union of the spans of polyhedral cones. Observing that

$$\mathcal{C}(\bar{x}) = \bigcup_{\beta \subset I^{00}_{00}(\bar{x}, \mu, \nu)} \mathcal{C}_\beta(\bar{x}, \lambda, \mu, \nu)$$

holds true provided $\bar{x}$ is S-stationary while $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{S}}(\bar{x})$ holds, the set $S(\bar{x}, \lambda, \mu, \nu)$ is closely related to the critical cone of (MPCC). In this regard, the name MPCC-SSOC in Definition 2.6 is quite reasonable. Observe that the inclusion $\mathcal{C}(\bar{x}) \subset S(\bar{x}, \lambda, \mu, \nu)$ holds for each S-stationary point $\bar{x}$ and each multiplier $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{S}}(\bar{x})$, i.e., MPCC-SSOC is slightly stronger than MPCC-SOSC in this situation.

2.3 Semismooth Newton methods

In this section, we collect some theory concerning the application of Newton methods for functions $F\colon \mathbb{R}^n \to \mathbb{R}^n$ which are not continuously differentiable. In the finite-dimensional case, one typically utilizes semismooth functions. Since semismooth functions are by definition locally Lipschitz continuous, this theory is not applicable to discontinuous functions. Hence, we exploit the concept of Newton differentiability, which is used in infinite-dimensional applications of Newton's method, see Chen, Nashed, Qi, 2000; Hintermüller, Ito, Kunisch, 2002; Ulbrich, 2002; Ito, Kunisch, 2008.

Definition 2.7. Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ and $DF\colon \mathbb{R}^n \to \mathbb{R}^{m \times n}$ be given. The function F is said to be Newton differentiable (with derivative DF) on a set $K \subset \mathbb{R}^n$ if

$$F(x+h) - F(x) - DF(x+h)\, h = o(\|h\|) \qquad \text{for } h \to 0$$

holds for all $x \in K$. For $\alpha \in (0, 1]$, the function F is Newton differentiable of order α, if

$$F(x+h) - F(x) - DF(x+h)\, h = \mathcal{O}(\|h\|^{1+\alpha}) \qquad \text{for } h \to 0$$

holds for all $x \in K$. Finally, F is said to be Newton differentiable of order ∞, if for all

$x \in K$ there is $\varepsilon_x > 0$ such that

$$F(x+h) - F(x) - DF(x+h)\,h = 0 \qquad \forall h \in \mathbb{R}^n : \|h\| < \varepsilon_x.$$

Clearly, if F is continuously differentiable, then $DF = F'$ is a Newton derivative. If F' is locally Lipschitz continuous, then F is Newton differentiable of order 1.

In the following example, we discuss the Newton differentiability of the minimum and maximum entry of a vector.

Example 2.8. For the nonsmooth functions $\min, \max\colon \mathbb{R}^n \to \mathbb{R}$, we establish the following convention for choosing Newton derivatives at arbitrary points $a \in \mathbb{R}^n$:

$$\begin{aligned} D\min(a_1,\ldots,a_n) &:= e_i^\top \text{ where } i = \min\{j \in \{1,\ldots,n\} \mid a_j = \min(a_1,\ldots,a_n)\}, \\ D\max(a_1,\ldots,a_n) &:= e_i^\top \text{ where } i = \min\{j \in \{1,\ldots,n\} \mid a_j = \max(a_1,\ldots,a_n)\}, \end{aligned} \tag{2.4}$$

i.e., we give priority to variables that appear first in a min or max expression. This choice ensures that min and max are indeed Newton differentiable of order ∞.

In order to find a solution $\bar{x}$ of $F(\bar{x}) = 0$ where $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is a Newton differentiable map, we use the iteration

$$x_{k+1} := x_k - DF(x_k)^{-1} F(x_k), \qquad k = 0, 1, \ldots$$

for an initial guess $x_0 \in \mathbb{R}^n$. As usual, we call this iteration scheme *semismooth Newton method* but emphasize that it applies to mappings F which are not semismooth in the classical sense.

Nowadays, the proof of the next theorem is classical, see, e.g., Chen, Nashed, Qi, 2000, proof of Theorem 3.4.

Theorem 2.9. Assume that $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is Newton differentiable on $K \subset \mathbb{R}^n$ with Newton derivative DF. Further assume that $\bar{x} \in K$ satisfies $F(\bar{x}) = 0$ and that the matrices $\{DF(x) \mid x \in B_\varepsilon(\bar{x})\}$ are uniformly invertible for some $\varepsilon > 0$. Then there exists $\delta > 0$ such that Newton's method is well defined and superlinearly convergent for any initial iterate $x_0 \in B_\delta(\bar{x})$. If F is additionally Newton differentiable of order 1, the convergence is quadratic, and we have convergence in one step if F is Newton differentiable of order ∞.

If the assumptions of Theorem 2.9 are satisfied, one obtains the equivalence of the known residuum $\|F(x)\|$ and the unknown distance $\|x - \bar{x}\|$.

Lemma 2.10. In addition to the assumptions of Theorem 2.9, suppose that the matrices $\{DF(x) \mid x \in B_\varepsilon(\bar{x})\}$ are bounded for some $\varepsilon > 0$. Then there exist constants $c, C, \delta > 0$

such that

$$\forall x \in B_\delta(\bar{x}): \quad c\,\|F(x)\| \le \|x - \bar{x}\| \le C\,\|F(x)\|.$$

The proof follows from

$$F(x) = F(x) - F(\bar{x}) = DF(x)\,(x - \bar{x}) + o(\|x - \bar{x}\|)$$

and the properties of $DF(x)$. For the reader's convenience, we provide the following chain rule.

Lemma 2.11. Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is Newton differentiable on $K \subset \mathbb{R}^n$ with derivative Df and that $g\colon \mathbb{R}^m \to \mathbb{R}^p$ is Newton differentiable on $f(K)$ with derivative Dg. Further, we assume that Df is bounded on a neighborhood of K and that Dg is bounded on a neighborhood of $f(K)$. Then $f \circ g$ is Newton differentiable on K with derivative given by $x \mapsto Dg(f(x))\,Df(x)$. If both f and g are Newton differentiable of order $\alpha \in (0,1] \cup \{\infty\}$, then $f \circ g$ is Newton differentiable of order α.

The chain rule can be shown along the lines of Clason, 2018, Theorem 9.3.

Example 2.12. Exploiting Example 2.8 and Lemma 2.11, the absolute value function $|\cdot|\colon \mathbb{R} \to \mathbb{R}$ is Newton differentiable of order ∞, since $|x| = \max(x, -x)$ for each $x \in \mathbb{R}$. Following the convention from Example 2.8, the associated Newton derivative is given by

$$\forall x \in \mathbb{R}: \quad D|\cdot|(x) = \begin{cases} 1 & \text{if } x \ge 0, \\ -1 & \text{if } x < 0. \end{cases} \tag{2.5}$$

3 M-Stationarity as a nonlinear system of equations

3.1 Preliminary considerations

As stated before, we want to reformulate the M-stationarity system (2.1) as an equation. To this end, we need to encode the complementarity conditions (2.1b) and the conditions (2.1c)–(2.1e), which depend on index sets, as the zero level set of a suitable function. For clarity of the presentation, we temporarily consider a simplified MPCC problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & 0 \le H(x) \perp G(x) \ge 0, \end{aligned} \tag{MPCC1}$$

with $G, H\colon \mathbb{R}^n \to \mathbb{R}$, i.e., there is only one complementarity constraint. This simplification will also simplify notation in this section. The results of this section will be transferred to the problem (MPCC) with p many complementarity conditions in Section 4. The M-stationarity system for (MPCC1) is given by

$$\nabla_x \mathcal{L}(x, \mu, \nu) = 0, \tag{3.1a}$$

$$0 \le H(x) \perp G(x) \ge 0, \tag{3.1b}$$
$$G(x)\mu = 0, \tag{3.1c}$$
$$H(x)\nu = 0, \tag{3.1d}$$
$$(\mu < 0 \wedge \nu < 0) \vee \mu\nu = 0. \tag{3.1e}$$

We want to find a function $\varphi\colon \mathbb{R}^4 \to \mathbb{R}^k$ such that (3.1) can be equivalently rewritten as $F(x, \mu, \nu) = 0$ where $F : \mathbb{R}^{n+1+1} \to \mathbb{R}^{n+k}$ has the form

$$F(x, \mu, \nu) := \begin{pmatrix} \nabla_x \mathcal{L}(x, \mu, \nu) \\ \varphi(G(x), H(x), \mu, \nu) \end{pmatrix}.$$

Since we want to apply a Newton method, we require $k = 2$. We also need that the associated Newton matrices $DF(\cdot)$ are invertible in a neighborhood of the solution of the system $F(x, \mu, \nu) = 0$. This invertibility will be guaranteed by certain properties of the Newton derivative $D\varphi$.

3.2 A nonlinear M-stationarity function

Recall that the set M from (1.1) corresponds to the M-stationarity conditions (3.1). We define the functions $\psi_1, \psi_2, \psi_3, \varphi_1 \colon \mathbb{R}^4 \to \mathbb{R}$ via

$$\psi_1(a, b, \mu, \nu) := \max(-a, |b|, |\mu|), \tag{3.2a}$$
$$\psi_2(a, b, \mu, \nu) := \max(-b, |a|, |\nu|), \tag{3.2b}$$
$$\psi_3(a, b, \mu, \nu) := \max(|a|, |b|, \mu, \nu), \tag{3.2c}$$
$$\varphi_1(a, b, \mu, \nu) := \min_{i=1,2,3} \psi_i(a, b, \mu, \nu). \tag{3.2d}$$

Next, we show that M is precisely the zero level set of φ_1. To this end, we note that

$$\begin{aligned} M = \{(a, b, \mu, \nu) \in \mathbb{R}^4 \mid a \ge 0,\ b = \mu = 0\} &\cup \{(a, b, \mu, \nu) \in \mathbb{R}^4 \mid b \ge 0,\ a = \nu = 0\} \\ &\cup \{(a, b, \mu, \nu) \in \mathbb{R}^4 \mid a = b = 0, \mu \le 0, \nu \le 0\}, \end{aligned}$$

i.e., M can be written as the union of three convex, closed sets.

Lemma 3.1. Let $(a, b, \mu, \nu) \in \mathbb{R}^4$ be given. Then $(a, b, \mu, \nu) \in M$ holds if and only if $\varphi_1(a, b, \mu, \nu) = 0$ is valid.

Proof. It is clear that $\varphi_1(a, b, \mu, \nu) \ge 0$. Hence, $\varphi_1(a, b, \mu, \nu) = 0$ if and only if one of the functions ψ_1, ψ_2, and ψ_3 vanishes at (a, b, μ, ν).

Now, we observe the equivalencies

$$\begin{aligned} \psi_1(a, b, \mu, \nu) = 0 \quad &\Leftrightarrow \quad -a \le 0,\ b = \mu = 0, \\ \psi_2(a, b, \mu, \nu) = 0 \quad &\Leftrightarrow \quad -b \le 0,\ a = \nu = 0, \\ \psi_3(a, b, \mu, \nu) = 0 \quad &\Leftrightarrow \quad a = b = 0,\ \mu, \nu \le 0. \end{aligned}$$

Thus, $\varphi_1(a, b, \mu, \nu) = 0$ holds if and only if one of the left hand sides is true and $(a, b, \mu, \nu) \in M$ if and only if one of the right hand sides is true. This shows the claim.

We remark that φ_1 describes the distance of a point (a, b, μ, ν) to the set M in the ℓ^∞-norm. This follows from the above representation of M and the fact that each of ψ_1, ψ_2, ψ_3 is the distance to one of the three convex subsets of M.

We choose the Newton derivative $D\varphi_1$ of φ_1 according to the conventions established in (2.4) and (2.5) together with the application of the chain rule. This choice for $D\varphi_1$ is fixed for the remainder of the article. It implies that

$$D\varphi_1(a, b, \mu, \nu) \in \left\{\pm e_1^\top, \pm e_2^\top, \pm e_3^\top, \pm e_4^\top\right\} \tag{3.3}$$

holds for all $(a, b, \mu, \nu) \in \mathbb{R}^4$.

Let us introduce the other component $\varphi_2 \colon \mathbb{R}^4 \to \mathbb{R}$, which is defined via

$$\varphi_2(a, b, \mu, \nu) := \begin{cases} \min(|b|, |\nu|) & \text{if } D\varphi_1(a, b, \mu, \nu) = \pm e_1^\top, \\ \min(|a|, |\mu|) & \text{if } D\varphi_1(a, b, \mu, \nu) = \pm e_2^\top, \\ |b| & \text{if } D\varphi_1(a, b, \mu, \nu) = \pm e_3^\top, \\ |a| & \text{if } D\varphi_1(a, b, \mu, \nu) = \pm e_4^\top, \end{cases} \tag{3.4}$$

where the cases are exhaustive due to (3.3). For the prospective Newton derivative $D\varphi_2$ of φ_2 we again use the convention established in (2.4) as well as (2.5) and use the same distinction of cases as in (3.4). The Newton differentiability of φ_2 will be shown in Lemma 3.3 below. Finally, let $\varphi \colon \mathbb{R}^4 \to \mathbb{R}^2$ be the function with components φ_1 and φ_2. The rows of the Newton derivative $D\varphi$ of φ are given by $D\varphi_1$ and $D\varphi_2$. Motivated by our arguments from Section 3.1, we call φ a nonlinear M-stationarity (NMS) function, see Lemma 3.3 below as well.

In the next lemma, we will look at the possible values of the Newton derivative of φ at points from M. This will be an important lemma in order to show that the Newton matrix $DF(\cdot)$ is invertible in a neighborhood of M, see Theorem 4.1. If φ_2 is chosen differently, one might obtain less tight estimates for the Newton matrices $D\varphi$, and this would result in more restrictive assumptions for the semismooth Newton method below, cf. the proof of Theorem 4.1. For that purpose, we define the sets of matrices

$$\forall i, j \in \{1, \ldots, 4\}\colon \quad J_{i,j} := \left\{\begin{bmatrix} \pm e_i^\top \\ \pm e_j^\top \end{bmatrix}\right\} \cup \left\{\begin{bmatrix} \pm e_j^\top \\ \pm e_i^\top \end{bmatrix}\right\} \subset \mathbb{R}^{2\times 4}.$$

Lemma 3.2. Let $\bar w = (\bar a, \bar b, \bar\mu, \bar\nu) \in M$ be given. Then there exists $\varepsilon > 0$ such that for all $w = (a, b, \mu, \nu) \in B_\varepsilon(\bar w)$, we have

$$\bar a > 0 \quad \Rightarrow \quad D\varphi(w) \in J_{2,3}, \tag{3.5a}$$

$$\bar b > 0 \quad \Rightarrow \quad D\varphi(w) \in J_{1,4}, \tag{3.5b}$$

$$\bar{\mu} \neq 0 \quad \Rightarrow \quad D\varphi(w) \in J_{1,2} \cup J_{1,4}, \tag{3.5c}$$
$$\bar{\nu} \neq 0 \quad \Rightarrow \quad D\varphi(w) \in J_{1,2} \cup J_{2,3}, \tag{3.5d}$$
$$\bar{w} = 0 \quad \Rightarrow \quad D\varphi(w) \in J_{1,2} \cup J_{2,3} \cup J_{1,4}. \tag{3.5e}$$

Proof. Due to the definition of φ_2, the possible values for $D\varphi(w)$ can only be in $J_{1,2} \cup J_{2,3} \cup J_{1,4}$ for all $w \in \mathbb{R}^4$. Clearly, the implication (3.5e) follows immediately.

Suppose that $\bar{a} > 0$ holds. Then we have $\bar{b} = \bar{\mu} = 0$. Therefore, there exists $\varepsilon > 0$ such that $\max(|b|, |\mu|) < a$ holds for all $w = (a, b, \mu, \nu) \in B_\varepsilon(\bar{w})$. It follows that $\varphi_1 = \psi_1 < \min(\psi_2, \psi_3)$ holds on $B_\varepsilon(\bar{w})$. Thus we obtain $D\varphi_1(w) \in \{\pm e_2^\top, \pm e_3^\top\}$. If we again consider that $|\mu| < |a|$ then the implication (3.5a) follows. The implication (3.5b) can be shown in a similar way.

Let us consider the case $\bar{\mu} \neq 0$ and $\bar{\nu} \neq 0$. Then we have $\bar{a} = \bar{b} = 0$ and also $\bar{\mu}, \bar{\nu} < 0$. Therefore, there exists $\varepsilon > 0$ such that $\max(|a|, |b|) < \min(-\mu, -\nu)$ holds for all $w = (a, b, \mu, \nu) \in B_\varepsilon(\bar{w})$. It follows that $\varphi_1 = \psi_3 < \min(\psi_1, \psi_2)$ holds on $B_\varepsilon(\bar{w})$. Thus we obtain $D\varphi_1(w) \in \{\pm e_1^\top, \pm e_2^\top\}$. If we consider that $|a| < |\mu|$ and $|b| < |\nu|$ then $D\varphi(w) \in J_{1,2}$ follows.

Next, we consider the case that $\bar{\mu} \neq 0$ but $\bar{\nu} = \bar{b} = 0$. Then we have $\bar{a} = 0$. Therefore, there exists $\varepsilon > 0$ such that $\max(|a|, |b|, |\nu|) < |\mu|$ holds for all $w = (a, b, \mu, \nu) \in B_\varepsilon(\bar{w})$. It follows that $\varphi_1 < \psi_1$ holds on $B_\varepsilon(\bar{w})$. By a distinction of cases we can obtain that $D\varphi_1(w) \in \{\pm e_1^\top, \pm e_2^\top, \pm e_4^\top\}$. If we consider (3.4) and that $|a| < |\mu|$ then $D\varphi(w) \in J_{1,2} \cup J_{1,4}$ follows.

For the case that $\bar{\mu} \neq 0$, $\bar{\nu} = 0$, but $\bar{b} > 0$ we already know from (3.5b) that $D\varphi(w) \in J_{1,2} \cup J_{1,4}$ holds as well. If we combine the previous cases, then we obtain (3.5c). The implication (3.5d) can be shown in a similar way.

We continue with some notable properties of φ. The first property is important because it allows us to characterize M-stationarity points as the solution set of a (nonsmooth) equation, and this is the essential property of an NMS-function.

Lemma 3.3.

(a) We have $\varphi(a, b, \mu, \nu) = 0$ if and only if $(a, b, \mu, \nu) \in M$.

(b) The function φ is Newton differentiable of order ∞ on M.

(c) The function φ is not continuous in any open neighborhood of M.

(d) The function φ is calm at every point $\bar{w} = (\bar{a}, \bar{b}, \bar{\mu}, \bar{\nu}) \in M$ with calmness modulus 1, i.e., there is a neighborhood U of $\bar{w}$ such that

$$\forall w \in U: \quad \|\varphi(w) - \varphi(\bar{w})\| \leq \|w - \bar{w}\|.$$

Proof. We start with part (a). Lemma 3.1 shows $(a, b, \mu, \nu) \in M$ if and only if $\varphi_1(a, b, \mu, \nu) = 0$. Thus, it remains to show that $\varphi_2(a, b, \mu, \nu) = 0$ for all points $(a, b, \mu, \nu) \in M$. Let $(a, b, \mu, \nu) \in M$ be given. We consider the case that $a > 0$. Then $b = \mu = 0$ follows. Due to (3.5a) we have $D\varphi_1(a, b, \mu, \nu) \in \{\pm e_2^\top, \pm e_3^\top\}$, which implies $\varphi_2(a, b, \mu, \nu) = 0$. For the case that $b > 0$ we can argue similarly. For the remaining case $a = b = 0$ the property $\varphi_2(a, b, \mu, \nu) = 0$ follows directly from the definition of φ_2.

For part (b), let us fix a point $\bar{w} = (\bar{a}, \bar{b}, \bar{\mu}, \bar{\nu}) \in M$. For φ_1, the Newton differentiability of order ∞ follows from the chain rule Lemma 2.11. Due to $\varphi_2(\bar{w}) = 0$, it suffices to show that

$$\varphi_2(w) = D\varphi_2(w)(w - \bar{w}) \tag{3.6}$$

holds in a neighborhood of $\bar{w}$. Let $\varepsilon > 0$ from Lemma 3.2 be given and consider $w = (a, b, \mu, \nu) \in B_\varepsilon(\bar{w})$. In case $D\varphi_2(w) = \pm e_1^\top$, (3.5a) implies $\bar{a} = 0$ and from the definition of $D\varphi_2$, we get $\varphi_2(w) = \pm a$. Hence, (3.6) follows. In case $D\varphi_2(w) = \pm e_3^\top$, (3.5c) implies $\bar{\mu} = 0$ and from the definition of $D\varphi_2$, we get $\varphi_2(w) = \pm\mu$. Again, (3.6) follows. The remaining cases follow analogously.

We continue with part (c). Any open neighborhood of M contains the point $w_t := (2t, 2t, t, 0)$ for some $t > 0$. It can be shown that

$$\varphi_2(2t, 2t, t, 0) = t \neq 0 = \lim_{s \downarrow 0} \varphi_2(2t - s, 2t, t, 0).$$

Hence φ_2 is not continuous at w_t.

In order to show part (d), we can utilize part (b), which implies that

$$\varphi(w) - \varphi(\bar{w}) = D\varphi(w)(w - \bar{w})$$

holds for all w in a neighborhood of $\bar{w}$. Since $\|D\varphi(w)\| = 1$ holds for all w in a neighborhood of $\bar{w}$ due to Lemma 3.2, we obtain

$$\|\varphi(w) - \varphi(\bar{w})\| \leq \|D\varphi(w)\|\|w - \bar{w}\| = \|w - \bar{w}\|.$$

The following lemma will be useful in order to interpret the semismooth Newton method as an active set strategy for (MPCC). The proof of this result follows by a simple but laborious distinction of cases and is, thus, omitted.

Lemma 3.4. Let $w = (a, b, \mu, \nu)$ and $\delta w = (\delta a, \delta b, \delta\mu, \delta\nu)$ be given. Then we have the equivalence

$$D\varphi(w)\,\delta w = -\varphi(w) \quad \Leftrightarrow \quad \begin{cases} \delta b = -b,\ \delta\mu = -\mu & \text{if } D\varphi(w) \in J_{2,3}, \\ \delta a = -a,\ \delta\nu = -\nu & \text{if } D\varphi(w) \in J_{1,4}, \\ \delta a = -a,\ \delta b = -b & \text{if } D\varphi(w) \in J_{1,2}. \end{cases}$$

4 Application of a semismooth Newton method

Using the NCP-function $\pi_{\min}\colon \mathbb{R}^2 \to \mathbb{R}$ as well as the NMS-function $\varphi\colon \mathbb{R}^4 \to \mathbb{R}^2$ constructed in Section 3, we introduce $F\colon \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^{2p}$ via

$$F(x, \lambda, \eta, \mu, \nu) := \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda, \eta, \mu, \nu) \\ [\pi_{\min}(-g_i(x), \lambda_i)]_{I^\ell} \\ h(x) \\ [\varphi(G_i(x), H_i(x), \mu_i, \nu_i)]_{I^p} \end{bmatrix}. \tag{4.1}$$

Clearly, by Lemma 3.3, a point $x \in \mathbb{R}^n$ is M-stationary for (MPCC) if and only if there is a quadruple $(\lambda, \eta, \mu, \nu)$ such that $F(x, \lambda, \eta, \mu, \nu) = 0$ holds. In this case, it holds $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{M}}(x)$. Observing that all the data functions f, g, h, G, and H are twice continuously differentiable, Lemmas 2.11 and 3.3 guarantee that F is Newton differentiable on the set of its roots. Thus, we may apply the semismooth Newton method from Section 4 in order to find the roots of F, i.e., M-stationary points of (MPCC).

In order to guarantee convergence of the Newton method to an M-stationary point $x \in \mathbb{R}^n$ of (MPCC) with associated multiplier $(\lambda, \eta, \mu, \nu) \in \Lambda^{\mathrm{M}}(x)$, we have to guarantee that the Newton derivative of F is uniformly invertible in a neighborhood of $z := (x, \lambda, \eta, \mu, \nu)$. Abstractly, we have

$$DF(z) = \begin{bmatrix} \nabla^2_{xx}\mathcal{L}(z) & g'(x)^\top & h'(x)^\top & G'(x)^\top & H'(x)^\top \\ A_1(z) & A_2(z) & 0 & 0 & 0 \\ h'(x) & 0 & 0 & 0 & 0 \\ A_3(z) & 0 & 0 & A_4(z) & A_5(z) \end{bmatrix}$$

for the Newton derivative of F at z were we used

$$\begin{aligned}
A_1(z) &:= [-D_a\pi_{\min}(-g_i(x), \lambda_i)\nabla g_i(x)^\top]_{I^\ell}, \\
A_2(z) &:= [D_b\pi_{\min}(-g_i(x), \lambda_i)e_i^\top]_{I^\ell}, \\
A_3(z) &:= [D_a\varphi(G_i(x), H_i(x), \mu_i, \nu_i)\nabla G_i(x)^\top + D_b\varphi(G_i(x), H_i(x), \mu_i, \nu_i)\nabla H_i(x)^\top]_{I^p}, \\
A_4(z) &:= [D_\mu\varphi(G_i(x), H_i(x), \mu_i, \nu_i)e_i^\top]_{I^p}, \\
A_5(z) &:= [D_\nu\varphi(G_i(x), H_i(x), \mu_i, \nu_i)e_i^\top]_{I^p}.
\end{aligned}$$

Theorem 4.1. Let $\bar{x} \in \mathbb{R}^n$ be an M-stationary point of (MPCC) where MPCC-LICQ holds. Furthermore, assume that MPCC-SSOC holds at $\bar{x}$ w.r.t. $(\bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu}) \in \Lambda^{\mathrm{M}}(\bar{x})$. Set $\bar{z} := (\bar{x}, \bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$ and observe that this point solves $F(\bar{z}) = 0$. Then there exist $\varepsilon > 0$ and $C > 0$ such that $\|DF(z)^{-1}\| \le C$ for all $z \in B_\varepsilon(\bar{z})$.

Proof. First, we provide a result for a linear system associated with the solution $\bar{z}$. To this end, let matrices $P_i \in \{(1,0), (0,1)\}$, $i \in I^\ell$, $Q_j \in J_{1,2} \cup J_{2,3} \cup J_{1,4}$, $j \in I^p$, be given such that

$$\left.\begin{array}{rclrcl} g_i(\bar{x}) < 0 & \Rightarrow & P_i = (0,1), & \bar{\lambda}_i > 0 & \Rightarrow & P_i = (1,0), \\ G_j(\bar{x}) > 0 & \Rightarrow & Q_j \in J_{2,3}, & H_j(\bar{x}) > 0 & \Rightarrow & Q_j \in J_{1,4}, \\ \bar{\mu}_j \neq 0 & \Rightarrow & Q_j \in J_{1,2} \cup J_{1,4}, & \bar{\nu}_j \neq 0 & \Rightarrow & Q_j \in J_{1,2} \cup J_{2,3} \end{array}\right\} \tag{4.2}$$

holds for all $i \in I^\ell$, $j \in I^p$, cf. (3.5). Associated with this choice of P_i, Q_j, we define the index sets

$$\begin{aligned} &I_1^\ell := \{i \in I^\ell \mid P_i = (1,0)\}, \quad I_2^\ell := \{i \in I^\ell \mid P_i = (0,1)\}, \\ &I_{1,2}^p := \{j \in I^p \mid Q_j \in J_{1,2}\}, \quad I_{1,4}^p := \{j \in I^p \mid Q_j \in J_{1,4}\}, \quad I_{2,3}^p := \{j \in I^p \mid Q_j \in J_{2,3}\}. \end{aligned}$$

Now, we consider the linear system with unknowns $\delta z = (\delta x, \delta\lambda, \delta\eta, \delta\mu, \delta\nu)$

$$\nabla^2_{xx}\mathcal{L}(\bar{z})\,\delta x + g'(\bar{x})^\top\,\delta\lambda + h'(\bar{x})^\top\,\delta\eta + G'(\bar{x})^\top\,\delta\mu + H'(\bar{x})^\top\,\delta\nu = r \tag{4.3a}$$

$$P_i \begin{pmatrix} -\nabla g_i(\bar{x})^\top \delta x \\ \delta\lambda_i \end{pmatrix} = s_i \quad \forall i \in I^\ell \tag{4.3b}$$

$$h'(\bar{x})\,\delta x = t \tag{4.3c}$$

$$Q_j \begin{pmatrix} \nabla G_j(\bar{x})^\top \delta x \\ \nabla H_j(\bar{x})^\top \delta x \\ \delta\mu_j \\ \delta\nu_j \end{pmatrix} = \begin{pmatrix} u_j \\ v_j \end{pmatrix} \quad \forall j \in I^p. \tag{4.3d}$$

Let us inspect the block (4.3b). In case $i \in I_2^\ell$, this block is equivalent to $\delta\lambda_i = s_i$. Hence, we can eliminate these variables. Now, we consider the last block (4.3d). In case $j \in I_{1,2}^p$, i.e., $Q_j \in J_{1,2}$, we can assume w.l.o.g. that

$$Q_j = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Hence, the jth component of the last block is equivalent to

$$\nabla G_j(\bar{x})^\top \delta x = u_j \quad \text{and} \quad \nabla H_j(\bar{x})^\top \delta x = v_j.$$

For $j \in I_{1,4}^p$, the last block is w.l.o.g. equivalent to

$$\nabla G_j(\bar{x})^\top \delta x = u_j \quad \text{and} \quad \delta\nu_j = v_j$$

and for $j \in I_{2,3}^p$ we get

$$\nabla H_j(\bar{x})^\top \delta x = v_j \quad \text{and} \quad \delta\mu_j = u_j.$$

Thus, the values $\delta\nu_j$ for $j \in I^p_{1,4}$ and $\delta\mu_j$ for $j \in I^p_{2,3}$ can be eliminated in the above system. With the index sets

$$I^p_\mu := I^p \setminus I^p_{2,3}, \qquad I^p_\nu := I^p \setminus I^p_{1,4}$$

we arrive at the reduced saddle-point system

$$\begin{aligned}
\nabla^2_{xx}\mathcal{L}(\bar z)\,\delta x + g'(\bar x)^\top_{I^\ell_1}\,\delta\lambda_{I^\ell_1} + h'(\bar x)^\top\,\delta\eta + G'(\bar x)^\top_{I^p_\mu}\,\delta\mu_{I^p_\mu} + H'(\bar x)^\top_{I^p_\nu}\,\delta\nu_{I^p_\nu} &= \tilde r\\
-g'(\bar x)_{I^\ell_1}\,\delta x &= s_{I^\ell_1}\\
h'(\bar x)\,\delta x &= t\\
G'(\bar x)_{I^p_\mu}\delta x &= u_{I^p_\mu}\\
H'(\bar x)_{I^p_\nu}\delta x &= u_{I^p_\nu}.
\end{aligned}$$

Therein, the modified right-hand side $\tilde r$ results from the elimination of some of the multipliers. It can be bounded by r, s, u, and v. Note that this reduced system is symmetric. Furthermore, it clearly holds

$$\{i \in I^g(\bar x) \,|\, \bar\lambda_i > 0\} \subset I^\ell_1 \subset I^g(\bar x) \tag{4.4a}$$
$$I^{0+}(\bar x) \cup I^{00}_{\pm\mathbb{R}}(\bar x, \bar\mu, \bar\nu) \subset I^p_\mu \subset I^{0+}(\bar x) \cup I^{00}(\bar x) \tag{4.4b}$$
$$I^{+0}(\bar x) \cup \; I^{00}_{\mathbb{R}\pm}(\bar x, \bar\mu, \bar\nu) \subset I^p_\nu \subset I^{+0}(\bar x) \cup I^{00}(\bar x) \tag{4.4c}$$

by definition of these index sets. Additionally, we have $I^p_\mu \cup I^p_\nu = I^p$ due to Lemma 3.2. By MPCC-LICQ, the constraint block of the reduced system is surjective and from MPCC-SSOC we get that the matrix $\nabla^2_{xx}\mathcal{L}(\bar z)$ is positive definite on the kernel of the constraint block. Now, Lemma 2.2 implies the invertibility of the system. By undoing the elimination of some of the multipliers, we find that the system (4.3) is invertible, i.e., there is a constant $c > 0$, such that the unique solution δz of (4.3) satisfies $\|\delta z\| \le c\,(\|r\| + \|s\| + \|t\| + \|u\| + \|v\|)$. Since there are only finitely many choices for the matrices P_i and Q_j, the constant c does not depend on the precise values of P_i and Q_j.

Now, we prove the uniform invertibility of the Newton matrix $DF(z)$ for all z in a neighborhood of $\bar z$. First, we can utilize Lemma 3.2 and the continuity of g, G, and H to obtain $\varepsilon > 0$ such that $P_i := D\pi_{\min}(-g_i(x), \lambda_i)$ and $Q_j := D\varphi(G_j(x), H_j(x), \mu_j, \nu_j)$ satisfy (4.2) for all $i \in I^\ell$ and $j \in I^p$. Note that we still use $\bar x$, $\bar\lambda$, $\bar\mu$, and $\bar\nu$ in (4.2). Thus, the Newton matrix $DF(z)$ is a perturbation of the system matrix from (4.3). Since f, g, h, G, and H are assumed to be twice continuously differentiable, the perturbation can be made arbitrarily small (by reducing ε if necessary). Thus, Lemma 2.3 ensures that we get a uniform bound for $DF(z)^{-1}$ for all $z \in B_\varepsilon(\bar z)$.

Now, we are in position to state a local convergence result for our nonsmooth Newton method based on the map F from (4.1). Its proof simply follows combining Theorems 2.9 and 4.1.

Theorem 4.2. Let $\bar{x} \in \mathbb{R}^n$ be an M-stationary point of (MPCC) where MPCC-LICQ holds. Furthermore, assume that MPCC-SSOC holds at $\bar{x}$ w.r.t. $(\bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu}) \in \Lambda^{\mathrm{M}}(\bar{x})$. Set $\bar{z} := (\bar{x}, \bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$. Then there exists $\delta > 0$ such that the nonsmooth Newton method applied to the mapping F from (4.1) is well defined and converges superlinearly to $\bar{z}$ for each initial iterate from $B_\delta(\bar{z})$. If, additionally, the second-order derivatives of the data functions f, g, h, G, and H are locally Lipschitz continous, then the convergence is quadratic.

Remark 4.3. In addition to the assumptions of Theorem 4.2, assume that the cost function f is quadratic while the constraint mappings g, h, G, and H are affine in (MPCC). Then Example 2.8, Lemma 2.11, and Lemma 3.3 guarantee that the mapping F from (4.1) is Newton differentiable of order ∞ on M. Thus, Theorem 2.9 guarantees one-step convergence for the associated nonsmooth Newton method if the initial iterate is sufficiently close to the reference point $\bar{z}$.

We note that the example from Fletcher et al., 2006, Section 7.3 satisfies our assumptions MPCC-LICQ and MPCC-SSOC at the origin which is an S-stationary point of the underlying complementarity-constrained optimization problem. Due to Theorem 4.2, local superlinear convergence of our nonsmooth Newton method is guaranteed. On the other hand, the SQP-method suggested in Fletcher et al., 2006 only converges linearly to the point of interest.

We mention that it is possible to interpret the Newton method as an active set strategy. To this end, one has to utilize Lemma 3.4. If the current iterate is denoted by $z_k = (x_k, \lambda_k, \eta_k, \mu_k, \nu_k)$, the next iterate $z_{k+1} = (x_{k+1}, \lambda_{k+1}, \eta_{k+1}, \mu_{k+1}, \nu_{k+1})$ solves the symmetric linear system

$$
\begin{aligned}
\nabla_x \mathcal{L}(z_k) &+ \nabla^2_{xx}\mathcal{L}(z_k)\,(x_{k+1} - x_k) \\
&+ g'(x_k)^\top_{I^\ell_1}\,(\lambda_{k+1} - \lambda_k)_{I^\ell_1} + h'(x_k)^\top\,(\eta_{k+1} - \eta_k) \\
&+ G'(x_k)^\top_{I^p_\mu}\,(\mu_{k+1} - \mu_k)_{I^p_\mu} + H'(x_k)^\top_{I^p_\nu}\,(\nu_{k+1} - \nu_k)_{I^p_\nu} = 0, \\
&\qquad g(x_k)_{I^\ell_1} + g'(x_k)_{I^\ell_1}\,(x_{k+1} - x_k)_{I^\ell_1} = 0, \qquad (\lambda_{k+1})_{I^\ell \setminus I^\ell_1} = 0, \\
&\qquad h(x_k) + h'(x_k)\,(x_{k+1} - x_k) = 0, \\
&\qquad G(x_k)_{I^p_\mu} + G'(x_k)_{I^p_\mu}\,(x_{k+1} - x_k)_{I^p_\mu} = 0, \qquad (\mu_{k+1})_{I^p \setminus I^p_\mu} = 0, \\
&\qquad H(x_k)_{I^p_\nu} + H'(x_k)_{I^p_\nu}\,(x_{k+1} - x_k)_{I^p_\nu} = 0, \qquad (\nu_{k+1})_{I^p \setminus I^p_\nu} = 0.
\end{aligned}
$$

Here, the index sets I^ℓ_1, I^p_μ, and I^p_ν are constructed similarly as in the proof of Theorem 4.1.

Let us briefly compare our algorithm with Izmailov, Solodov, 2008, Alg. 2.2. Therein, the authors use an identification procedure to obtain the active sets $I^{+0}(\bar{x})$, $I^{0+}(\bar{x})$, and $I^{00}(\bar{x})$ and, afterwards, $\bar{x}$ is approximated by an active set strategy. This approach is very similar to our suggestion. For the convergence theory, however, they presume validity of

MPCC-LICQ and MPCC-SOSC at a given local minimizer of (MPCC) which, thus, is an S-stationary point (observe that MPCC-SOSC is called *piecewise SOSC* in Izmailov, Solodov, 2008, and take notice of Izmailov, Solodov, 2008, pages 1006-1007). Recall that MPCC-SOSC is slightly weaker than MPCC-SSOC, which is required in our Theorem 4.2. However, the algorithm from Izmailov, Solodov, 2008 is designed for the computation of S-stationary points and cannot approximate M-stationary points (which are not already strongly stationary).

5 Globalization

A possible idea for the globalization of the nonsmooth Newton method from Section 2.3 is to exploit the squared residual of F as a merit function. Unfortunately, it can be easily checked that the resulting map $z \mapsto \frac{1}{2}\|F(z)\|^2$ is not smooth. Exploiting the well-known fact that the square of the Fischer–Burmeister function π_{FB} is smooth, see, e.g., Facchinei, Soares, 1997, Proposition 3.4, we are, however, in position to construct a smooth merit function. Therefore, let us first mention that the function φ_1 has the equivalent representation

$$\theta_1(a,b,\mu,\nu) := |\min(a,b)|, \tag{5.1a}$$
$$\theta_2(a,b,\mu,\nu) := \min(|a|,|\mu|), \tag{5.1b}$$
$$\theta_3(a,b,\mu,\nu) := \min(|b|,|\nu|), \tag{5.1c}$$
$$\theta_4(a,b,\mu,\nu) := \max(0,\min(\mu,|\nu|),\min(\nu,|\mu|)), \tag{5.1d}$$
$$\varphi_1(a,b,\mu,\nu) = \max_{i=1,\ldots,4} \theta_i(a,b,\mu,\nu). \tag{5.1e}$$

Indeed, one can check that the zeros of the function $\theta(a,b,\mu,\nu) := \max_{i=1,\ldots,4}\theta_i(a,b,\mu,\nu)$ coincide with the set M and, by construction, θ is 1-Lipschitz continuous w.r.t. the ℓ^∞-norm. Based on these observations and an elementary distinction of cases, it is now possible to exploit the particular structure of θ in order to show that this function equals the ℓ^∞-distance to M. Noting that φ_1 from (3.2d) has the same property, θ and φ_1 actually need to coincide. This motivates the definition of

$$\theta_{1,\mathrm{FB}}(a,b,\mu,\nu) := |\pi_{\mathrm{FB}}(a,b)|, \tag{5.2a}$$
$$\theta_{2,\mathrm{FB}}(a,b,\mu,\nu) := \pi_{\mathrm{FB}}(|a|,|\mu|), \tag{5.2b}$$
$$\theta_{3,\mathrm{FB}}(a,b,\mu,\nu) := \pi_{\mathrm{FB}}(|b|,|\nu|), \tag{5.2c}$$
$$\theta_{4,\mathrm{FB}}(a,b,\mu,\nu) := \begin{cases} 0 & \text{if } \mu,\nu \le 0, \\ \pi_{\mathrm{FB}}(|\mu|,|\nu|) & \text{else,} \end{cases} \tag{5.2d}$$
$$\theta_{\mathrm{FB}}(a,b,\mu,\nu) := [\theta_{i,\mathrm{FB}}(a,b,\mu,\nu)]_{i=1,\ldots,4}. \tag{5.2e}$$

Now, we introduce the modified residual $F_{\mathrm{FB}}\colon \mathbb{R}^n\times\mathbb{R}^\ell\times\mathbb{R}^m\times\mathbb{R}^p\times\mathbb{R}^p \to \mathbb{R}^n\times\mathbb{R}^\ell\times\mathbb{R}^m\times\mathbb{R}^{4p}$ as stated below for arbitrary $x\in\mathbb{R}^n$, $\lambda\in\mathbb{R}^\ell$, $\eta\in\mathbb{R}^m$, and $\mu,\nu\in\mathbb{R}^p$:

$$F_{\mathrm{FB}}(x,\lambda,\eta,\mu,\nu) := \begin{bmatrix} \nabla_x\mathcal{L}(x,\lambda,\eta,\mu,\nu) \\ [\pi_{\mathrm{FB}}(-g_i(x),\lambda_i)]_{I^\ell} \\ h(x) \\ [\theta_{\mathrm{FB}}(G_i(x),H_i(x),\mu_i,\nu_i)]_{I^p} \end{bmatrix}. \tag{5.3}$$

In the next lemma, we show that the squared residuals of F and F_{FB} are, in some sense, equivalent.

Lemma 5.1. There exist constants $c, C > 0$ with

$$c\,\|F_{\mathrm{FB}}(x,\lambda,\eta,\mu,\nu)\|^2 \le \|F(x,\lambda,\eta,\mu,\nu)\|^2 \le C\,\|F_{\mathrm{FB}}(x,\lambda,\eta,\mu,\nu)\|^2$$

for all $x\in\mathbb{R}^n$, $\lambda\in\mathbb{R}^\ell$, $\eta\in\mathbb{R}^m$, and $\mu,\nu\in\mathbb{R}^p$.

Proof. Throughout the proof, $w := (a,b,\mu,\nu)\in\mathbb{R}^4$ is arbitrarily chosen. Due to Tseng, 1996, Lemma 3.1, the functions $\pi_{\min}$ and π_{FB} are equivalent in the sense

$$\frac{2}{2+\sqrt{2}}\,|\pi_{\min}(a,b)| \le |\pi_{\mathrm{FB}}(a,b)| \le (2+\sqrt{2})\,|\pi_{\min}(a,b)|. \tag{5.4}$$

Thus, keeping in mind the definitions of F and F_{FB} from (4.1) and (5.3), we only need to show the equivalence of φ and θ_{FB}.

The relation (5.4) yields

$$\frac{2}{2+\sqrt{2}}\,|\theta_i(w)| \le |\theta_{i,\mathrm{FB}}(w)| \le (2+\sqrt{2})\,|\theta_i(w)|$$

for $i = 1,\ldots,4$. Together with the estimate

$$\varphi_1(w) = \max_{i=1,\ldots,4}\theta_i(w) \le \left(\sum_{i=1}^4 \theta_i^2(w)\right)^{1/2} \le 2\,\varphi_1(w),$$

we get equivalence of the functions $|\varphi_1|$ and $\|\theta_{\mathrm{FB}}\|$.

In order to complete the proof, we only need to show

$$|\varphi_2(w)| \le |\varphi_1(w)| \tag{5.5}$$

for all w, since this already yields the equivalence of φ and θ_{FB}. Let us distinguish some cases. If we have $D\varphi_1(w)\in\{\pm e_1^\top, \pm e_2^\top\}$, then it clearly holds

$$|\varphi_2(w)| \le \max(\theta_2(w),\theta_3(w)) \le \max_{i=1,\ldots,4}\theta_i(w) = \varphi_1(w).$$

Now, suppose that $D\varphi_1(w) = \pm e_3^\top$ holds. Since the Newton derivative of φ_1 is evaluated based on the representation of φ_1 given in (3.2d), we obtain $\varphi_1(w) = \min(\psi_1(w), \psi_3(w)) = |\mu|$. This implies

$$\varphi_2(w) = |b| \leq \min(\psi_1(w), \psi_3(w)) = \varphi_1(w).$$

The case $D\varphi_1(w) = \pm e_4^\top$ can be handled analogously. This shows (5.5) for arbitrary w and the proof is complete.

For the globalization of our nonsmooth Newton method, we make use of the merit function $\Phi_{\mathrm{FB}} \colon \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ given by

$$\Phi_{\mathrm{FB}}(z) := \frac{1}{2} \|F_{\mathrm{FB}}(z)\|^2$$

for all $z = (x, \lambda, \eta, \mu, \nu) \in \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^p$. This function is continuously differentiable: First, recall that the square of the function π_{FB} is continuously differentiable. The gradient of $(a, b) \mapsto \pi_{\mathrm{FB}}(a, b)^2$ vanishes on the complementarity angle $\{(a, b) \mid 0 \leq a \perp b \geq 0\}$. This implies the continuous differentiability of $(a, b) \mapsto \pi_{\mathrm{FB}}(|a|, |b|)^2$. Similar arguments can be used to check the continuous differentiability of the function $\theta_{4,\mathrm{FB}}^2$.

Now, we can utilize the globalization idea from De Luca, Facchinei, Kanzow, 2000, Section 3 and Izmailov, Solodov, 2008, Algorithm 3.2: If the Newton step d_k can be computed and satisfies

$$\frac{\Phi_{\mathrm{FB}}(z_k + d_k)}{\Phi_{\mathrm{FB}}(z_k)} \leq q \tag{5.6}$$

(with a fixed parameter $q \in (0, 1)$), we perform the Newton step $z_{k+1} = z_k + d_k$. If the Newton system is not solvable or if its solution d_k violates the angle test, we instead use $d_k := -\nabla\Phi_{\mathrm{FB}}(z_k)$. Afterwards, we use an Armijo line search to obtain the step length α_k and update the iterate via $z_{k+1} = z_k + \alpha_k\, d_k$. This globalization strategy is described in Algorithm 1.

Due to Lemmas 2.10 and 5.1, the ratio test (5.6) is satisfied (and, consequently, the Newton steps are performed) for all z_k in the neighborhood of solutions satisfying the assumptions of Theorem 4.1. Consequently, the convergence guarantees of Algorithm 1 follow along the lines of De Luca, Facchinei, Kanzow, 2000, Section 3, Izmailov, Solodov, 2008, Thms 3.4, 3.5. In particular, we obtain:

- All accumulation points of the iterates z_k are stationary points of Φ_{FB}.
- If (5.6) is satisfied infinitely often, then $\Phi_{\mathrm{FB}}(z_k) \to 0$. In this case, any accumulation point is M-stationary.
- If an accumulation point $\bar{z}$ of z_k satisfies the assumptions of Theorem 4.1, then the entire sequence converges superlinearly/quadratically towards $\bar{z}$.

Data: parameters $q, \tau_{\text{abs}}, \rho \in (0,1)$, starting point $z_0 \in \mathbb{R}^{n+l+m+2p}$
Set $k = 0$;
while $\|F(z_k)\| > \tau_{abs}$ **do**
  Solve $DF(z_k)d_k = -F(z_k)$ for d_k;
  if *ratio test* (5.6) *is satisfied* **then**
    Set $z_{k+1} = z_k + d_k$;
  else
    if $\nabla\Phi_{\text{FB}}(z_k)^\top d_k > -\rho\|d_k\|\|\nabla\Phi_{\text{FB}}(z_k)\|$ **then**
      Set $d_k = -\nabla\Phi_{\text{FB}}(z_k)$;
    end
    Determine $z_{k+1} = z_k + \alpha_k d_k$ using an Armijo line search for Φ_{FB};
  end
  Set $k = k + 1$;
end

Algorithm 1: Globalization of the semismooth Newton method.

6 Convergence for linear-quadratic problems beyond MPCC-LICQ

We consider the linear-quadratic case, i.e., we assume that the function f is quadratic and that the mappings g, h, G, and H are affine. Due to the complementarity constraints, the solution of (MPCC) is still very challenging. On the other hand, it follows from Flegel, Kanzow, 2005b, Theorem 3.5, Proposition 3.8 that local minimizers of the associated problem (MPCC) are M-stationary without further assumptions. This makes the search for M-stationary points even more attractive. Our goal is to verify that one-step convergence of a modification of our Newton method is possible under a weaker constraint qualification than MPCC-LICQ.

Let an M-stationary point $\bar{x} \in \mathbb{R}^n$ of (MPCC) with multiplier $(\bar{\lambda}, \bar{\rho}, \bar{\mu}, \bar{\nu}) \in \Lambda^{\text{M}}(\bar{x})$ be given and set $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$. We require that the matrix

$$\begin{bmatrix} g'(\bar{x})_{I^g_+(\bar{x},\bar{\lambda})} \\ h'(\bar{x}) \\ G'(\bar{x})_{I^{0+}(\bar{x}) \cup I^{00}_{\pm\mathbb{R}}(\bar{x},\bar{\mu},\bar{\nu})} \\ H'(\bar{x})_{I^{+0}(\bar{x}) \cup I^{00}_{\mathbb{R}\pm}(\bar{x},\bar{\mu},\bar{\nu})} \end{bmatrix} \tag{6.1}$$

possesses full row rank, where we used the multiplier-dependent index sets from (2.3) and

$$I^g_+(\bar{x}, \bar{\lambda}) := \{i \in I^g(\bar{x}) \mid \bar{\lambda}_i > 0\}.$$

Clearly, this condition is, in general, weaker than MPCC-LICQ. Further, we assume that MPCC-SSOC holds at $\bar{x}$ w.r.t. $(\bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$.

Let $(x, \lambda, \eta, \mu, \nu)$ denote the current iterate. We will assume that it is close to the solution $(\bar{x}, \bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$. Then arguing as in the proof of Theorem 4.1, using the active-set interpretation from the end of Section 4 as well as the linear-quadratic structure of the problem, the next iterate $(x^+, \lambda^+, \eta^+, \mu^+, \nu^+)$ is given by the solution of the linear system

$$\nabla_x \mathcal{L}(x^+, \lambda^+, \eta^+, \mu^+, \nu^+) = 0, \tag{6.2a}$$
$$g(x^+)_{I_1^\ell} = 0, \qquad (\lambda^+)_{I^\ell \setminus I_1^\ell} = 0, \tag{6.2b}$$
$$h(x^+) = 0, \tag{6.2c}$$
$$G(x^+)_{I_\mu^p} = 0, \qquad (\mu^+)_{I^p \setminus I_\mu^p} = 0, \tag{6.2d}$$
$$H(x^+)_{I_\nu^p} = 0, \qquad (\nu^+)_{I^p \setminus I_\nu^p} = 0. \tag{6.2e}$$

Here, the index sets I_1^ℓ, I_μ^p, and I_ν^p are constructed similarly as in the proof of Theorem 4.1.

Using the estimates (4.4) for the index sets, one can check that $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\eta}, \bar{\mu}, \bar{\nu})$ solves the above system. In particular, we have the implications

$$\text{the sets } I_1^\ell,\ I_\mu^p,\ I_\nu^p \text{ satisfy (4.4)} \;\Rightarrow\; \bar{z} \text{ solves (6.2)}, \tag{6.3a}$$
$$\left.\begin{aligned} I_+^g(\bar{x}, \bar{\lambda}) &= I_1^\ell \\ I^{0+}(\bar{x}) \cup I_{\pm\mathbb{R}}^{00}(\bar{x}, \bar{\mu}, \bar{\nu}) &= I_\mu^p \\ I^{+0}(\bar{x}) \cup I_{\mathbb{R}\pm}^{00}(\bar{x}, \bar{\mu}, \bar{\nu}) &= I_\nu^p \end{aligned}\right\} \;\Rightarrow\; \text{(6.2) is uniquely solvable.} \tag{6.3b}$$

Note that (6.3b) follows from the assumption that (6.1) has full row rank. However, system (6.2) is, in general, not uniquely solvable. If the system has multiple solutions, Lemma 2.2 and MPCC-SSOC imply that the matrix

$$\begin{bmatrix} g'(\bar{x})_{I_1^\ell} \\ h'(\bar{x}) \\ G'(\bar{x})_{I_\mu^p} \\ H'(\bar{x})_{I_\nu^p} \end{bmatrix} \tag{6.4}$$

does not possess full row rank. Note that this matrix might possess more rows than (6.1).

We will see that it is possible to remove one index from one of the index sets I_1^ℓ, I_μ^p, I_ν^p such that the inclusions (4.4) are still satisfied, as long as (6.2) is not uniquely solvable. Thus, (6.3) will imply that we can find the solution $\bar{z}$ using this strategy repeatedly.

We can filter out the linear dependent rows from the knowledge of our current iterate. The possible linear dependent rows in (6.4) correspond to the index sets

$$I_1^\ell \setminus I_+^g(\bar{x}, \bar{\lambda}), \qquad I_\mu^p \setminus \big(I^{0+}(\bar{x}) \cup I_{\pm\mathbb{R}}^{00}(\bar{x}, \bar{\mu}, \bar{\nu})\big), \qquad I_\nu^p \setminus \big(I^{+0}(\bar{x}) \cup I_{\mathbb{R}\pm}^{00}(\bar{x}, \bar{\mu}, \bar{\nu})\big). \tag{6.5}$$

Using (4.4), these indices are contained in

$$\begin{aligned} I^g_0(\bar x, \bar\lambda) &:= I^g(\bar x) \setminus I^g_+(\bar x, \bar\lambda), \\ I^{00}_{0\mathbb{R}}(\bar x, \bar\mu, \bar\nu) &:= I^{00}(\bar x) \setminus I^{00}_{\pm\mathbb{R}}(\bar x, \bar\mu, \bar\nu), \\ I^{00}_{\mathbb{R}0}(\bar x, \bar\mu, \bar\nu) &:= I^{00}(\bar x) \setminus I^{00}_{\mathbb{R}\pm}(\bar x, \bar\mu, \bar\nu), \end{aligned}$$

respectively. Now, we use the following procedure: First, we sort the indices

$$\begin{aligned} &i \in I^\ell_1 && \text{according to } \lambda_i \\ &j \in I^p_\mu && \text{according to } \max(|\mu_j|, |H_j(x)|) \\ &j \in I^p_\nu && \text{according to } \max(|\nu_j|, |G_j(x)|) \end{aligned}$$

in increasing order. If the current iterate is sufficiently close to the solution, the above list will contain the problematic indices from (6.5) at the top. Then we can remove the indices one-by-one from the corresponding index sets, until the system (6.2) becomes uniquely solvable. Note that this modification of the index sets ensures that (4.4) stays valid. Hence, the solution $\bar z$ remains to be a solution of (6.2), cf. (6.3a). Since the matrix (6.1) has full row rank, this process stops if all indices from (6.5) are removed (or earlier).

Example 6.1. We consider the classical example Scheel, Scholtes, 2000, Example 3 with a quadratic regularization term. That is, we have $n = 3$, $\ell = 2$, $m = 0$, as well as $p = 1$, and the functions are given by

$$f(x) = x_1 + x_2 - x_3 + \frac{c}{2}\,\|x\|^2,\; g(x) = (-4\,x_1 + x_3, -4\,x_2 + x_3)^\top,\; G(x) = x_1,\; H(x) = x_2,$$

where $c > 0$ is the regularization parameter. The global minimizer is $\bar x = 0$, and this point is M-stationary with multipliers $\bar\lambda = (3/4, 1/4)^\top$, $\bar\mu = 2$, $\bar\nu = 0$. An alternate set of multipliers is $\tilde\lambda = (1/4, 3/4)^\top$, $\tilde\mu = 0$, $\tilde\nu = 2$. Since all four constraints are active in $\bar x$, MPCC-LICQ cannot be satisfied. However, the matrix (6.1) is given by

$$\begin{pmatrix} g'(\bar x)_{\{1,2\}} \\ G'(\bar x) \end{pmatrix} = \begin{pmatrix} -4 & 0 & 1 \\ 0 & -4 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and this matrix possesses full row rank. Since $\nabla^2 f(\bar x)$ is positive definite, MPCC-SSOC holds as well. Hence, the above theory applies and we obtain one-step convergence if the initial guess is sufficiently close to $(\bar x, \bar\lambda, \bar\mu, \bar\nu)$, see Remark 4.3 as well.

The application of this idea to problems which are not linear-quadratic is subject to future research. We expect that the above idea can be generalized easily to problems with affine constraints. In this situation, every local minimizer is M-stationary. Hence, it is suitable to solve such problems with an algorithm capable to find M-stationary points.

7 Application to the control of a discretized obstacle problem

We consider a discretized version of the optimization problem from Wachsmuth, 2014, Section 6.1. This is an infinite dimensional MPCC for which strong stationarity does not hold at the uniquely determined minimizer. We will see that the same property holds for its discretization.

Let us fix a discretization parameter $N \in \mathbb{N}$. The optimization variable $x \in \mathbb{R}^{3N}$ is partitioned as $x = (y, u, \xi)$. The discretized problem uses the data functions

$$f(x) = \tfrac{1}{2}\|y\|^2 + e^\top y + \tfrac{1}{2}\|u\|^2, \qquad g(x) = -u, \qquad h(x) = A\,y - u + \xi,$$
$$G(x) = -y, \qquad H(x) = \xi,$$

where $e := (1, \ldots, 1) \in \mathbb{R}^N$ is the all-ones vector and the matrix A is given entrywise:

$$\forall i, j \in \{1, \ldots, N\}\colon \quad A_{i,j} := \begin{cases} 2 & \text{if } i = j, \\ -1 & \text{if } i = j \pm 1, \\ 0 & \text{else.} \end{cases}$$

Up to the scaling, this matrix arises from the finite-difference discretization of the one-dimensional Laplacian.

One can check that $\bar{x} = 0$ is the unique global minimizer. Since all constraints are affine, $\bar{x}$ is an M-stationary point of this program. Furthermore, there are no weakly stationary points (i.e., feasible points satisfying (2.1a)–(2.1d)) different from $\bar{x}$. The M-stationary multipliers $(\lambda, \eta, \mu, \nu)$ associated with $\bar{x}$ are not unique. Indeed, for every diagonal matrix $D \in \mathbb{R}^{N\times N}$ with diagonal entries from $\{0, 1\}$, one can check that the solution of the system

$$\begin{pmatrix} A & I \\ I - D & D \end{pmatrix} \begin{pmatrix} \nu \\ \mu \end{pmatrix} = \begin{pmatrix} e \\ 0 \end{pmatrix}, \; \nu = -\eta = \lambda$$

yields M-stationary multipliers, and all multipliers are obtained by this construction. Let us mention that none of these multipliers solves the associated system of S-stationarity since for each $i \in \mathbb{N}$, either μ_i or ν_i is positive. Thus, $\bar{x}$ is not an S-stationary point. For all these multipliers, the matrix (6.1) possesses full row rank. On the other hand, MPCC-LICQ is clearly violated at $\bar{x}$ since this point is a local minimizer of the given MPCC which is not S-stationary. Finally, one can check that MPCC-SSOC is valid at $\bar{x}$ w.r.t. all the multipliers characterized above.

Numerical experiments on this discretized obstacle problem were carried out using MATLAB (R2019b). Therefore, we implemented Algorithm 1. For the parameters, we chose $q = 0.999$, $\tau_{\text{abs}} = 10^{-11}$, $\rho = 10^{-3}$, and $N = 256$. We ran Algorithm 1 1000 times with starting points $z_0 = (x_0, \lambda_0, \eta_0, \mu_0, \nu_0) \in \mathbb{R}^{n+l+m+2p} = \mathbb{R}^{7N}$ which were chosen from a uniform distribution on $[-1000, 1000]^{7N}$. In each run, the solution $\bar{x}$ has been found, while the computed associated multipliers differ from run to run. That behavior, however, had to be expected since we know from above that the associated set $\Lambda^{\mathrm{M}}(\bar{x})$ is not a singleton. The average number of iterations across all runs was 13.7. We could also observe the local one-step convergence that we expected from Remark 4.3.

8 Conclusion and outlook

We demonstrated that the M-stationarity system of a mathematical problem with complementarity constraints can be reformulated as a system of nonsmooth and even discontinuous equations, see Section 3. It has been shown that this system can be solved with the aid of a semismooth Newton method. Local fast convergence to M-stationary points can be guaranteed under validity of MPCC-SSOC and MPCC-LICQ, see Theorem 4.2, where the latter assumption can be weakened in case of linear-quadratic problems, see Section 6. Furthermore, we provided a reasonable globalization strategy, see Section 5. There is some hope that similar to Section 6, it is possible to weaken MPCC-LICQ in the setting of nonlinear constraints if only validity of a suitable constant rank assumption can be guaranteed. Clearly, the fundamental ideas of this paper are not limited to mathematical programs with complementarity constraints but can be adjusted in order to suit other problem classes from disjunctive programming such as vanishing-constrained, switching-constrained, or cardinality-constrained programs. It remains an open question to what extent the theory of this paper can be generalized to infinite-dimensional complementarity-constrained optimization problems.

Acknowledgment

This work is supported by the DFG Grant *Bilevel Optimal Control: Theory, Algorithms, and Applications* (Grant No. WA 3636/4-2) within the Priority Program SPP 1962 (Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization).

References

Benko, M.; Gfrerer, H. (2016). “An SQP method for mathematical programs with complementarity constraints with strong convergence properties”. *Kybernetika* 52.2, pp. 169–208. DOI: 10.14736/kyb-2016-2-0169.

Chen, X.; Nashed, Z.; Qi, L. (2000). “Smoothing Methods and Semismooth Methods for Nondifferentiable Operator Equations”. *SIAM Journal on Numerical Analysis* 38.4, pp. 1200–1216. DOI: 10.1137/s0036142999356719.

Clason, C. (2018). *Nonsmooth Analysis and Optimization.* arXiv: 1708.04180v2.

De Luca, T.; Facchinei, F.; Kanzow, C. (1996). "A semismooth equation approach to the solution of nonlinear complementarity problems". *Mathematical Programming* 75.3, pp. 407–439. DOI: 10.1007/BF02592192.

De Luca, T.; Facchinei, F.; Kanzow, C. (2000). "A Theoretical and Numerical Comparison of Some Semismooth Algorithms for Complementarity Problems". *Computational Optimization and Applications* 16.2, pp. 173–205. DOI: 10.1023/A:1008705425484.

Facchinei, F.; Soares, J. (1997). "A New Merit Function For Nonlinear Complementarity Problems And A Related Algorithm". *SIAM Journal on Optimization* 7.1, pp. 225–247. DOI: 10.1137/S1052623494279110.

Flegel, M. L.; Kanzow, C. (2005a). "On the Guignard constraint qualification for mathematical programs with equilibrium constraints". *Optimization* 54.6, pp. 517–534. DOI: 10.1080/02331930500342591.

Flegel, M. L.; Kanzow, C. (2005b). "On M-stationary points for mathematical programs with equilibrium constraints". *Journal of Mathematical Analysis and Applications* 310.1, pp. 286–302. DOI: 10.1016/j.jmaa.2005.02.011.

Fletcher, R.; Leyffer, S.; Ralph, D.; Scholtes, S. (2006). "Local Convergence of SQP Methods for Mathematical Programs with Equilibrium Constraints". *SIAM Journal on Optimization* 17.1, pp. 259–286. DOI: 10.1137/S1052623402407382.

Fukushima, M.; Tseng, P. (2002). "An Implementable Active-Set Algorithm for Computing a B-Stationary Point of a Mathematical Program with Linear Complementarity Constraints". *SIAM Journal on Optimization* 12.3, pp. 724–739. DOI: 10.1137/S1052623499363232.

Galántai, A. (2012). "Properties and construction of NCP functions". *Computational Optimization and Applications* 52.3, pp. 805–824. DOI: 10.1007/s10589-011-9428-9.

Gfrerer, H. (2014). "Optimality Conditions for Disjunctive Programs Based on Generalized Differentiation with Application to Mathematical Programs with Equilibrium Constraints". *SIAM Journal on Optimization* 24.2, pp. 898–931. DOI: 10.1137/130914449.

Guo, L.; Lin, G.-H.; Ye, J. J. (2013). "Second-Order Optimality Conditions for Mathematical Programs with Equilibrium Constraints". *Journal of Optimization Theory and Applications* 158.1, pp. 33–64. DOI: 10.1007/s10957-012-0228-x.

Guo, L.; Lin, G.-H.; Ye, J. J. (2015). "Solving Mathematical Programs with Equilibrium Constraints". *Journal of Optimization Theory and Applications* 166.1, pp. 234–256. DOI: 10.1007/s10957-014-0699-z.

Hintermüller, M.; Ito, K.; Kunisch, K. (2002). "The Primal-Dual Active Set Strategy as a Semismooth Newton Method". *SIAM Journal on Optimization* 13.3, pp. 865–888. DOI: 10.1137/s1052623401383558.

Hoheisel, T.; Kanzow, C.; Schwartz, A. (2013). "Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints". *Mathematical Programming* 137.1, pp. 257–288. DOI: 10.1007/s10107-011-0488-5.

Hu, X. M.; Ralph, D. (2004). "Convergence of a Penalty Method for Mathematical Programming with Complementarity Constraints". *Journal of Optimization Theory and Applications* 123.2, pp. 365–390. DOI: 10.1007/s10957-004-5154-0.

Huang, X. X.; Yang, X. Q.; Zhu, D. L. (2006). "A Sequential Smooth Penalization Approach to Mathematical Programs with Complementarity Constraints". *Numerical Functional Analysis and Optimization* 27.1, pp. 71–98. DOI: 10.1080/01630560500538797.

Ito, K.; Kunisch, K. (2008). *Lagrange Multiplier Approach to Variational Problems and Applications*. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9780898718614.

Izmailov, A. F.; Pogosyan, A. L.; Solodov, M. V. (2012). "Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints". *Computational Optimization and Applications* 51.1, pp. 199–221. DOI: 10.1007/s10589-010-9341-7.

Izmailov, A. F.; Solodov, M. V. (2008). "An Active-Set Newton Method for Mathematical Programs with Complementarity Constraints". *SIAM Journal on Optimization* 19.3, pp. 1003–1027. DOI: 10.1137/070690882.

Júdice, J. J.; Sherali, H. D.; Ribeiro, I. M.; Faustino, A. M. (2007). "Complementarity Active-Set Algorithm for Mathematical Programming Problems with Equilibrium Constraints". *Journal of Optimization Theory and Applications* 134.3, pp. 467–481. DOI: 10.1007/s10957-007-9231-z.

Kanzow, C.; Yamashita, N.; Fukushima, M. (1997). "New NCP-Functions and Their Properties". *Journal of Optimization Theory and Applications* 94.1, pp. 115–135. DOI: 10.1023/A:1022659603268.

Leyffer, S.; López-Calva, G.; Nocedal, J. (2006). "Interior Methods for Mathematical Programs with Complementarity Constraints". *SIAM Journal on Optimization* 17.1, pp. 52–77. DOI: 10.1137/040621065.

Luo, Z.-Q.; Pang, J.-S.; Ralph, D. (1996). *Mathematical Programs with Equilibrium Constraints*. Cambridge: Cambridge University Press.

Luo, Z.-Q.; Pang, J.-S.; Ralph, D. (1998). "Piecewise Sequential Quadratic Programming for Mathematical Programs with Nonlinear Complementarity Constraints". *Multilevel Optimization: Algorithms and Applications*. Ed. by Migdalas, A.; Pardalos, P. M.; Värbrand, P. Boston: Springer, pp. 209–229. DOI: 10.1007/978-1-4613-0307-7_9.

Mehlitz, P. (2019). "On the linear independence constraint qualification in disjunctive programming". *Optimization*, pp. 1–37. DOI: 10.1080/02331934.2019.1679811.

Outrata, J. V.; Kočvara, M.; Zowe, J. (1998). *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Dordrecht: Kluwer Academic.

Ralph, D.; Wright, S. J. (2004). "Some properties of regularization and penalization schemes for MPECs". *Optimization Methods and Software* 19.5, pp. 527–556. DOI: 10.1080/10556780410001709439.

Scheel, S.; Scholtes, S. (2000). "Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity". *Mathematics of Operations Research* 25.1, pp. 1–22. DOI: 10.1287/moor.25.1.1.15213.

Sun, D.; Qi, L. (1999). "On NCP-Functions". *Computational Optimization and Applications* 13.1, pp. 201–220. DOI: 10.1023/A:1008669226453.

Tseng, P. (1996). "Growth behavior of a class of merit functions for the nonlinear complementarity problem". *Journal of Optimization Theory and Applications* 89.1, pp. 17–37. DOI: 10.1007/bf02192639.

Ulbrich, M. (2002). “Semismooth Newton Methods for Operator Equations in Function Spaces”. *SIAM Journal on Optimization* 13.3, pp. 805–841. DOI: 10.1137/s1052623400371569.

Wachsmuth, G. (2014). “Strong Stationarity for Optimal Control of the Obstacle Problem with Control Constraints”. *SIAM Journal on Optimization* 24.4, pp. 1914–1932. DOI: 10.1137/130925827.

Ye, J. J. (2005). “Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints”. *Journal of Mathematical Analysis and Applications* 307.1, pp. 350–369. DOI: 10.1016/j.jmaa.2004.10.032.