

# QUASI-VARIATIONAL INEQUALITIES IN BANACH SPACES: THEORY AND AUGMENTED LAGRANGIAN METHODS\*

CHRISTIAN KANZOW<sup>†</sup> AND DANIEL STECK<sup>†</sup>

**Abstract.** This paper deals with quasi-variational inequality problems (QVIs) in a generic Banach space setting. We provide a theoretical framework for the analysis of such problems which is based on two key properties: the pseudomonotonicity (in the sense of Brezis) of the variational operator and a Mosco-type continuity of the feasible set mapping. We show that these assumptions can be used to establish the existence of solutions and their computability via suitable approximation techniques. In addition, we provide a practical and easily verifiable sufficient condition for the Mosco-type continuity property in terms of suitable constraint qualifications. Based on the theoretical framework, we construct an algorithm of augmented Lagrangian type which reduces the QVI to a sequence of standard variational inequalities. A full convergence analysis is provided which includes the existence of solutions of the subproblems as well as the attainment of feasibility and optimality. Applications and numerical results are included to demonstrate the practical viability of the method.

**Key words.** quasi-variational inequality, Mosco convergence, augmented Lagrangian method, global convergence, existence of solutions, Lagrange multiplier, constraint qualification

**AMS subject classifications.** 49J, 49K, 49M, 65K, 90C

**DOI.** 10.1137/18M1230475

**1. Introduction.** Let  $X$  be a real Banach space with continuous dual  $X^*$ , and let  $F : X \rightarrow X^*$ ,  $\Phi : X \rightrightarrows X$  be given mappings. The purpose of this paper is to analyze the *quasi-variational inequality (QVI)* which consists of finding  $x \in X$  such that

$$(1) \quad x \in \Phi(x), \quad \langle F(x), d \rangle \geq 0 \quad \forall d \in \mathcal{T}_{\Phi(x)}(x),$$

where  $\mathcal{T}_{\Phi(x)}(x)$  is the (Bouligand) tangent cone to  $\Phi(x)$  at  $x$  (see section 2). If  $\Phi$  is *convex-valued*, i.e., if  $\Phi(x)$  is convex for all  $x$ , then (1) can equivalently be stated as

$$(2) \quad x \in \Phi(x), \quad \langle F(x), y - x \rangle \geq 0 \quad \forall y \in \Phi(x).$$

The QVI was first defined in [7] and has since become a standard tool for the modeling of various equilibrium-type scenarios in the natural sciences. The resulting applications include game theory [25], solid and continuum mechanics [8, 26, 43, 51], economics [30, 31], probability theory [41], transportation [13, 20, 56], biology [24], and stationary problems in superconductivity, thermoplasticity, or electrostatics [3, 27, 28, 44, 54]. For further information, we refer the reader to the corresponding papers, the monographs [5, 43, 48], and the references therein.

Despite the abundant applications of QVIs, there is little in terms of a general theory or algorithmic approach for these problems, particularly in infinite dimensions. The present paper is an attempt to fill this gap. Observe that, in most applications,

\*Received by the editors December 3, 2018; accepted for publication (in revised form) September 23, 2019; published electronically December 17, 2019.

<https://doi.org/10.1137/18M1230475>

**Funding:** This research was supported by the German Research Foundation (DFG) within the priority program “Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization” (SPP 1962) under grant KA 1296/24-1.

<sup>†</sup>Institute of Mathematics, University of Würzburg, Campus Hubland Nord, 97074 Würzburg, Germany (kanzow@mathematik.uni-wuerzburg.de, daniel.steck@mathematik.uni-wuerzburg.de).

the feasible set mapping  $\Phi$  which governs the QVI can be cast into the general framework

$$(3) \quad \Phi(x) = \{y \in C : G(x, y) \in K\},$$

where  $C \subseteq X$  and  $K \subseteq Y$  are nonempty closed convex sets,  $Y$  is a real Banach space, and  $G : X^2 \rightarrow Y$  a given mapping. One can think of  $C$  as the set of “simple” or nonparametric constraints, whereas  $G$  describes the parametric constraints which turn the problem into a proper QVI. This decomposition is of course not unique.

In this paper, we provide a thorough analysis of QVIs in general and for the specific case where the feasible set is given by (3). Our analysis is based on two key properties: the pseudomonotonicity of  $F$  (in the sense of Brezis) and a weak sequential continuity of  $\Phi$  involving the notion of Mosco convergence (see section 2). These properties provide a general framework which includes many application examples, and they can be used to establish multiple desirable characteristics of QVIs such as the existence of solutions and their computability via suitable approximation techniques. In addition, we provide a systematic approach to Karush–Kuhn–Tucker (KKT)-type optimality conditions for QVIs and give a sufficient condition for the Mosco-type continuity of  $\Phi$  in terms of suitable constraint qualifications. To the best of our knowledge, this is one of the first generic and practically verifiable sufficient conditions for Mosco convergence in the QVI literature.

After the theoretical investigations, we turn our attention to an algorithmic approach based on augmented Lagrangian techniques. Recall that the augmented Lagrangian method (ALM) is one of the standard approaches for the solution of constrained optimization problems and is contained in almost any textbook on optimization [9, 10, 19, 49]. In recent years, ALMs have seen a certain resurgence [4, 11] in the form of safeguarded methods which use a slightly different update of the Lagrange multiplier estimate and turn out to have very strong global convergence properties [11]. A comparison of the classical ALM and its safeguarded analogue can be found in [35]. Moreover, the safeguarded ALM has been extended to QVIs in finite dimensions [32, 36, 52] and to constrained optimization problems and variational inequalities in Banach spaces [33, 37, 38, 40].

In the present paper, we will continue these developments and present a variant of the ALM for QVIs in the general framework described above. Using the decomposition (3) of the feasible set, we use a combined multiplier-penalty approach to eliminate the parametric constraint  $G(x, y) \in K$  and therefore reduce the QVI to a sequence of standard variational inequalities (VIs) involving the set  $C$ . The resulting algorithm generates a primal-dual sequence with certain asymptotic optimality properties.

The paper is organized as follows. In section 2, we establish and recall some theoretical background for the analysis of QVIs, including elements of convex and functional analysis. Section 3 contains a formal approach to the KKT conditions of the QVI, and we continue with an existence result for solutions in section 4. Starting with section 5, we turn our attention to the augmented Lagrangian method. After a formal statement of the algorithm, we continue with a global convergence analysis in section 6 and a primal-dual convergence theory for the nonconvex case in section 7. We then provide some applications of the algorithm in section 8 and conclude with some final remarks in section 9.

**Notation.** Throughout this paper,  $X$  and  $Y$  are always real Banach spaces, and their duals are denoted by  $X^*$  and  $Y^*$ , respectively. We write  $\rightarrow$ ,  $\rightharpoonup$ , and  $\rightharpoonup^*$  for strong, weak, and weak- $^*$  convergence, respectively, and denote by  $B_r^X$  the closed  $r$ -ball around zero in  $X$  (similarly for  $Y$ , etc.). Moreover, partial derivatives are

denoted by  $D_x, D_y$ , and so on. In particular, given the mapping  $G$  from (3), the symbol  $D_y G(x, x)$  denotes the derivative of the mapping  $G$  with respect to its second argument, evaluated at the point  $(x, x)$ . Throughout this manuscript, we make the implicit assumption that these derivatives exist whenever these symbols occur.

**2. Preliminaries.** We begin with some preliminary definitions. If  $S$  is a non-empty closed subset of some space  $Z$ , then  $S^\circ := \{\psi \in Z^* : \langle \psi, s \rangle \leq 0 \text{ for all } s \in S\}$  is the *polar cone* of  $S$ . Moreover, if  $z \in S$  is a given point, we denote by

$$\mathcal{T}_S(z) := \{d \in Z : \exists z^k \rightarrow z, t_k \downarrow 0 \text{ such that } z^k \in S \text{ and } (z^k - z)/t_k \rightarrow d\}$$

the *tangent cone* of  $S$  in  $z$ . If  $S$  is additionally convex, we also define the *radial* and *normal cones*

$$\mathcal{R}_S(z) := \{\alpha(s - z) : \alpha \geq 0, s \in S\}, \quad \mathcal{N}_S(z) := \{\psi \in Z^* : \langle \psi, s - z \rangle \leq 0 \forall s \in S\}.$$

For convex sets, it is well-known that  $\mathcal{T}_S(z) = \overline{\mathcal{R}_S(z)}$  and  $\mathcal{N}_S(z) = \mathcal{T}_S(z)^\circ = \mathcal{R}_S(z)^\circ = (S - z)^\circ$ . Clearly, if  $Z$  is a Hilbert space, we may treat  $\mathcal{N}_S(z)$  and other polars of sets in  $Z$  as subsets of  $Z$  instead of  $Z^*$ .

Throughout this paper, we will also need various notions of continuity.

**DEFINITION 2.1.** Let  $X, Y$  be Banach spaces and  $T : X \rightarrow Y$  an operator. We say that  $T$  is

- (a) bounded if it maps bounded sets to bounded sets;
- (b) weakly sequentially continuous if  $x^k \rightharpoonup x$  implies  $T(x^k) \rightarrow T(x)$ ;
- (c) weak-\* sequentially continuous if  $Y = W^*$  for some Banach space  $W$ , and  $x^k \rightharpoonup x$  implies  $T(x^k) \rightharpoonup^* T(x)$  in  $W^*$ ;
- (d) completely continuous if  $x^k \rightharpoonup x$  implies  $T(x^k) \rightarrow T(x)$ .

Clearly, complete continuity implies ordinary continuity as well as weak sequential continuity. Moreover, the latter implies weak-\* sequential continuity.

**2.1. Weak topologies and Mosco convergence.** For the treatment of the space  $X$ , we will need some topological concepts from Banach space theory. The *weak topology* on  $X$  is defined as the weakest (coarsest) topology for which all  $f \in X^*$  are continuous [17]. Accordingly, we call a set weakly closed (open, compact) if it is closed (open, compact) in the weak topology. Note that the notion of convergence induced by the weak topology is precisely that of weak convergence [55].

In many cases, e.g., when applying results from the literature which are formulated in a topological framework, it is desirable to use sequential notions of closedness, continuity, etc., instead of their topological counterparts. For instance, in the special case of the weak topology, it is well-known that every weakly closed set is weakly sequentially closed, but the converse is not true in general [6].

A possible remedy to this situation is to define a slightly different topology, which we call the *weak sequential topology* on  $X$ . This is the topology induced by weak convergence; more precisely, a set is called open in this topology if its complement is weakly sequentially closed. It is easy to see that this induces a topology (see [12, 21]). Moreover, since every weakly closed set is weakly sequentially closed, the weak sequential topology is finer than the weak topology, and like the latter [17, Prop. 3.3] it is also a Hausdorff topology. We will make use of the weak sequential topology in section 3.

For the analysis of the QVI (1), we will inevitably need certain continuity properties of the set-valued map  $\Phi : X \rightrightarrows X$ . Since we are dealing with a possibly infinite-

dimensional space  $X$ , these properties should take into account the weak (sequential) topology on  $X$ . An important notion in this context is that of Mosco convergence.

**DEFINITION 2.2** (Mosco convergence). *Let  $S$  and  $S_k$ ,  $k \in \mathbb{N}$ , be subsets of  $X$ . We say that  $\{S_k\}$  Mosco converges to  $S$ , and write  $S_k \xrightarrow{M} S$ , if*

- (i) *for every  $y \in S$ , there is a sequence  $y^k \in S_k$  such that  $y^k \rightarrow y$ , and*
- (ii) *whenever  $y^k \in S_k$  for all  $k$  and  $y$  is a weak limit point of  $\{y^k\}$ , then  $y \in S$ .*

The concept of Mosco convergence plays a key role in multiple aspects of the analysis of QVIs such as existence [44, 48], approximation [45], or the convergence of algorithms [27]. It is typically used as part of a continuity property of the mapping  $\Phi$ .

**DEFINITION 2.3** (weak Mosco-continuity). *Let  $\Phi : X \rightrightarrows X$  be a set-valued mapping and  $x \in X$ . We say that  $\Phi$  is weakly Mosco-continuous in  $x$  if  $x^k \rightharpoonup x$  implies  $\Phi(x^k) \xrightarrow{M} \Phi(x)$ . If this holds for every  $x \in X$ , we simply say that  $\Phi$  is weakly Mosco-continuous.*

The two conditions defining weak Mosco-continuity are occasionally referred to as complete inner and weak outer semicontinuity. Observe moreover that, if  $\Phi$  is weakly Mosco-continuous, then  $\Phi(x)$  is weakly closed for all  $x \in X$ .

Note that we will give a sufficient condition for the weak Mosco-continuity of the feasible set mapping in terms of certain constraint qualifications in section 3.

**2.2. Pseudomonotone operators.** The purpose of this section is to discuss the concept of *pseudomonotone operators* in the sense of Brezis [16] (see also [53, 61]). Note that there is another property in the literature, introduced by Karamardian [39], which is also sometimes referred to as pseudomonotonicity. We stress that the two concepts are distinct and that, throughout the remainder of this paper, pseudomonotonicity will always refer to the property below.

**DEFINITION 2.4** (pseudomonotonicity). *We say that an operator  $T : X \rightarrow X^*$  is pseudomonotone if, for every sequence  $\{x^k\} \subseteq X$  such that  $x^k \rightharpoonup x \in X$  and  $\limsup_{k \rightarrow \infty} \langle T(x^k), x^k - x \rangle \leq 0$ , we have  $\langle T(x), x - y \rangle \leq \liminf_{k \rightarrow \infty} \langle T(x^k), x^k - y \rangle$  for all  $y \in X$ .*

Despite its peculiar appearance, pseudomonotone operators will prove immensely useful because they provide a unified approach to monotone and nonmonotone operators. It should be noted that, despite its name suggesting otherwise, pseudomonotonicity can also be viewed as a rather weak kind of continuity. In fact, many operators are pseudomonotone simply by virtue of satisfying certain continuity properties. Some important examples are summarized in the following lemma.

**LEMMA 2.5** (sufficient conditions for pseudomonotonicity). *Let  $X$  be a Banach space and  $T, U : X \rightarrow X^*$  given operators. Then the following hold:*

- (i) *If  $T$  is monotone and continuous, then  $T$  is pseudomonotone.*
- (ii) *If  $T$  is completely continuous, then  $T$  is pseudomonotone.*
- (iii) *If  $T$  is continuous and  $\dim(X) < +\infty$ , then  $T$  is pseudomonotone.*
- (iv) *If  $T$  and  $U$  are pseudomonotone, then  $T + U$  is pseudomonotone.*

*Proof.* This is essentially [61, Prop. 27.6]. Note that [61] defines pseudomonotonicity only on reflexive Banach spaces, but this property is not used in the proof of the result.  $\square$

It follows from the lemma above that, in particular, every continuous operator  $T : X \rightarrow X^*$  on a finite-dimensional space  $X$  is pseudomonotone, even if it is not monotone.

The following result shows that bounded pseudomonotone operators automatically enjoy some kind of continuity. Note that the result generalizes [61, Prop. 27.7(b)] since we do not assume the reflexivity of  $X$ .

**LEMMA 2.6.** *Let  $F : X \rightarrow X^*$  be a bounded pseudomonotone operator. Then  $F$  is demicontinuous, i.e., it maps strongly convergent sequences to weak-\* convergent sequences. In particular, if  $\dim(X) < +\infty$ , then  $F$  is continuous.*

*Proof.* Let  $\{x^k\} \subseteq X$  be a sequence with  $x^k \rightarrow x$  for some  $x \in X$ . Observe that  $\{F(x^k)\}$  is bounded in  $X^*$  and hence

$$|\langle F(x^k), x^k - x \rangle| \leq \|F(x^k)\|_{X^*} \|x^k - x\|_X \rightarrow 0.$$

Thus, by pseudomonotonicity, we obtain

$$(4) \quad \langle F(x), x - y \rangle \leq \liminf_{k \rightarrow \infty} \langle F(x^k), x^k - y \rangle = \liminf_{k \rightarrow \infty} \langle F(x^k), x - y \rangle$$

for all  $y \in X$ , where we used the boundedness of  $\{F(x^k)\}$  and the fact that  $x^k \rightarrow x$ . Inserting  $\tilde{y} := 2x - y$  for an arbitrary  $y \in X$ , we also obtain

$$(5) \quad \langle F(x), y - x \rangle = \langle F(x), x - \tilde{y} \rangle \leq \liminf_{k \rightarrow \infty} \langle F(x^k), x^k - \tilde{y} \rangle = \liminf_{k \rightarrow \infty} \langle F(x^k), y - x \rangle,$$

where the last equality uses the fact that  $\{F(x^k)\}$  is bounded and that  $x^k - \tilde{y} = y - x + o(1)$ . Putting (4) and (5) together, it follows that  $\langle F(x^k), x - y \rangle \rightarrow \langle F(x), x - y \rangle$  for all  $y \in X$ . This implies  $F(x^k) \rightharpoonup^* F(x)$ , and the proof is done.  $\square$

In the context of QVIs, the pseudomonotonicity of the mapping  $F$  plays a key role since it ensures, together with the weak Mosco-continuity of  $\Phi$ , that weak limit points of sequences of approximate solutions of the QVI are exact solutions.

**PROPOSITION 2.7.** *Let  $F$  be a bounded pseudomonotone operator and let  $\Phi$  be weakly Mosco-continuous. Assume that  $\{x^k\} \subseteq X$  converges weakly to  $\bar{x}$ , that  $\bar{x} \in \Phi(\bar{x})$ , and that there are null sequences  $\{\delta_k\}, \{\varepsilon_k\} \subseteq \mathbb{R}$  (possibly negative) such that*

$$(6) \quad \langle F(x^k), y - x^k \rangle \geq \delta_k + \varepsilon_k \|y - x^k\|_X \quad \forall y \in \Phi(x^k)$$

for all  $k$ . Then  $\bar{x}$  is a solution of the QVI.

*Proof.* By Mosco-continuity, there is a sequence  $\bar{x}^k \in \Phi(x^k)$  such that  $\bar{x}^k \rightarrow \bar{x}$ . Inserting  $\bar{x}^k$  into (6) yields  $\liminf_{k \rightarrow \infty} \langle F(x^k), \bar{x}^k - x^k \rangle \geq 0$  and, since  $\{F(x^k)\}$  is bounded,  $\liminf_{k \rightarrow \infty} \langle F(x^k), \bar{x} - x^k \rangle \geq 0$ . The pseudomonotonicity of  $F$  therefore implies that

$$(7) \quad \langle F(\bar{x}), y - \bar{x} \rangle \geq \limsup_{k \rightarrow \infty} \langle F(x^k), y - x^k \rangle \quad \forall y \in X.$$

To show that  $\bar{x}$  solves the QVI, let  $y \in \Phi(\bar{x})$ . Using the Mosco-continuity of  $\Phi$ , we obtain a sequence  $y^k \in \Phi(x^k)$  such that  $y^k \rightarrow y$ . By (6), we have  $\liminf_{k \rightarrow \infty} \langle F(x^k), y^k - x^k \rangle \geq 0$ , hence  $\liminf_{k \rightarrow \infty} \langle F(x^k), y - x^k \rangle \geq 0$ , and (7) implies that  $\langle F(\bar{x}), y - \bar{x} \rangle \geq 0$ .  $\square$

The above result will play a key role in our subsequent analysis. Note that the assumption (6) can be relaxed; in fact, we only need the right-hand side to converge to zero whenever  $y^k \in \Phi(x^k)$  and  $\{y^k\}$  remains bounded. However, for our purposes, the formulation in (6) is sufficient.

Let us also remark that, if  $\Phi(x) \equiv \Phi$  is constant, then the Mosco-continuity in Proposition 2.7 is satisfied trivially and we obtain a stability result under pseudomonotonicity alone. In this case, it is easy to see that the boundedness of  $F$  can be omitted.

**2.3. Convexity of the feasible set mapping.** As mentioned in the introduction, an important distinction in the context of QVIs is whether the feasible sets  $\Phi(x)$ ,  $x \in X$ , are convex or not. This question is particularly important if the feasible set has the form (3), since we need to clarify which requirements on the mapping  $G$  are sufficient for the convexity of  $\Phi(x)$ . Ideally, these conditions should be easy and also yield some useful analytical properties.

Assume for the moment that  $K$  is a closed convex cone. Then  $K$  induces an order relation,  $y \leq_K z$  if and only if  $z - y \in K$ , and  $K$  itself can be regarded as the nonnegative cone with respect to  $\leq_K$ . Hence, it is natural to assume that the mapping  $G$  is concave with respect to this ordering.

If  $K$  is not a cone, then the appropriate order relation turns out to be induced by the recession cone  $K_\infty := \{y \in Y : y + K \subseteq K\}$ . Note that  $K_\infty$  is always a nonempty closed convex cone [15] and therefore induces an order relation as outlined above.

**DEFINITION 2.8** ( $K_\infty$ -concavity). *We say that  $G$  is  $K_\infty$ -concave with respect to  $y$  if the mapping  $G(x, \cdot)$  is concave with respect to the order relation induced by  $K_\infty$ , or equivalently*

$$G(x, (1 - \alpha)y_1 + \alpha y_2) - (1 - \alpha)G(x, y_1) - \alpha G(x, y_2) \in K_\infty$$

for all  $x, y_1, y_2 \in X$  and  $\alpha \in [0, 1]$ .

Some important consequences of  $K_\infty$ -concavity are formulated in the following lemma. A proof can be found in [37, Lem. 2.1] for the case where  $Y$  is a Hilbert space, and the general case is completely analogous.

**LEMMA 2.9** (properties of  $K_\infty$ -concavity). *Let  $G$  be  $K_\infty$ -concave with respect to  $y$ , and let  $x \in X$ . Then (i) the function  $y \mapsto \langle \lambda, G(x, y) \rangle$  is convex for all  $\lambda \in K_\infty^\circ$ , (ii) the function  $y \mapsto d_K(G(x, y))$  is convex, and (iii) the set  $\Phi(x)$  is convex.*

For later use, we will also need an analogue of the standard gradient inequality for convex functions. This result is contained in the following proposition.

**PROPOSITION 2.10** (property of differentiable  $K_\infty$ -concavity). *Let  $G$  be  $K_\infty$ -concave and differentiable with respect to  $y$ . Then  $G(x, y_1) \leq_K G(x, y_2) + D_y G(x, y_2)(y_1 - y_2)$  for all  $x, y_1, y_2 \in X$ .*

*Proof.* Let  $x, y_1, y_2 \in X$  and  $d := y_1 - y_2$ . For all  $t \in (0, 1)$ , the  $K_\infty$ -concavity yields

$$G(x, y_2 + td) = G(x, (1 - t)y_2 + ty_1) \in (1 - t)G(x, y_2) + tG(x, y_1) + K_\infty.$$

Rearranging this inclusion, dividing by  $t$ , and recalling that  $K_\infty$  is a cone yields

$$G(x, y_2) + \frac{G(x, y_2 + td) - G(x, y_2)}{t} \in G(x, y_1) + K_\infty.$$

For  $t \downarrow 0$ , we obtain  $G(x, y_2) + D_y G(x, y_2)d \geq_K G(x, y_1)$ . Hence, the result follows.  $\square$

**3. Regularity and optimality conditions.** The purpose of this section is to provide a formal approach to first-order optimality conditions involving Lagrange

multipliers for the QVI. As is commonly done in the context of QVIs, we say that a point  $x \in X$  is *feasible* if  $x \in \Phi(x)$ .

The key observation for the first-order optimality conditions is the following: a point  $\bar{x}$  solves the QVI if and only if  $\bar{d} := 0$  minimizes the function  $d \mapsto \langle F(\bar{x}), d \rangle$  over  $d \in \mathcal{T}_{\Phi(\bar{x})}(\bar{x})$ . Under a suitable constraint qualification (see below), one can show that  $\mathcal{T}_{\Phi(\bar{x})}(\bar{x}) = \{d \in \mathcal{T}_C(\bar{x}) : D_y G(\bar{x}, \bar{x})d \in \mathcal{T}_K(G(\bar{x}, \bar{x}))\}$ . Hence, this problem reduces to

$$(8) \quad \min \langle F(\bar{x}), d \rangle \quad \text{s.t.} \quad d \in \mathcal{T}_C(\bar{x}), \quad D_y G(\bar{x}, \bar{x})d \in \mathcal{T}_K(G(\bar{x}, \bar{x})).$$

The KKT system of the QVI is now obtained by applying the standard theory of KKT conditions to this optimization problem in  $\bar{d} = 0$ . This motivates the definition

$$(9) \quad \mathcal{L} : X \times Y^* \rightarrow X^*, \quad \mathcal{L}(x, \lambda) := F(x) + D_y G(x, x)^* \lambda,$$

as the Lagrange function of the QVI. The corresponding first-order optimality conditions are given as follows.

DEFINITION 3.1 (KKT conditions). *A tuple  $(\bar{x}, \bar{\lambda}) \in X \times Y^*$  is a KKT point of (1), (3) if*

$$(10) \quad -\mathcal{L}(\bar{x}, \bar{\lambda}) \in \mathcal{N}_C(\bar{x}) \quad \text{and} \quad \bar{\lambda} \in \mathcal{N}_K(G(\bar{x}, \bar{x})).$$

We call  $\bar{x}$  a *stationary point* if  $(\bar{x}, \bar{\lambda})$  is a KKT point for some multiplier  $\bar{\lambda} \in Y^*$ , and denote by  $\Lambda(\bar{x}) \subseteq Y^*$  the set of such multipliers.

As mentioned before, a constraint qualification is necessary to obtain the assertion that every solution of the QVI admits a Lagrange multiplier. To this end, we apply the standard Robinson constraint qualification to (8).

DEFINITION 3.2 (Robinson constraint qualification). *Let  $x \in X$  be an arbitrary point. We say that  $x$  satisfies*

(i) *the extended Robinson constraint qualification (ERCQ) if*

$$(11) \quad 0 \in \text{int}[G(x, x) + D_y G(x, x)(C - x) - K],$$

(ii) *the Robinson constraint qualification (RCQ) if  $x$  is feasible and satisfies (11).*

For nonlinear programming-type constraints, RCQ reduces to the Mangasarian–Fromovitz constraint qualification (MFCQ), see [15, p. 71]. The extended RCQ is similar to the so-called *extended MFCQ*, which generalizes MFCQ to (possibly) infeasible points.

The connection between the QVI and its KKT conditions is much stronger than it is for optimization problems. The reason behind this is that, in a way, the QVI itself is already a problem formulation tailored toward “first-order optimality.”

PROPOSITION 3.3. *If  $(\bar{x}, \bar{\lambda})$  is a KKT point, then  $\bar{x}$  is a solution of the QVI. Conversely, if  $\bar{x}$  is a solution of the QVI and RCQ holds in  $\bar{x}$ , then there exists a multiplier  $\bar{\lambda} \in Y^*$  such that  $(\bar{x}, \bar{\lambda})$  is a KKT point, and the corresponding multiplier set is bounded.*

*Proof.* Throughout the proof, let  $G(x, y) := (y, G(x, y))$  and  $K := C \times K$ , so that

$$(12) \quad \Phi(x) = \{y \in C : G(x, y) \in K\} = \{y \in X : G(x, y) \in K\} \quad \forall x \in X.$$

An easy calculation shows that the KKT conditions remain invariant under this reformulation of the constraint system. By [15, Lem. 2.100], the same holds for RCQ.

Now, let  $(\bar{x}, \bar{\lambda})$  be a KKT point of the QVI. Using (12) and [37, Thm. 2.4], it follows that  $\bar{x}$  is a solution of the variational inequality  $\bar{x} \in S$ ,  $\langle F(\bar{x}), d \rangle \geq 0$  for all  $d \in \mathcal{T}_S(\bar{x})$ , where  $S := \Phi(\bar{x})$  is considered fixed. But this means that  $\bar{x}$  solves the QVI.

Conversely, if  $\bar{x}$  solves the QVI and RCQ holds in  $\bar{x}$ , then (12) and [15, Cor. 2.91] imply that  $\mathcal{T}_{\Phi(\bar{x})}(\bar{x}) = \{d \in \mathcal{T}_C(\bar{x}) : D_y G(\bar{x}, \bar{x})d \in \mathcal{T}_K(G(\bar{x}, \bar{x}))\}$ , so that  $\bar{d} = 0$  is a solution of (8). The (ordinary) Robinson constraint qualification [15, Def. 2.86] for this problem takes on the form

$$0 \in \text{int}[D_y G(\bar{x}, \bar{x})\mathcal{T}_C(\bar{x}) - \mathcal{T}_K(G(\bar{x}, \bar{x}))].$$

Since  $C - \bar{x} \subseteq \mathcal{T}_C(\bar{x})$  and  $K - G(\bar{x}, \bar{x}) \subseteq \mathcal{T}_K(G(\bar{x}, \bar{x}))$ , this condition is implied by (11). The result now follows by applying a standard KKT theorem to (8), such as [15, Thm. 3.9].  $\square$

We now turn to another consequence of RCQ which is a certain metric regularity of the feasible set(s). This property can also be used to deduce the weak Mosco-continuity of  $\Phi$  (see Corollary 3.5). The main idea is that, given some reference point  $\bar{x}$ , we can view the parameter  $x$  in the feasible set mapping  $\Phi(x)$  as a perturbation parameter and use the following result from perturbation theory.

**LEMMA 3.4.** *Let  $\bar{x} \in X$  and  $\bar{y} \in \Phi(\bar{x})$ . Assume that  $G$  and  $D_y G$  are continuous on  $U \times X$ , where  $U$  is the space  $X$  equipped with an arbitrary topology, and that*

$$0 \in \text{int}[G(\bar{x}, \bar{y}) + D_y G(\bar{x}, \bar{y})(C - \bar{y}) - K].$$

*Then there are  $c > 0$  and a neighborhood  $N$  of  $(\bar{x}, \bar{y})$  in  $U \times X$  such that  $\text{dist}(y, \Phi(x)) \leq c \text{dist}(G(x, y), K)$  for all  $(x, y) \in N$  with  $y \in C$ .*

*Proof.* Similar to above, let  $\mathbf{G} : U \times X \rightarrow X \times Y$  be the mapping  $\mathbf{G}(x, y) := (y, G(x, y))$ , and define the set  $\mathbf{K} := C \times K$ . Observe that  $\mathbf{G}$  is Fréchet-differentiable with respect to  $y$ , that both  $\mathbf{G}$  and  $D_y \mathbf{G}$  are continuous on  $U \times X$ , and that  $\Phi(x) = \{y \in X : \mathbf{G}(x, y) \in \mathbf{K}\}$  for all  $x \in X$ . By [15, Lem. 2.100], we have  $0 \in \text{int}[\mathbf{G}(\bar{x}, \bar{y}) + D_y \mathbf{G}(\bar{x}, \bar{y})X - \mathbf{K}]$ . Hence, by [15, Thm. 2.87], there are  $c > 0$  and a neighborhood  $N$  of  $(\bar{x}, \bar{y})$  in  $U \times X$  such that

$$\text{dist}(y, \Phi(x)) \leq c \text{dist}(\mathbf{G}(x, y), \mathbf{K})$$

for all  $(x, y) \in N$ . Clearly, if  $y \in C$ , then  $\text{dist}(\mathbf{G}(x, y), \mathbf{K}) = \text{dist}(G(x, y), K)$ .  $\square$

As mentioned before, we can use Lemma 3.4 to prove the weak Mosco-continuity of  $\Phi$ . To this end, we only need to apply the lemma in the special case where  $U$  is the space  $X$  equipped with the weak sequential topology.

**COROLLARY 3.5.** *Let  $\Phi(x) = \{y \in C : G(x, y) \in K\}$  and let  $\bar{x}$  be a feasible point. Assume that  $G$  is  $K_\infty$ -concave with respect to  $y$ , that  $G$  and  $D_y G$  satisfy the continuity property*

$$(13) \quad x^k \rightharpoonup x, \quad y^k \rightarrow y \quad \implies \quad G(x^k, y^k) \rightarrow G(x, y), \quad D_y G(x^k, y^k) \rightarrow D_y G(x, y)$$

*for all  $x, y \in X$ , and that RCQ holds in  $\bar{x}$ . Then  $\Phi$  is weakly Mosco-continuous in  $\bar{x}$ .*

*Proof.* We first show the weak outer semicontinuity of  $\Phi$ . Let  $x^k \rightharpoonup \bar{x}$  and  $y^k \in \Phi(x^k)$ ,  $y^k \rightharpoonup \bar{y}$ . Then  $\{y^k\} \subseteq C$ , which implies  $\bar{y} \in C$ . Furthermore, Proposition 2.10 yields

$$G(x^k, \bar{y}) + D_y G(x^k, \bar{y})(y^k - \bar{y}) \in G(x^k, y^k) + K_\infty \subseteq K$$



for all  $k$ . Taking the limit  $k \rightarrow \infty$  and using (13), it follows that the left-hand side converges weakly to  $G(\bar{x}, \bar{y})$ . Hence,  $G(\bar{x}, \bar{y}) \in K$  and  $\bar{y} \in \Phi(\bar{x})$ .

For the inner semicontinuity, let  $x^k \rightharpoonup \bar{x}$  and  $\bar{y} \in \Phi(\bar{x})$ . As in the proof of Proposition 3.3, we may assume that  $C = X$ . By assumption, the constraint system  $y \in C$ ,  $G(\bar{x}, y) \in K$  satisfies the (ordinary) RCQ in  $\bar{x}$ ; thus, by convexity, this constraint system satisfies RCQ in every  $y \in \Phi(\bar{x})$  (see, e.g., [15, Thms. 2.83 and 2.104]), in particular for  $y := \bar{y}$ . Now, let  $U$  denote the space  $X$  equipped with the weak sequential topology. Then  $G$  and  $D_y G$  are continuous on  $U \times X$ . By Lemma 3.4, there exists  $c > 0$  such that

$$\text{dist}(\bar{y}, \Phi(x^k)) \leq c \text{dist}(G(x^k, \bar{y}), K)$$

for  $k \in \mathbb{N}$  sufficiently large. Since  $G(x^k, \bar{y}) \rightarrow G(\bar{x}, \bar{y}) \in K$  by assumption, it follows that the right-hand side converges to zero as  $k \rightarrow \infty$ . Hence, we can choose points  $y^k \in \Phi(x^k)$  with  $\|y^k - \bar{y}\|_X \rightarrow 0$ . This completes the proof.  $\square$

**4. An existence result for QVIs.** The existence of solutions to QVIs is a rather delicate topic. Many results, especially for infinite-dimensional problems, either deal with specific problem settings [2, 44] or consider general QVIs under rather long lists of assumptions, often including monotonicity [5, 23, 48]. Two interesting exceptions are the papers [42, 57], which deal with quite general classes of QVIs and prove existence results under suitable compactness and continuity assumptions. However, the results contained in these papers actually require the *complete continuity* of the mapping  $F$ . This is a very restrictive assumption which cannot be expected to hold in many applications; for instance, it does not even hold if  $X$  is a Hilbert space and  $F(x) = x$ . An analogous comment applies if  $F$  involves additional summands, e.g., if  $F$  arises from the derivative of an optimal control-type objective function with a Tikhonov regularization parameter.

In this paper, we pursue a different approach, which is based on a combination of the weak Mosco-continuity from section 2.1 and the Brezis-type pseudomonotonicity from section 2.2. A particularly intuitive idea is given by Proposition 2.7, which suggests that we can tackle the QVI by solving a sequence of approximating problems and then using a suitable limiting argument to obtain a solution of the problem of interest. For the precise implementation of this idea, we will need some auxiliary results.

The first result we need is a slight modification of an existence theorem of Brezis, Nirenberg, and Stampacchia for VI-type equilibrium problems (see [18, Thm. 1]). Note that the result in [18] is formulated in a rather general setting and uses filters instead of sequences; however, when applied to the Banach space setting, one can dispense with filters by using, for instance, Day's lemma [46, Lem. 2.8.5]. We shall not demonstrate the resulting proof here, mainly for the sake of brevity and since it is basically identical to that given in [18]. The interested reader will also find the complete proof in the dissertation [58].

**PROPOSITION 4.1 (Brezis–Nirenberg–Stampacchia).** *Let  $A \subseteq X$  be a nonempty, convex, weakly compact set, and  $\Psi : A \times A \rightarrow \mathbb{R}$  a mapping such that*

- (i)  $\Psi(x, x) \leq 0$  for all  $x \in A$ ,
- (ii) for every  $x \in A$ , the function  $\Psi(x, \cdot)$  is (quasi-)concave,
- (iii) for every  $y \in A$  and every finite-dimensional subspace  $L$  of  $X$ , the function  $\Psi(\cdot, y)$  is lsc on  $A \cap L$ , and
- (iv) whenever  $x, y \in A$ ,  $\{x^k\} \subseteq A$  converges weakly to  $x$ , and  $\Psi(x^k, (1-t)x + ty) \leq 0$  for all  $t \in [0, 1]$  and  $k \in \mathbb{N}$ , then  $\Psi(x, y) \leq 0$ .

*Then there exists  $\hat{x} \in A$  such that  $\Psi(\hat{x}, y) \leq 0$  for all  $y \in A$ .*

We will mainly need Proposition 4.1 to obtain the existence of solutions to the approximating problems in our existence result for QVIs. For this purpose, it will be useful to present a slightly more tangible corollary of the result.

**COROLLARY 4.2.** *Let  $A \subseteq X$  be a nonempty, convex, weakly compact set,  $F : X \rightarrow X^*$  a bounded pseudomonotone operator, and  $\varphi : A^2 \rightarrow \mathbb{R}$  a mapping such that*

- (i)  $\varphi(x, x) = 0$  for all  $x \in A$ ,
- (ii) for every  $x \in A$ , the function  $\varphi(x, \cdot)$  is concave, and
- (iii) for every  $y \in A$ , the function  $\varphi(\cdot, y)$  is weakly sequentially lsc.

*Then there exists  $\hat{x} \in A$  such that  $\langle F(\hat{x}), \hat{x} - y \rangle + \varphi(\hat{x}, y) \leq 0$  for all  $y \in A$ .*

*Proof.* We claim that the mapping  $\Psi : A^2 \rightarrow \mathbb{R}$ ,  $\Psi(x, y) := \langle F(x), x - y \rangle + \varphi(x, y)$ , satisfies the assumption of Proposition 4.1. Clearly,  $\Psi(x, x) \leq 0$  for every  $x \in A$ , and  $\Psi$  is (quasi-)concave with respect to the second argument. Moreover, by the properties of pseudomonotone operators (Lemma 2.6),  $\Psi$  is lsc with respect to the first argument on  $A \cap L$  for any finite-dimensional subspace  $L$  of  $X$ . Finally, let  $x, y \in A$ , let  $\{x^k\} \subseteq A$  be a sequence converging weakly to  $x$ , and assume that

$$(14) \quad \Psi(x^k, (1-t)x + ty) \leq 0 \quad \forall t \in [0, 1], \quad \forall k \in \mathbb{N}.$$

We need to show that  $\Psi(x, y) \leq 0$ . By (14), we have in particular that  $\Psi(x^k, x) \leq 0$  and  $\Psi(x^k, y) \leq 0$  for all  $k$ . The first of these conditions implies that

$$\begin{aligned} 0 &\geq \limsup_{k \rightarrow \infty} \Psi(x^k, x) \geq \limsup_{k \rightarrow \infty} \langle F(x^k), x^k - x \rangle + \liminf_{k \rightarrow \infty} \varphi(x^k, x) \\ &\geq \limsup_{k \rightarrow \infty} \langle F(x^k), x^k - x \rangle, \end{aligned}$$

where we used the weak sequential lower semicontinuity of  $\varphi$  with respect to  $x$  and the fact that  $\varphi(x, x) = 0$ . Hence, by the pseudomonotonicity of  $F$ , we obtain

$$\begin{aligned} \Psi(x, y) &= \langle F(x), x - y \rangle + \varphi(x, y) \\ &\leq \liminf_{k \rightarrow \infty} [\langle F(x^k), x^k - y \rangle + \varphi(x^k, y)] = \liminf_{k \rightarrow \infty} \Psi(x^k, y) \leq 0. \end{aligned}$$

Therefore,  $\Psi$  satisfies all the requirements of Proposition 4.1, and the result follows.  $\square$

Apart from the above result, we will also need some information on the behavior of the “parametric” distance function  $(x, y) \mapsto d_{\Phi(x)}(y)$ . Here, the weak Mosco-continuity of  $\Phi$  plays a key role and allows us to prove the following lemma.

**LEMMA 4.3.** *Let  $\Phi : X \rightrightarrows X$  be weakly Mosco-continuous. Then, for every  $y \in X$ , the distance function  $x \mapsto d_{\Phi(x)}(y)$  is weakly sequentially upper semicontinuous on  $X$ .*

*If, in addition, there are nonempty subsets  $A, B \subseteq X$  such that  $\Phi(A) \subseteq B$  and  $B$  is weakly compact, then the function  $x \mapsto d_{\Phi(x)}(x)$  is weakly sequentially lsc on  $A$ .*

*Proof.* Let  $y \in X$  be a fixed point and let  $\{x^k\} \subseteq X$ ,  $x^k \rightharpoonup x \in X$ . If  $w \in \Phi(x)$  is an arbitrary point, then there is a sequence  $w^k \in \Phi(x^k)$  such that  $w^k \rightarrow w$ . It follows that

$$\|y - w\|_X = \lim_{k \rightarrow \infty} \|y - w^k\|_X \geq \limsup_{k \rightarrow \infty} d_{\Phi(x^k)}(y).$$

Since  $w \in \Phi(x)$  was arbitrary, this implies that  $d_{\Phi(x)}(y) \geq \limsup_{k \rightarrow \infty} d_{\Phi(x^k)}(y)$ .

We now prove the second assertion. Let  $\{x^k\} \subseteq A$  be a sequence with  $x^k \rightharpoonup x \in A$ . Without loss of generality, let  $d_{\Phi(x^k)}(x^k) \rightarrow \liminf_{k \rightarrow \infty} d_{\Phi(x^k)}(x^k)$ , and let  $\Phi(x^k)$  be nonempty for all  $k$ . Choose points  $w^k \in \Phi(x^k)$  such that  $\|x^k - w^k\|_X \leq d_{\Phi(x^k)}(x^k) +$

$1/k$ . By assumption, the sequence  $\{w^k\}$  is contained in the weakly compact set  $B$ , and thus there is an index set  $I \subseteq \mathbb{N}$  such that  $w^k \rightharpoonup_I w$  for some  $w \in B$ . Since  $w^k \in \Phi(x^k)$  for all  $k$ , the weak Mosco-continuity of  $\Phi$  implies  $w \in \Phi(x)$ . It follows that

$$d_{\Phi(x)}(x) \leq \|x - w\|_X \leq \liminf_{k \in I} \|x^k - w^k\|_X = \liminf_{k \rightarrow \infty} d_{\Phi(x^k)}(x^k).$$

This completes the proof.  $\square$

We now turn to the main existence result for QVIs.

**THEOREM 4.4.** *Consider a QVI of the form (1). Assume that (i)  $F$  is bounded and pseudomonotone, (ii)  $\Phi$  is weakly Mosco-continuous, and (iii) there is a nonempty, convex, weakly compact set  $A \subseteq X$  such that, for all  $x \in A$ ,  $\Phi(x)$  is nonempty, closed, convex, and contained in  $A$ . Then the QVI admits a solution  $\bar{x} \in A$ .*

*Proof.* For  $k \in \mathbb{N}$ , let  $\Psi_k : A^2 \rightarrow \mathbb{R}$  be the bifunction

$$\Psi_k(x, y) := \langle F(x), x - y \rangle + k[d_{\Phi(x)}(x) - d_{\Phi(x)}(y)].$$

By Lemma 4.3 and Corollary 4.2, there exist points  $x^k \in A$  such that  $\Psi_k(x^k, y) \leq 0$  for all  $y \in A$ . Since  $A$  is weakly compact, the sequence  $\{x^k\}$  admits a weak limit point  $\bar{x} \in A$ ; without loss of generality, assume that the entire sequence  $\{x^k\}$  converges weakly to  $\bar{x}$ . Moreover, by assumption, there are points  $y^k \in \Phi(x^k) \subseteq A$  for all  $k$ . For these points, we obtain

$$0 \geq \Psi_k(x^k, y^k) = \langle F(x^k), x^k - y^k \rangle + kd_{\Phi(x^k)}(x^k).$$

By the boundedness of  $A$  and  $F$ , the first term is bounded. Hence, dividing by  $k$ , we obtain  $d_{\Phi(x^k)}(x^k) \rightarrow 0$ , thus  $d_{\Phi(\bar{x})}(\bar{x}) = 0$  by Lemma 4.3, and hence  $\bar{x} \in \Phi(\bar{x})$ .

Finally, we claim that  $\bar{x}$  solves the QVI. Observe that, for all  $k$  and  $y \in \Phi(x^k)$ ,

$$0 \geq \Psi_k(x^k, y) = \langle F(x^k), x^k - y \rangle + kd_{\Phi(x^k)}(x^k) \geq \langle F(x^k), x^k - y \rangle.$$

Thus, by Proposition 2.7, it follows that  $\bar{x}$  is a solution of the QVI.  $\square$

In the finite-dimensional case with, say,  $X = \mathbb{R}^n$ , the mapping  $F$  being bounded and pseudomonotone is equivalent to the continuity of  $F$ . Hence it is easy to see that Theorem 4.4 reduces to the existence result from [52] in this particular situation.

The applicability of the above theorem depends most crucially on the weak Mosco-continuity of  $\Phi$  and the existence of the weakly compact set  $A$ . It should be possible to modify the theorem by requiring some form of coercivity instead, but this is outside the scope of the present paper.

**Example 4.5.** This example is based on [44]. Let  $\Omega \subseteq \mathbb{R}^d$ ,  $d \geq 2$ , be a bounded domain, let  $X := H_0^1(\Omega)$ , and consider the QVI given by the functions

$$F(u) := -\Delta u - f, \quad \Phi(u) := \{v \in H_0^1(\Omega) : \|\nabla v\| \leq \Psi(u)\},$$

where  $\|\cdot\|$  is the Euclidean norm,  $f \in H^{-1}(\Omega)$ , and  $\Psi : H_0^1(\Omega) \rightarrow L^\infty(\Omega)$  is completely continuous. Observe that  $F$  is pseudomonotone by Lemma 2.5(i). Assume now that  $c_1 \leq \Psi(u) \leq c_2$  for all  $u$  and some  $c_1, c_2 > 0$ . Then  $\Phi$  is weakly Mosco-continuous by [44, Lem. 1]. Moreover,  $0 \in \Phi(u)$  for all  $u \in H_0^1(\Omega)$ , and the Poincaré inequality implies that there is an  $R > 0$  such that  $\Phi(u) \subseteq B_R^X$  for all  $u \in H_0^1(\Omega)$ . We conclude that all the requirements of Theorem 4.4 are satisfied; hence, the QVI admits a solution.

Alternative existence results, in a Hilbert space setting, are based on the fixed point formulation

$$x = P_{\Phi(x)}(x - \rho F(x))$$

for some scalar  $\rho > 0$ , where  $P$  denotes the projection operator in the given Hilbert space norm. Existence results based on this characterization require the following condition:

$$(15) \quad \|P_{\Phi(x)}(z) - P_{\Phi(y)}(z)\| \leq \gamma \|x - y\| \quad \forall x, y, z \in H,$$

together with strong monotonicity and global Lipschitz continuity of  $F$  (cf. [50]). Condition (15) holds with  $\gamma = 1$  if  $\Phi(x) = \Phi_0$  for a constant set  $\Phi_0$ , and can also be verified, under some assumptions, in the “moving set case” where  $\Phi(x) = c(x) + \Phi_0$  for a suitable mapping  $c : H \rightarrow H$  and a constant set  $\Phi_0$  (nonempty, closed, and convex).

The following result relates (15) to the Mosco convergence and shows that (15) almost implies the Mosco convergence.

**PROPOSITION 4.6.** *Let  $H$  be a Hilbert space, and assume  $\Phi : H \rightrightarrows H$  is a multifunction with  $\Phi(x)$  nonempty, closed, and convex for all  $x \in H$  and such that (15) holds. Then the following statements hold:*

- (a) *Given  $x^k \rightarrow x$  and  $y \in \Phi(x)$ , there exists a sequence  $y^k \in \Phi(x^k)$  such that  $y^k \rightarrow y$ .*
- (b) *Given  $x^k \rightarrow x$  and a sequence  $y^k \rightarrow y$  with  $y^k \in \Phi(x^k)$  for all  $k$ , it holds that  $y \in \Phi(x)$ .*

*Proof.* (a) Let  $x^k \rightarrow x$  and  $y \in \Phi(x)$ . Define  $y^k := P_{\Phi(x^k)}(y)$ . Then (15) implies

$$\|P_{\Phi(x^k)}(y) - P_{\Phi(x)}(y)\| \leq \gamma \|x^k - x\| \rightarrow 0.$$

The definition of  $y^k$  together with the fact that  $y = P_{\Phi(x)}(y)$  therefore yields  $y^k \rightarrow y$ .

(b) Let  $x^k \rightarrow x, y^k \in \Phi(x^k)$  for all  $k$  and  $y^k \rightarrow y$ . Then it follows from (15) that

$$\|y^k - P_{\Phi(x)}(y^k)\| = \|P_{\Phi(x^k)}(y^k) - P_{\Phi(x)}(y^k)\| \leq \gamma \|x^k - x\| \rightarrow 0.$$

On the other hand,  $y^k \rightarrow y$  and the continuity of the projection operator (onto the fixed set  $\Phi(x)$ ) implies

$$\|y^k - P_{\Phi(x)}(y^k)\| \rightarrow \|y - P_{\Phi(x)}(y)\|.$$

Consequently, we obtain  $y = P_{\Phi(x)}(y)$ , and hence  $y \in \Phi(x)$ .  $\square$

Note that Proposition 4.6 considers strongly convergent sequences  $\{x^k\}$ , whereas in the definition of Mosco convergence, this sequence is assumed to be weakly convergent. The other assumptions used by the two types of existence results are quite different: The one based on the projection formulation assumes that  $F$  is strongly monotone and globally Lipschitz continuous, conditions which are obviously not necessary in our Theorem 4.4. On the other hand,  $F$  being strongly monotone and Lipschitz continuous does not necessarily imply that we can find a suitable weakly compact set  $A$  with the properties stated in Theorem 4.4 (take, e.g.,  $X = \mathbb{R}^n$  and  $F(x) := x + 1$ ).

**5. The augmented Lagrangian method.** We now present the augmented Lagrangian method for the QVI (1). The main approach is to penalize the function  $G$  and therefore reduce the QVI to a sequence of standard VIs. Throughout the remainder of the paper, we assume that  $i : Y \hookrightarrow H$  densely for some real Hilbert space  $H$  and that  $\mathcal{K}$  is a closed convex subset of  $H$  with  $i^{-1}(\mathcal{K}) = K$ . Note that we will usually suppress this embedding for the sake of notational simplicity.

Note that, in many applications, including those to be presented in our numerical section, this scenario is naturally satisfied. Moreover, this embedding into a Hilbert space  $H$  enables us to apply the augmented Lagrangian technique since it allows the application of projections which exist and are (often) easy to compute in the Hilbert space setting. Moreover, the square of the distance function is smooth in this framework, a property that is important at least when the augmented Lagrangian approach is applied to optimization problems. Nevertheless, the existence of the above embedding should be viewed as an assumption and might not be easy to obtain in certain situations.

Consider now the augmented Lagrangian  $\mathcal{L}_\rho : X \times H \rightarrow X^*$  given by

$$(16) \quad \mathcal{L}_\rho(x, \lambda) := F(x) + \rho D_y G(x, x)^* \left[ G(x, x) + \frac{\lambda}{\rho} - P_{\mathcal{K}} \left( G(x, x) + \frac{\lambda}{\rho} \right) \right],$$

where  $P_{\mathcal{K}}$  denotes the projection operator onto the closed convex set  $\mathcal{K}$  with respect to the norm induced by the Hilbert space  $H$ . Note that if  $\mathcal{K}$  is a cone, then we can simplify the above formula to  $\mathcal{L}_\rho(x, \lambda) = F(x) + D_y G(x, x)^* P_{\mathcal{K}^\circ}(\lambda + \rho G(x, x))$  by using Moreau's decomposition [6, 47].

For the construction of our algorithm, we will need a means of controlling the penalty parameters. To this end, we define the utility function

$$(17) \quad V(x, \lambda, \rho) := \left\| G(x, x) - P_{\mathcal{K}} \left( G(x, x) + \frac{\lambda}{\rho} \right) \right\|_H.$$

The function  $V$  is a composite measure of feasibility and complementarity; it arises from an inherent slack variable transformation which is often used to define the augmented Lagrangian for inequality or cone constraints.

**ALGORITHM 5.1** (augmented Lagrangian method). *Let  $(x^0, \lambda^0) \in X \times H$ ,  $\rho_0 > 0$ ,  $\gamma > 1$ ,  $\tau \in (0, 1)$ , let  $B \subseteq H$  be a bounded set, and set  $k := 0$ .*

Step 1. *If  $(x^k, \lambda^k)$  satisfies a suitable termination criterion: STOP.*

Step 2. *Choose  $w^k \in B$  and compute an inexact solution (see below)  $x^{k+1}$  of the VI*

$$(18) \quad x \in C, \quad \langle \mathcal{L}_{\rho_k}(x, w^k), y - x \rangle \geq 0 \quad \forall y \in C.$$

Step 3. *Update the vector of multipliers to*

$$(19) \quad \lambda^{k+1} := \rho_k \left[ G(x^{k+1}, x^{k+1}) + \frac{w^k}{\rho_k} - P_{\mathcal{K}} \left( G(x^{k+1}, x^{k+1}) + \frac{w^k}{\rho_k} \right) \right].$$

Step 4. *If  $k = 0$  or*

$$(20) \quad V(x^{k+1}, w^k, \rho_k) \leq \tau V(x^k, w^{k-1}, \rho_{k-1})$$

*holds, set  $\rho_{k+1} := \rho_k$ ; otherwise, set  $\rho_{k+1} := \gamma \rho_k$ .*

Step 5. *Set  $k \leftarrow k + 1$  and go to Step 1.*

Let us make some simple observations. First, regardless of the primal iterates  $\{x^k\}$ , the multipliers  $\{\lambda^k\}$  always lie in the polar cone  $\mathcal{K}_\infty^\circ$  by [37, Lem. 2.2]. Moreover, if  $\mathcal{K}$  is a cone, then the well-known Moreau decomposition for closed convex cones implies that  $\lambda^{k+1} = P_{\mathcal{K}^\circ}(w^k + \rho_k G(x^{k+1}, x^{k+1}))$ .

Second, we note that Algorithm 5.1 uses a safeguarded multiplier sequence  $\{w^k\}$  in certain places where classical augmented Lagrangian methods use the sequence  $\{\lambda^k\}$ . This bounding scheme goes back to [4, 52] and is crucial to establishing good global convergence results for the method [4, 11, 38]. In practice, one usually tries to keep  $w^k$  as “close” as possible to  $\lambda^k$ , e.g., by defining  $w^k := P_B(\lambda^k)$ , where  $B$  (the bounded set from the algorithm) is chosen suitably to allow cheap projections.

Our final observation concerns the definition of  $\lambda^{k+1}$ . Regardless of the manner in which  $x^{k+1}$  is computed (exactly or inexactly), we always have the equality

$$(21) \quad \mathcal{L}_{\rho_k}(x^{k+1}, w^k) = \mathcal{L}(x^{k+1}, \lambda^{k+1}) \quad \forall k \in \mathbb{N},$$

which follows directly from the definition of  $\mathcal{L}_\rho$  and the multiplier updating scheme (19). This equality is the main motivation for the definition of  $\lambda^{k+1}$ .

We now prove a lemma which essentially asserts some kind of “approximate normality” of  $\lambda^k$  and  $G(x^k, x^k)$ . Recall that the KKT conditions of the QVI require that  $\bar{\lambda} \in \mathcal{N}_K(G(\bar{x}, \bar{x}))$ . However, we have to take into account that  $G(x^k, x^k)$  is not necessarily an element of  $K$ .

**LEMMA 5.2.** *There is a null sequence  $\{r_k\} \subseteq [0, \infty)$  such that  $(\lambda^k, y - G(x^k, x^k)) \leq r_k$  for all  $y \in \mathcal{K}$  and  $k \in \mathbb{N}$ .*

*Proof.* Let  $y \in \mathcal{K}$  and define the sequence  $s^{k+1} := P_{\mathcal{K}}(G(x^{k+1}, x^{k+1}) + w^k/\rho_k)$ . Then  $s^{k+1} \in \mathcal{K}$  and  $\lambda^{k+1} \in \mathcal{N}_{\mathcal{K}}(s^{k+1})$  by [6, Prop. 6.46]. Moreover, we have

$$(22) \quad G(x^{k+1}, x^{k+1}) = \frac{\lambda^{k+1} - w^k}{\rho_k} + s^{k+1}.$$

This yields

$$(23) \quad \begin{aligned} (\lambda^{k+1}, y - G(x^{k+1}, x^{k+1})) &= \left( \lambda^{k+1}, y - \frac{1}{\rho_k}(\lambda^{k+1} - w^k) - s^{k+1} \right) \\ &\leq \frac{1}{\rho_k} \left[ (\lambda^{k+1}, w^k) - \|\lambda^{k+1}\|_H^2 \right], \end{aligned}$$

where we used  $\lambda^{k+1} \in \mathcal{N}_{\mathcal{K}}(s^{k+1})$  for the last inequality. We now show that the sequence  $\{r_k\}$  given by the right-hand side satisfies  $\limsup_{k \rightarrow \infty} r_k \leq 0$ . This yields the desired result (by replacing  $r_k$  with  $\max\{0, r_k\}$ ). If  $\{\rho_k\}$  is bounded, then (20) and (22) imply  $\|\lambda^{k+1} - w^k\|_H/\rho_k \rightarrow 0$  and therefore  $\|\lambda^{k+1} - w^k\|_H \rightarrow 0$ . This yields the boundedness of  $\{\lambda^{k+1}\}$  in  $H$  as well as  $(\lambda^{k+1}, w^k) - \|\lambda^{k+1}\|_H^2 = (\lambda^{k+1}, w^k - \lambda^{k+1}) \rightarrow 0$ . Hence,  $r_k \rightarrow 0$ . We now assume that  $\rho_k \rightarrow \infty$ . Note that (23) is a quadratic function in  $\lambda$ . A simple calculation therefore shows that  $r_k \leq \|w^k\|_H^2/(4\rho_k)$  and, hence,  $\limsup_{k \rightarrow \infty} r_k \leq 0$ .  $\square$

Let us point out that the inequality in the above lemma is uniform since the sequence  $\{r_k\}$  does not depend on the point  $y \in \mathcal{K}$ . Moreover, we remark that the proof uses only the definition of  $\lambda^{k+1}$  and does not make any assumption on the sequence  $\{x^k\}$ .

**6. Global convergence for convex constraints.** In this section, we analyze the convergence properties of Algorithm 5.1 for QVIs where the set  $\Phi(x)$  is convex for all  $x$ . In the situation where  $\Phi(x) = \{y \in C : G(x, y) \in K\}$  as in (3), the natural analytic notion of convexity is the  $\mathcal{K}_\infty$ -concavity of  $G$  with respect to  $y$  (see section 2). To reflect this, we make the following set of assumptions.

**ASSUMPTION 6.1.** We assume that (i)  $F$  is bounded and pseudomonotone, (ii)  $\Phi$  is weakly Mosco-continuous, (iii)  $G$  is  $\mathcal{K}_\infty$ -concave with respect to  $y$ , (iv)  $d_{\mathcal{K}} \circ G$  is weakly sequentially lsc, and (v)  $x^{k+1} \in C$  and  $\varepsilon^{k+1} - \mathcal{L}_{\rho_k}(x^{k+1}, w^k) \in \mathcal{N}_C(x^{k+1})$  for all  $k$ , where  $\varepsilon^k \rightarrow 0$ .

Note that (i) and (ii) were also used in Theorem 4.4. Moreover, the assumption on the sequence  $\{x^k\}$  is just an inexact version of the VI subproblem (18).

We continue by proving the feasibility and optimality of weak limit points of the sequence  $\{x^k\}$ . The first result in this direction is the following.

**LEMMA 6.2.** Let Assumption 6.1 hold and let  $\bar{x}$  be a weak limit point of  $\{x^k\}$ . If  $\Phi(\bar{x})$  is nonempty, then  $\bar{x}$  is feasible.

*Proof.* Let  $I \subseteq \mathbb{N}$  be an index set such that  $x^{k+1} \rightarrow_I \bar{x}$ . Observe first that  $\bar{x} \in C$  since  $C$  is closed and convex, and hence weakly sequentially closed. It therefore remains to show that  $G(\bar{x}, \bar{x}) \in K$ . If  $\{\rho_k\}$  remains bounded, then the penalty updating scheme (20) implies

$$d_{\mathcal{K}}(G(x^{k+1}, x^{k+1})) \leq \left\| G(x^{k+1}, x^{k+1}) - P_{\mathcal{K}}\left(G(x^{k+1}, x^{k+1}) + \frac{w^k}{\rho_k}\right) \right\|_H \rightarrow 0.$$

Since  $d_{\mathcal{K}} \circ G$  is weakly sequentially lsc, we obtain  $d_{\mathcal{K}}(G(\bar{x}, \bar{x})) = 0$  and thus  $G(\bar{x}, \bar{x}) \in K$ . Assume now that  $\rho_k \rightarrow \infty$  and that  $G(\bar{x}, \bar{x}) \notin K$ , or equivalently  $d_{\mathcal{K}}^2(G(\bar{x}, \bar{x})) > 0$ . Since  $\Phi(\bar{x})$  is nonempty, we can choose an  $y \in \Phi(\bar{x})$ , and by inner semicontinuity there exists a sequence  $y^{k+1} \in \Phi(x^{k+1})$  such that  $y^{k+1} \rightarrow_I y$ . Now, let  $h_k(x, y) := d_{\mathcal{K}}(G(x, y) + w^k/\rho_k)$ . Observe that  $h_k$  is convex in  $y$  by Lemma 2.9 and that  $h_k^2$  is continuously differentiable in  $y$  by [6, Cor. 12.30]. Moreover, since the distance function is nonexpansive, it follows that  $h_k(x^{k+1}, y^{k+1}) \leq \|w^k\|_H/\rho_k \rightarrow 0$ . Thus, using Assumption 6.1(iv), we obtain

$$\liminf_{k \rightarrow \infty} [h_k(x^{k+1}, x^{k+1}) - h_k(x^{k+1}, y^{k+1})] = \liminf_{k \rightarrow \infty} h_k(x^{k+1}, x^{k+1}) \geq d_{\mathcal{K}}(G(\bar{x}, \bar{x})) > 0.$$

Hence, there is a constant  $c_1 > 0$  such that  $h_k^2(x^{k+1}, x^{k+1}) - h_k^2(x^{k+1}, y^{k+1}) \geq c_1$  for all  $k \in I$  sufficiently large. The convexity of  $h_k$  (and of  $h_k^2$ ) now yields

$$\langle D_y(h_k^2)(x^{k+1}, x^{k+1}), y^{k+1} - x^{k+1} \rangle \leq h_k^2(x^{k+1}, y^{k+1}) - h_k^2(x^{k+1}, x^{k+1}) \leq -c_1.$$

Furthermore, by (i), there is a constant  $c_2 \in \mathbb{R}$  such that  $\langle F(x^{k+1}), x^{k+1} - y^{k+1} \rangle \geq c_2$  for all  $k \in I$ . Now, let  $\{\varepsilon^k\}$  be the sequence from Assumption 6.1. Observe that  $\mathcal{L}_{\rho_k}(x^{k+1}, w^k) = F(x^{k+1}) + (\rho_k/2)D_y(h_k^2)(x^{k+1}, x^{k+1})$ . Therefore,

$$\langle \varepsilon^{k+1}, y^{k+1} - x^{k+1} \rangle \leq \langle \mathcal{L}_{\rho_k}(x^{k+1}, w^k), y^{k+1} - x^{k+1} \rangle \leq -\frac{\rho_k c_1}{2} - c_2 \rightarrow -\infty,$$

which contradicts  $\varepsilon^{k+1} \rightarrow 0$ .  $\square$

Since the augmented Lagrangian method is, at its heart, a penalty-type algorithm, the attainment of feasibility is paramount to the success of the algorithm. The above

lemma gives us some information in this direction since it guarantees that the weak limit point  $\bar{x}$  is “as feasible as possible” in the sense that it is feasible if (and only if)  $\Phi(\bar{x})$  is nonempty. In many particular examples of QVIs (e.g., for moving-set problems), we know a priori that  $\Phi(x)$  is nonempty for all  $x \in X$ , and this directly yields the feasibility of  $\bar{x}$ .

**THEOREM 6.3.** *Let Assumption 6.1 hold and let  $\bar{x}$  be a weak limit point of  $\{x^k\}$ . If  $\Phi(\bar{x})$  is nonempty, then  $\bar{x}$  is feasible and a solution of the QVI.*

*Proof.* The feasibility follows from Lemma 6.2. For the optimality part, we apply Proposition 2.7. To this end, let  $y^k \in \Phi(x^k)$ , let  $\{\varepsilon^k\}$  be the sequence from Assumption 6.1, and recall that  $\mathcal{L}_{\rho_k}(x^{k+1}, w^k) = \mathcal{L}(x^{k+1}, \lambda^{k+1})$ . Then

$$\begin{aligned} & \langle \varepsilon^{k+1}, y^{k+1} - x^{k+1} \rangle \\ & \leq \langle F(x^{k+1}) + D_y G(x^{k+1}, x^{k+1})^* \lambda^{k+1}, y^{k+1} - x^{k+1} \rangle \\ & = \langle F(x^{k+1}), y^{k+1} - x^{k+1} \rangle + \langle \lambda^{k+1}, D_y G(x^{k+1}, x^{k+1})(y^{k+1} - x^{k+1}) \rangle \\ & \leq \langle F(x^{k+1}), y^{k+1} - x^{k+1} \rangle + \langle \lambda^{k+1}, G(x^{k+1}, y^{k+1}) - G(x^{k+1}, x^{k+1}) \rangle, \end{aligned}$$

where we used the convexity of  $y \mapsto \langle \lambda^{k+1}, G(x^{k+1}, y) \rangle$  which follows from Lemma 2.9. Since  $G(x^{k+1}, y^{k+1}) \in \mathcal{K}$ , the last term is bounded from above by  $r_{k+1}$ , where  $\{r_k\}$  is the null sequence from Lemma 5.2. The result therefore follows from Proposition 2.7.  $\square$

The above theorem guarantees the optimality of any weak limit point of the sequence generated by Algorithm 5.1. Despite this, it should be pointed out that the result is purely “primal” in the sense that no assertions are made for the multiplier sequence. We will investigate the dual (or, more precisely, primal-dual) behavior of the augmented Lagrangian method in more detail in section 7.

If the mapping  $F$  is strongly monotone, then we obtain strong convergence of the iterates.

**COROLLARY 6.4.** *Let Assumption 6.1 hold and assume that there is a  $c > 0$  such that*

$$\langle F(x) - F(y), x - y \rangle \geq c \|x - y\|_X^2 \quad \forall x, y \in X.$$

*If  $x^k \rightharpoonup_I \bar{x}$  for some subset  $I \subseteq \mathbb{N}$  and  $\bar{x}$  is a solution of the QVI, then  $x^k \rightarrow_I \bar{x}$ .*

*Proof.* By the weak Mosco-continuity of  $\Phi$ , there is a sequence  $\bar{x}^k \in \Phi(x^k)$  such that  $\bar{x}^k \rightarrow \bar{x}$ . Recalling the proof of Theorem 6.3, we have  $\liminf_{k \in I} \langle F(x^k), \bar{x}^k - x^k \rangle \geq 0$  and therefore  $\liminf_{k \in I} \langle F(x^k), \bar{x} - x^k \rangle \geq 0$ . The strong monotonicity of  $F$  yields

$$c \|x^k - \bar{x}\|_X^2 \leq \langle F(x^k) - F(\bar{x}), x^k - \bar{x} \rangle = \langle F(x^k), x^k - \bar{x} \rangle - \langle F(\bar{x}), x^k - \bar{x} \rangle.$$

But the limsup of the first term is less than or equal to zero, and the second term converges to zero since  $x^k \rightharpoonup_I \bar{x}$ . Hence,  $\|x^k - \bar{x}\|_X \rightarrow 0$ , and the proof is complete.  $\square$

**7. Primal-dual convergence analysis.** The purpose of this section is to establish a formal connection between the primal-dual sequence  $\{(x^k, \lambda^k)\}$  generated by the augmented Lagrangian method and the KKT conditions of the QVI. The main motivation for this is the relationship (21) between the augmented Lagrangian and the ordinary Lagrangian, and the approximate normality of  $\lambda^k$  and  $G(x^k, x^k)$  from Lemma 5.2. These two components essentially constitute an asymptotic version of the KKT conditions and suggest that a careful analysis of the primal-dual sequence  $\{(x^k, \lambda^k)\}$  could lead to suitable optimality assertions.



ASSUMPTION 7.1. We assume that

- (i)  $F$  bounded and pseudomonotone,
- (ii) the mappings  $G$  and  $D_y G$  are completely continuous, and
- (iii)  $x^{k+1} \in C$  and  $\varepsilon^{k+1} - \mathcal{L}_{\rho_k}(x^{k+1}, w^k) \in \mathcal{N}_C(x^{k+1})$  for all  $k$ , where  $\varepsilon^k \rightarrow 0$ .

The above assumption is certainly natural in the sense that  $x^{k+1}$  is an approximate solution of the corresponding VI subproblem and that the degree of inexactness vanishes as  $k \rightarrow \infty$ . However, we obviously need to clarify whether this assumption can be satisfied in practice. To this end, we obtain the following result.

LEMMA 7.2. Let Assumption 7.1(i)–(ii) hold and assume that  $C$  is weakly compact. Then the VI subproblems (18) admit solutions for every  $k \in \mathbb{N}$ .

*Proof.* Observe that, for all  $k \in \mathbb{N}$ , the function  $x \mapsto \mathcal{L}_{\rho_k}(x, w^k)$  is pseudomonotone by Lemma 2.5. Hence, by Corollary 4.2, the corresponding VIs admit solutions in  $C$ .  $\square$

Our next result deals with the feasibility of the iterates. As observed in the previous section, the attainment of feasibility is crucial to the success of penalty-type methods such as the augmented Lagrangian method. This aspect is even more important for QVIs due to the inherently difficult structure of the constraints.

LEMMA 7.3. Let Assumption 7.1 hold and let  $\bar{x}$  be a weak limit point of  $\{x^k\}$ . Then  $\bar{x} \in C$  and  $-D_y(d_{\mathcal{K}}^2 \circ G)(\bar{x}, \bar{x}) \in \mathcal{N}_C(\bar{x})$ . If  $\bar{x}$  satisfies ERCQ, then  $\bar{x}$  is feasible.

*Proof.* Clearly,  $\bar{x} \in C$  since  $C$  is weakly sequentially closed. If  $\{\rho_k\}$  is bounded, then (20) implies  $d_{\mathcal{K}}(G(x^{k+1}, x^{k+1})) \rightarrow 0$ , which yields  $G(\bar{x}, \bar{x}) \in \mathcal{K}$  since  $G$  is completely continuous. Hence, there is nothing to prove. Now, let  $\rho_k \rightarrow \infty$ , let  $x^{k+1} \rightarrow_I \bar{x}$  on some subset  $I \subseteq \mathbb{N}$ , and let  $\{\varepsilon^k\} \subseteq X^*$  be the sequence from Assumption 7.1. Then

$$\varepsilon^{k+1} - F(x^{k+1}) - D_y G(x^{k+1}, x^{k+1})^* \lambda^{k+1} \in \mathcal{N}_C(x^{k+1})$$

for all  $k \in \mathbb{N}$ . We now divide this inclusion by  $\rho_k$  and use the definition of  $\lambda^{k+1}$  and the fact that  $\mathcal{N}_C(x^{k+1})$  is a cone. It follows that

$$\begin{aligned} -D_y G(x^{k+1}, x^{k+1})^* \left[ G(x^{k+1}, x^{k+1}) + \frac{w^k}{\rho_k} - P_{\mathcal{K}} \left( G(x^{k+1}, x^{k+1}) + \frac{w^k}{\rho_k} \right) \right] \\ + \frac{\varepsilon^{k+1} - F(x^{k+1})}{\rho_k} \in \mathcal{N}_C(x^{k+1}). \end{aligned}$$

Note that  $G(x^{k+1}, x^{k+1}) \rightarrow G(\bar{x}, \bar{x})$  and  $D_y G(x^{k+1}, x^{k+1}) \rightarrow D_y G(\bar{x}, \bar{x})$  by complete continuity. Thus, taking the limit  $k \rightarrow_I \infty$  and using a standard closedness property of the normal cone mapping, we obtain

$$(24) \quad -D_y G(\bar{x}, \bar{x})^* [G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x}))] \in \mathcal{N}_C(\bar{x}),$$

which is the first claim. Assume now that ERCQ holds in  $\bar{x}$ , and let  $r > 0$  be such that  $B_r^Y \subseteq G(\bar{x}, \bar{x}) + D_y G(\bar{x}, \bar{x})(C - \bar{x}) - K$ . Then, for any  $y \in B_r^Y$ , there are  $z \in C$  and  $w \in K$  such that  $y = G(\bar{x}, \bar{x}) + D_y G(\bar{x}, \bar{x})(z - \bar{x}) - w$ . In particular, we have

$$\begin{aligned} \langle G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x})), y \rangle &= \langle D_y G(\bar{x}, \bar{x})^* [G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x}))], z - \bar{x} \rangle \\ &\quad + \langle G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x})), G(\bar{x}, \bar{x}) - w \rangle. \end{aligned}$$

The first term is nonnegative by (24), and so is the second term by standard projection inequalities. Hence,  $\langle G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x})), y \rangle \geq 0$  for all  $y \in B_r^Y$ , which implies

$\langle G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x})), y \rangle = 0$  for all  $y \in B_r^Y$  and, since  $Y$  is dense in  $H$ , it follows that  $G(\bar{x}, \bar{x}) - P_{\mathcal{K}}(G(\bar{x}, \bar{x})) = 0$ . This completes the proof.  $\square$

The above is our main feasibility result for this section. Note that the assertion  $-D_y(d_{\mathcal{K}}^2 \circ G)(\bar{x}, \bar{x}) \in \mathcal{N}_C(\bar{x})$  has a rather natural interpretation: the function  $d_{\mathcal{K}}^2 \circ G$  is a measure of infeasibility, and the lemma above states that any weak limit point  $\bar{x}$  of  $\{x^k\}$  has a (partial) minimization property for this function on the set  $C$ .

The observation that an extended form of RCQ yields the actual feasibility of the point  $\bar{x}$  is motivated by similar arguments for penalty-type methods in finite dimensions. In fact, a common condition used in the convergence theory of such methods is the extended MFCQ (see section 2), which is a generalization of the ordinary MFCQ to points which are not necessarily feasible.

Having dealt with the feasibility aspect, we now turn to the main primal-dual convergence result.

**THEOREM 7.4.** *Let Assumption 7.1 hold, let  $x^k \rightarrow_I \bar{x}$  on some subset  $I \subseteq \mathbb{N}$ , and assume that  $\bar{x}$  satisfies ERCQ. Then  $\bar{x}$  is feasible, the sequence  $\{\lambda^k\}_{k \in I}$  is bounded in  $Y^*$ , and each of its weak-\* accumulation points is a Lagrange multiplier corresponding to  $\bar{x}$ .*

*Proof.* The feasibility of  $\bar{x}$  follows from Lemma 7.3. Observe furthermore that  $G(x^k, x^k) \rightarrow_I G(\bar{x}, \bar{x})$  and  $D_y G(x^k, x^k) \rightarrow_I D_y G(\bar{x}, \bar{x})$ . If  $\{r_k\}$  and  $\{\varepsilon^k\}$  are the sequences from Lemma 5.2 and Assumption 7.1, respectively, then we have

$$(25) \quad \varepsilon^k - \mathcal{L}(x^k, \lambda^k) \in \mathcal{N}_C(x^k) \quad \text{and} \quad \langle \lambda^k, y - G(x^k, x^k) \rangle \leq r_k \quad \forall y \in \mathcal{K}$$

for all  $k \geq 1$ . To prove the boundedness of  $\{\lambda^k\}_{k \in I}$ , we now proceed as follows. Applying the generalized open mapping theorem [15, Thm. 2.70] to the multifunction  $\Psi(u) := G(\bar{x}, \bar{x}) + D_y G(\bar{x}, \bar{x})u - K$  on the domain  $C - \bar{x}$ , we see that there is an  $r > 0$  such that

$$B_r^Y \subseteq G(\bar{x}, \bar{x}) + D_y G(\bar{x}, \bar{x})[(C - \bar{x}) \cap B_1^X] - K.$$

By the definition of the dual norm, we can choose a sequence  $\{y^k\} \subseteq Y$  of unit vectors such that  $\langle \lambda^k, y^k \rangle \geq \frac{1}{2} \|\lambda^k\|_{Y^*}$ . For every  $k$ , we can write

$$-ry^k = G(\bar{x}, \bar{x}) + D_y G(\bar{x}, \bar{x})(v^k - \bar{x}) - z^k$$

with  $\{v^k\} \subseteq C$  bounded and  $\{z^k\} \subseteq K$ . It follows that  $ry^k = z^k - G(x^k, x^k) - D_y G(x^k, x^k)(v^k - x^k) + \delta^k$  with  $\delta^k \rightarrow 0$  as  $k \rightarrow_I \infty$ . Assume now that  $k$  is large enough so that  $\|\delta^k\|_Y \leq r/4$ . Then, by (25),

$$\begin{aligned} \frac{r}{2} \|\lambda^k\|_{Y^*} &\leq \langle \lambda^k, ry^k \rangle \leq \langle \lambda^k, z^k - G(x^k, x^k) \rangle - \langle \lambda^k, D_y G(x^k, x^k)(v^k - x^k) \rangle + \frac{r}{4} \|\lambda^k\|_{Y^*} \\ &\leq \langle \lambda^k, z^k - G(x^k, x^k) \rangle + \langle F(x^k) - \varepsilon^k, v^k - x^k \rangle + \frac{r}{4} \|\lambda^k\|_{Y^*}. \end{aligned}$$

By (25), the first two terms are bounded from above (for  $k \in I$ ) by some constant  $c > 0$ . Reordering the above inequality yields  $\frac{r}{4} \|\lambda^k\|_{Y^*} \leq c$ , and the result follows.

Finally, let us show that every weak-\* limit point of  $\{\lambda^k\}_{k \in I}$  is a Lagrange multiplier. Without loss of generality, we assume that  $\lambda^k \rightharpoonup^* \bar{\lambda}$  for some  $\bar{\lambda} \in Y^*$  on the same subset  $I \subseteq \mathbb{N}$ , where  $\{x^k\}$  converges weakly to  $\bar{x}$ . By (25) and standard properties of the normal cone, we have  $\bar{\lambda} \in \mathcal{N}_K(G(\bar{x}, \bar{x}))$ . Now, let  $y \in C$ . Then, by (25),

$$(26) \quad \langle \varepsilon^k, y - x^k \rangle \leq \langle F(x^k), y - x^k \rangle + \langle \lambda^k, D_y G(x^k, x^k)(y - x^k) \rangle.$$

By complete continuity, we have  $D_y G(x^k, x^k) \rightarrow_I D_y G(\bar{x}, \bar{x})$  and  $D_y G(\bar{x}, \bar{x})(y - x^k) \rightarrow_I D_y G(\bar{x}, \bar{x})(y - \bar{x})$  (see [22, Thm. 1.5.1]). Hence,  $D_y G(x^k, x^k)(y - x^k) \rightarrow_I D_y G(\bar{x}, \bar{x})(y - \bar{x})$ . We now argue similarly to Proposition 2.7: inserting  $y := \bar{x}$  into (26) yields  $\liminf_{k \in I} \langle F(x^k), \bar{x} - x^k \rangle \geq 0$ . Using the pseudomonotonicity of  $F$  and the fact that  $\lambda^k \rightharpoonup_I^* \bar{\lambda}$ , we obtain that

$$\begin{aligned} & \langle F(\bar{x}), y - \bar{x} \rangle + \langle \bar{\lambda}, D_y G(\bar{x}, \bar{x})(y - \bar{x}) \rangle \\ & \geq \limsup_{k \in I} [\langle F(x^k), y - x^k \rangle + \langle \lambda^k, D_y G(x^k, x^k)(y - x^k) \rangle]. \end{aligned}$$

Since  $y \in C$  was arbitrary, this means that  $-\mathcal{L}(\bar{x}, \bar{\lambda}) \in \mathcal{N}_C(\bar{x})$ . Hence,  $(\bar{x}, \bar{\lambda})$  is a KKT point of the QVI.  $\square$

**8. Applications and numerical results.** In this section, we present some numerical applications of our theoretical framework. The main idea is to show the applicability of our approach to a variety of different problems. A comparison with other methods, possibly developed for a more specialized setting, is not provided here. We stress, however, that the paper [38] contains a comparison of augmented Lagrangian methods with the related Moreau–Yosida regularization technique for the class of infinite-dimensional optimization problems. We also note that our algorithm generalizes various methods from finite dimensions, e.g., for QVIs [32, 36, 52] or generalized Nash equilibrium problems [34]. Hence, any of the applications in those papers remain valid for the present one.

However, the much more interesting case is that of QVIs which are fundamentally infinite-dimensional. In the following, we will present three such examples, discuss their theoretical background, and then apply the augmented Lagrangian method to discretized versions of the problems. The implementation was done in MATLAB and uses the algorithmic parameters

$$\lambda^0 := 0, \quad \rho_0 := 1, \quad \tau = 0.1, \quad \gamma := 10, \quad B := [-10^6, 10^6],$$

where  $B$  is understood in the pointwise sense. Due to the occurrence of the nonsmooth projection operator, the VI subproblems arising in the algorithm are solved by a semi-smooth Newton-/Levenberg–Marquardt-type method. Specifically, given a nonsmooth system of equations  $H(x) = 0$ , the new iterate  $x^{k+1}$  is given by  $x^k + d^k$  (in the successful case), where  $d^k$  solves the regularized system  $(V_k^T V_k + \gamma_k I)d = -V_k^T H(x^k)$  for an element  $V_k$  of the generalized Jacobian of  $H$  at the current iterate  $x^k$ . The update of the nonnegative regularization parameter  $\gamma_k$  is done in a trust-region-like way. The outer and inner iterations are terminated when the residual of the corresponding first-order system drops below a certain threshold; the corresponding tolerances are chosen as  $10^{-4}$  ( $10^{-6}$ ) for the outer (inner) iterations in Examples 1 and 3, and  $10^{-6}$  ( $10^{-8}$ ) in Example 2.

Since our examples are defined in function spaces, we will typically use the notation  $(u, v)$  instead of  $(x, y)$  for the variable pairs in the space  $X^2$ . This should be rather clear from the context and hopefully does not introduce any confusion. Moreover, for a function space  $Y$ , we will denote by  $Y_+$  the nonnegative cone in  $Y$ .

**8.1. An implicit Signorini problem.** The application presented here is an implicit Signorini-type problem [7, 48]. Let  $\Omega \subseteq \mathbb{R}^2$  be a bounded domain with smooth boundary  $\Gamma$ , and let  $X$  denote the space

$$X := \{u \in H^1(\Omega) : \Delta u \in L^2(\Omega)\} \quad \text{with norm } \|u\|_X := \|u\|_{H^1(\Omega)} + \|\Delta u\|_{L^2(\Omega)}.$$

Recall that the trace operator  $\tau$  maps  $H^1(\Omega)$  into  $H^{1/2}(\Gamma)$ , that  $H^{1/2}(\Gamma)^* =: H^{-1/2}(\Gamma)$  [1], and that the normal derivative  $\partial_n : X \rightarrow H^{-1/2}(\Gamma)$  is well-defined and continuous [59]. For fixed elements  $h_0, \phi \in H^{1/2}(\Gamma)$  with  $\phi \geq 0$ , consider the set-valued mapping

$$\Phi(u) := \{v \in X : \tau v \geq h(u) \text{ on } \Gamma\}, \quad h(u) := h_0 - \langle \phi, \partial_n u \rangle,$$

where  $h : H^1(\Omega) \rightarrow H^{1/2}(\Gamma)$  and the duality pairing is understood between  $H^{1/2}(\Gamma)$  and  $H^{-1/2}(\Gamma)$ . The problem in question now is the QVI

$$u \in \Phi(u), \quad \langle Au - f, v - u \rangle \geq 0 \quad \forall v \in \Phi(u),$$

where  $A : X \rightarrow X^*$  is a monotone differential operator and  $f \in H^{-1}(\Omega)$ . This problem can be cast into our general framework by choosing

$$\begin{aligned} X &= \{u \in H^1(\Omega) : \Delta u \in L^2(\Omega)\}, \quad C := X, \quad F(u) := Au - f, \\ Y &:= H^{1/2}(\Gamma), \quad G(u, v) := \tau v - h(u), \quad K := Y_+. \end{aligned}$$

We claim that the implicit Signorini problem satisfies Assumption 6.1. The pseudomonotonicity of  $F$  follows from Lemma 2.5, and the mapping  $G$  is weakly sequentially continuous. Moreover, the Mosco-continuity of  $\Phi$  is rather straightforward to prove since  $u \mapsto \langle \phi, \partial_n u \rangle$  maps into a one-dimensional subspace of  $Y$ . Assumption 6.1 (iii) holds since  $G$  is linear in this example, and (iv) is satisfied since  $d_K$  is convex (hence weakly sequentially lsc) and  $G$  is linear and bounded (hence maps weakly convergent sequences to weakly convergent ones), which implies that the composition  $d_K \circ G$  is weakly sequentially lsc.

It follows from the theory in section 6 that every weak limit point of the sequence  $\{u^k\}$  generated by Algorithm 5.1 is a solution of the QVI. (Note that  $\Phi(u)$  is nonempty for all  $u \in X$ , as was assumed in Lemma 6.2.) Observe furthermore that, since  $C = X$ , the sequences  $\{u^k\}$  and  $\{\lambda^k\}$  generated by the algorithm satisfy

$$0 \leftarrow \langle F(u^k), h \rangle + \langle \tau^* \lambda^k, h \rangle = \langle F(u^k), h \rangle + \langle \lambda^k, \tau h \rangle$$

for all  $h \in X$ . Since the image  $\tau(X)$  of the trace operator contains  $H^{3/2}(\Gamma)$  [1], it follows that a subsequence of  $\{\lambda^k\}$  converges weak-\* in  $H^{3/2}(\Gamma)^*$ .

We now present some numerical results for the domain  $\Omega := (0, 1)^2$  and the differential operator  $Au := u - \Delta u$ . For the implementation of the method, we set  $H := L^2(\Gamma)$ ,  $K := H_+$ , and discretize the domain  $\Omega$  by means of a uniform grid with  $n \in \mathbb{N}$  points per row or column (including boundary points), i.e.,  $n^2$  points in total. The remaining problem parameters are given by  $f \equiv -1$  and  $\phi = h_0 \equiv 1$ .

$n$	16	32	64	128	256
outer it.	9	9	9	10	10
inner it.	42	41	42	47	52
$\rho_{\max}$	$10^4$	$10^4$	$10^5$	$10^5$	$10^5$

We observe that the method scales rather well with increasing dimension  $n$ . In particular, the outer iteration numbers and final penalty parameters remain nearly constant, and the increase in terms of inner iteration numbers is very moderate. (Here and in the subsequent tables, inner it. refers to the cumulated number of inner iterations used by our semismooth Newton-type method.) Figure 1 shows the resulting solution for one discretization.

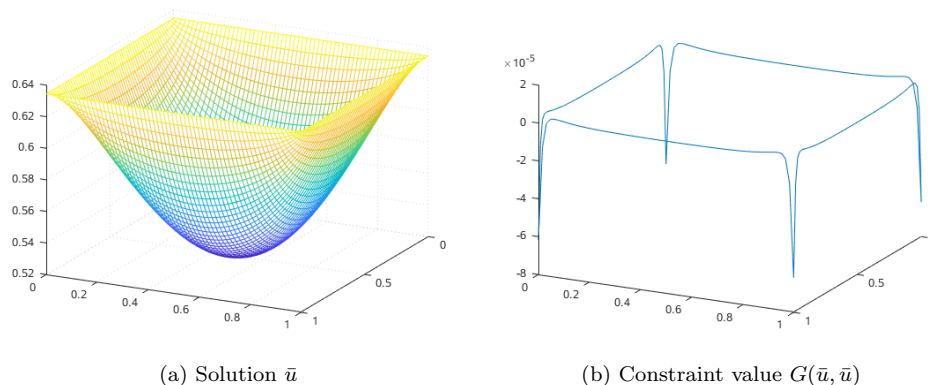


FIG. 1. Numerical results for the implicit Signorini problem with  $n = 64$ .

**8.2. Parametric gradient constraints.** This example is based on the theoretical framework in [27, 44] and represents a QVI with pointwise gradient-type constraints. Note that we have already given a brief discussion of such examples in section 4. Let  $d, p \geq 2$  and  $X := W_0^{1,p}(\Omega)$  for some bounded domain  $\Omega \subseteq \mathbb{R}^d$ . Consider the set-valued mapping

$$(27) \quad \Phi(u) := \{v \in X : \|\nabla v\| \leq \Psi(u)\},$$

where  $\|\cdot\|$  is the Euclidean norm in  $\mathbb{R}^d$  and  $\Psi : X \rightarrow L^\infty(\Omega)$ , as well as the resulting QVI

$$u \in \Phi(u), \quad \langle -\Delta_p u - f, v - u \rangle \geq 0 \quad \forall v \in \Phi(u),$$

where  $f \in X^*$  and  $\Delta_p : X \rightarrow X^*$  is the  $p$ -Laplacian defined by

$$\langle \Delta_p u, v \rangle := - \int_{\Omega} \|\nabla u(x)\|^{p-2} \nabla u(x)^T \nabla v(x) \, dx.$$

Hence, we have  $F(u) := -\Delta_p u - f$ . Observe that  $F$  is monotone, bounded, and continuous [27], and hence pseudomonotone by Lemma 2.5(i). Assume now that  $\Psi$  is completely continuous and satisfies  $\Psi u \geq c_1$  for all  $u$  and some  $c_1 > 0$ . Then  $\Phi$  is weakly Mosco-continuous by [44, Lem. 1].

When applying the augmented Lagrangian method to the above problem, a significant challenge lies in the analytical formulation of the feasible set. Observe that the original formulation in (27) is nonsmooth. This issue is probably not critical since the nonsmoothness is a rather “mild” one, but nevertheless an alternative formulation is necessary to formally apply our algorithmic framework. To this end, we first discretize the problem and then reformulate the (finite-dimensional) gradient constraint as  $G(u, v) \leq 0$  with  $G(u, v) = \|\nabla v\|^2 - \Psi(u)^2$ .

For the discretized problems, both Assumptions 6.1 and 7.1 are satisfied. Moreover, the feasible sets  $\Phi(u)$  are nonempty for all  $u \in X$ . Hence, it follows from Theorem 6.3 that every limit point  $\bar{u}$  of the (finite-dimensional) sequence  $\{u^k\}$  is a solution of the QVI. For the boundedness of the multiplier sequence, it remains to verify that RCQ holds in  $\bar{u}$ . In fact, RCQ (which, in this case, is just MFCQ) holds everywhere, since the point zero is a Slater point of the mapping  $v \mapsto G(u, v)$  for all  $u \in X$ .

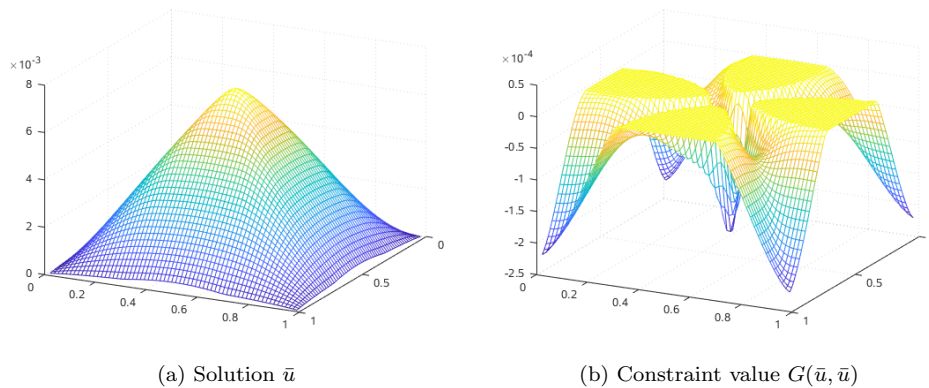


FIG. 2. Numerical results for the gradient-constrained QVI with  $n = 64$ .

As a numerical application, we consider a slightly modified version of [27, Ex. 3] on the unit square  $\Omega := (0, 1)^2$ . Note that the original example in this reference is actually solved by the solution of the  $p$ -Laplace equation  $-\Delta_p u - f = 0$ ,  $u \in W_0^{1,p}(\Omega)$ . Hence, we have modified the example to use the same function  $f$  as in [27] but with the constraint function replaced by  $\Psi(u) := 0.01 + 2|\int_{\Omega} u(x) dx|$ .

Similarly to [27], we discretize the problem with  $n \in \mathbb{N}$  interior points per row or column and use backward differences to approximate the gradient and  $p$ -Laplace operators.

$n$	16	32	64	128	256
outer it.	8	7	7	7	7
inner it.	63	52	63	73	103
$\rho_{\max}$	$10^4$	$10^4$	$10^4$	$10^4$	$10^5$

As before, the method scales rather well with increasing dimension  $n$ , and the outer iteration numbers and final penalty parameters remain nearly constant. The solution and the constraint value are depicted in Figure 2.

**8.3. Generalized Nash equilibrium problems.** The example presented in this section is a generalized Nash equilibrium problem (GNEP) arising from optimal control. Let  $N \in \mathbb{N}$  be the number of players, each in control of a variable  $u^\nu \in X_\nu$  for some real Banach space  $X_\nu$ . We set  $X := X_1 \times \cdots \times X_N$  and write  $u = (u^\nu, u^{-\nu})$ ,  $X = X_\nu \times X_{-\nu}$ , to emphasize the role of player  $\nu$  in the vector  $u$  or the space  $X$ . The GNEP consists of  $N$  optimization problems of the form

$$\min_{u^\nu \in X_\nu} J_\nu(u) \quad \text{s.t.} \quad u^\nu \in \Phi_\nu(u^{-\nu}),$$

where  $J_\nu : X \rightarrow \mathbb{R}$  is the objective function of player  $\nu$  and  $\Phi_\nu : X_{-\nu} \rightrightarrows X_\nu$  the corresponding feasible set. Throughout this section, we consider a special type of GNEP arising from optimal control. To this end, let  $\Omega \subseteq \mathbb{R}^d$ ,  $d \in \{2, 3\}$ , be a bounded Lipschitz domain, and set  $X := L^2(\Omega)^N$ . Furthermore, let  $S : L^2(\Omega) \rightarrow H_0^1(\Omega) \cap C(\bar{\Omega})$  be the solution operator of the standard Poisson equation [60], let  $\psi \in C(\bar{\Omega})$ ,  $f, y_d^\nu \in L^2(\Omega)$ ,  $\alpha_\nu > 0$  for all  $\nu$ , and let  $y(u) := S(u^1 + \cdots + u^N + f)$ . Consider the objective functions

$$J_\nu(u) := \frac{1}{2} \|y(u) - y_d^\nu\|_{L^2(\Omega)}^2 + \frac{\alpha_\nu}{2} \|u^\nu\|_{L^2(\Omega)}^2,$$

as well as the feasible set mappings

$$\Phi_\nu(u^{-\nu}) := \{u^\nu \in U_{\text{ad}}^\nu : y(u^\nu, u^{-\nu}) \geq \psi\},$$

where  $U_{\text{ad}}^\nu \subseteq L^2(\Omega)$  is the set of admissible controls for player  $\nu$ .

The existence of solutions of the GNEP can be shown as in [29]. The GNEP can also be rewritten as a QVI by defining  $F(u) := (D_{u^\nu} J_\nu)_{\nu=1}^N$  and  $\Phi(u) := \prod_{\nu=1}^N \Phi_\nu(u^{-\nu})$ . The mapping  $F$  is continuous and strongly monotone [33], and hence pseudomonotone. Observe that  $\Phi$  can be written in analytic form by setting  $C := \prod_{\nu=1}^N U_{\text{ad}}^\nu$ ,  $Y := C(\bar{\Omega})^N$ ,  $K := Y_+$ , and

$$G(u, v) := (y(u^\nu, v^{-\nu}) - \psi)_{\nu=1}^N.$$

Using the structure of  $G$ , it is now easy to see that the QVI-tailored RCQ from section 3 holds in a feasible point  $\bar{u} \in X$  if and only if, for each  $\nu$ , the Robinson constraint qualification holds in  $\bar{u}^\nu$  for the optimization problem

$$(28) \quad \min_{u^\nu \in U_{\text{ad}}^\nu} J_\nu(u^\nu, \bar{u}^{-\nu}) \quad \text{s.t.} \quad y(u^\nu, \bar{u}^{-\nu}) \geq \psi.$$

This is a standard assumption in the optimal control context which is implied, for instance, by a Slater-type condition [38]. Assume now that the QVI-tailored RCQ holds, and observe that both  $G$  and  $D_v G$  (which is constant) are completely continuous by standard results on partial differential equations [60]. It then follows from Corollary 3.5 that the feasible set mapping  $\Phi$  is weakly Mosco-continuous in  $\bar{u}$ .

We now apply the augmented Lagrangian method and choose  $H := L^2(\Omega)^N$ ,  $\mathcal{K} := H_+$ . Keeping in mind the discussion above, we obtain from Theorem 6.3 that every weak limit point  $\bar{u}$  of the sequence  $\{u^k\}$  where RCQ holds is a generalized Nash equilibrium and that the corresponding subsequence of  $\{u^k\}$  actually converges strongly to  $\bar{u}$  by Corollary 6.4. The weak-\* convergence of the multiplier sequence follows from Theorem 7.4.

As a numerical application, we consider the first problem presented in [33]. As before, we discretize the domain  $\Omega := (0, 1)^2$  by means of a uniform grid with  $n \in \mathbb{N}$  interior points per row or column. The problem in question is a four-player game where  $f := 1$ ,  $U_{\text{ad}}^\nu := \{u^\nu \in X : -12 \leq u^\nu \leq 12\}$ , and  $\alpha := (2.8859, 4.3374, 2.5921, 3.9481)$ . The state constraint  $\psi$  is given by

$$\psi(x_1, x_2) := \cos(5\sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}) + 0.1,$$

and the desired states  $y_d^\nu$  are defined as  $y_d^\nu := \xi_\nu - \xi_{5-\nu}$ , where

$$\xi_\nu(x_1, x_2) := 10^3 \max\{0, 1 - 4 \max\{|x_1 - z_\nu^1|, |x_2 - z_\nu^2|\}\}$$

with  $z^1 := (0.25, 0.75, 0.25, 0.75)$  and  $z^2 := (0.25, 0.25, 0.75, 0.75)$ .

The results of the iteration are displayed in Figure 3 and agree with those in [33]. The corresponding iteration numbers are given as follows.

$n$	16	32	64	128	256
outer it.	10	12	12	12	12
inner it.	24	29	34	39	43
$\rho_{\max}$	$10^8$	$10^{10}$	$10^{10}$	$10^{10}$	$10^{10}$

Once again, we observe that the iteration numbers and final penalty parameters remain nearly constant with increasing  $n$ .

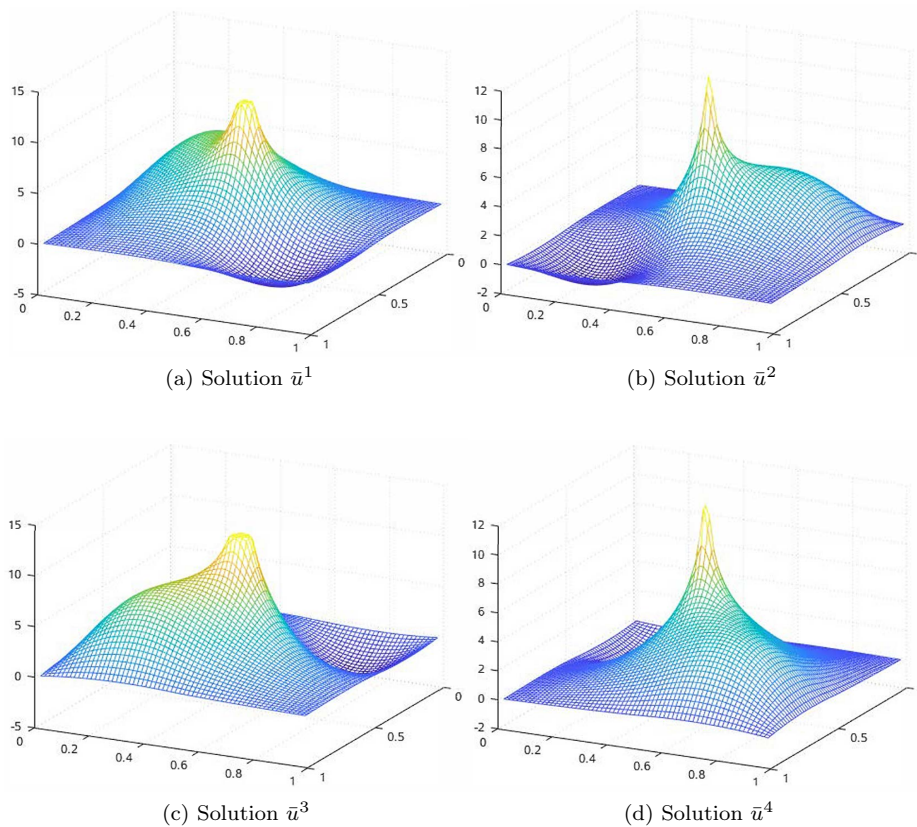


FIG. 3. Numerical results for the optimal control-type GNEP with  $n = 64$ .

**9. Final remarks.** We have presented a theoretical analysis of QVIs in infinite dimensions and isolated some key properties for their numerical treatment, namely, the pseudomonotonicity of the mapping  $F$  and the weak Mosco-continuity of the feasible set mapping. Based on these conditions, we have also given an existence result for a rather general type of QVI in Banach spaces. In addition, we have given a formal approach to the first-order optimality (KKT) conditions of QVIs by relating them to those of constrained optimization problems.

Based on the theoretical investigations, we have provided an algorithm of augmented Lagrangian type which possesses powerful global convergence properties. In particular, every weak limit point of the primal sequence is a solution of the QVI under standard assumptions, and the convergence of the multiplier sequence can be established under a suitable constraint qualification (see also section 8.1).

Despite the rather comprehensive character of the paper, there are still multiple aspects which could lead to further developments and research. One important generalization would be to consider not only QVIs of the first kind, but also quasi-variational problems of the second kind or, even more generally, so-called quasi-equilibrium problems (see [14]). In addition, it would be interesting to consider local convergence properties of the augmented Lagrangian method.



## REFERENCES

- [1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, 2nd ed., Pure Appl. Math. (Amst.) 140, Elsevier/Academic Press, Amsterdam, 2003.
- [2] S. ADLY, M. BERGOUNIOUX, AND M. AIT MANSOUR, *Optimal control of a quasi-variational obstacle problem*, J. Global Optim., 47 (2010), pp. 421–435, <https://doi.org/10.1007/s10898-008-9366-y>.
- [3] A. ALPHONSE, M. HINTERMÜLLER, AND C. N. RAUTENBERG, *Directional differentiability for elliptic quasi-variational inequalities of obstacle type*, Calc. Var. Partial Differential Equations, 58 (2019), <https://doi.org/10.1007/s00526-018-1473-0>.
- [4] R. ANDREANI, E. G. BIRGIN, J. M. MARTÍNEZ, AND M. L. SCHUVERDT, *On augmented Lagrangian methods with general lower-level constraints*, SIAM J. Optim., 18 (2007), pp. 1286–1309, <https://doi.org/10.1137/060654797>.
- [5] C. BAIocchi AND A. CAPELO, *Variational and Quasivariational Inequalities*, Wiley, New York, 1984.
- [6] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011, <https://doi.org/10.1007/978-1-4419-9467-7>.
- [7] A. BENSOUSSAN AND J.-L. LIONS, *Impulse Control and Quasivariational Inequalities*,  $\mu$ , Gauthier-Villars, Montrouge, 1984.
- [8] P. BEREMLIJSKI, J. HASLINGER, M. KOČVARA, AND J. OUTFRATA, *Shape optimization in contact problems with Coulomb friction*, SIAM J. Optim., 13 (2002), pp. 561–587, <https://doi.org/10.1137/S1052623401395061>.
- [9] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [10] D. P. BERTSEKAS, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.
- [11] E. G. BIRGIN AND J. M. MARTÍNEZ, *Practical Augmented Lagrangian Methods for Constrained Optimization*, SIAM, Philadelphia, 2014, <https://doi.org/10.1137/1.9781611973365>.
- [12] G. BIRKHOFF, *On the combination of topologies*, Fund. Math., 26 (1936), pp. 156–166, <http://eudml.org/doc/212830>.
- [13] M. C. BLIEMER AND P. H. BOVY, *Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem*, Transportation Res. Part B, 37 (2003), pp. 501–519.
- [14] E. BLUM AND W. OETTLI, *From optimization and variational inequalities to equilibrium problems*, Math. Student, 63 (1994), pp. 123–145.
- [15] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res., Springer, New York, 2000, <https://doi.org/10.1007/978-1-4612-1394-9>.
- [16] H. BREZIS, *Équations et inéquations non linéaires dans les espaces vectoriels en dualité*, Ann. Inst. Fourier (Grenoble), 18 (1968), pp. 115–175, [http://www.numdam.org/item?id=AIF-1968\\_\\_18\\_1\\_115\\_0](http://www.numdam.org/item?id=AIF-1968__18_1_115_0).
- [17] H. BREZIS, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer, New York, 2011.
- [18] H. BREZIS, L. NIRENBERG, AND G. STAMPACCHIA, *A remark on Ky Fan's minimax principle*, Boll. Unione Mat. Ital. (4), 6 (1972), pp. 293–300.
- [19] A. R. CONN, N. I. M. GOULD, AND P. L. TOINT, *Trust-Region Methods*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, 2000, <https://doi.org/10.1137/1.9780898719857>.
- [20] M. DE LUCA AND A. MAUGERI, *Discontinuous Quasi-Variational Inequalities and Applications to Equilibrium Problems*, in Nonsmooth Optimization: Methods and Applications (Erice, 1991), Gordon and Breach, Montreux, 1992, pp. 70–75.
- [21] R. M. DUDLEY, *On sequential convergence*, Trans. Amer. Math. Soc., 112 (1964), pp. 483–507, <https://doi.org/10.2307/1994157>.
- [22] S. V. EMEL'YANOV, S. K. KOROVIN, N. A. BOBYLEV, AND A. V. BULATOV, *Homotopy of Extremal Problems. Theory and Applications*, De Gruyter Ser. Nonlinear Anal. Appl. 11, De Gruyter, Berlin, 2007, <https://doi.org/10.1515/9783110893014>.
- [23] F. FLORES-BAZÁN, *Existence theorems for generalized noncoercive equilibrium problems: The quasi-convex case*, SIAM J. Optim., 11 (2001), pp. 675–690, <https://doi.org/10.1137/S1052623499364134>.
- [24] P. HAMMERSTEIN AND R. SELTEN, *Game Theory and Evolutionary Biology*, in Handbook of Game Theory with Economic Applications, Vol. II, Handbooks in Econom. 11, North-Holland, Amsterdam, 1994, pp. 929–993.
- [25] P. T. HARKER, *Generalized Nash games and quasi-variational inequalities*, European J. Oper. Res., 54 (1991), pp. 81–94.
- [26] J. HASLINGER AND P. D. PANAGIOTOPOULOS, *The reciprocal variational approach to the Signorini problem with friction. Approximation results*, Proc. Roy. Soc. Edinburgh Sect. A, 98 (1984), pp. 365–383, <https://doi.org/10.1017/S0308210500013536>.

- [27] M. HINTERMÜLLER AND C. N. RAUTENBERG, *A sequential minimization technique for elliptic quasi-variational inequalities with gradient constraints*, SIAM J. Optim., 22 (2012), pp. 1224–1257, <https://doi.org/10.1137/110837048>.
- [28] M. HINTERMÜLLER AND C. N. RAUTENBERG, *Parabolic quasi-variational inequalities with gradient-type constraints*, SIAM J. Optim., 23 (2013), pp. 2090–2123, <https://doi.org/10.1137/120874308>.
- [29] M. HINTERMÜLLER AND T. SUROWIEC, *A PDE-constrained generalized Nash equilibrium problem with pointwise control and state constraints*, Pac. J. Optim., 9 (2013), pp. 251–273.
- [30] T. ICHIIISHI, *Game Theory for Economic Analysis*, Econom. Theory Econom. Math. Econom., Academic Press, New York, 1983.
- [31] W. JING-YUAN AND Y. SMEERS, *Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices*, Oper. Res., 47 (1999), pp. 102–112.
- [32] C. KANZOW, *On the multiplier-penalty-approach for quasi-variational inequalities*, Math. Program., 160 (2016), pp. 33–63, <https://doi.org/10.1007/s10107-015-0973-3>.
- [33] C. KANZOW, V. KARL, D. STECK, AND D. WACHSMUTH, *The multiplier-penalty method for generalized Nash equilibrium problems in Banach spaces*, SIAM J. Optim., 29 (2019), pp. 767–793, <https://doi.org/10.1137/17M114114X>.
- [34] C. KANZOW AND D. STECK, *Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems*, SIAM J. Optim., 26 (2016), pp. 2034–2058, <https://doi.org/10.1137/16M1068256>.
- [35] C. KANZOW AND D. STECK, *An example comparing the standard and safeguarded augmented Lagrangian methods*, Oper. Res. Lett., 45 (2017), pp. 598–603, <https://doi.org/10.1016/j.orl.2017.09.005>.
- [36] C. KANZOW AND D. STECK, *Augmented Lagrangian and exact penalty methods for quasi-variational inequalities*, Comput. Optim. Appl., 69 (2018), pp. 801–824, <https://doi.org/10.1007/s10589-017-9963-0>.
- [37] C. KANZOW AND D. STECK, *On error bounds and multiplier methods for variational problems in Banach spaces*, SIAM J. Control Optim., 56 (2018), pp. 1716–1738, <https://doi.org/10.1137/17M1146518>.
- [38] C. KANZOW, D. STECK, AND D. WACHSMUTH, *An augmented Lagrangian method for optimization problems in Banach spaces*, SIAM J. Control Optim., 56 (2018), pp. 272–291, <https://doi.org/10.1137/16M1107103>.
- [39] S. KARAMARDIAN, *Complementarity problems over cones with monotone and pseudomonotone maps*, J. Optim. Theory Appl., 18 (1976), pp. 445–454, <https://doi.org/10.1007/BF00932654>.
- [40] V. KARL AND D. WACHSMUTH, *An augmented Lagrange method for elliptic state constrained optimal control problems*, Comput. Optim. Appl., 69 (2018), pp. 857–880, <https://doi.org/10.1007/s10589-017-9965-y>.
- [41] I. KHARROUBI, J. MA, H. PHAM, AND J. ZHANG, *Backward SDEs with constrained jumps and quasi-variational inequalities*, Ann. Probab., 38 (2010), pp. 794–840, <https://doi.org/10.1214/09-AOP496>.
- [42] W. K. KIM, *Remark on a generalized quasivariational inequality*, Proc. Amer. Math. Soc., 103 (1988), pp. 667–668, <https://doi.org/10.2307/2047197>.
- [43] A. S. KRAVCHUK AND P. J. NEITTAANMÄKI, *Variational and Quasi-Variational Inequalities in Mechanics*, Solid Mech. Appl. 147, Springer, New York, 2007, <https://doi.org/10.1007/978-1-4020-6377-0>.
- [44] M. KUNZE AND J. F. RODRIGUES, *An elliptic quasi-variational inequality with gradient constraints and some of its applications*, Math. Methods Appl. Sci., 23 (2000), pp. 897–908, [https://doi.org/10.1002/1099-1476\(20000710\)23:10<897::AID-MMA141>3.0.CO;2-H](https://doi.org/10.1002/1099-1476(20000710)23:10<897::AID-MMA141>3.0.CO;2-H).
- [45] M. B. LIGNOLA AND J. MORGAN, *Convergence of solutions of quasi-variational inequalities and applications*, Topol. Methods Nonlinear Anal., 10 (1997), pp. 375–385, <https://doi.org/10.12775/TMNA.1997.038>.
- [46] R. E. MEGGINSON, *An Introduction to Banach Space Theory*, Grad. Texts in Math. 183, Springer, New York, 1998, <https://doi.org/10.1007/978-1-4612-0603-3>.
- [47] J.-J. MOREAU, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci. Paris, 255 (1962), pp. 238–240.
- [48] U. MOSCO, *Implicit variational problems and quasi variational inequalities*, Lecture Notes in Math. 543, (1976), pp. 83–156.
- [49] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.
- [50] M. A. NOOR, *Existence results for quasi variational inequalities*, Banach J. Math. Anal., 1 (2007), pp. 186–194, <https://doi.org/10.15352/bjma/1240336215>.

- [51] J. OUTRATA, M. KOČVARA, AND J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. Theory, Applications and Numerical Results*, Nonconvex Optim. Appl. 28, Kluwer Academic, Dordrecht, 1998, <https://doi.org/10.1007/978-1-4757-2825-5>.
- [52] J.-S. PANG AND M. FUKUSHIMA, *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Comput. Manag. Sci., 2 (2005), pp. 21–56, <https://doi.org/10.1007/s10287-004-0010-0>.
- [53] M. RENARDY AND R. C. ROGERS, *An Introduction to Partial Differential Equations*, 2nd ed., Texts Appl. Math. 13, Springer, New York, 2004.
- [54] J. F. RODRIGUES AND L. SANTOS, *A parabolic quasi-variational inequality arising in a superconductivity model*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (4), 29 (2000), pp. 153–169.
- [55] W. RUDIN, *Functional Analysis*, 2nd ed., Internat. Ser. Pure Appl. Math., McGraw-Hill, New York, 1991.
- [56] L. SCRIMALI, *Quasi-variational inequalities in transportation networks*, Math. Models Methods Appl. Sci., 14 (2004), pp. 1541–1560, <https://doi.org/10.1142/S0218202504003714>.
- [57] M. H. SHIH AND K.-K. TAN, *Generalized quasi-variational inequalities in locally convex topological vector spaces*, J. Math. Anal. Appl., 108 (1985), pp. 333–343, [https://doi.org/10.1016/0022-247X\(85\)90029-0](https://doi.org/10.1016/0022-247X(85)90029-0).
- [58] D. STECK, *Lagrange Multiplier Methods for Constrained Optimization and Variational Problems in Banach Spaces*, Ph.D. thesis, University of Würzburg, 2018.
- [59] L. TARTAR, *An Introduction to Sobolev Spaces and Interpolation Spaces*, Lect. Notes Unione Mat. Ital. 3, Springer, New York, 2007.
- [60] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations*, American Mathematical Society, Providence, RI, 2010, <https://doi.org/10.1090/gsm/112>.
- [61] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications. II/B: Nonlinear Monotone Operators*, Springer, New York, 1990, <https://doi.org/10.1007/978-1-4612-0985-0>.