ON THE MULTIPLIER-PENALTY-APPROACH FOR QUASI-VARIATIONAL INEQUALITIES

Christian Kanzow

Preprint 329 February 2015

University of Würzburg
Institute of Mathematics
Emil-Fischer-Str. 30
97074 Würzburg
Germany
e-mail: kanzow@mathematik.uni-wuerzburg.de

February 19, 2015

Abstract

The multiplier-penalty approach is one of the classical methods for the solution of constrained optimization problems. This method was generalized to the solution of quasi-variational inequalities by Pang and Fukushima (*Computational Management Science 2, 2005, pp. 21-56*). Based on the recent improvements achieved for the multiplier-penalty approach for optimization, we generalize the method by Pang and Fukushima for quasi-variational inequalities in several respects: a) We allow to compute inexact KKT-points of the resulting subproblems; b) We improve the existing convergence theory; c) We investigate some special classes of quasi-variational inequalities where the resulting subproblems turn out to be easy to solve. Some numerical results indicate that the corresponding method works quite reliable in practice.

Key Words: Quasi-variational inequalities; Global convergence; Multiplier-penalty method; Augmented Lagrangian; Extended Mangarasarian-Fromovitz constraint qualification; Constant positive linear independence constraint qualification; Monotone mappings.

1 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}^n$, $g^P : \mathbb{R}^{2n} \to \mathbb{R}^m$ and $g^I : \mathbb{R}^n \to \mathbb{R}^l$ be given vector-valued functions, and let $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be the set-valued mapping defined by

$$K(x) := \left\{ y \in \mathbb{R}^n \mid g^P(y,x) \leq 0,\ g^I(y) \leq 0 \right\}, \tag{1}$$

where g^P denotes the parameterized constraints and g^I the independent (independent of the parameter x) or individual ones. Note that also equality constraints can be included, but to keep the notation simple, we consider only inequality constraints. Then the finite-dimensional *quasi-variational inequality problem* QVI(F, g^P, g^I) or, simply, QVI consists of finding a point $x \in K(x)$ such that

$$F(x)^T(y - x) \geq 0 \quad \forall y \in K(x). \tag{2}$$

If the set $K(x)$ is independent of x, i.e. $K(x) = \mathbf{K}$ for all $x \in \mathbb{R}^n$ with the constant set

$$\mathbf{K} := \left\{ y \in \mathbb{R}^n \mid g^I(y) \leq 0 \right\}, \tag{3}$$

then the QVI reduces to the standard variational inequality (VI) problem, cf. [23] for a comprehensive treatment.

Historically, the QVI was introduced in the paper [7] by Bensoussan, Goursat, and Lions in the context of impulse control problems. Bensoussan and Lions provide some further material in their subsequent papers [8, 9]. Soon after, the QVI turned out to be a powerful tool to model several complex equilibrium situations arising in different fields like generalized Nash equilibrium problems [6, 27], mechanics [5, 10, 30, 37, 44, 45], economics [33, 53], statistics [36], transportation [13, 17, 50], biology [26], or stationary problems in superconductivity, thermoplasticity, and electrostatics [31, 32, 38]. The reader is also referred to the two monographs [39] by Mosco and [5] by Baiocchi and Capelo for a more comprehensive analysis of QVIs.

Due to the complicated structure of QVIs, the numerical solution of (finite-dimensinal) QVIs is a very challenging problem. To the best of our knowledge, the first approach is due to Chan and Pang [14], where a projection-type method is used to solve the special class of QVIs in which $K(x) = c(x) + Q$ with a fixed, convex set $Q \subseteq \mathbb{R}^n$ and $c : \mathbb{R}^n \to \mathbb{R}^n$ a suitable mapping. The majority of the subsequent algorithmic works consider similar projection or fixed-point approaches, again mainly for the above class of QVIs, cf. [40, 42, 43, 49, 51]. Fukushima [25] presents a reformulation of QVIs as an optimization problem based on the idea of gap functions. Improvements and modifications of this approach can be found in [28, 29]. But also the gap function idea is capable to deal with special classes of QVIs only, and usually leads to either a nonsmooth optimization problem or an optimization problem where the objective function is expensive to evaluate.

Pang and Fukushima [48] were the first who provide a convergence theory for a general class of QVIs. Their method is based on the multiplier-penalty or augmented Lagrangian-type idea known from optimization, but lacks from the fact that, in general, the resulting subproblems are difficult to solve. Finally, there is the idea of solving general QVIs based on the corresponding KKT or related optimality conditions. This point of view has been taken by Facchinei et al. [22] where a potential reduction interior-point approach with a quite satisfactory global convergence theory has been developed, as well as in [20] where a semismooth Newton-type method is investigated which, under suitable assumptions, is globally and locally fast convergent. Purely locally convergent Newton methods are also considered by Outrata and co-workers [44, 45, 46]. In general, however, one can say that the local convergence theory requires relatively strong assumptions.

In this paper, we follow the multiplier-penalty approach by Pang and Fukushima [48]. These authors include the difficult parameterized constraints to the original function F and leave the individual constraints as they are, hence they solve a sequence of standard variational inequalities. This idea is taken from the field of optimization, where the multiplier-penalty method is one of the standard algorithms for constrained optimization problem, cf. [11, 41]. Successful implementations of the multiplier-penalty method are provided by LANCELOT [16] and, more recently, ALGECAN [12].

From a theoretical point of view, there are some recent contributions that improve the existing convergence theory of multiplier-penalty methods for constrained optimization problems. The classical assumption for global convergence is the extended Mangasarian-Fromovitz constraint qualification (EMFCQ) that was also used by Pang and Fukushima [48] in their QVI-counterpart of the augmented Lagrangian approach. The paper [1] by Andreani et al. generalizes the convergence theory by replacing EMFCQ by the constant positive linear dependence (CPLD) constraint qualifcation. Subsequently, Andreani et al. [2, 3] show that CPLD can be replaced even by some weaker assumptions. All these conditions imply a local error bound. Based on this observation, Izmailov et al. [34] then prove a convergence result for augmented Lagrangian-type methods using an error bound condition, but they also require an additional technical assumption to get convergence.

Note, however, that some of these papers (have to) assume that a limit point is feasible for the optimization problem, though a characterization of infeasible points can also be given, see, for example, the book [12] which summarizes some of the recent contributions on augmented Lagrangian methods for constrained optimization problems.

Our first aim is to improve Pang and Fukushima's multiplier-penalty approach for the solution of QVIs [48] by adapting the more recent results known for optimization problems to the setting of QVIs. The second aim is to get a better understanding of some classes of QVIs for which it can be guaranteed that they can be solved by the multiplier-penalty scheme. In particular, the following are the main contributions of this paper:

- We only need to compute KKT points (not necessarily solutions) of the resulting VIs that arise at each iteration.
- We need to compute inexact KKT points only.
- We generalize the convergence theory provided by Pang and Fukushima [48] by using weaker assumptions and allowing the penalty parameter to stay finite.
- We investigate some special classes of QVIs for which the resulting VIs are "simple" in the sense that, e.g., they yield monotone VIs.

Some numerical results will also be presented to illustrate the behaviour of the augmented Lagrangian approach.

The paper is organized in the following way: Section 2 states some background material, in particular on suitable constraint qualifications that will play an essential role in the subsequent analysis. Section 3 gives a precise statement of the multiplier-penalty method for QVIs together with a refined convergence analysis that generalizes the corresponding result from [48]. Section 4 then investigates several special classes of QVIs, as introduced in the paper [22], and provides assumptions under which, e.g., these classes of QVIs yield monotone VI-subproblems in our augmented Lagrangian approach. Some numerical results are presented in Section 5, and we close with some final remarks in Section 6.

Notation: The symbol $\|\cdot\|$ denotes the Euclidean vector norm or its associated matrix norm, $B_r(x)$ is the (closed) Euclidean ball of radius $r > 0$ around a given point x; on the other hand, $\|\cdot\|_\infty$ indicates the maximum norm. For a differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, we denote its Jacobian at x by $F'(x)$ or $JF(x)$, whereas $\nabla F(x)$ stands for the transposed Jacobian. Given a smooth mapping $g(y, x)$ of two arguments (like g^P), say $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$, we write $\nabla_y g(y, x)$ and $\nabla_x g(y, x)$ for the corresponding partial transposed Jacobians with respect to y and x, respectively.

2 Preliminaries

This section recalls some standard terminology, states some known results, and introduces the notion of an ε-stationary point for variational inequalities. We begin with some constraint qualifications that will play an essential role in our subsequent analysis.

Definition 2.1 *Consider a set $X := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ (i = 1, \ldots, l)\}$ described by a finite number of inequality constraints, where each function $g_i : \mathbb{R}^n \to \mathbb{R}$ is assumed to be continuously differentiable.*

(a) Given a point $x^* \in \mathbb{R}^n$ and a subset $I \subseteq \{1, \ldots, l\}$, we call the gradients $\nabla g_i(x^*)$ $(i \in I)$ positively linearly independent *if the implication*

$$\Big[\sum_{i \in I} \lambda_i \nabla g_i(x^*) = 0, \ \lambda_i \geq 0 \ (i \in I)\Big] \Longrightarrow \lambda_i = 0 \ (i \in I)$$

holds; otherwise, the gradients are called positively linearly dependent.

(b) A feasible point $x^* \in X$ satisfies the Mangasarian-Fromovitz constraint qualification *(MFCQ) if the gradients $\nabla g_i(x^*)$ $(i : g_i(x^*) = 0)$ of the active constraints are positively linearly independent.*

(c) A (not necessarily feasible) point $x^* \in \mathbb{R}^n$ satisfies the extended Mangasarian-Fromovitz constraint qualification *(EMFCQ) if the gradients $\nabla g_i(x^*)$ $(i : g_i(x^*) \geq 0)$ of the active and violated constraints are positively linearly independent.*

(d) A feasbile point $x^* \in X$ satisfies the constant positive linear dependence condition *(CPLD) if for any subset $I \subseteq \{i \mid g_i(x^*) = 0\}$ such that the gradients*

$\nabla g_i(x^*)$ $(i \in I)$ are positively linearly dependent, there exists a neighbourhood $N(x^*)$ such that the gradients $\nabla g_i(x)$ $(i \in I)$ are linearly dependent for all $x \in N(x^*)$.

MFCQ is a standard assumption used in the optimization literature. The closely related EMFCQ condition is often used to guarantee feasibility of limit points in the context of penalty-type methods. The CPLD condition is less known and goes back to [47], where it was used to investigate the convergence properties of SQP-type methods. The fact that CPLD is a constraint qualification was shown later in [4]; note that CPLD is obviously weaker than MFCQ, moreover it covers the CRCQ (constant rank constraint qualification) condition. There are some recent generalizations of CPLD, cf. [2, 3] for more details. In principle, it should be possible to replace the CPLD condition by these weaker conditions in our convergence analysis, but one of these weaker conditions is actually equivalent to CPLD since, for notational convenience, we consider inequality constraints only, and the other would require some more technical overhead so that we prefer to keep dealing with CPLD.

Our convergence analysis needs the following Carathéodory-type result whose proof may be found, e.g., in [12, Lem. 3.1] and [52, Lem A.1].

Lemma 2.2 *Assume that a given vector $w \in \mathbb{R}^n$ has a representation of the form*

$$w = \sum_{j=1}^{p} \mu_j u^j + \sum_{i=1}^{m} \lambda_i v^i$$

with $u^j, v^i \in \mathbb{R}^n, \lambda_i \geq 0$ and $\mu_j \in \mathbb{R}$ for all $i = 1, \dots, m$ and $j = 1, \dots, p$. Then there exist index sets $I \subseteq \{1, \dots, m\}$ and $J \subseteq \{1, \dots, p\}$ as well as scalars $\tilde{\lambda}_i \geq 0$ $(i \in I)$ and $\tilde{\mu}_j \in \mathbb{R}$ $(j \in J)$ such that

$$w = \sum_{j \in J} \tilde{\mu}_j u^j + \sum_{i \in I} \tilde{\lambda}_i v^i$$

and such that the vectors u^j $(j \in J)$, v^i $(i \in I)$ are linearly independent.

We next consider the QVI defined by (2) with the set $K(x)$ given by (1). The following smoothness assumptions are required for the remaining part of this paper.

Assumption 2.3

(a) The function F is continuous on $\mathbb{R}^n$.

(b) The function $g^I : \mathbb{R}^n \to \mathbb{R}^l$ is continuously differentiable on $\mathbb{R}^n$.

(c) The function $g^P : \mathbb{R}^{2n} \to \mathbb{R}^m$ is continuously differentiable on $\mathbb{R}^{2n}$.

To derive the KKT conditions of the QVI from (2), (1), let x^* be a solution of this QVI. Then we have $x^* \in K(x^*)$ and

$$F(x^*)^T (x - x^*) \geq 0 \quad \forall x \in K(x^*).$$

This means that

$$F(x^*)^T (x - x^*) \geq 0 \quad \forall x : g^P(x, x^*) \leq 0, \; g^I(x) \leq 0.$$

Setting $f_*(x) := F(x^*)^T(x - x^*)$, it therefore follows that x^* is a solution of the (parameterized) optimization problem

$$\min_x \; f_*(x) \quad \text{s.t.} \quad \begin{aligned} & g_i^P(x, x^*) \le 0 \quad \forall i = 1, \ldots, m, \\ & g_j^I(x) \le 0 \quad \forall j = 1, \ldots, l. \end{aligned} \tag{4}$$

Assuming that a suitable constraint qualification (like MFCQ, CPLD etc.) holds at the solution x^*, it follows that there exist some Lagrange multipliers such that the triple (x^*, λ^*, μ^*) satisfies the following KKT conditions:

$$\begin{aligned} \nabla f_*(x^*) + \sum_{i=1}^m \lambda_i^* \nabla_y g_i^P(x^*, x^*) + \sum_{j=1}^l \mu_j^* \nabla g_j^I(x^*) = 0, \\ \lambda_i^* \ge 0, \; g_i^P(x^*, x^*) \le 0, \; \lambda_i^* g_i^P(x^*, x^*) = 0 \quad \forall i = 1, \ldots, m, \\ \mu_j^* \ge 0, \; g_j^I(x^*) \le 0, \; \mu_j^* g_j^I(x^*) = 0 \quad \forall j = 1, \ldots, l. \end{aligned}$$

Since $\nabla f_*(x^*) = F(x^*)$, this justifies the following terminology.

Definition 2.4 *Consider the QVI defined by* (2) *with* $K(x)$ *given by* (1)*. Then the system*

$$\begin{aligned} F(x) + \sum_{i=1}^m \lambda_i \nabla_y g_i^P(x, x) + \sum_{j=1}^l \mu_j \nabla g_j^I(x) = 0, \\ \lambda_i \ge 0, \; g_i^P(x, x) \le 0, \; \lambda_i g_i^P(x, x) = 0 \quad \forall i = 1, \ldots, m, \\ \mu_j \ge 0, \; g_j^I(x) \le 0, \; \mu_j g_j^I(x) = 0 \quad \forall j = 1, \ldots, l \end{aligned}$$

is called the KKT conditions *of the underlying QVI. Every triple* (x^*, λ^*, μ^*) *satisfying these KKT conditions is called a* KKT point *of the QVI.*

Note that there is a very close relation between the QVI and the corresponding KKT conditions. This comes from the observation that the objective function in the previous derivation of the KKT conditions is always linear (hence convex), so that the KKT conditions are automatically sufficient optimality conditions if, in addition, the constraints are convex. This follows from the standard theory for nonlinear programs. We summarize our observations in the following result.

Theorem 2.5 *The following statements hold:*

(a) If x^* *is a solution of the QVI defined by* (2), (1) *and any standard constraint qualification holds at* x^**, then there exist multipliers* λ^*, μ^* *such that* (x^*, λ^*, μ^*) *is a KKT point of the QVI.*

(b) Conversely, if (x^*, λ^*, μ^*) *is a KKT point of the QVI* (2), (1) *such that the constraints* $g_i^P(\cdot, x)$ *are convex for any fixed* $x \in \mathbb{R}^n$ *and all components* $i = 1, \ldots, m$ *and such that* $g_j^I(\cdot)$ *are also convex for all* $j = 1, \ldots, l$*, then* x^* *is a solution of the QVI.*

This close relationship between the QVI and its KKT points plays a central role in our approach since, in the end, we will try to compute KKT points.

More precisely, our aim is to compute such a KKT point by solving a related sequence of standard variational inequalities (VIs). In order to save computation time and to be more realistic, we allow inexact solutions of the VIs. To this end, consider the VI

$$F(x^*)^T(x - x^*) \geq 0 \quad \forall x \in X, \tag{5}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and the feasible set $X := \{x \in \mathbb{R}^n \mid g_j(x) \leq 0 \ (j = 1, \ldots, l)\}$ is described by a finite number of continuously differentiable inequality constraints. The following definition introduces our notion of an ε-stationary point of this VI.

Definition 2.6 *Consider the VI from* (5)*, and let* $\varepsilon \geq 0$*. We call* $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l$ *an* ε-inexact KKT point *of the VI if the following inequalities hold:*

$$\Big\| F(\bar{x}) + \sum_{j=1}^{l} \bar{\mu}_j \nabla g_j(\bar{x}) \Big\|_\infty \leq \varepsilon,$$
$$\bar{\mu}_j \geq -\varepsilon, \quad g_j(\bar{x}) \leq \varepsilon, \quad \big|\bar{\mu}_j g_j(\bar{x})\big| \leq \varepsilon, \quad \forall j = 1, \ldots, l.$$

Note that for $\varepsilon = 0$ an ε-inexact KKT point of a VI is a standard KKT point. We further note that this notion of an ε-inexact KKT point is very general, and many times it is possible to find an approximate KKT point which satisfies all the above conditions with, in addition, $\bar{\mu}_j \geq 0$ for all $j = 1, \ldots, l$. The more general condition from Definition 2.6 turns out to be sufficient to prove a global convergence result under the EMFCQ assumption, whereas our refined convergence analysis requires that, in addition, we have the nonnegativity of all $\bar{\mu}_j$.

3 Algorithm and Convergence

This section is divided into three parts: We first derive our algorithm in Section 3.1. A global convergence result under the EMFCQ condition is provided in Section 3.2. A more refined analysis under the CPLD condition is then given in Section 3.3.

3.1 Statement of Algorithm

Consider the QVI from (2) with the set $K(x)$ defined by (1). Throughout this section, we suppose that Assumption 2.3 holds.

In order to solve the QVI, we follow the approach from [48] where the authors try to compute a solution of the QVI by solving a sequence of suitable VIs. The idea comes from the augmented Lagrangian/multiplier-penalty approach which is a prominent tool for optimization problems, cf. [41].

To motivate the algorithm, let x^* be a (fixed) solution of the QVI from (2), (1). As indicated in the previous section, this implies that x^* is also a solution of the optimization problem (4) with $f_*(x) := F(x^*)^T(x - x^*)$. Applying the standard multiplier-penalty approach to the parameterized constraints and leaving the functions g_j^I explicitly in the constraints yields the optimization problem

$$\min_x \quad L_a(x,u,\rho) := f_*(x) + \frac{1}{2\rho}\sum_{i=1}^m \left[\max^2\{0, u_i + \rho g_i^P(x,x^*)\} - (u_i)^2\right] \quad \text{s.t.} \quad x \in \mathbf{K}, \tag{6}$$

where $\mathbf{K}$ is defined by (3), $\rho > 0$ is a suitable penalty parameter, and u denotes an estimate for the multipliers λ_i associated to the functions g_i^P. Following the classical multiplier-penalty idea, we therefore try to compute a solution x^* of the QVI by minimizing (6). The corresponding (primal) optimality conditions lead to the standard variational inequality

$$\Big[F(x) + \sum_{i=1}^m \max\{0, u_i^k + \rho_k g_i^P(x,x)\}\nabla_y g_i^P(x,x)\Big]^T (y-x) \geq 0, \quad \forall y \in \mathbf{K}.$$

The successive solution of these problems gives us the following multiplier-penalty or sequential variational inequality approach for the solution of QVIs. It differs from [48] especially by allowing inexact KKT points at each subproblem.

Algorithm 3.1 (Sequential Variational Inequality Approach for QVIs)

(S.0) Let $u^{\max} \in \mathbb{R}^m, \tau \in (0,1), \gamma > 1$. Choose $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l$, $u^0 \in [0, u^{\max}]$, $\varepsilon_0 > 0$, $\rho_0 > 0$, and set $k := 0$.

(S.1) If (x^k, λ^k, μ^k) is a KKT point of the QVI (2): STOP.

(S.2) Compute an ε_k-inexact KKT point (x^{k+1}, μ^{k+1}) of the VI_k, which is to find $x \in \mathbf{K}$ such that

$$\Big[F(x) + \sum_{i=1}^m \max\{0, u_i^k + \rho_k g_i^P(x,x)\}\nabla_y g_i^P(x,x)\Big]^T (y-x) \geq 0, \quad \forall y \in \mathbf{K}, \tag{7}$$

and define $\lambda_i^{k+1} := \max\big\{0, u_i^k + \rho_k g_i^P(x^{k+1}, x^{k+1})\big\}$ for all $i = 1, \ldots, m$.

(S.3) If

$$\big\|\min\big\{-g^P(x^{k+1}, x^{k+1}), \lambda^{k+1}\big\}\big\|_\infty \leq \tau \big\|\min\big\{-g^P(x^k, x^k), \lambda^k\big\}\big\|_\infty,$$

then set $\rho_{k+1} := \rho_k$, else set $\rho_{k+1} := \gamma\rho_k$.

(S.4) Choose $u^{k+1} \in [0, u^{\max}]$, $\varepsilon_{k+1} \leq \varepsilon_k$, set $k \leftarrow k+1$, and go to (S.1).

Note that, by construction, the sequence $\{u^k\}$ generated by Algorithm 3.1 is nonnegative and bounded. A natural choice would be something like $u^{k+1} := \min\{\lambda^{k+1}, u^{\max}\}$, i.e. it is a kind of safe-guarded Lagrange multiplier update to keep u^k bounded. This boundedness will play an essential role in the subsequent convergence analysis, see also Example 3.3 below which illustrates the necessity of the boundedness of the sequence $\{u^k\}$.

We further note that our updating of the penalty parameter ρ_k is different from most other papers in the optimization context. A classical updating of ρ_k typically depends on the size of the infeasibility $\max\{0, g^P(x^{k+1}, x^{k+1})\}$ only. The description of the ALGECAN software in [12] as well as the underlying theoretical papers by Andreani et al. [1, 2, 3] use $g^P(x^{k+1}, x^{k+1})$ and u^{k+1} instead of $g^P(x^{k+1}, x^{k+1})$ and λ^{k+1} in our context. The only paper which seems to have an identical updating rule in the optimization context is Izmailov et al. [34].

Note also that the multiplier-penalty method for QVIs described by Pang and Fukushima [48] has no particular updating rule for ρ_k. These authors assume from the very beginning that a sequence $\{\rho_k\}$ is given that converges to infinity. Dealing with the possibility that $\{\rho_k\}$ remains bounded, however, is extremely important from a numerical point of view in order to avoid an unnecessary ill-conditioning of the VI-subproblems.

3.2 Convergence Properties under EMFCQ

In the following theorem, we show that every accumulation point of a sequence generated by Algorithm 3.1 is a KKT point of the QVI provided that an EMFCQ condition holds at this limit point.

Theorem 3.2 *Let Assumption 2.3 hold, let $\{x^k\}$ be a sequence generated by Algorithm 3.1 with $\varepsilon_k \downarrow 0$, and let x^* be an accumulation point of $\{x^k\}$ such that the following EMFCQ condition holds at x^*:*

$$\left.\begin{aligned} \sum_{i\in\alpha} \lambda_i \nabla_y g_i^P(x^*, x^*) + \sum_{j\in\gamma} \mu_j \nabla g_j^I(x^*) = 0, \\ \lambda_i \geq 0, \quad \forall i \in \alpha, \\ \mu_j \geq 0, \quad \forall j \in \gamma \end{aligned}\right\} \Rightarrow \lambda_i = \mu_j = 0, \quad \forall i \in \alpha, \forall j \in \gamma, \tag{8}$$

where

$$\alpha := \{i \mid g_i^P(x^*, x^*) \geq 0\}, \quad \gamma := \{j \mid g_j^I(x^*) = 0\}. \tag{9}$$

Then there exist multipliers $(\lambda^*, \mu^*) \in \mathbb{R}^m \times \mathbb{R}^l$ such that (x^*, λ^*, μ^*) is a KKT point of the QVI.

Proof: By construction, (x^{k+1}, μ^{k+1}) is an ε_k-inexact stationary point of VI_k for each $k \in \mathbb{N}$. Together with the definition of λ^{k+1}, we therefore have

$$\Big\| F(x^{k+1}) + \sum_{i=1}^m \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j=1}^l \mu_j^{k+1} \nabla g_j^I(x^{k+1}) \Big\|_\infty \leq \varepsilon_k \tag{10}$$

and

$$\mu_j^{k+1} \geq -\varepsilon_k, \quad g_j^I(x^{k+1}) \leq \varepsilon_k, \quad \big|\mu_j^{k+1} g_j^I(x^{k+1})\big| \leq \varepsilon_k \quad \forall j = 1, \ldots, l. \tag{11}$$

Let us define the index sets

$$\alpha_k := \big\{i \mid g_i^P(x^k, x^k) \geq 0\big\} \quad \text{and} \quad \gamma_k := \big\{j \mid g_j^I(x^k) = 0\big\}.$$

Furthermore, let $K \subseteq \mathbb{N}$ be an infinite subset such that the subsequence $\{x^{k+1}\}_{k\in K}$ converges to x^*. Then

$$\alpha_{k+1} \subseteq \alpha \quad \text{and} \quad \gamma_{k+1} \subseteq \gamma \quad \forall k \in K \text{ sufficiently large;}$$

indeed, since $x^{k+1} \to_K x^*$, for each $i \notin \alpha$, we have $g_i^P(x^*, x^*) < 0$, and the continuity of g_i^P therefore implies $g_i^P(x^{k+1}, x^{k+1}) < 0$ for all $k \in K$ sufficiently large. This means that $i \notin \alpha_{k+1}$ and shows that the inclusion $\alpha_{k+1} \subseteq \alpha$ holds. In a similar way, we can show that $\gamma_{k+1} \subseteq \gamma$ holds for all $k \in K$ large enough.

We now consider two cases.

Case 1: The sequence $\{\rho_k\}$ remains bounded. Then (S.3) implies that

$$\min\big\{ - g_i^P(x^{k+1}, x^{k+1}), \lambda_i^{k+1}\big\} \to 0 \quad \forall i = 1, \dots, m, \quad k \to \infty. \tag{12}$$

Since $\{x^{k+1}\}_K \to x^*$ and $g_i(x^*, x^*) < 0$ for all $i \notin \alpha$, we therefore get $\lambda_i^{k+1} \to 0 =: \lambda_i^*$ $(i \notin \alpha)$. Using $\big|\mu_j^{k+1} g_j^I(x^{k+1})\big| \leq \varepsilon_k$, we also see that $\big|\mu_j^{k+1}\big| \leq \varepsilon_k / \big|g_j^I(x^{k+1})\big| \to_K 0 =: \mu_j^*$ $(j \notin \gamma)$. We claim that also the remaining sequence of multipliers

$$\big\{\big(\big(\lambda_i^{k+1}\big)_{i\in\alpha}, \big(\mu_j^{k+1}\big)_{j\in\gamma}\big)\big\}_{k\in K}$$

is bounded. Suppose the contrary, say

$$\big\|\big(\big(\lambda_i^{k+1}\big)_{i\in\alpha}, \big(\mu_j^{k+1}\big)_{j\in\gamma}\big)\big\| \to_K \infty.$$

Then we may assume without loss of generality that the corresponding normalized sequence converges:

$$\frac{\big(\big(\lambda_i^{k+1}\big)_{i\in\alpha}, \big(\mu_j^{k+1}\big)_{j\in\gamma}\big)}{\big\|\big(\big(\lambda_i^{k+1}\big)_{i\in\alpha}, \big(\mu_j^{k+1}\big)_{j\in\gamma}\big)\big\|} \to_K \big(\big(\overline{\lambda}_i\big)_{i\in\alpha}, \big(\overline{\mu}_j\big)_{j\in\gamma}\big)$$

for some nonzero and nonnegative limit point $\big(\big(\overline{\lambda}_i\big)_{i\in\alpha}, \big(\overline{\mu}_j\big)_{j\in\gamma}\big)$. Hence, dividing (10) by $\big\|\big(\big(\lambda_i^{k+1}\big)_{i\in\alpha}, \big(\mu_j^{k+1}\big)_{j\in\gamma}\big)\big\|$ and taking the limit $k \to_K \infty$, we obtain

$$\sum_{i\in\alpha} \overline{\lambda}_i \nabla_y g_i^P(x^*, x^*) + \sum_{j\in\gamma} \overline{\mu}_j \nabla g_j^I(x^*) = 0$$

with $\overline{\lambda}_i \geq 0$ $(i \in \alpha)$ and $\overline{\mu}_j \geq 0$ $(j \in \gamma)$. Using (8), we therefore get $\overline{\lambda}_i = 0$ $(i \in \alpha)$ and $\overline{\mu}_j = 0$ $(j \in \gamma)$, but this contradicts the fact that $\big\|\big(\big(\overline{\lambda}_i\big)_{i\in\alpha}, \big(\overline{\mu}_j\big)_{j\in\gamma}\big)\big\| = 1$. Consequently, the sequences $\{\lambda_i^{k+1}\}_K$ $(i \in \alpha)$ and $\{\mu_j^{k+1}\}_K$ $(j \in \gamma)$ are bounded. Subsequencing if necessary, we may assume that they converge, say

$$\lambda_i^{k+1} \to_K \lambda_i^* \ (i \in \alpha) \quad \text{and} \quad \mu_j^{k+1} \to_K \mu_j^* \ (j \in \gamma).$$

Hence, taking the limit $k \to_K \infty$ in (10), (11), and (12), and using the fact that $\lambda_i^* = 0$ $(i \notin \alpha), \mu_j^* = 0$ $(j \notin \gamma)$, we obtain

$$\begin{aligned} F(x^*) + \sum_{i=1}^m \lambda_i^* \nabla_y g_i^P(x^*, x^*) + \sum_{j=1}^l \mu_j^* \nabla g_j^I(x^*) = 0, \\ \mu_j^* \geq 0, \ g_j^I(x^*) \leq 0, \ \mu_j^* g_j^I(x^*) = 0 \quad \forall j = 1, \dots, l, \\ \min\big\{ - g_i^P(x^*, x^*), \lambda^*\big\} = 0 \quad \forall i = 1, \dots, m. \end{aligned}$$

Since the last equations are equivalent to $\lambda_i^* \geq 0, g_i^P(x^*, x^*) \leq 0$ and $\lambda_i^* g_i^P(x^*, x^*) = 0$ for all $i = 1, \dots, m$, it follows that (x^*, λ^*, μ^*) is a KKT point of the QVI (2).

Case 2: The sequence $\{\rho_k\}$ is unbounded, hence $\{\rho_k\} \to \infty$ for $k \to \infty$. Since $\{x^{k+1}\} \to_K x^*, g_i^P(x^*, x^*) < 0$ for all $i \notin \alpha$, $\{u_i^k\}$ is bounded, and $\rho_k \to \infty$ for $k \to \infty$, it follows that $u_i^k + \rho_k g_i^P(x^{k+1}, x^{k+1}) \leq 0$ for all $i \notin \alpha$ and all $k \in K$ sufficiently large. This yields $\lambda_i^{k+1} = 0$ for all $i \notin \alpha$ and all $k \in K$ large enough. Furthermore, similar to Case 1, we obtain $\left|\mu_j^{k+1}\right| \leq \varepsilon_k / \left|g_j^I(x^{k+1})\right| \to 0$ for $k \to_K \infty$ and for all $j \notin \gamma$. Exploiting EMFCQ, we therefore obtain in the same way as in Case 1 that the remaining sequence of multipliers $\left\{\left(\left(\lambda_i^{k+1}\right)_{i\in\alpha}, \left(\mu_j^{k+1}\right)_{j\in\gamma}\right)\right\}$ is bounded. Once again, we may therefore assume that this subsequence converges on the set K, say

$$\left\{\lambda_i^{k+1}\right\} \to_K \lambda_i^* \ (i \in \alpha) \quad \text{and} \quad \left\{\mu_j^{k+1}\right\} \to_K \mu_j^* \ (j \in \gamma).$$

But the convergence of $\{\lambda_i^{k+1}\}$ for all $i = 1, \dots, m$ (recall that $\lambda_i^{k+1} \to_K 0 =: \lambda_i^*$ for all $i \notin \alpha$) together with the definition $\lambda_i^{k+1} = \max\left\{0, u_i^k + \rho_k g_i^P(x^{k+1}, x^{k+1})\right\}$ and the boundedness of $\{u_i^k\}$ implies that $g_i^P(x^*, x^*) \leq 0$ for all $i = 1, \dots, m$. Since we already noted that $\mu_j^{k+1} \to_K 0 =: \mu_j^* \ (j \notin \gamma)$, we obtain from (10) and (11) together with our previous discussion that

$$F(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_y g_i^P(x^*, x^*) + \sum_{j=1}^{l} \mu_j^* \nabla g_j^I(x^*) = 0$$

as well as

$$\begin{aligned} \mu_j^* \geq 0,\ g_j^I(x^*) \leq 0,\ \mu_j^* g_j^I(x^*) = 0 \quad \forall j = 1, \dots, l, \\ \lambda_i^* \geq 0,\ g_i^P(x^*, x^*) \leq 0,\ \lambda_i^* g_i^P(x^*, x^*) = 0 \quad \forall i = 1, \dots, m. \end{aligned}$$

This shows that (x^*, λ^*, μ^*) is a KKT point of the QVI (2) also in the second case. □

Note that the previous result generalizes the corresponding convergence theorem from [48]. The latter corresponds to the case where $\varepsilon_k = 0$ for all $k \in \mathbb{N}$ (and $\rho_k \to \infty$ for $k \to \infty$). Since the subproblems VI_k are nonlinear, it is unrealistic to compute an exact KKT point of these subproblems; in addition, it might be much less time consuming to compute only inexact KKT points, especially in the first iterations.

The convergence theorem exploits the fact that the sequence $\{u^k\}$ is bounded. This assumption is not crucial at all since it is up to the user to choose this sequence. Nevertheless, it is interesting to see that the boundedness of $\{u^k\}$ is crucial for the convergence analysis. This is illustrated by the following counterexample.

Example 3.3 Consider the QVI with $n = 2$, $m = 1$, $l = 0$, and

$$F(x) := x, \quad g^P(y, x) := y_2 - x_1.$$

Note that we write g^P instead of g_1^P since there is just a single parameterized constraint in this example. Further note that Assumption 2.3 holds, and EMFCQ is satisfied at any point since the partial gradient $\nabla_y g^P(x, x) = (0, 1)^T$ is linearly independent. A simple calculation shows that $x^* := (0, 0)^T$ together with the multiplier $\lambda^* := 0$ is the unique KKT point of the given QVI.

Consider the subproblem VI_k. Since there are no individual constraints, this subproblem reduces to a nonlinear system of equations $F_k(x) = 0$, where F_k is given by

$$F_k(x) := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \max\left\{0, u^k + \rho_k(x_2 - x_1)\right\} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Given $\rho_k > 0$, and choosing $u^k := 1+\rho_k$, it is not difficult to see that $x^{k+1} := (0, -1)^T$ is a solution of $F_k(x) = 0$. Obviously, for $k \to \infty$, this sequence converges to $(0, -1)^T$, but this limit point does not correspond to a KKT point of the given QVI. This observation does not contradict Theorem 3.2 since, for the particular choice of u^k in this example, we have $u^k \to \infty$ for $\rho_k \to \infty$. ♢

The next example is used to illustrate a couple of different properties of multiplier-penalty methods and its inexact counterparts.

Example 3.4 Consider again a QVI with $n = 2$, $m = 1$, and $l = 0$, defined by

$$F(x) := \begin{pmatrix} x_1 \\ (x_2 + 1)^2 \end{pmatrix}, \quad g^P(y, x) := y_2 - x_1.$$

The unique KKT point of this problem is given by $x^* := (0, -1)^T$ and $\lambda^* := 0$.

On the other hand, the corresponding subproblem VI_k reduces to the nonlinear system of equations $F_k(x) = 0$, where

$$F_k(x) := \begin{pmatrix} x_1 \\ (x_2 + 1)^2 \end{pmatrix} + \max\left\{0, u^k + \rho_k(x_2 - x_1)\right\} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This yields the following observations:

(a) The subproblem $F_k(x) = 0$ has no solution for any $\rho_k < u^k$ although the underlying QVI has a unique solution.

(b) On the other hand, for $\rho_k \geq u^k$, the subproblem $F_k(x) = 0$ has a unique solution $x^{k+1} = (0, -1)^T$ which is already a solution of the QVI; hence a finite value of ρ_k leads to a solution of the QVI.

(c) The existence of ε_k-inexact KKT points can be guaranteed for smaller values of ρ_k; in particular, an elementary calculation shows that, for example, $x^{k+1} := (\varepsilon_k, -1)^T$ is an ε_k-inexact KKT point for all $\rho_k \geq \frac{u^k - \varepsilon_k}{1+\varepsilon_k}$. ♢

An interesting question is now, if there exists a *finite* penalty parameter ρ, for which the solution of the VI is already a solution of the QVI. The proof of Theorem 3.2 as well as the previous example seem to indicate that such a result might hold. However, the following counterexample shows that this is not true even in very favourable situations where all functions are linear and F is strongly monotone.

Example 3.5 Let $n = 2, m = 1, l = 0$ and consider the corresponding QVI given by

$$F(x) := \begin{pmatrix} x_1 \\ x_2 - 1 \end{pmatrix}, \quad g^P(y, x) := y_2 - x_1.$$

Note that this example satisfies the implication (8). An elementary calculation shows that there is a unique KKT point given by $x^* := (0, 0)^T$ with corresponding Lagrange multiplier $\lambda^* := 1$.

Since there are no individual constraints, the resulting VI_k-subproblem reduces to the nonlinear system of equations $F_k(x) = 0$, where

$$F_k(x) := \begin{pmatrix} x_1 \\ x_2 - 1 \end{pmatrix} + \max\left\{0, u^k + \rho_k(x_2 - x_1)\right\}\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Taking $u^k := 0$, it follows that $x^k := \left(0, \frac{1}{1+\rho_k}\right)^T$ is the unique solution of this VI_k-subproblem. Hence we have $x^k \to x^*$ for $k \to \infty$, but $x^k \neq x^*$ for all $k \in \mathbb{N}$. $\Diamond$

The existence of a limit point of the sequence $\{x^k\}$ is guaranteed, for example, if the set $\mathbf{K}$ is bounded (hence compact due to the continuity of the functions g_j^I). The same condition also guarantees the existence of (exact) solutions of the subproblems VI_k. However, since we compute (inexact) KKT points, we still need a condition which yields the existence of Lagrange multipliers. The following result therefore provides a simple CPLD-type condition which shows that any solution of VI_k gives a KKT point. In particular, this result therefore guarantees the existence of inexact KKT points.

Proposition 3.6 *Let x^* be feasible for the QVI, i.e. $x^* \in K(x^*)$, and suppose that the gradients*

$$\nabla g_j^I(x^*) \ (j \in \gamma) \text{ are positively linearly independent,} \tag{13}$$

where γ is defined by (9). Then there exists an $\varepsilon > 0$ such that the gradients $\nabla g_j^I(x)$ $(j \in \gamma(x))$ remain positively linearly independent for all $x \in B_\varepsilon(x^*)$ feasible for VI_k (with arbitrary $k \in \mathbb{N}$ since the feasible set of VI_k is independent of the particular k), where $\gamma(x) := \{j \mid g_j^I(x) = 0\}$.

Proof: Consider VI_k with a fixed index k. As in the previous proof, we can show that $\gamma(x) \subseteq \gamma$ holds for all x feasible with respect to VI_k. Suppose that the statement does not hold. Then there exists a sequence $\{x^k\} \to x^*$, with each x^k being feasible for VI_k, as well as a sequence of multipliers $\{\mu_j^k\}_{j \in \gamma(x^k)}$ such that

$$\sum_{j \in \gamma(x^k)} \mu_j^k \nabla g_j^I(x^k) = 0, \quad \mu_j^k \geq 0 \ (j \in \gamma(x^k))$$

and, without loss of generality, $\|(\mu_j^k)_{j \in \gamma(x^k)}\| = 1$ for all $k \in \mathbb{N}$. Since $\gamma(x^k) \subseteq \gamma$ (at least for all k sufficiently large), we may add some zero multipliers and therefore get

$$\sum_{j \in \gamma} \mu_j^k \nabla g_j^I(x^k) = 0 \tag{14}$$

and $\|(\mu_j^k)_{j \in \gamma}\| = 1$. Subsequencing if necessary, we may assume that $\mu_j^k \to \mu_j^*$ for all $j \in \gamma$. Taking the limit $k \to \infty$, possibly only on a subsequence, we obtain from (14) that

$$\sum_{j \in \gamma} \mu_j^* \nabla g_j^I(x^*) = 0$$

for some nonzero and nonnegative vector $(\mu_j^*)_{j\in\gamma}$, but this contradicts our assumption from (13). $\square$

Note that assumption (13) holds, in particular, under the EMFCQ condition from (8). We further note that the existence of Lagrange multipliers can also be shown under weaker or different constraint qualifications.

3.3 Refined Convergence Analysis

Recall that the EMFCQ condition in Theorem 3.2 guarantees, in particular, that any accumulation point of a sequence generated by Algorithm 3.1 is feasible for the given QVI, and this property is independent of the behaviour of the sequence of penalty parameters $\{\rho_k\}$, which might be bounded or unbounded.

The following result shows that we automatically get feasibility of an accumulation point in the case where the sequence $\{\rho_k\}$ remains bounded. Note that this result is a consequence of the particular updating rule for the penalty parameter.

Lemma 3.7 *Let Assumption 2.3 hold, let $\{x^k\}$ be a sequence generated by Algorithm 3.1 with $\varepsilon_k \downarrow 0$, and let x^* be an accumulation point of $\{x^k\}$. Then x^* is feasible for the QVI provided that the sequence of penalty parameters $\{\rho_k\}$ is bounded.*

Proof: Since ρ_k is bounded, we have $\rho_k = \rho_{k_0}$ for all $k \geq k_0$ for some index $k_0 \in \mathbb{N}_0$ and, therefore,

$$\min\big\{ -g_i^P(x^{k+1}, x^{k+1}), \lambda_i^{k+1})\big\} \to 0 \quad \forall i = 1, \dots, m \text{ for } k \to \infty. \tag{15}$$

Let x^* be an accumulation point and $\{x^{k+1}\}_{k\in K}$ be a subsequence converging to x^*. Suppose that there exists an index $i \in \{1, \dots, m\}$ such that $g_i^P(x^*, x^*) > 0$. Then, by continuity, there exists a constant $\delta > 0$ such that $g_i^P(x^{k+1}, x^{k+1}) \geq \delta$ for all $k \in K$ sufficiently large. We therefore obtain from (15) that

$$0 \leftarrow \min\big\{ -g_i^P(x^{k+1}, x^{k+1}), \lambda_i^{k+1})\big\} \leq -g_i^P(x^{k+1}, x^{k+1}) \leq -\delta < 0,$$

and this contradiction completes the proof. $\square$

The previous result allows us to state a global convergence result under the assumption that the sequence $\{\rho_k\}$ is bounded and that we can compute ε_k-KKT points (x^k, μ^k) satisfying $\mu^k \geq 0$. Note that we still need a constraint qualification to guarantee the existence of Lagrange multipliers, but that the constraint qualification used in the subsequent result is significantly weaker than EMFCQ.

Theorem 3.8 *Let Assumption 2.3 hold, let $\{x^k\}$ be a sequence generated by Algorithm 3.1 with $\varepsilon_k \downarrow 0$, let x^* be an accumulation point of $\{x^k\}$, and assume that ε_k-stationarity is satisfied with nonnegative multipliers μ^k. Suppose that $\{\rho_k\}$ is bounded, and that the gradient vectors $\nabla g_j^I(x^*)$ $(j \in \gamma)$ satisfy CPLD, where $\gamma := \{j \mid g_j^I(x^*) = 0\}$. Then there exist multipliers $(\lambda^*, \mu^*) \in \mathbb{R}^m \times \mathbb{R}^l$ such that (x^*, λ^*, μ^*) is a KKT point of the QVI.*

Proof: Let us write

$$L_k := F(x^{k+1}) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla g_j^I(x^{k+1}). \tag{16}$$

Then the ε_k-stationarity of (x^{k+1}, μ^{k+1}) together with the definition of λ^{k+1} and the additional assumption that $\mu^{k+1} \geq 0$ yields

$$\|L_k\|_\infty \leq \varepsilon_k, \quad \lambda_i^{k+1} \geq 0 \ \ \forall i = 1, \ldots, m \tag{17}$$

and

$$\mu_j^{k+1} \geq 0, \quad g_j^I(x^{k+1}) \leq \varepsilon_k, \quad \left|\mu_j^{k+1} g_j^I(x^{k+1})\right| \leq \varepsilon_k \ \ \forall j = 1, \ldots, l. \tag{18}$$

Let $\{x^{k+1}\}_{k\in K}$ be a subsequence converging to x^*, and recall that x^* is feasible in view of Lemma 3.7. Since $\{u^k\}$ is bounded by construction, we may assume without loss of generality that $\{u^k\}_{k\in K} \to u^*$ for some limit point $u^* \in \mathbb{R}^m$. Moreover, the assumed boundedness of $\{\rho_k\}$ implies that $\rho_k = \rho_{k_0}$ for all $k \geq k_0$ and a sufficiently large index $k_0 \in \mathbb{N}_0$. Altogether, the definition of λ^{k+1} therefore implies that the subsequence $\{\lambda^{k+1}\}_{k\in K}$ is convergent: For all $i = 1, \ldots, m$, we have

$$\lambda_i^{k+1} = \max\left\{0, u_i^k + \rho_k g_i^P(x^{k+1}, x^{k+1})\right\} \to_K \max\left\{0, u_i^* + \rho_{k_0} g_i^P(x^*, x^*)\right\} =: \lambda_i^*.$$

We now define another sequence $\{\hat{\mu}_j^{k+1}\}_{k\in\mathbb{N}}$ by

$$\hat{\mu}_j^{k+1} := \begin{cases} \mu_j^{k+1}, & \text{if } j \in \gamma, \\ 0, & \text{if } j \notin \gamma \end{cases} \tag{19}$$

as well as

$$\begin{aligned} \hat{L}_k &:= F(x^{k+1}) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j=1}^{l} \hat{\mu}_j^{k+1} \nabla g_j^I(x^{k+1}) \\ &= F(x^{k+1}) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j\in\gamma} \mu_j^{k+1} \nabla g_j^I(x^{k+1}). \end{aligned} \tag{20}$$

Since $\left|\mu_j^{k+1} g_j^I(x^{k+1})\right| \leq \varepsilon_k$ for all $j = 1, \ldots, l$, it follows that $\mu_j^{k+1} \to_K 0$ for all $j \notin \gamma$, which in turn implies

$$\sum_{j\notin\gamma} \mu_j^{k+1} \nabla g_j^I(x^{k+1}) \to_K 0.$$

Taking into account that $L_k \to 0$, we therefore also obtain $\hat{L}_k \to 0$. Furthermore, we have

$$\hat{L}_k - F(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i^P(x^{k+1}, x^{k+1}) = \sum_{j\in\gamma} \mu_j^{k+1} \nabla g_j^I(x^{k+1}).$$

In view of Lemma 2.2, we may find a subset $J_k \subseteq \gamma$ as well as a sequence $\tilde{\mu}_j^{k+1} \geq 0$ $(j \in J_k)$ such that, for each $k \in K$,

$$\hat{L}_k - F(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_j^P(x^{k+1}, x^{k+1}) = \sum_{j\in J_k} \tilde{\mu}_j^{k+1} \nabla g_j^I(x^{k+1})$$

and such that the gradients $\nabla g_j^I(x^{k+1})$ are linearly independent. Since there are only finitely many subsets J_k, we may assume once again without loss of generality that $J_k = J$ for all $k \in K$ with a fixed subset $J \subseteq \gamma$. Hence we have

$$\hat{L}_k - F(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_j^P(x^{k+1}, x^{k+1}) = \sum_{j \in J} \tilde{\mu}_j^{k+1} \nabla g_j^I(x^{k+1}). \tag{21}$$

We claim that the sequence $\{(\tilde{\mu}_j^{k+1})_{j \in J}\}_{k \in K}$ is bounded. Otherwise, suppose that $\big\|\{(\tilde{\mu}_j^{k+1})_{j \in J}\}\big\| \to_K \infty$. Since the corresponding normalized sequence is bounded, we may assume that it converges, say

$$\frac{(\tilde{\mu}_j^{k+1})_{j \in J}}{\big\|\{(\tilde{\mu}_j^{k+1})_{j \in J}\}\big\|} \to_K (\bar{\mu}_j)_{j \in J},$$

where the limit has norm one and is therefore nonzero. Furthermore, since $\tilde{\mu}_j^{k+1} \geq 0$ for all $k \in \mathbb{N}_0$, it follows immediately that $\bar{\mu}_j \geq 0$ for all $j = 1, \ldots, l$. Dividing (21) by $\big\|\{(\tilde{\mu}_j^{k+1})_{j \in J}\}\big\|$, taking the limit $k \to_K \infty$ and taking into account that all expressions on the left-hand side of (21) are bounded, we obtain

$$0 = \sum_{j \in J} \bar{\mu}_j \nabla g_j^I(x^*).$$

Since $(\bar{\mu}_j)_{j \in J}$ is nonzero, it follows that the gradients $\nabla g_j^I(x^*)$ $(j \in J)$ are positively linearly dependent. In view of CPLD, it follows that the gradients $\nabla g_j^I(x^{k+1})$ $(j \in J)$ are linearly dependent for all $k \in K$ sufficiently large, a contradiction to the choice of the index set $J = J_k$. Hence the sequence $\{\tilde{\mu}_j^{k+1}\}_{k \in K}$ converges for each $j \in J$, say

$$\tilde{\mu}_j^{k+1} \to_K \mu_j^* \geq 0 \quad \forall j \in J.$$

Using the definition of $\hat{L}_k$ together with $\tilde{L}_k \to 0$, it follows that

$$0 = \lim_{k \in K} \hat{L}_k = F(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_y g_i^P(x^*, x^*) + \sum_{j \in J} \mu_j^* \nabla g_j^I(x^*).$$

Since the boundedness of the penalty parameter also implies

$$0 = \lim_{k \in K} \min\big\{ -g_i^P(x^{k+1}, x^{k+1}), \lambda_i^{k+1} \big\} = \min\big\{ -g_i^P(x^*, x^*), \lambda_i^* \big\} \quad \forall i = 1, \ldots, m,$$

it follows that $\lambda_i^* \geq 0, g_i^P(x^*, x^*) \leq 0$, and $\lambda_i^* g_i^P(x^*, x^*) = 0$ for all $i = 1, \ldots, m$. Setting $\mu_j^* := 0$ for all $j \notin J$, it also follows that $\mu_j^* g_j^I(x^*) = 0$ for all $j = 1, \ldots, l$, and this completes the proof. $\square$

The following counterexample indicates that the EMFCQ condition is really necessary to get feasibility in the limit. Any other condition which is only slightly weaker or different does not guarantee feasibility. To this end, note that the subsequent example violates (E)MFCQ, but satisfies CPLD and even the stronger CRCQ condition.

Example 3.9 Consider the QVI with $n = 2, m = 2, l = 2$ defined by

$$F(x) := \begin{pmatrix} x_1 \\ x_2^2 \end{pmatrix},$$
$$g_1^P(y, x) := y_2 - x_1, \quad g_2^P(y, x) := x_1 - y_2,$$
$$g_1^I(y) := -y_1 - 1, \quad g_2^I(y) := -y_2 - 1.$$

Then all KKT points (x^*, λ^*, μ^*) are of the following form: $x^* = (0, 0)^T$ is unique, $\mu^* = (0, 0)^T$ is also unique, but the components of λ^* are nonunique, they only have to satisfy the relation $\lambda_1^* = \lambda_2^* \geq 0$.

Now consider the corresponding VI_k-subproblem. Computing the KKT points of VI_k using $u^k := 0$ and $\rho_k > 1$, a simple calculation shows that both $x^k := (0, 0)^T$ and $x^k := (0, -1)^T$ yield KKT points. While the former immediately yields the KKT point of the given QVI, the latter converges to the point $(0, -1)^T$ for $k \to \infty$, but this limit point is infeasible for the underlying QVI.

Note that this example satisfies CPLD (even the stronger CRCQ since we have linear constraints only), whereas (E)MFCQ is violated (there is a hidden equality constraint). $\Diamond$

The following result is the counterpart of Theorem 3.8 with an unbounded sequence of penalty parameters. Note that, motivated by the previous discussion and, in particular, by Example 3.9, it explicitly assumes feasibility of an accumulation point.

Theorem 3.10 *Let Assumption 2.3 hold, let $\{x^k\}$ be a sequence generated by Algorithm 3.1 with $\varepsilon_k \downarrow 0$, let x^* be a feasible accumulation point of $\{x^k\}$, and assume that ε_k-stationarity is satisfied with nonnegative multipliers μ^k. Suppose that $\{\rho_k\}$ is unbounded, and that the gradient vectors $\nabla g_i^P(x^*, x^*)$ $(i \in \alpha)$, $\nabla g_j^I(x^*)$ $(j \in \gamma)$ satisfy CPLD, where $\alpha := \{i \mid g_i^P(x^*, x^*) = 0\}$ and $\gamma := \{j \mid g_j^I(x^*) = 0\}$. Then there exist multipliers $(\lambda^*, \mu^*) \in \mathbb{R}^m \times \mathbb{R}^l$ such that (x^*, λ^*, μ^*) is a KKT point of the QVI.*

Proof: Most parts of the proof are similar to the one of Theorem 3.8. In particular, let L_k denote once again the expression defined in (16). Then ε_k-stationarity of (x^{k+1}, μ^{k+1}), together with the definition of λ^{k+1}, shows that both (17) and (18) hold. Let $\{x^{k+1}\}_{k \in K}$ be a subsequence converging to x^*, and recall that x^* is assumed to be feasible for the underlying QVI.

Since $\rho_k \to \infty$ by assumption, $\{u^k\}$ is bounded by construction, g_i^P is continuous and $\{x^{k+1}\}_{k \in K}$ converges to x^*, we obtain for all $k \in K$ sufficiently large that

$$\lambda_i^{k+1} = \max\left\{0, u_i^k + \rho_k g_i^P(x^{k+1}, x^{k+1})\right\} = 0 \quad \forall i \notin \alpha.$$

Hence we have

$$L_k - F(x^{k+1}) = \sum_{i \in \alpha} \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j=1}^{l} \mu_j^{k+1} \nabla g_j^I(x^{k+1})$$

for all $k \in K$ large enough. We next define a sequence of modified multipliers $\{\hat{\mu}_j^{k+1}\}_{k \in \mathbb{N}}$ as in (19) and the related sequence $\{\hat{L}_k\}_{k \in \mathbb{N}}$ as in (20). The proof of Theorem 3.8 shows that $\hat{L}_k \to_K 0$. Since

$$\hat{L}_k - F(x^{k+1}) = \sum_{i \in \alpha} \lambda_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j \in \gamma} \mu_j^{k+1} \nabla g_j^I(x^{k+1})$$

with $\lambda_i^{k+1} \geq 0$ $(i \in \alpha)$ (by construction) and $\mu_j^{k+1} \geq 0$ $(j \in \gamma)$ (by assumption), it follows from Lemma 2.2 that, for each $k \in \mathbb{N}_0$, there exist index sets $I_k \subseteq \alpha$ and $J_k \subseteq \gamma$ as well as multipliers $\tilde{\lambda}_i^{k+1} \geq 0$ $(i \in I_k)$ and $\tilde{\mu}_j^{k+1}$ $(j \in J_k)$ such that

$$\hat{L}_k - F(x^{k+1}) = \sum_{i \in I_k} \tilde{\lambda}_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j \in J_k} \tilde{\mu}_j^{k+1} \nabla g_j^I(x^{k+1}),$$

and such that the gradient vectors $\nabla_y g_i^P(x^{k+1}, x^{k+1})$ $(i \in I_k)$, $\nabla g_j^I(x^{k+1})$ are linearly independent for all $k \in \mathbb{N}$. Since there are only finitely many possible subsets I_k and J_k, we may assume without loss of generality that $I = I_k$ and $J = J_k$ for all $k \in K$ with some fixed index sets $I \subseteq \alpha$ and $J \subseteq \gamma$. Hence, for all $k \in K$, we have

$$\hat{L}_k - F(x^{k+1}) = \sum_{i \in I} \tilde{\lambda}_i^{k+1} \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j \in J} \tilde{\mu}_j^{k+1} \nabla g_j^I(x^{k+1}), \qquad (22)$$

and the corresponding gradient vectors are linearly independent. Again following the proof of Theorem 3.8 and exploiting the current CPLD assumption, we can see that the sequences $\{\tilde{\lambda}_i^{k+1}\}_{k \in K}$ $(i \in I)$ and $\{\tilde{\mu}_j^{k+1}\}_{k \in K}$ $(j \in J)$ are bounded. Without loss of generality, we may therefore assume that they converge, say

$$\tilde{\lambda}_i^{k+1} \to_K \lambda_i^* \ (i \in I) \quad \text{and} \quad \tilde{\mu}_j^{k+1} \to_K \mu_j^* \ (j \in J)$$

with some nonnegative limits λ_i^* $(i \in I)$, μ_j^* $(j \in J)$. Setting $\lambda_i^* := 0$ $(i \notin I)$ and $\mu_j^* := 0$ $(j \notin J)$, taking the limit $k \to_K \infty$ in (22), exploiting the convergence of $x^{k+1} \to_K x^*$, the continuity of all functions involved as well as the fact that $\hat{L}_k \to_K 0$, we obtain

$$F(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla_y g_i^P(x^*, x^*) + \sum_{j=1}^{l} \mu_j^* \nabla g_j^I(x^*) = 0.$$

Now, it is not difficult to see that this implies that (x^*, λ^*, μ^*) is a KKT point. □

The following example shows that the assumptions used in Theorem 3.10 are indeed weaker than the EMFCQ condition exploited in Theorem 3.2.

Example 3.11 Consider the QVI with $n = 2, m = 2, l = 0$ defined by

$$F(x) := \begin{pmatrix} x_1 \\ x_2 + 1 \end{pmatrix}, \quad g_1^P(y, x) := y_2 - x_1 \leq 0, \quad g_2^P(y, x) := x_1 - y_2 \leq 0.$$

Then an elementary calculation shows that the set of KKT points (x^*, λ^*) is given by $x^* := (0, 0)^T$ (which is unique) and $\lambda^* := (\lambda_1^*, \lambda_2^*)^T$ with $\lambda_1^* \geq 0$ arbitrary and $\lambda_2^* = 1 + \lambda_1^*$, hence there is an unbounded set of multipliers. Note that this assumption satisfies the CPLD condition, but not (E)MFCQ.

Since there are no individual constraints, the corresponding subproblem VI_k reduces to the nonlinear system of equations

$$\begin{pmatrix} x_1 \\ x_2 + 1 \end{pmatrix} + \max\left\{0, u_1^k + \rho_k(x_2 - x_1)\right\} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \max\left\{0, u_2^k + \rho_k(x_1 - x_2)\right\} \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Taking, for simplicity, $u^k = (0,0)^T$ for all $k \in \mathbb{N}_0$, it follows that the unique solution of VI_k is given by

$$x^k = \left(0, \frac{-1}{1+\rho_k}\right)^T \quad \forall k \in \mathbb{N}_0.$$

For $\rho_k \to \infty$, we see that $x^k \to x^* = (0,0)^T$, i.e., in the limit, we obtain a feasible point which is indeed the KKT point (hence solution) of the QVI, as guaranteed by Theorem 3.10. $\Diamond$

Note that the CPLD condition required in our refined convergence analysis is satisfied automatically, e.g., for the case of linear constraints (whereas MFCQ or EMFCQ might be violated for linear constraints).

We close this section by pointing out a difference between the convergence result under the EMFCQ condition in Theorem 3.2 and under the CPLD condition in Theorems 3.8, 3.10: Given a (sub-) sequence of $\{x^k\}$ converging to x^*, the former result shows that the corresponding sequence of multipliers $\{(\lambda^k, \mu^k)\}$ generated by Algorithm 3.1 converges to a vector (λ^*, μ^*) such that (x^*, λ^*, μ^*) satisfies the KKT conditions; hence our algorithm automatically produces suitable estimates of the optimal Lagrange multipliers. This is different in the latter situation, where the convergence proofs show that the sequence $\{(\lambda^k, \mu^k)\}$ generated by our algorithm is not necessarily convergent, since these multipliers had to be changed within the corresponding proofs. This observation plays some role from a numerical point of view: The most natural stopping criterion would be to check the KKT conditions for the current triple (x^k, λ^k, μ^k). However, our convergence theorems only guarantee (under certain assumptions) that the x-parts of this sequence convergence to a suitable stationary point x^*. In order to get appropriate Lagrange multiplier estimates, however, one might have to compute different approximate multipliers in order to satisfy this termination criterion. A natural choice would be to (re-) compute $(\lambda^{k+1}, \mu^{k+1})$ as a solution of the constrained linear least squares problem

$$\min \left\| F(x^{k+1}) + \sum_{i=1}^{m} \lambda_i \nabla_y g_i^P(x^{k+1}, x^{k+1}) + \sum_{j=1}^{l} \mu_j \nabla g_j^I(x^{k+1}) \right\|^2 \quad \text{s.t.} \quad \lambda \geq 0, \; \mu \geq 0. \tag{23}$$

Our actual implementation of Algorithm 3.1 will exploit this rule.

4 Solution of VI-Subproblems

The previous convergence theory for the multiplier-penalty method works, in principle, for general QVIs provided that we are able to find (at least approximately) a KKT point of the resulting VI_k that arises at each iteration. The purpose of this section is to identify some classes of QVIs where these VI-subproblems are "easy" to solve in the sense that they yield monotone or even strictly/strongly monotone VIs. To this end, we begin with a general discussion in Section 4.1. Motivated by the subclasses of QVIs introduced in [22], we then proceed by considering four different classes of QVIs for which suitable conditions can be given such that the resulting VIs are "easy".

4.1 General Discussion

Here we discuss the question under which conditions the resulting VI-subproblems from Step (S.2) of Algorithm 3.1 can be solved. To this end, we assume throughout this section that the feasible set $\mathbf{K}$ from (3) is convex. The latter is essentially a convexity assumption for the mappings g_j^I $(j = 1, \dots, l)$ and often satisfied or easy to verify.

The more difficult part is due to the mapping

$$F_k(x) := F(x) + \sum_{i=1}^{m} \max\left\{0, u_i^k + \rho_k g_i^P(x,x)\right\} \nabla_y g_i^P(x,x) \tag{24}$$

that arises in our VI$_k$-subproblems. In order to investigate the properties of this mapping, we first state the following (stronger) smoothness assumption.

Assumption 4.1 (a) The function F is continuously differentiable on $\mathbb{R}^n$.

(b) The function $g^I : \mathbb{R}^n \to \mathbb{R}^l$ is twice continuously differentiable $\mathbb{R}^n$.

(c) The function $g^P : \mathbb{R}^{2n} \to \mathbb{R}^m$ is twice continuously differentiable on $\mathbb{R}^{2n}$.

This smoothness condition is assumed to hold throughout this section.

We now begin our investigation of the mapping F_k from (24). Due to the max-term, this function is not continuously differentiable everywhere. However, it is locally Lipschitz, and we can compute (or estimate) its generalizd Jacobian (in the sense of Clarke [15]).

Proposition 4.2 *Suppose that Assumption 4.1 holds. Then the generalized Jacobian of F_k at a point $x \in \mathbb{R}^n$ satisfies the inclusion $\partial F_k(x) \subseteq F'(x) + G_k(x)$ with*

$$\begin{aligned} G_k(x) \;\subseteq\; & \sum_{i=1}^{m} \max\left\{0, u_i^k + \rho_k g_i^P(x,x)\right\} \left[\nabla_{yy}^2 g_i^P(x,x) + \nabla_{yx}^2 g_i^P(x,x)\right] \\ & + \rho_k \sum_{i=1}^{m} s_i^k \nabla_y g_i^P(x,x) \left[\nabla_y g_i^P(x,x) + \nabla_x g_i^P(x,x)\right]^T, \end{aligned}$$

where

$$s_i^k \begin{cases} = 1, & \text{if } u_i^k + \rho_k g_i^P(x,x) > 0, \\ \in [0,1], & \text{if } u_i^k + \rho_k g_i^P(x,x) = 0, \\ = 0, & \text{if } u_i^k + \rho_k g_i^P(x,x) < 0. \end{cases}$$

Proof: Note that the only nonsmoothness in the definition of the mapping F_k comes from the max-terms. Since these nonsmooth terms are compositions of a smooth and a convex (hence regular mapping in the sense of Clarke [15]) function, we can apply the corresponding sum, product, and chain rules for the Clarke generalized gradient and Jacobian, cf. [15]. This immediately yields the desired formula. □

In view of the previous result, the generalized Jacobian of F_k can be estimated by $\partial F_k(x) \subseteq F'(x) + G_k(x)$, i.e., we have the sum of the (unique) Jacobian $F'(x)$ and the (multivalued) matrix $G_k(x)$. The remainder of this section will concentrate on conditions which guarantee that the elements of $G_k(x)$ are positive semidefinite. Then the positive semidefiniteness of $F'(x)$, a standard assumption for variational inequalities, yields the monotonicity of the mapping F_k due to a result in [35]. This, in turn, means that there are plenty of methods available for the solution of the resulting VI-subproblem within our multiplier-penalty approach. Other methods might require the stronger assumption that the elements in $\partial F_k(x)$ are positive definite, but this is also implied by the positive semidefiniteness of $G_k(x)$ provided that the Jacobian $F'(x)$ is positive definite. Therefore, the question under which we get "easy" VI-subproblems within the multiplier-penalty method boils down to the question under which the elements in $G_k(x)$ are positive semidefinite. This is precisely the question we want to answer in our subsequent analysis.

A first and general result in this direction can be obtained by writing

$$h_i(x) := g_i^P(x,x) \quad \forall i = 1,\dots,m. \tag{25}$$

Then we may simplify the previous estimate of the set $G_k(x)$ as

$$\begin{aligned} G_k(x) \;\subseteq\; & \sum_{i=1}^m \max\big\{0, u_i^k + \rho_k h_i(x)\big\} J_x\big(\nabla_y g_i^P(x,x)\big) \\ & + \rho_k \sum_{i=1}^m s_i^k \nabla_y g_i^P(x,x) J h_i(x) \end{aligned}$$

since

$$\begin{aligned} Jh_i(x) &= \big(\nabla_y g_i^P(x,x) + \nabla_x g_i^P(x,x)\big)^T \quad \text{and} \\ J_x\big(\nabla_y g_i^P(x,x)\big) &= \nabla^2_{yy} g_i^P(x,x) + \nabla^2_{yx} g_i^P(x,x), \end{aligned}$$

where the meaning of the expression $\nabla^2_{yx} g_i^P(x,x)$ should be clear from the context. This yields the following result.

Theorem 4.3 *Suppose that Assumption 4.1 holds. Assume further that the matrices*

$$\nabla_y g_i^P(x,x) J h_i(x) \quad \forall i : u_i^k + \rho_k h_i(x) \geq 0$$

and

$$J_x\big(\nabla_y g_i^P(x,x)\big) \quad \forall i : u_i^k + \rho_k h_i(x) > 0$$

are positive semidefinite. Then all elements in $G_k(x)$ are positive semidefinite.

Proof: The above representation of $G_k(x)$ together with our assumptions imply that each element of $G_k(x)$ is a nonnegative sum of positive semidefinite matrices and, therefore, positive semidefinite itself. □

The following subsections consider special classes of QVIs and verify either directly or by using the condition from Theorem 4.3 that the matrices in $G_k(x)$ are positive semidefinite under suitable assumptions.

4.2 Jointly Convex GNEPs

We now examine the case where we get a QVI by reformulation of a generalized Nash equilibrium problem (GNEP). In such a GNEP, we have the players $\nu = 1, \dots N$, each controlling his variables $x^\nu \in \mathbb{R}^{n_\nu}$ and trying to solve their personal optimization problem

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t.} \quad g^\nu(x^\nu, x^{-\nu}) \leq 0, \; h^\nu(x^\nu) \leq 0 \tag{26}$$

with given functions $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$, $g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, and $h^\nu : \mathbb{R}^n \to \mathbb{R}^{p_\nu}$ for $\nu = 1, \dots, N$, where $n := n_1 + \dots + n_N$ and $x^{-\nu}$ subsumes the variables of all players except x^ν. This GNEP is called *player-convex* if all component functions h_i^ν are convex, and $\theta_\nu(\cdot, x^{-\nu})$ as well as each of the components $g_i^\nu(\cdot, x^{-\nu})$ are convex in x^ν (for every fixed $x^{-\nu}$). The GNEP is called *jointly-convex* if $g^1 = g^2 = \dots = g^N =: g^{NE}$ and g^{NE} is convex in the entire variable x (NE stands for Nash equilibrium).

The player-convex GNEP is the most general type of Nash equilibrium problems that one typically finds in the literature. It is known to be equivalent to the QVI defined by

$$F(x) := \begin{pmatrix} \nabla_{x^1}\theta_1(x)) \\ \vdots \\ \nabla_{x^N}\theta_N(x) \end{pmatrix} \tag{27}$$

and

$$K(x) := K_1(x^{-1}) \times K_2(x^{-2}) \times \dots \times K_N(x^{-N}),$$

where

$$K_\nu(x^{-\nu}) := \left\{ x^\nu \mid g^\nu(x^\nu, x^{-\nu}) \leq 0, \; h^\nu(x^\nu) \leq 0 \right\}$$

denotes the (parameterized) feasible set of player ν, $\nu = 1, \dots, N$, cf. [27] and the survey papers [19, 24] on GNEPs. Hence the mappings g^P and g^I in our general QVI-notation are given by

$$g^P(y, x) := \begin{pmatrix} g^1(y^1, x^{-1}) \\ \vdots \\ g^N(y^N, x^{-N}) \end{pmatrix}, \qquad g^I(y) := \begin{pmatrix} h^1(y^1) \\ \vdots \\ h^N(y^N) \end{pmatrix},$$

repectively. In particular, in the case of a jointly-convex GNEP, we therefore have

$$g^P(y, x) := \begin{pmatrix} g^{NE}(y^1, x^{-1}) \\ \vdots \\ g^{NE}(y^N, x^{-N}) \end{pmatrix} \quad \text{with} \quad g^{NE}(x) := \begin{pmatrix} g_1^{NE}(x) \\ \vdots \\ g_m^{NE}(x) \end{pmatrix} \tag{28}$$

with g^{NE} being repeated N times.

For the rest of this section, we consider only this jointly-convex case. Our aim is to show that the corresponding VI-subproblems arising in Algorithm 3.1 are monotone in this case. First note that g^P is a mapping from $\mathbb{R}^{2n}$ to $\mathbb{R}^{mN}$ in this case, where $m := m_1 = \dots = m_N$. The partial (with respect to y) transposed Jacobian of g^P is a matrix in $\mathbb{R}^{n \times mN}$ and has the following block structure:

$$\nabla_y g^P(x,x) = \left(\begin{array}{ccc|c|ccc} \nabla_{x^1} g_1^{NE}(x) & \cdots & \nabla_{x^1} g_m^{NE}(x) & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & \vdots & & \vdots \\ \vdots & & \vdots & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \nabla_{x^N} g_1^{NE}(x) & \cdots & \nabla_{x^N} g_m^{NE}(x) \end{array} \right), \tag{29}$$

where

$$\nabla_{x^\nu} g_i^{NE}(x) \in \mathbb{R}^{n_\nu \times 1} \quad \forall i = 1, \ldots, m, \; \forall \nu = 1, \ldots, N.$$

Exploiting this structure, we get the following result.

Proposition 4.4 *Consider the jointly-convex GNEP, let g^P be defined by (28), $\rho > 0$ an arbitrary parameter, and assume that $u_i = u_{m+i} = \ldots = u_{(N-1)m+i}$ holds for all $i = 1, \ldots, m$. Then the function*

$$G(x) := \sum_{i=1}^{mN} \max\left\{0, u_i + \rho g_i^P(x,x)\right\} \nabla_y g_i^P(x,x)$$

is monotone on $\mathbb{R}^n$.

Proof: Exploiting the representation (29), we can rewrite the mapping G in the following way:

$$\begin{aligned} G(x) &= \sum_{i=1}^{mN} \max\left\{0, u_i + \rho g_i^P(x,x)\right\} \nabla_y g_i^P(x,x) \\ &= \nabla_y g^P(x,x) \max\left\{0, u + \rho g^P(x,x)\right\} \\ &= \begin{pmatrix} \sum_{i=1}^m \nabla_{x^1} g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} \\ \vdots \\ \sum_{i=1}^m \nabla_{x^N} g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} \end{pmatrix}, \end{aligned}$$

where the maximum in the second line is taken component-wise, and the third line exploits the assumption regarding the vector u. This immediately yields

$$\begin{aligned} &\big(G(x) - G(z)\big)^T (x - z) \\ &= \begin{pmatrix} \sum_{i=1}^m \nabla_{x^1} g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} - \sum_{i=1}^m \nabla_{x^1} g_i^{NE}(z) \max\left\{0, u_i + \rho g_i^{NE}(z)\right\} \\ \vdots \\ \sum_{i=1}^m \nabla_{x^N} g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} - \sum_{i=1}^m \nabla_{x^N} g_i^{NE}(z) \max\left\{0, u_i + \rho g_i^{NE}(z)\right\} \end{pmatrix}^T (x - z) \\ &= \left(\sum_{i=1}^m \nabla g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} - \sum_{i=1}^m \nabla g_i^{NE}(z) \max\left\{0, u_i + \rho g_i^{NE}(z)\right\} \right)^T (x - z) \\ &= \sum_{i=1}^m \left(\nabla g_i^{NE}(x) \max\left\{0, u_i + \rho g_i^{NE}(x)\right\} - \nabla g_i^{NE}(z) \max\left\{0, u_i + \rho g_i^{NE}(z)\right\} \right)^T (x - z) \end{aligned}$$

$$\geq \quad 0$$

for all $x, z \in \mathbb{R}^n$. Here, the nonnegativity comes from the following observation: Since each g_i^{NE} is convex by assumption, it follows that $\max\{0, u_i + \rho g_i^{NE}(x)\}$ is nonnegative and convex, too. Hence $\frac{1}{2\rho} \max^2 \{0, u_i + \rho g_i^{NE}(x)\}$ is still convex, and differentiable due to the assumed differentiability of g_i^{NE}. Hence the gradient of this convex function is a montone mapping; but this gradient is given by

$$\nabla\left(\frac{1}{2\rho}\max^2\{0, u_i + \rho g_i^{NE}(x)\}\right) = \max\{0, u_i + \rho g_i^{NE}(x)\}\nabla g_i^{NE}(x),$$

and this concludes the proof. □

A simple consequence of this proposition is the following monotonicity result.

Theorem 4.5 *Consider the jointly-convex GNEP with F and g^P being defined by (27) and (28), respectively. Then the corresponding mapping F_k from (24) with an arbitrary $\rho_k > 0$ is monotone provided that $F'(x)$ is positive semidefinite and $u_i^k = u_{m+i}^k = \ldots = u_{(N-1)m+i}^k$ holds for all $i = 1, \ldots, N$.*

Note that the positive semidefiniteness of the matrix $F'(x)$, for which the name "Jaco-Hessian" was coined in [18], is a very common condition in the framework of (jointly-convex) GNEPs. Furthermore, the assumption on the choice of the parameters u^k is rather natural, it just requires those components to be equal for which the corresponding components of the mapping g^P from (28) are identical, too.

The following counterexample shows that monotonicity of F_k cannot be expect for GNEPs which are not jointly convex. More precisely, this counterexample shows that the monotonicity of F_k may not hold even in the case where the functions g^ν are convex in the entire variable x.

Example 4.6 Consider a GNEP with $N = 2$ players, and let x_i denote the (single) variable of player i. Let the feasible sets of the two players be given by

$$\begin{aligned} K_1(x_2) &:= \{x_1 \mid g^1(x) := x_1 + x_2 \leq 0\}, \\ K_2(x_1) &:= \{x_2 \mid g^2(x) := x_1 - x_2 \leq 0\}. \end{aligned}$$

Then the mapping g^P associated to the corresponding QVI is given by

$$g^P(y, x) := \begin{pmatrix} g^1(y^1, x^{-1}) \\ g^2(y^2, x^{-2}) \end{pmatrix} = \begin{pmatrix} y_1 + x_2 \\ x_1 - y_2 \end{pmatrix},$$

hence we have

$$\nabla_y g^P(x, x) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We verify the nonmonotonicity of the mapping

$$G(x) \;=\; \sum_{i=1}^{m} \max\{0, u_i + \rho g_i^P(x, x)\}\nabla_y g_i^P(x, x)$$

$$= \nabla_y g^P(x,x) \max\big\{0, u + \rho g^P(x,x)\big\}.$$

Take $x = 0, u = 0, \rho > 0$ arbitrary and let $z \in \mathbb{R}^2$ be unspecified for the moment. Then

$$\begin{aligned} G(x) &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{since both max-terms vanish, and} \\ G(z) &= \rho \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \max\{0, z_1 + z_2\} \\ \max\{0, z_1 - z_2\} \end{pmatrix} = \begin{pmatrix} \rho \max\{0, z_1 + z_2\} \\ -\rho \max\{0, z_1 - z_2\} \end{pmatrix}. \end{aligned}$$

This implies

$$\begin{aligned} \big(G(x) - G(z)\big)^T (x - z) &= G(z)^T z \\ &= \rho z_1 \max\{0, z_1 + z_2\} - \rho z_2 \max\{0, z_1 - z_2\}. \end{aligned}$$

In particular, for $z = (-1, 2)^T$, we therefore obtain

$$\big(G(x) - G(z)\big)^T (x - z) = -\rho \quad \forall \rho > 0,$$

hence $F_k(x) = F(x) + G(x)$ is nonmonotone regardless of the properties of F (e.g., even for strongly monotone mappings F), since ρ can be chosen arbitrarily large. $\Diamond$

The following example also shows that Theorem 4.5 does not hold for jointly convex GNEPs without the additional assumption on the choice of the vector u^k.

Example 4.7 Consider again a GNEP with two players, and let x_i denote the (single) variable of player i. Let the feasible sets be given by

$$\begin{aligned} K_1(x_2) &:= \big\{x_1 \mid g^1(x) := x_1 + x_2 \leq 0\big\}, \\ K_2(x_1) &:= \big\{x_2 \mid g^2(x) := x_1 + x_2 \leq 0\big\}. \end{aligned}$$

Since $g := g^1 = g^2$ is a linear function, it is convex in all variables, furthermore, the same function is used for both players, hence we are in the situation of a jointly convex GNEP. The corresponding QVI-mapping g^P is given by

$$g^P(y,x) = \begin{pmatrix} g(y^1, x^{-1}) \\ g(y^2, x^{-2}) \end{pmatrix} = \begin{pmatrix} y_1 + x_2 \\ x_1 + y_2 \end{pmatrix}.$$

This implies

$$\nabla_y g^P(y,x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This, in turn, yields

$$\begin{aligned} G(x) &= \nabla_y g^P(x,x) \max\big\{0, u + \rho g^P(x,x)\big\} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \max\{0, u_1 + \rho(x_1 + x_2)\} \\ \max\{0, u_2 + \rho(x_1 + x_2)\} \end{pmatrix} \\ &= \begin{pmatrix} \max\{0, u_1 + \rho(x_1 + x_2)\} \\ \max\{0, u_2 + \rho(x_1 + x_2)\} \end{pmatrix}. \end{aligned}$$

Let $x := (0,0)^T, z := (1,-2)^T, \rho > 0$ arbitrary and $u = (u_1, u_2)^T$ unspecified for the moment. Then

$$G(x) = \begin{pmatrix} \max\{0, u_1\} \\ \max\{0, u_2\} \end{pmatrix} \quad \text{and} \quad G(z) = \begin{pmatrix} \max\{0, u_1 - \rho\} \\ \max\{0, u_2 - \rho\} \end{pmatrix}.$$

This implies

$$\begin{aligned} &\big(G(x) - G(z)\big)^T (x - z) \\ = &\begin{pmatrix} \max\{0, u_1\} - \max\{0, u_1 - \rho\} \\ \max\{0, u_2\} - \max\{0, u_2 - \rho\} \end{pmatrix}^T \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\ = &-\max\{0, u_1\} + \max\{0, u_1 - \rho\} + 2\max\{0, u_2\} - 2\max\{0, u_2 - \rho\}. \end{aligned}$$

In particular, for any vector $u = (u_1, u_2)^T$ with $0 < u_1 < \rho$ and $u_2 = 0$, we obtain

$$\big(G(x) - G(z)\big)^T (x - z) = -u_1 < 0.$$

This shows that G is not monotone, hence $F_k = F + G$ is also nonmonotone in general. ◊

4.3 The Moving Set Case

One of the most studied instances of QVIs considers feasible sets of the form

$$K(x) := c(x) + Q$$

for a continuously differentiable function $c : \mathbb{R}^n \to \mathbb{R}^n$ and a fixed set $Q \subseteq \mathbb{R}^n$ which we assume to be given by

$$Q := \{x \in \mathbb{R}^n \mid q(x) \leq 0\},$$

where $q : \mathbb{R}^n \to \mathbb{R}^m$ is twice continuously differentiable with each q_i being convex; in particular, Q is then a convex set. This class of QVIs is sometimes called the "moving set case" since, geometrically, the fixed set Q moves along the trajectory defined by the mapping c.

Monotonicity of the corresponding subproblems in Algorithm 3.1 can be guaranteed under the conditions of the following result.

Theorem 4.8 *Consider a QVI of the moving set case, and let $x \in \mathbb{R}^n$. Further assume that c is given by $c(x) = \big(c_1(x_1), \dots, c_m(x_m)\big)^T$ such that $c_i'(x_i) < 1$ for all $i = 1, \dots, m$. Then all elements from the set $G_k(x)$ (cf. Theorem 4.2) are positive semidefinite.*

Proof: We prove the assertion by verifying the sufficient conditions from Theorem 4.3. To this end, we first observe that the moving set case corresponds to our general setting using the functions

$$g_i^P(y, x) := q_i(y - c(x)), \quad h_i(x) := q_i(x - c(x)).$$

We therefore have

$$\nabla_y g_i^P(x,x) = \nabla q_i(x - c(x)) \quad \text{and} \quad \nabla_x g_i^P(x,x) = -Jc(x)^T \nabla q_i(x - c(x)).$$

This implies

$$\begin{aligned} Jh_i(x) &= \big(\nabla_y g_i^P(x,x) + \nabla_x g_i^P(x,x)\big)^T \\ &= \nabla q_i(x - c(x))^T - \nabla q_i(x - c(x))^T Jc(x) \\ &= \nabla q_i(x - c(x))^T \big(I - Jc(x)\big). \end{aligned}$$

Consequently, we get

$$\nabla_y g_i^P(x,x) Jh_i(x) = \nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T \big(I - Jc(x)\big).$$

In view of our assumptions, $S := I - Jc(x)$ is a positive definite diagonal matrix which we can factorize as $S = DD$ with another positive definite diagonal matrix D. We then obtain

$$\begin{aligned} v^T \nabla_y g_i^P(x,x) Jh_i(x) v &= v^T \nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T DDv \\ &= v^T DD^{-1} \nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T DDv \\ &\overset{w:=Dv}{=} w^T D^{-1} \nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T Dw \\ &\geq 0 \end{aligned}$$

for all $v \in \mathbb{R}^n$ since $\nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T$ and, therefore, also the similar matrix $D^{-1} \nabla q_i(x - c(x)) \nabla q_i(x - c(x))^T D$ are symmetric positive semidefinite.

A direct computation also shows that

$$J_x\big(\nabla_y g_i^P(x,x)\big) = \big(I - Jc(x)\big)^T \nabla^2 q_i(x - c(x)).$$

Since q_i is convex, its Hessian is symmetric positive semidefinite everywhere. Hence, similar to the previous case, one can also show that $J_x\big(\nabla_y g_i^P(x,x)\big)$ is positive semidefinite. The statement therefore follows directly from Theorem 4.3. $\square$

We note that the previous proof goes through for the case where $I - Jc(x)$ is symmetric positive definite, in particular, the functions c_i can depend on the whole vector x, but the "off-diagonal elements" must be sufficiently small. However, the symmetry plays a central role here. This is indicated by the following counterexample.

Example 4.9 Let $n = 2$, $q(x) := x_1$, and $c(x) := (\epsilon(x_1 + x_2), 0)^T$ with $\epsilon > 0$. Then we have

$$I - Jc(x) = I - \begin{pmatrix} \varepsilon & \varepsilon \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1-\varepsilon & -\varepsilon \\ 0 & 1 \end{pmatrix}.$$

Hence this matrix may be viewed as an arbitrary small perturbation of a diagonal matrix. Moreover, it turns out to be positive definite for all sufficiently small $\varepsilon > 0$; more precisely, for $\varepsilon < 2/3$, we obtain for all nonzero $v \in \mathbb{R}^2$

$$\begin{aligned} v^T\big(I - Jc(x)\big)v &= (1-\varepsilon)v_1^2 - \varepsilon v_1 v_2 + v_2^2 \\ &\geq (1-\varepsilon)v_1^2 + (1-\varepsilon)v_2^2 - \varepsilon |v_1|\, |v_2| \end{aligned}$$

$$
\begin{aligned}
&= (1-\varepsilon)\big(v_1^2 + v_2^2 - \frac{\varepsilon}{1-\varepsilon}|v_1|\,|v_2|\big) \\
&\geq (1-\varepsilon)\big(v_1^2 + v_2^2 - 2|v_1|\,|v_2|\big) \\
&= (1-\varepsilon)\big(|v_1| - |v_2|\big)^2 \\
&\geq 0,
\end{aligned}
$$

and it is not difficult to see that at least one of the inequalities must be strict for every $v \neq 0$.

Nevertheless, the nonsymmetry of $Jc(x)$ destroys the monotonicity of F_k. To see this, it suffices to verify the nonmonotonicity of the corresponding mapping

$$
G(x) := \sum_{i=1}^{m} \max\big\{0, u_i^k + \rho_k g_i^P(x,x)\big\}\nabla_y g_i^P(x,x).
$$

In this example, we have $m = 1$ and

$$
g^P(y,x) = q(y - c(x)) = y_1 - \varepsilon x_1 - \varepsilon x_2 \quad \Longrightarrow \quad \nabla_y g^P(x,x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}
$$

which, in turn, yields the representation

$$
G(x) = \max\big\{0, u_1^k + \rho_k((1-\varepsilon)x_1 - \varepsilon x_2)\big\}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.
$$

Using

$$
\varepsilon < 1,\ \rho_k = 1,\ u_1^k = 0,\ x = (1,0)^T,\ z = (0, z_2)^T \text{ with } z_2 < \frac{\varepsilon - 1}{\varepsilon}\ (< 0),
$$

we obtain

$$
\begin{aligned}
&\big(G(x) - G(z)\big)^T(x - z) \\
&\quad= \big(\max\big\{0, u_1^k + \rho_k((1-\varepsilon)x_1 - \varepsilon x_2)\big\} - \max\big\{0, u_1^k + \rho_k((1-\varepsilon)z_1 - \varepsilon z_2)\big\}\big)(x_1 - z_1) \\
&\quad= \max\{0, 1-\varepsilon\} - \max\{0, -\varepsilon z_2\} \\
&\quad= 1 - \varepsilon + \varepsilon z_2 \\
&\quad< 0,
\end{aligned}
$$

where the last inequality follows from the choice of z_2. This indicates that we cannot expect F_k to be monotone in this case. $\Diamond$

The corresponding result in [22] is somewhat different from our result: Both are similar since, in some sense, they require $\|Jc(x)\|$ to be small. On the other hand, [22] does not need a symmetry assumption for this matrix, but requires $F'(x)$ to be nonsingular in order to provide conditions for the nonsingularity of a suitable matrix related to the interior-point-type method discussed in that reference.

4.4 Bilinear Case

In this subsection, we consider the case where the constraints g^P and g^I are defined as follows

$$g^P(y,x) := \begin{pmatrix} x^TQ_1y - \gamma_1 \\ \vdots \\ x^TQ_my - \gamma_m \end{pmatrix} \quad \text{and} \quad g^I(y) := \begin{pmatrix} q_1(y) \\ \vdots \\ q_l(y) \end{pmatrix}, \tag{30}$$

where each mapping $q_i : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and convex, $Q_i \in \mathbb{R}^{n\times n}$ are symmetric positive semidefinite matrices for all $i = 1, \ldots, m$, and $\gamma_i \in \mathbb{R}$ are given real numbers. In this case, we get the following result.

Theorem 4.10 *Let $x \in \mathbb{R}^n$ arbitrarily given, and let g^P be defined as in* (30)*. Then all elements in $G_k(x)$ are positive semidefinite.*

Proof: For an arbitrary $i \in \{1, \ldots, m\}$, we have

$$\nabla_y g_i^P(x,x) Jh_i(x) = (Q_ix)(2Q_ix)^T \quad \text{and} \quad J_x(\nabla_y g_i^P(x,x)) = Q_i.$$

Since all these matrices are positive semidefinite (either as a dyadic product of a vector with itself or by assumption), the statement follows immediately from Theorem 4.3. □

4.5 Binary Constraints

In this subclass, the constraints g_i^P each depend only on a single pair (y_j, x_j) of variables for some $j \in \{1, \ldots, n\}$. Since this index j depends on the particular component i, we write $j(i)$ thoughout this section. This means that the QVIs discussed here are defined by a feasible set $K(x)$ given as follows:

$$K(x) := \{y \in \mathbb{R}^n \mid g_i^P(y_{j(i)}, x_{j(i)}) \leq 0,\ i = 1, \ldots, m\}. \tag{31}$$

Individual constraints are also allowed, but are suppressed here since they play no role for the monotonicity of the operator F_k.

At first we will give a general result for this class of problems and then we investigate some special cases.

Theorem 4.11 *Let g^P be defined as in* (31)*, and let $x \in \mathbb{R}^n$ be given. Assume that*

$$\nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)}) \cdot h_i'(x_{j(i)}) \geq 0 \quad \forall i = 1, \ldots, m \tag{32}$$

and

$$\nabla_{x_{j(i)}} \Big(\nabla_{y_{j(i)}} g_i(x_{j(i)}, x_{j(i)})\Big) \geq 0 \quad \forall i = 1, \ldots, m \tag{33}$$

hold. Then all elements in $G_k(x)$ are positive semidefinite.

Proof: Again, we verify the statement using the general criterion from Theorem 4.3 to prove the statement. First note that the gradient of g_i^P with respect to y is given by

$$\nabla_y g_i^P(x,x) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)}) \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where the only possibly nonzero entry is in position $j(i)$. Hence the Jacobian $J_x\big(\nabla_y g_i^P(x,x)\big)$ is a matrix with the entry $\nabla_{x_{j(i)}}\big(\nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)})\big)$ at the diagonal position $(j(i), j(i))$ and all other entries equal to zero. In view of (33), this matrix is positive semidefinite.

Furthermore, we obtain

$$Jh_i(x) = \big(0, \dots, 0, \nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)}) + \nabla_{x_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)}), 0, \dots, 0\big)$$

with the nonzero entry also at position $j(i)$. Hence the product $\nabla_y g_i^P(x,x) Jh_i(x)$ is again a diagonal matrix with all elements being zero except the one at position $j(i)$ which is given by

$$\nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)})\big(\nabla_{y_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)}) + \nabla_{x_{j(i)}} g_i^P(x_{j(i)}, x_{j(i)})\big).$$

But this expression is nonnegative due to (32). Altogether, the statement therefore follows from Theorem 4.3. □

We now consider a special case of constraints with variable right hand side that are also binary constraints, namely a feasible set $K(x)$ that is defined by

$$K(x) := \big\{ y \in \mathbb{R}^n \,|\, g_i^P(y,x) := q_i(y_{j(i)}) - c_i(x_{j(i)}) \le 0\,,\; i = 1, \dots, m \big\}\,, \qquad (34)$$

where all functions q_i are convex and twice continuously differentiable, whereas the functions c_i are assumed to be continuously differentiable only. Then we have

$$\begin{aligned} \nabla_{y_{j(i)}} g_i^P(y,x) &= q_i'(y_{j(i)}), \\ \nabla_{x_{j(i)}}\big(\nabla_{y_{j(i)}} g_i(x_{j(i)}, x_{j(i)})\big) &= q_i''(x_{j(i)}), \\ h_i'(x_{j(i)}) &= q_i'(x_{j(i)}) - c_i'(x_{j(i)}). \end{aligned}$$

Since $q_i''(x_{j(i)}) \ge 0$ by the assumed convexity of q_i, we immediately obtain the following result from Theorem 4.11.

Corollary 4.12 *Let g_i be defined as in (34), and let $x \in \mathbb{R}^n$ be arbitrarily given. Assume that*

$$q_i'(x_{j(i)})\big(q_i'(x_{j(i)}) - c_i'(x_{j(i)})\big) \ge 0 \quad \forall i = 1, \dots, m. \qquad (35)$$

Then all elements in $G_k(x)$ are positive semidefinite.

Note that condition (35) holds automatically if q_i is monotonically increasing with $q_i' \geq c_i'$ for all $i = 1, \ldots, m$.

There is a whole class of problems, where the feasible set is of the form (34), namely problems with so called box constraints (with variable right hand side). For this class of problems, the feasible set is defined by

$$K(x) := \{y \in \mathbb{R}^n \mid y_i - \alpha_i x_i - \gamma_i \leq 0\,,\; i = 1, \ldots, m := n\}\,. \tag{36}$$

Taking into account the very special structure of these constraints, we directly obtain the following consequence of Corollary 4.12.

Corollary 4.13 *Let g_i^P be defined as in* (36)*, and let $x \in \mathbb{R}^n$ be arbitrarily given. Assume that $\alpha_i \leq 1$ for all $i = 1, \ldots, n$. Then all elements in $G_k(x)$ are positive semidefinite.*

5 Numerical Results

In this section, we want to present some numerical results for the multiplier-penalty method, also compared to some other possible methods. The aim is mainly to verify the reliability of the method, not to show its efficiency. In general, since we have to solve VI-subproblems at each iteration, the method takes more time than other methods which only need to solve linear systems of equations. Of course, one could improve the efficiency significantly by using a fancy solver for the resulting VI-subproblems, but this solver then depends very much on the particular class of QVIs which we are solving. We eventually decided to apply a (nonmonotone) semismooth Newton method in order to compute an (inexact) KKT point of the VI-subproblems.

The parameters chosen for Algorithm 3.1 are

$$\gamma := 5,\;\; \tau := 0.9,\;\; u^{\max} := 10^{10} e,$$

where $e := (1, \ldots, 1)^T$. Furthermore, we take $\varepsilon_k := 10^{-8}$ for all $k \in \mathbb{N}_0$, hence the inexact KKT points are computed with a relatively high accuracy at each iteration. Our termination criterion is given by

$$\left\| \begin{pmatrix} F(x^k) + \sum_{i=1}^m \lambda_i^k \nabla_y g_i^P(x^k, x^k) + \sum_{j=1}^l \mu_j^k \nabla g_j^I(x^k) \\ \min\left\{\mu^k, -g^I(x^k)\right\} \\ \min\left\{\lambda^k, -g^P(x^k, x^k)\right\} \end{pmatrix} \right\|_\infty \leq \text{tol} := 10^{-4},$$

where the min-operator is taken component-wise, i.e., we check whether the KKT conditions are satisfied within a certain accuracy. Since these KKT conditions depend on the current multiplier estimates, we do not use the multipliers from the statement of Algorithm 3.1, but, for reasons explained at the end of Section 3.3, re-compute these multipliers by solving the linear least-squares problem (23) at the end of each outer iteration.

The starting point is $(x^0, \lambda^0, \mu^0) := (0, 0, 0)$ for all test examples. We further use the initial penalty parameter $\rho_0 := 1$ and use $u^0 := 0$, while u^{k+1} in (S.4) is updated by $u^{k+1} := \min\{\lambda^{k+1}, u^{\max}\}$. The test examples themselves are chosen from the QVILIB test problem collection [21], their names and dimension are given in Table 1. This table contains the results (in terms of number of iterations) of four different methods:

(a) semismooth Newton method: **Semi**

(b) multiplier-penalty method: **Mult-Pen**

(c) multiplier-penalty method with semismooth Newton preprocessor: **Mult-Pen-Pre**

(d) potential-reduction algorithm: **PRA** (note that this solver uses a different starting vector, cf. [22] for more details).

The semismooth Newton method is a simple implementation of the algorithm presented in [20]; this method applies the standard semismooth Newton method to a nonsmooth reformulation of the KKT conditions of the QVI based on the Fischer-Burmeister function. The method is globalized by an Armijo-type line search. Method (b) is the multiplier-penalty method as described before. The third method is again our multiplier-penalty algorithm, but with the semismooth Newton method as a preprocessor. This means that we apply at most 20 iterations of the semismooth Newton method to each test problem; some (easy) test problems are actually solved by this preprocessor. We switch from this preprocessor to the multiplier-penalty method if either the preprocessor generates a stepsize that is too small, or when there is no significant progress (in terms of the reduction of the function value of the corresponding merit function) or if the maximum number of 20 iterations for this preprocessor is reached. Method (d) is the potential-reduction method from [22] which we view as the best method that is currently available for the solution of QVIs.

Table 1 shows that the semismooth Newton method **Semi** is not a reliable choice and leads to many failures. We admit that the implementation is a very simple one, with no enhancements like a nonmonotonicity strategy etc. Nevertheless, the overall picture is similar to the one given in [20]. However, it is interesting to see that the semismooth Newton method has no problems in solving the two test examples `BiLin*` for which the multiplier-penalty algorithm has difficulties.

The reliability of the other three methods **Mult-Pen**, **Mult-Pen-Pre**, and **PRA** is much better; these methods produce 3, 2, and 4 failures on the test set, respectively, hence from that point of view, the two multiplier-penalty approaches are even better than the potential-reduction interior-point algorithm from [22]. Despite this fact, it is interesting to see that these methods sometimes have a completely different behaviour on various test examples: The multiplier-penalty technique has difficulties in solving the `BiLin*` examples, which can be solved easily by the potential-reduction code, whereas the latter has severe problems in solving the four `RHS*` examples which turn out to be extremely easy for the multiplier-penalty technique.

We also note that we start the multiplier-penalty phase within the method **Mult-Pen-Pre** by using the final point provided from the preprocessing phase. This might not be a good idea in general when the preprocessing was stopped due to some difficulties in the semismooth Newton method, because this indicates some

test problem	n	m	l	**Semi**	**Mult-Pen**	**Mult-Pen-Pre**	**PRA**
BiLin1A	5	3	10	17	–	17 + 0	13
BiLin1B	5	3	10	36	–	–	10
Box1A	5	10	0	6	37	6 + 0	8
Box1B	5	10	0	–	–	–	–
Box2A	500	1000	0	–	11	4 + 13	19
Box2B	500	1000	0	75	13	20 + 13	23
Box3A	500	1000	0	–	11	20 + 29	18
Box3B	500	1000	0	–	12	6 + 15	25
KunR11	2500	2500	0	–	11	0 + 11	14
KunR12	4900	4900	0	–	11	0 + 11	22
KunR21	2500	2500	0	–	1	0 + 1	21
KunR22	4900	4900	0	–	1	0 + 1	23
KunR31	2500	2500	0	–	23	0 + 23	–
KunR32	4900	4900	0	–	21	0 + 21	–
MovSet1A	5	1	0	8	31	8 + 0	10
MovSet1B	5	1	0	–	31	2 + 41	16
MovSet2A	5	1	0	7	36	7 + 0	12
MovSet2B	5	1	0	44	44	20 + 53	36
MovSet3A1	1000	1	0	10	3	10 + 0	11
MovSet3A2	2000	1	0	11	3	11 + 0	11
MovSet3B1	1000	1	0	13	3	13 + 0	11
OutKZ31	62	62	62	–	7	0 + 7	18
OutKZ41	82	82	82	–	8	0 + 8	20
OutZ40	2	2	4	5	1	5 + 0	8
OutZ41	2	2	4	5	1	5 + 0	18
OutZ42	4	4	4	16	4	16 + 0	8
OutZ43	4	4	0	34	4	20 + 4	8
OutZ44	4	4	0	12	4	12 + 0	8
RHS1A1	200	199	0	1	1	1 + 0	87
RHS1B1	200	199	0	1	1	1 + 0	–
RHS2A1	200	199	0	1	1	1 + 0	71
RHS2B1	200	199	0	1	1	1 + 0	84
Scrim21	2400	2400	2400	–	24	20 + 24	17
Scrim22	4800	4800	4800	–	24	20 + 24	17

Table 1: Number of iterations for four different methods. The entries of the form $a + b$ for two numbers $a, b \in \mathbb{N}_0$ in column **Mult-Pen-Pre** indicate the number of preprocessing semismooth Newton steps and the subsequent number (if any) of multiplier-penalty iterations, respectively.

singularity-type problems, hence, from a practical point of view, it might be better to start the multiplier-penalty phase with the original starting point. In our table, however, we would re-obtain the results from the column **Mult-Pen** and therefore would not get new insights. Apart from this, it is interesting to observe that, many times, when we start the multiplier-penalty method using the final point from the semismooth Newton preprocessor, the multiplier-penalty technique has no difficulties in solving the resulting VI_k-subproblem which indicates that these subproblems have significantly better properties than the linearized problems arising from the semismooth Newton method. On the other hand, there exist also a couple of examples where it was not easy to solve the VI_k-subproblems. In particular, all $(3 + 2 = 5)$ failures within the two multiplier-penalty methods are due to the fact that a minimum step size was reached within the solution procedure for the VI_k-subproblems.

Altogether, it follows that the multiplier-penalty technique yields a very reliable solver for QVIs that should at least be viewed as a promising alternative to existing methods, especially since its behaviour is sometimes very much different from the one of existing codes.

6 Final Remarks

In this paper, we have studied the global convergence properties of a multiplier-penalty method for the solution of quasi-variational inequalities. Based on some recent developments for this class of methods within the framework of optimization problems, we were able to improve the existing results for QVIs considerably. In principle, it is possible to slightly improve our results by using another constraint qualification introduced in [3]. From our point of view, however, a more interesting future project would be to generalize the convergence theory under an error bound condition like the one used in [34] for optimization problems. There are two questions which arise in this context: a) Besides the error bound itself, the paper [34] requires an additional technical condition which needs to be translated to and interpreted for QVIs in a suitable way; b) Is it possible to get rid of this additional condition and thus to prove global convergence under the error bound condition alone?

References

[1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuvert: *Augmented Lagrangian methods under the constant positive linear dependence constraint qualification.* Mathematical Programming 111, 2008, pp. 5–32.

[2] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva: *A relaxed constant positive linear dependence constraint qualification applied to an augmented Lagrangian method.* Mathematical Programming 135, 2012, pp. 255–273.

[3] R. Andreani, G. Haeser, M.L. Schuverdt, and P.J.S. Silva: *Two new weak constraint qualifications and applications.* SIAM Journal on Optimization 22, 2012, pp. 1109–1135.

[4] R. Andreani, J.M. Martínez, and M.L. Schuverdt: *The CPLD condition of Qi and Wei implies the quasinormality constraint qualification.* Journal of Optimization Theory and Applications 125, 2005, pp. 473–485.

[5] C. Baiocchi and A. Capelo: *Variational and Quasivariational Inequalities: Applications to Free Boundary Problems.* Wiley, New York, 1984.

[6] A. Bensoussan: *Points de Nash dans le cas de fonctionnelles quadratiques et jeux differentiels lineaires a N personnes.* SIAM Journal on Control 12, 1974, pp. 460–499.

[7] A. Bensoussan, M. Goursat, and J.-L. Lions: *Contrôle impulsionnel et inéquations quasi-variationnelles stationnaires.* C. R. Acad. Sci. Paris Sér. A 276, 1973, pp. 1279–1284.

[8] A. Bensoussan and J.-L. Lions: *Nouvelle formulation de problèmes de contrôle impulsionnel et applications.* C. R. Acad. Sci. Paris Sér. A 276, 1973, pp. 1189–1192.

[9] A. Bensoussan and J.-L. Lions: *Nouvelles méthodes en contrôle impulsionnel.* Applied Mathematics and Optimization 1, 1975, pp. 289–312.

[10] P. Beremlijski, J. Haslinger, M. Kocvara, and J.V. Outrata: *Shape optimization in contact problems with Coulomb friction.* SIAM Journal on Optimization 13, 2002, pp. 561–587.

[11] D.P. Bertsekas: *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, 1982.

[12] E.G. Birgin and J.M. Martínez: *Practical Augmented Lagrangian Methods for Constrained Optimization.* SIAM, Philadelphia, PA, 2014.

[13] M. Bliemer and P. Bovy: *Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem.* Transportation Research B 37, 2003, pp. 501–519.

[14] D. Chan and J.-S. Pang: *The generalized quasi-variational inequality problem.* Mathematics of Operations Research 7, 1982, pp. 211–222.

[15] F.H. Clarke: *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983 (reprinted by SIAM, Philadelphia, PA, 1990).

[16] A.R. Conn, N.I.M. Gould, and Ph.L. Toint: *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A).* Springer Series in Computational Mathematics 17, Springer, New York, 1992.

[17] M. De Luca and A. Maugeri: *Discontinuous quasi-variational inequalities and applications to equilibrium problems.* In: F. Giannessi (ed.): *Nonsmooth Optimization: Methods and Applications.* Gordon and Breach, Amsterdam, 1992, pp. 70–75.

[18] F. Facchinei and C. Kanzow: *Penalty methods for the solution of generalized Nash equilibrium problems.* SIAM Journal on Optimization 20, 2010, pp. 2228-2253.

[19] F. Facchinei and C. Kanzow: *Generalized Nash equilibrium problems.* Annals of Operations Research 175, 2010, pp. 177–211.

[20] F. Facchinei, C. Kanzow, S. Karl, and S. Sagratella: *The semismooth Newton method for the solution of quasi-variational inequalities.* Computational Optimization and Applications, to appear.

[21] F. Facchinei, C. Kanzow, and S. Sagratella: *QVILIB: A library of quasi-variational inequality test problems.* Pacific Journal of Optimization 9, 2013, pp. 225–250.

[22] F. Facchinei, C. Kanzow, and S. Sagratella: *Solving quasi-variational inequalities via their KKT-conditions.* Mathematical Programming 144, 2014, pp. 369-412.

[23] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems. Volumes I + II.* Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[24] A. Fischer, M. Herrich, and K. Schönefeld: *Generalized Nash equilibrium problems – Recent advances and challenges.* Pesquisa Operacional, to appear.

[25] M. Fukushima: *A class of gap functions for quasi-variational inequality problems.* Journal of Industrial and Management Optimization 3, 2007, pp. 165–171.

[26] P. Hammerstein and R. Selten: *Game theory and evolutionary biology.* In: R.J. Aumann and S. Hart (eds.): *Handbook of Game Theory with Economic Applications, Vol. 2.* North-Holland, Amsterdam, 1994, pp. 929–993.

[27] P.T. Harker: *Generalized Nash games and quasi-variational inequalities.* European Journal on Operational Research 54, 1991, pp. 81–94.

[28] N. Harms, T. Hoheisel, and C. Kanzow: *On a smooth dual gap function for a class of quasi-variational inequalities.* Journal of Optimization Theory and Applications 163, 2014, pp. 413–438.

[29] N. Harms, C. Kanzow, and O. Stein: *Smoothness properties of a regularized gap function for quasi-variational inequalities.* Optimization Methods and Software 29, 2014, pp. 720–750.

[30] J. Haslinger and P.D. Panagiotopoulos: *The reciprocal variational approach to the Signorini problem with friction. Approximation results.* Proc. Royal Soc. Edinburgh: Sect. A Math. 98, 1984, pp. 365–383.

[31] M. Hintermüller and C.N. Rautenberg: *A sequential minimization technique for elliptic quasi-variational inequalities with gradient constraints.* SIAM Journal on Optimization 22, 2012, pp. 1224–1257.

[32] M. Hintermüller and C.N. Rautenberg: *Parabolic quasi-variational inequalities with gradient-type constraints.* SIAM Journal on Optimization 23, 2013, pp. 2090–2123.

[33] T. Ichiishi: *Game Theory for Economic Analysis.* Academic Press, New York, 1983.

[34] A.F. Izmailov, M.V. Solodov, and E.I. Uskov: *Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints.* SIAM Journal on Optimization 22, 2012, pp. 1579–1606.

[35] H. Jiang and L. Qi: *Local uniqueness and convergence of iterative methods for nonsmooth variational inequalities.* Journal of Mathematical Analysis and Applications 196, 1995, pp. 314–331.

[36] I. Kharroubi, J. Ma, H. Pham, and J. Zhang: *Backward SDEs with constrained jumps and quasi-variational inequalities.* Annals of Probability 38, 2010, pp. 794–840.

[37] A.S. Kravchuk and P.J. Neittaanmaki: *Variational and Quasi-Variational Inequalities in Mechanics.* Springer, Dordrecht, 2007.

[38] M. Kunze and J.F. Rodrigues: *An elliptic quasi-variational inequality with gradient constraints and some of its applications.* Mathematical Models in the Applied Sciences 23, 2000, pp. 897–908.

[39] U. Mosco: *Implicit variational problems and quasi variational inequalities.* In J. Gossez, E. Lami Dozo, J. Mawhin, L. Waelbroeck (eds.): *Nonlinear Operators and the Calculus of Variations.* Lecture Notes in Mathematics Vol. 543, Springer, Berlin, 1976, pp. 83–156.

[40] Y. Nesterov and L. Scrimali: *Solving strongly monotone variational and quasi-variational inequalities.* CORE Discussion Paper 2006/107, Catholic University of Louvain, Center for Operations Research and Econometrics, 2006.

[41] J. Nocedal and S.J. Wright: *Numerical Optimization.* Springer, 2nd edition, 2006.

[42] M.A. Noor: *On general quasi-variational inequalities.* Journal King Saud University 24, 2012, pp. 81–88.

[43] M.A. Noor: *An iterative scheme for a class of quasi variational inequalities.* Journal of Mathematical Analysis and Applications 110, 1985, pp. 463–468.

[44] J.V. Outrata and M. Kocvara: *On a class of quasi-variational inequalities.* Optimization Methods and Software 5, 1995, pp. 275–295.

[45] J.V. Outrata, M. Kocvara, and J. Zowe: *Nonsmooth approach to optimization problems with equilibrium constraints.* Kluwer Academic Publishers, Dordrecht and Boston, 1998.

[46] J.V. Outrata and J. Zowe: *A Newton method for a class of quasi-variational inequalities.* Computational Optimization and Applications 4, 1995, pp. 5–21.

[47] L. Qi and Z. Wei: *On the constant positive linear dependence condition and its applications to SQP methods.* SIAM Journal on Optimization 10, 2000, pp. 963–981.

[48] J.-S. Pang and M. Fukushima: *Quasi-variational inequalities, generalized Nash equilibria and multi-leader-follower games.* Computational Management Science 2, 2005, pp. 21-56 (Erratum: ibid 6, 2009, pp. 373-375).

[49] I.P. Ryazantseva: *First-order methods for certain quasi-variational inequalities in Hilbert space.* Computational Mathematics and Mathematical Physics 47, 2007, pp. 183–190.

[50] L. Scrimali: *Quasi-variational inequality formulation of the mixed equilibrium in mixed multiclass routing games.* CORE Discussion Paper 2007/7, Catholic University of Louvain, Center for Operations Research and Econometrics, 2007.

[51] A.H. Siddiqi and Q.H. Ansari: *Strongly nonlinear quasivariational inequalities.* Journal of Mathematical Analysis and Applications 149, 1990, pp. 444–450.

[52] S. Steffensen and M. Ulbrich: *A new relaxation scheme for mathematical programs with equilibrium constraints.* SIAM Journal on Optimization 20, 2010, pp. 2504–2539.

[53] J.Y. Wei and Y. Smeers: *Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices.* Operations Research 47, 1999, pp. 102–112.