Augmented Lagrangian and Exact Penalty Methods for Quasi-Variational Inequalities

Christian Kanzow and Daniel Steck*

June 13, 2016

Abstract

A variant of the classical augmented Lagrangian method was recently proposed in [29, 38] for the solution of quasi-variational inequalities (QVIs). In this paper, we describe an improved convergence analysis to the method. In particular, we introduce a secondary QVI as a new optimality concept for quasi-variational inequalities and discuss several special classes of QVIs within our theoretical framework. Finally, we present a modification of the augmented Lagrangian method which turns out to be an exact penalty method, and also give detailed numerical results illustrating the performance of both methods.

1 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ and $c : \mathbb{R}^{2n} \to \mathbb{R}^r$ be given functions, and let $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be the set-valued mapping defined by

$$K(x) = \{y \in \mathbb{R}^n \mid c(y, x) \leq 0\}. \tag{1}$$

With these definitions, we consider the quasi-variational inequality problem, denoted QVI(F, K), which consists of finding a point $x \in K(x)$ such that

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

Note that we could have included equality constraints in the definition of K, but we chose to omit these for the sake of notational simplicity.

The QVI was first introduced in [3]. Note that, if c is a function of y only, then $K(x) = K$ and (2) reduces to the standard variational inequality problem (VI), which consists of finding an $x \in K$ such that

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

*University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; e-mails: kanzow@mathematik.uni-wuerzburg.de, daniel.steck@mathematik.uni-wuerzburg.de

Hence, the QVI encompasses the VI and, by extension, convex optimization problems. However, the true strength of the QVI lies in its ability to model significantly more complex problems such as generalized Nash equilibrium problems [2, 23]. Some other areas where QVIs arise include mechanics [1, 4, 24, 32, 37, 36], economics [27, 28], statistics [31], transportation [10, 7, 39], biology [22], or stationary problems in superconductivity, thermoplasticity, and electrostatics [25, 26, 33]. For a more comprehensive description of QVIs and their properties, we refer the reader to the two monographs by Baiocchi [1] and Mosco [35].

In this paper, we follow the augmented Lagrangian approach which was first used in [38] and later improved in [29]. Our aim is to give a refined convergence theory which is both simple and expressive, and includes the convergence theorems from [29] as special cases. This is done by introducing a secondary QVI called *Feasibility QVI* as a new optimality concept and, hence, splitting the convergence theory into a discussion of feasibility and optimality.

The separate treatment of feasibility and optimality turns out to be the main ingredient to obtain stronger or new convergence results for several special classes of QVIs. To this end, recall that, for optimization problems, feasibility is closely linked to optimality, in particular for penalty-type schemes which includes the augmented Lagrangian method (ALM). This motivates to have a closer look at the feasibility issue. We note that we did a similar analysis for the generalized Nash equilibrium problem [30] before which will be shown to be a special case of our current setting.

As a further extension to the algorithm, we describe a modification which is an exact penalty method. The basic approach behind this goes back to [11] (for optimization problems) and was extended to standard variational inequalities in [9]. Moreover, it turns out that our exact penalty method integrates very well with the theoretical framework of the "standard" augmented Lagrangian method.

This paper is organized as follows. In Section 2, we give some preliminary theoretical background which forms the basis for our subsequent analysis. Section 3 contains a precise statement of the augmented Lagrangian method. We continue with a compact convergence analysis in Section 4, where we introduce the aforementioned *Feasibility QVI*. In Section 5, we deal with properties of the Feasibility QVI and consider some special classes of QVIs including the generalized Nash equilibrium problem and the well-known moving set case. Finally, Section 6 describes a modification of the augmented Lagrangian method which is an exact penalty method, and we conclude with numerical results and some final remarks in Section 7 and 8, respectively.

Notation: Given a scalar α, we define $\alpha_+ := \max\{0, \alpha\}$. Similarly, for a vector x, we write x_+ for the vector where the plus-operator is applied to each component. A vector-valued function $q : \mathbb{R}^n \to \mathbb{R}^m$ is called convex if all component functions are convex. Finally, for a continuously differentiable function $c : \mathbb{R}^{2n} \to \mathbb{R}^r$ in the variables (y, x), we denote the partial (with respect to y) transposed Jacobian by $\nabla_y c(y, x)$. Hence, for the i-th component, $\nabla_y c_i(y, x)$ is the gradient, viewed as a column vector.

2 Preliminaries

This section contains some theoretical background on QVIs. We first recall the definition of a KKT point.

Definition 2.1. *A tuple (x, λ) is called a KKT point of* (2) *if the system*

$$F(x) + \nabla_y c(x,x)\lambda = 0 \quad \text{and} \quad \min\{-c(x,x), \lambda\} = 0 \tag{3}$$

is satisfied.

Note that the condition $\min\{-c(x,x), \lambda\} = 0$ is equivalent to $c(x,x) \le 0$, $\lambda \ge 0$ and $c(x,x)^T\lambda = 0$. Any vector $x \in \mathbb{R}^n$ satisfying $x \in K(x)$, i.e. $c(x,x) \le 0$, will be called *feasible* for the QVI from (2), (1). Using an idiom which is common in optimization theory, we will also call a point $x \in \mathbb{R}^n$ a KKT point of the QVI if there is a multiplier vector λ such that (3) holds.

The relation between the QVI and its KKT conditions is well-established [15] and is, essentially, a generalization of the well-known correspondence for classical optimization problems. Note that, in particular, the KKT conditions are sufficient optimality conditions if the set $K(x)$ is convex for every x.

We next introduce some constraint qualifications. Note that we call a collection of vectors $v_1, \ldots, v_k$ *positively linearly dependent* if the system $\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$, $\lambda \ge 0$, has a nontrivial solution. Otherwise, the vectors are called *positively linearly independent.*

Definition 2.2. *Consider a constraint function $c : \mathbb{R}^{2n} \to \mathbb{R}^r$ and a point $\bar{x} \in \mathbb{R}^n$. We say that*

(a) *LICQ holds in $\bar{x}$ if the set of gradients $\nabla_y c_i(\bar{x}, \bar{x})$ with $c_i(\bar{x}, \bar{x}) = 0$ is linearly independent.*

(b) *MFCQ holds in $\bar{x}$ if the set of gradients $\nabla_y c_i(\bar{x}, \bar{x})$ with $c_i(\bar{x}, \bar{x}) = 0$ is positively linearly independent.*

(c) *EMFCQ holds in $\bar{x}$ if the set of gradients $\nabla_y c_i(\bar{x}, \bar{x})$ with $c_i(\bar{x}, \bar{x}) \ge 0$ is positively linearly independent.*

(d) *CPLD holds in $\bar{x}$ if, for every $I \subset \{i \mid c_i(\bar{x}, \bar{x}) = 0\}$ such that the vectors $\nabla_y c_i(\bar{x}, \bar{x})$ $(i \in I)$ are positively linearly dependent, there is a neighbourhood of $\bar{x}$ where the gradients $\nabla_y c_i(x, x)$ $(i \in I)$ are linearly dependent.*

In the simple case where c does not depend on a second argument, it is clear that the above conditions are equivalent to their well-known classical counterparts. The only difference is that we define, for example, LICQ at an arbitrary point $\bar{x}$ that is not necessarily feasible for the underlying QVI. This slightly more general definition is required in our exact penalty approach, cf. Section 6.

Based on the CPLD constraint qualification, which is the weakest among the above conditions, we now prove the following theorem which describes the relationship between the KKT conditions and an asymptotic analogue thereof.

Theorem 2.3. *Let (x^k) be a sequence converging to $\bar{x}$ and let $(\lambda^k) \subset \mathbb{R}^r$ be a sequence of multipliers such that*

$$F(x^k) + \nabla_y c(x^k, x^k)\lambda^k \to 0 \quad \text{and} \quad \min\{-c(x^k, x^k), \lambda^k\} \to 0.$$

Then, if CPLD holds in $\bar{x}$, it follows that $\bar{x}$ is a KKT point of the QVI.

Proof. First note that our assumption implies that the limit $\bar{x}$ is feasible with respect to the constraints $c(x, x) \le 0$. Moreover, we point out that the assumption remains valid if we replace λ^k by λ^k_+, hence we may assume, without loss of generality, that $\lambda^k \ge 0$ for all k. Since $\lambda^k_i \to 0$ holds for every i with $c_i(\bar{x}, \bar{x}) < 0$, we obtain

$$F(x^k) + \sum_{c_i(\bar{x},\bar{x})=0} \lambda_i^k \nabla_y c_i(x^k, x^k) \to 0.$$

Using a Carathéodory-type result, cf. [6, Lem. 3.1], we can choose subsets $I^k \subset \{i \mid c_i(\bar{x}, \bar{x}) = 0\}$ such that the gradients $\nabla_y c_i(x^k, x^k)$, $i \in I^k$, are linearly independent and we can write

$$\sum_{c_i(\bar{x},\bar{x})=0} \lambda_i^k \nabla_y c_i(x^k, x^k) = \sum_{i \in I^k} \hat{\lambda}_i^k \nabla_y c_i(x^k, x^k)$$

with suitable vectors $\hat{\lambda}^k \ge 0$. Subsequencing if necessary, we may assume that $I^k = I$ for every k, i.e. we get

$$F(x^k) + \sum_{i \in I} \hat{\lambda}_i^k \nabla_y c_i(x^k, x^k) \to 0. \tag{4}$$

Hence, to conclude the proof, it suffices to show that $(\hat{\lambda}^k)$ is bounded. If this were not the case, we could divide (4) by $\|\hat{\lambda}^k\|$, take the limit $k \to \infty$ on a suitable subsequence and obtain a nontrivial positive linear combination of the gradients $\nabla_y c_i(\bar{x}, \bar{x})$, $i \in I$, which vanishes. Hence, by CPLD, these gradients should be linearly dependent in a neighbourhood of $\bar{x}$, which is a contradiction. □

Theorem 2.3 is a natural analogue of [6, Thm. 3.6] and will be a central building block for our subsequent analysis. Note that it is easy to carry out a similar proof under MFCQ and conclude that, in this case, the sequence (λ^k) itself must be bounded. Due to the similarity of the two proofs, we have omitted this additional result. Note also that Theorem 2.3 is slightly more general than the corresponding result on approximate KKT points of optimization problems in [6] since we do not require the intermediate multipliers λ^k to be nonnegative.

3 The Augmented Lagrangian Method

Throughout the remainder of this paper, we consider QVIs where the constraint function c has the decomposition

$$c(y, x) = \begin{pmatrix} g(y, x) \\ h(y, x) \end{pmatrix}$$

with two continuously differentiable functions $g : \mathbb{R}^{2n} \to \mathbb{R}^m$ and $h : \mathbb{R}^{2n} \to \mathbb{R}^p$. The purpose of this approach is to account for the possibility of partial penalization: the constraints defined by g will be penalized, whereas h is an (optional) constraint function which will stay as a constraint in the penalized subproblems. We stress that this framework is very general and gives us some flexibility to deal with different situations. The most natural choices are probably the following ones:

1. Penalize all contraints. This full penalization approach is the simplest and most straightforward approach where, formally, we set $p = 0$. The resulting subproblems are unconstrained and therefore become nonlinear equations.

2. Another natural splitting is the case where h covers the non-parametric constraints (i.e. those which do not depend on x), whereas g subsumes the remaining constraints. The resulting penalized problems then become standard VIs and are therefore easier to solve than the original QVI since the (presumably) difficult constraints are moved into the objective function. This is the approach taken in [29, 38].

3. Finally, for certain problems, it might make sense to include some of the parametric constraints into h. In this case, the subproblems themselves are QVIs, but might still be easier to solve than the original QVI, for example, in the particular case where the penalized subproblems belong to a special class of QVIs such as the moving-set class.

Note that the decomposition of c into g and h also entails two multiplier vectors, which we will usually refer to as λ and μ, respectively. For instance, the KKT conditions from Definition 2.1 take on the form

$$F(x) + \nabla_y g(x,x)\lambda + \nabla_y h(x,x)\mu = 0,$$
$$\min\{-g(x,x), \lambda\} = 0, \quad \text{and} \quad \min\{-h(x,x), \mu\} = 0.$$

In order to formally describe the partial penalization scheme, we now consider the set-valued mapping $K_h : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which is given by

$$K_h(x) = \{y \in \mathbb{R}^n \mid h(y,x) \le 0\}. \tag{5}$$

Finally, we define the (partial) Lagrangian and augmented Lagrangian as

$$\begin{aligned} L(x,\lambda) &= F(x) + \nabla_y g(x,x)\lambda, \\ L_\rho(x,\lambda) &= F(x) + \nabla_y g(x,x)(\lambda + \rho g(x,x))_+. \end{aligned} \tag{6}$$

Here, $\rho > 0$ is a given penalty parameter and $\lambda \in \mathbb{R}^m$ is a multiplier. A basic ALM considers a sequence of QVIs, each defined by the augmented Lagrangian (for given values ρ and λ) and the set-valued mapping K_h. The following is a precise statement of our augmented Lagrangian method for the solution of QVIs. Note that the method is identical to the one used in [29].

Algorithm 3.1. *(Augmented Lagrangian method for QVIs)*

(S.0) Let $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^{n+m+p}$, $\rho_0 > 0$, $u^{\max} \ge 0$, $\gamma > 1$, $\tau \in (0,1)$, *and set* $k = 0$.

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

(S.2) *Choose* $u^k \in [0, u^{\max}]^m$ *and compute an approximate KKT point (see below)* (x^{k+1}, μ^{k+1}) *of the penalized QVI, which consists of finding an* $x \in K_h(x)$ *such that*

$$L_{\rho_k}(x, u^k)^T(y - x) \geq 0 \quad \forall y \in K_h(x). \tag{7}$$

(S.3) *Set* $\lambda^{k+1} = \big(u^k + \rho_k g(x^{k+1}, x^{k+1})\big)_+$. *If*

$$\|\min\{-g(x^{k+1}, x^{k+1}), \lambda^{k+1}\}\| \leq \tau \|\min\{-g(x^k, x^k), \lambda^k\}\| \tag{8}$$

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

(S.4) *Set* $k \leftarrow k + 1$ *and go to (S.1).*

Note that we have deliberately left some aspects of the algorithm unspecified. For instance, we do not give an explicit formula for the sequence (u^k), which is meant to be a "bounded analogue" of the multipliers. The most natural choice [6, 29] is $u^k = \min\{\lambda^k, u^{\max}\}$, which is just the projection of λ^{k+1} onto the m-dimensional hypercube $[0, u^{\max}]^m$.

Secondly, we have not specified what constitutes an approximate KKT point in Step 2. To this end, we make the following assumption about the variables x^{k+1} and μ^{k+1}. The main idea is that there is no need to solve the penalized QVIs in (S.2) exactly, but with increasing accuracy for $k \to \infty$.

Assumption 3.2. *At Step 2 of Algorithm 3.1, we obtain* x^{k+1} *and* μ^{k+1} *such that*

$$L_{\rho_k}(x^{k+1}, u^k) + \nabla_y h(x^{k+1}, x^{k+1})\mu^{k+1} \to 0$$
$$\min\{-h(x^{k+1}, x^{k+1}), \mu^{k+1}\} \to 0$$

holds for $k \to \infty$.

This is a very natural assumption which asserts that the pair (x^{k+1}, μ^{k+1}) satisfies an approximate KKT condition for the subproblems, and the degree of inexactness converges to zero for $k \to \infty$. However, it should be noted that the multipliers μ^{k+1} which arise from the subproblems are allowed to be negative. Hence, Assumption 3.2 is actually weaker than classical approximate KKT conditions which have been used in similar contexts [6, 29]. In particular, it permits the use of standard algorithms such as Newton-type methods for the (approximate) solution of the subproblems, which may yield a negative multiplier estimate.

4 Convergence Analysis

We proceed by considering the convergence properties for Algorithm 3.1. Our analysis is split into a discussion of feasibility and optimality. Regarding the former, note that Assumption 3.2 already implies that every limit point of (x^k) satisfies the h-constraints. For the discussion of feasibility with respect to g, we introduce an auxiliary QVI which consists of finding $x \in K_h(x)$ such that

$$\left(\nabla_y \|g_+(x, x)\|^2\right)^T (y - x) \geq 0 \quad \forall y \in K_h(x). \tag{9}$$

This QVI will simply be referred to as the *Feasibility QVI*. Note that the function $\|g_+(y,x)\|^2$ is continuously differentiable, so the Feasibility QVI is well-defined. It turns out that the behaviour of Algorithm 3.1 is closely related to this auxiliary problem. We proceed by formally proving this connection and then giving a possible interpretation.

Theorem 4.1. *Assume that (x^k) is a sequence generated by Algorithm 3.1 and $\bar{x}$ is a limit point of (x^k). Then, if h satisfies CPLD in $\bar{x}$, it follows that $\bar{x}$ is a KKT point of the Feasibility QVI.*

Proof. Let $K \subset \mathbb{N}$ be such that $x^{k+1} \to_K \bar{x}$. If (ρ_k) is bounded, then $\bar{x}$ is feasible and there is nothing to prove. Hence, we assume that $\rho_k \to \infty$. By Assumption 3.2, we have

$$L_{\rho_k}(x^{k+1}, u^k) + \nabla_y h(x^{k+1}, x^{k+1})\mu^{k+1} \to 0$$

for some multiplier sequence (μ^k) which satisfies $\min\{-h(x^{k+1}, x^{k+1}), \mu^{k+1}\} \to 0$. From (6), we obtain

$$F(x^{k+1}) + \nabla_y g(x^{k+1}, x^{k+1})(u^k + \rho_k g(x^{k+1}, x^{k+1}))_+ + \nabla_y h(x^{k+1}, x^{k+1})\mu^{k+1} \to 0.$$

Dividing by ρ_k and using the boundedness of (u^k) and $(F(x^{k+1}))_K$, it follows that

$$\nabla_y g(x^{k+1}, x^{k+1}) g_+(x^{k+1}, x^{k+1}) + \nabla_y h(x^{k+1}, x^{k+1}) \frac{\mu^{k+1}}{\rho_k} \to_K 0.$$

Since $\nabla_y \|g_+(x,x)\|^2 = 2\nabla_y g(x,x) g_+(x,x)$, the result follows from Theorem 2.3. □

The above theorem establishes the aforementioned connection between the Feasibility QVI (9) and Algorithm 3.1. Note that the Feasibility QVI has a very natural interpretation. If $\bar{x}$ is a solution, then we have

$$\left(\nabla_y \|g_+(\bar{x}, \bar{x})\|^2\right)^T (y - \bar{x}) \geq 0 \quad \forall y \in K_h(\bar{x}),$$

which means that, roughly speaking, we cannot find a descent direction for the constraint violation $\|g_+(x,x)\|^2$ which does not harm the feasibility with respect to h. Equivalently, $\bar{x}$ satisfies the first-order necessary conditions of the optimization problem

$$\min \ \|g_+(x, \bar{x})\|^2 \quad \text{s.t.} \quad x \in K_h(\bar{x}).$$

One of the most important special cases of Theorem 4.1 arises if we omit the function h. In other words, we are performing a full penalization. In this case, the CPLD condition is superfluous, and we obtain

$$\nabla_y \|g_+(\bar{x}, \bar{x})\|^2 = 0,$$

which means that $\bar{x}$ is a stationary point of the constraint violation $\|g_+(\cdot, \bar{x})\|^2$.

We now turn to the optimality of limit points of Algorithm 3.1. To this end, note that we can restate Assumption 3.2 as

$$\begin{gathered} F(x^{k+1}) + \nabla_y g(x^{k+1}, x^{k+1})\lambda^{k+1} + \nabla_y h(x^{k+1}, x^{k+1})\mu^{k+1} \to 0 \\ \min\{-h(x^{k+1}, x^{k+1}), \mu^{k+1}\} \to 0, \end{gathered} \tag{10}$$

which already suggests that the sequence of triples (x^k, λ^k, μ^k) satisfies an approximate KKT condition for the QVI (2). In fact, the only missing ingredient is the asymptotic complementarity of g and λ.

Theorem 4.2. *Let (x^k) be a sequence generated by Algorithm 3.1 and $\bar{x}$ be a limit point of (x^k). Assume that one of the following conditions holds:*

(a) *$\bar{x}$ is feasible and the function (g, h) satisfies CPLD in $\bar{x}$.*

(b) *The function (g, h) satisfies EMFCQ in $\bar{x}$.*

Then $\bar{x}$ is a KKT point of the QVI (2).

Proof. Due to Theorem 5.1 (which we will encounter later), it suffices to consider (a). To this end, let $K \subset \mathbb{N}$ be such that $x^{k+1} \to_K \bar{x}$. By (10) and Theorem 2.3, we only need to show that

$$\min\{-g(x^{k+1}, x^{k+1}), \lambda^{k+1}\} \to_K 0$$

holds. If (ρ_k) is bounded, this readily follows from (8). If, on the other hand, $\rho_k \to \infty$ and i is an index such that $g_i(\bar{x}, \bar{x}) < 0$, the updating scheme from Step 3 of Algorithm 3.1 together with the boundedness of (u_i^k) implies

$$\lambda_i^{k+1} = \left(u_i^k + \rho_k g_i(x^{k+1}, x^{k+1})\right)_+ = 0$$

for sufficiently large k. This completes the proof. □

The above theorem shows that attaining feasibility is a crucial aspect of Algorithm 3.1. Since we know that, by Theorem 4.1, every limit point of the sequence (x^k) is a KKT point of the Feasibility QVI, it is natural to ask whether an implication of the form

$$\bar{x} \text{ is a KKT point of the Feasibility QVI} \overset{?}{\Longrightarrow} \bar{x} \text{ is feasible}$$

holds. Judging by the interpretation of the Feasibility QVI given earlier in this section, it is natural to expect this implication to hold under certain assumptions on the constraint functions g and h. This is the subject of the following section.

5 Properties of the Feasibility QVI

We now discuss properties of the Feasibility QVI (9). The first result in this direction deals with general nonlinear constraints using EMFCQ (cf. Definition 2.2) and therefore completes the proof of Theorem 4.2.

Theorem 5.1. *Let $\bar{x}$ be a KKT point of the Feasibility QVI and assume that the function (g, h) satisfies EMFCQ in $\bar{x}$. Then $\bar{x}$ is feasible for the QVI* (2).

Proof. By assumption, there is a multiplier $\lambda \in \mathbb{R}^p$ such that

$$\nabla_y \|g_+(\bar{x}, \bar{x})\|^2 + \nabla_y h(\bar{x}, \bar{x})\lambda = 0 \quad \text{and} \quad \min\{-h(\bar{x}, \bar{x}), \lambda\} = 0.$$

Expanding the sums and omitting some vanishing terms, we obtain

$$2 \sum_{g_i(\bar{x}, \bar{x}) > 0} \nabla_y g_i(\bar{x}, \bar{x}) g_i(\bar{x}, \bar{x}) + \sum_{h_j(\bar{x}, \bar{x}) = 0} \nabla_y h_j(\bar{x}, \bar{x}) \lambda^j = 0.$$

Hence, by EMFCQ, it follows that the first sum must be empty, and $\bar{x}$ is feasible. □

Since Theorem 5.1 guarantees the feasibility of $\bar{x}$, the EMFCQ assumption boils down to standard MFCQ which is known to imply CPLD, hence all conditions from Theorem 4.2 (a) hold.

We see that our convergence theory includes, as a special case, the approach from [29, 38] which directly uses the extended MFCQ to prove corresponding optimality results. However, our approach has the advantage that it allows us to conduct a dedicated analysis for special instances of QVIs. To this end, we consider some of the classes of QVIs discussed in [17].

5.1 The Moving Set Case

Possibly the most prominent special class of QVIs is the moving set case, where we have $K(x) = c(x) + Q$ for some closed and convex set Q. Usually, Q is given by

$$Q = \{x \in \mathbb{R}^n \mid q(x) \leq 0\} \tag{11}$$

with a convex mapping $q : \mathbb{R}^n \to \mathbb{R}^m$. Recalling the constraint system (1) of the general QVI, we see that the moving set case can be modeled by the function $g(y, x) = q(y - c(x))$. For the sake of simplicity, we perform a full penalization and omit the function h. Note that

$$\nabla_y \|g_+(y, x)\|^2 = 2\nabla q(y - c(x))q_+(y - c(x))$$

holds for every $x, y \in \mathbb{R}^n$. Hence, if $\bar{x}$ is a KKT point of the Feasibility QVI (9), it follows that

$$0 = \nabla_y \|g_+(\bar{x}, \bar{x})\|^2 = 2\nabla q(\bar{x} - c(\bar{x}))q_+(\bar{x} - c(\bar{x})).$$

In other words, the point $\bar{x} - c(\bar{x})$ is a stationary point of the function $\|q_+(x)\|^2$. But this is a convex function and, hence, $\bar{x} - c(\bar{x})$ is a global minimum. This immediately yields the following result.

Theorem 5.2. *Consider the moving set case with Q given by* (11) *and assume that Q is nonempty. Then, if $\bar{x}$ is a KKT point of the Feasibility QVI, $\bar{x}$ is feasible.*

Proof. Like above, we see that $\bar{x} - c(\bar{x})$ is a global minimum of $\|q_+(x)\|^2$. But if Q is nonempty, the global minimum of this function is zero and we obtain $\bar{x} - c(\bar{x}) \in Q$ or, equivalently, $\bar{x} \in K(\bar{x})$. □

As a direct consequence of the previous results, we obtain that every limit point of Algorithm 3.1 is a KKT point of the QVI in the moving set case. The precise statement is as follows.

Corollary 5.3. *Consider the moving set case with nonempty Q given by* (11). *Then every limit point $\bar{x}$ of Algorithm 3.1 is a KKT point of the QVI provided that the corresponding function g satisfies CPLD in $\bar{x}$.*

Proof. Recall that there are no h-constraints. Hence Theorem 4.1 implies, without any further assumptions, that $\bar{x}$ is a KKT point of the Feasibility QVI. Therefore, Theorem 5.2 yields that $\bar{x}$ is feasible. The result then follows from Theorem 4.2. □

Note that we could have carried out a similar proof in the slightly more general setting where

$$g(y,x) = q_1(y - c(x)) \quad \text{and} \quad h(y,x) = q_2(y - c(x)).$$

This corresponds to the case where both g and h describe sets moving along the trajectory c, and the feasible set is given by $K(x) = c(x) + Q_1 \cap Q_2$ with Q_i defined as in (11).

5.2 Generalized Nash Equilibrium Problems

The well-known generalized Nash equilibrium problem [13, 20], which we simply refer to as GNEP, consists of $N \in \mathbb{N}$ players, each in control of a vector $x^\nu \in \mathbb{R}^{n_\nu}$. Furthermore, every player attempts to solve his respective optimization problem

$$\min_{x^\nu} \ \theta_\nu(x) \quad \text{s.t.} \quad g^\nu(x) \le 0, \quad h^\nu(x) \le 0, \tag{12}$$

where $x = (x^1, \ldots, x^N)$ denotes the vector of variables of all players. Note that we also write $x = (x^\nu, x^{-\nu})$ for the vector, where $x^{-\nu}$ subsumes the variables of all players except player ν. This notation is particularly useful to stress the importance of the block vector x^ν within x.

In order to account for the possibility of partial penalization, we have equipped each player with two constraint functions g^ν and h^ν. It is well-established [23] that, under certain convexity assumptions, the GNEP (12) is equivalent to the quasi-variational inequality problem QVI(F, K) with

$$F(x) = \begin{pmatrix} \nabla_{x^1}\theta_1(x) \\ \vdots \\ \nabla_{x^N}\theta_N(x) \end{pmatrix} \tag{13}$$

and

$$K(x) = \left\{ y \mid g^\nu(y^\nu, x^{-\nu}) \le 0, \ h^\nu(y^\nu, x^{-\nu}) \le 0 \text{ for every } \nu \right\}. \tag{14}$$

In fact, this equivalence follows from the simple observation that the KKT system of the GNEP (which is just the concatenation of the KKT systems for every player) is identical to the KKT system of the QVI(F, K).

With regard to (14), it follows that the GNEP can be modeled as a QVI by using the constraint functions

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

It is easily verified that, in this setting, Algorithm 3.1 is basically equivalent to the augmented Lagrangian method for GNEPs which was recently considered in [30]. In fact, this equivalence goes even further: a closer look at the Feasibility QVI (9) shows that

$$\nabla_y \|g_+(y,x)\|^2 = \begin{pmatrix} \nabla_{y^1}\|g_+^1(y^1, x^{-1})\|^2 \\ \vdots \\ \nabla_{y^N}\|g_+^N(y^N, x^{-N})\|^2 \end{pmatrix}.$$

Hence, the Feasibility QVI is nothing but the QVI reformulation of the GNEP where player ν's optimization problem is given by

$$\min_{x^\nu} \; \|g_+^\nu(x)\|^2 \quad \text{s.t.} \quad h^\nu(x) \leq 0.$$

This is the *Feasibility GNEP* from [30] and is used to model the feasibility properties of the respective players. A simple corollary of this analysis is the following, which corresponds to the well-known *jointly-convex* GNEP as a special case.

Corollary 5.4. *Consider a GNEP of the form* (12) *where $g^\nu = \tilde{g}$ is a joint constraint and h^ν depends on x^ν only. Then every solution of the Feasibility GNEP is a feasible point.*

Proof. See [30, Thm. 4.4]. □

Note that, in the GNEP setting, the CPLD constraint qualification from Definition 2.2 becomes the *GNEP-CPLD* condition from [30]. Furthermore, it goes without saying that a direct consequence of Corollary 5.4 is that every limit point of Algorithm 3.1 is a KKT point of the GNEP.

5.3 The Bilinear Case

Here, we have

$$g(y,x) = \begin{pmatrix} x^T Q_1 y - \gamma_1 \\ \vdots \\ x^T Q_m y - \gamma_m \end{pmatrix} \quad \text{and} \quad h(y,x) = \begin{pmatrix} q_1(y) \\ \vdots \\ q_p(y) \end{pmatrix}, \tag{15}$$

where the matrices $Q_i \in \mathbb{R}^{n \times n}$ are symmetric positive semidefinite, $\gamma_i \in \mathbb{R}$ are given numbers and the functions q_i are convex. It is easy to see that

$$\nabla_y \|g_+(y,x)\|^2 = 2\nabla_y g(y,x) g_+(y,x) = 2 \sum_{i=1}^m \max\{0, g_i(y,x)\} Q_i x$$

and the full gradient of $x \mapsto \|g_+(x,x)\|^2$ is given by

$$\nabla \left(\|g_+(x,x)\|^2 \right) = 4 \sum_{i=1}^m \max\{0, g_i(x,x)\} Q_i x.$$

Moreover, since h depends on y only, it therefore follows that every KKT point of the Feasibility QVI is a KKT point (and, hence, a global solution) of the convex optimization problem

$$\min \; \|g_+(x,x)\|^2 \quad \text{s.t.} \quad h(x,x) \leq 0. \tag{16}$$

We therefore get the following result.

Theorem 5.5. *Consider the bilinear case with g given by* (15) *and assume that the feasible set is nonempty. Then, if $\bar{x}$ is a KKT point of the Feasibility QVI, $\bar{x}$ is feasible.*

Proof. In this case, $\bar{x}$ is a global minimum of (16). But the minimum of this problem is zero and, hence, $\bar{x}$ is feasible. □

As in the previous cases, it follows that every limit point of Algorithm 3.1 is a KKT point of the QVI.

5.4 Linear Constraints with Variable Right-hand Side

For this class of QVIs, we have

$$g(y,x) = Ay - b - c(x), \tag{17}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^n$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ is a given function. For the sake of simplicity, we assume that there is no additional constraint function h. It follows that

$$\nabla_y \|g_+(y,x)\|^2 = 2\nabla_y g(y,x) g_+(y,x) = 2A^T g_+(y,x).$$

Hence, if $\bar{x}$ is a KKT point of the Feasibility QVI (9), we have

$$0 = \nabla_y \|g_+(\bar{x},\bar{x})\|^2 = 2A^T g_+(\bar{x},\bar{x}).$$

This motivates the following theorem.

Theorem 5.6. *Consider a QVI where g is given by* (17) *and* $\operatorname{rank}(A) = m$. *Then, if $\bar{x}$ is a KKT point of the Feasibility QVI, $\bar{x}$ is feasible.*

Proof. The above formulas show that $A^T g_+(\bar{x},\bar{x}) = 0$ and, hence, $g_+(\bar{x},\bar{x}) = 0$. □

As in the previous cases, it follows that every limit point of Algorithm 3.1 is a KKT point of the QVI.

5.5 Box Constraints

Here, we have

$$g(y,x) = \begin{pmatrix} \ell(x) - y \\ y - u(x) \end{pmatrix}, \tag{18}$$

where $\ell, u : \mathbb{R}^n \to \mathbb{R}^n$ are given mappings which describe lower and upper bounds on the variable y, where these bounds are allowed to depend on the vector x. Note that this is actually a particular case of linear constraints with variable right-hand side. More precisely, it corresponds to the general setting (17) by means of the definitions

$$A = \begin{pmatrix} -I \\ I \end{pmatrix} \quad \text{and} \quad c(x) = \begin{pmatrix} \ell(x) \\ -u(x) \end{pmatrix}.$$

Following the same argument as in Section 5.4, we see that, if $\bar{x}$ is a KKT point of the Feasibility QVI, then

$$0 = A^T g_+(\bar{x},\bar{x}) = (\bar{x} - u(\bar{x}))_+ - (\ell(\bar{x}) - \bar{x})_+.$$

In particular, it follows that $\bar{x}$ is feasible if the constraint functions ℓ and u satisfy $\ell(\bar{x}) \le u(\bar{x})$.

Theorem 5.7. *Consider a QVI where g is given by* (18). *Then, if $\bar{x}$ is a KKT point of the Feasibility QVI and $\ell(\bar{x}) \le u(\bar{x})$ holds, $\bar{x}$ is feasible.*

Proof. This follows from the calculations above. □

As in the previous cases, it follows that every limit point of Algorithm 3.1 is a KKT point of the QVI.

6 An Exact Penalty Method

We consider a modification of the augmented Lagrangian method (Algorithm 3.1) which, under suitable conditions, is an exact penalty method. This approach only works if we are performing a full penalization of the constraints. Hence, we will omit the function h throughout this section.

The basic approach is to remove the explicit multipliers in the augmented Lagrangian and replace them by a multiplier function which is dependent on x. More precisely, for a given x, we compute λ as a solution of the minimization problem

$$\min_{\lambda} \|F(x) + \nabla_y g(x,x)\lambda\|^2 + \|G(x)\lambda\|^2, \tag{19}$$

where $G(x) = \operatorname{diag}(g_1(x,x), \ldots, g_m(x,x))$. This is a linear least-squares problem, and the following lemma precisely states when it has a unique solution.

Lemma 6.1. *The multiplier problem* (19) *has a unique solution for every $x \in \mathbb{R}^n$ such that g satisfies LICQ in x. In this case, the solution vector $\lambda(x)$ is given by*

$$\lambda(x) = -M_0(x)^{-1}\nabla_y g(x,x)^T F(x), \tag{20}$$

where $M_0(x)$ is the positive definite matrix

$$M_0(x) = \nabla_y g(x,x)^T \nabla_y g(x,x) + G(x)^2.$$

As a logical consequence of the above lemma, we make the blanket assumption that g satisfies LICQ at every point $x \in \mathbb{R}^n$. This implies that $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ is a well-defined and continuously differentiable function. With this in mind, we consider the following basic algorithm for the realization of the exact penalty approach.

Algorithm 6.2. *(Exact penalty method for QVIs)*

(S.0) Choose $x^0 \in \mathbb{R}^n$, $\rho_0 > 0$, $\gamma > 1$, and set $k = 0$.

(S.1) If $(x^k, \lambda(x^k))$ is a KKT point of the QVI: STOP.

(S.2) Compute a zero x^{k+1} of the function $L_{\rho_k}(x, \lambda(x))$.

(S.3) Set $\rho_{k+1} = \gamma\rho_k$, set $k \leftarrow k+1$, and go to (S.1).

It turns out that this method is very similar to Algorithm 3.1. To see this, note that, if x^{k+1} is a zero of $L_{\rho_k}(x, \lambda(x))$, then it is also a zero of the function

$$L_{\rho_k}(x, u^k), \quad \text{with} \quad u^k = \lambda(x^{k+1}).$$

Furthermore, if (x^{k+1}) converges to $\bar{x}$ on some subsequence $K \subset \mathbb{N}$, then $(u^k)_K$ is bounded. This implies that Algorithm 6.2 inherits many convergence properties of Algorithm 3.1.

Theorem 6.3. *Assume that Algorithm 6.2 does not terminate finitely and $\bar{x}$ is a limit point of the sequence (x^k). Then $\bar{x}$ satisfies*

$$\nabla_y \|g_+(\bar{x}, \bar{x})\|^2 = 0. \tag{21}$$

Proof. The proof is identical to that of Theorem 4.1, with h omitted. □

It should be noted that (21) is nothing but the Feasibility QVI from (9), which takes on the above form in the case of full penalization.

We now turn to the central property of Algorithm 6.2, which is the exactness property. To this end, we first prove a technical lemma.

Lemma 6.4. *Let x be a zero of $L_\rho(x, \lambda(x))$ and let*

$$s = \min\left\{-g(x,x), \frac{\lambda(x)}{\rho}\right\}, \quad t = \max\left\{-g(x,x), \frac{\lambda(x)}{\rho}\right\}.$$

Then $Ms = 0$, where M is the matrix $M = \nabla_y g(x,x)^T \nabla_y g(x,x) - G(x)\operatorname{diag}(t)$. Furthermore, if $s = 0$, then $(x, \lambda(x))$ is a KKT point of the QVI.

Proof. From the definition of $\lambda(x)$, we obtain

$$\nabla_y g(x,x)^T L(x, \lambda(x)) = -G(x)^2 \lambda(x).$$

But a simple case-by-case analysis shows that

$$-g_i(x,x)\lambda_i(x) = \rho s_i t_i \quad \text{for every } i = 1, \dots, m.$$

Hence, it follows that $\nabla_y g(x,x)^T L(x, \lambda(x)) = \rho G(x)\operatorname{diag}(t)s$. Since

$$0 = L_\rho(x, \lambda(x)) = L(x, \lambda(x)) - \rho \nabla_y g(x,x)s, \tag{22}$$

we finally obtain

$$0 = \nabla_y g(x,x)^T L_\rho(x, \lambda(x)) = \rho\left[G(x)\operatorname{diag}(t) - \nabla_y g(x,x)^T \nabla_y g(x,x)\right]s.$$

This proves the first part. If $s = 0$, then $\min\{-g(x,x), \lambda(x)\} = 0$ and (22) implies that $(x, \lambda(x))$ is a KKT point of the QVI. □

The following is the central property of Algorithm 6.2. Due to Lemma 6.4, the proof is rather short.

Theorem 6.5. *Assume that the iterates (x^k) generated by Algorithm 6.2 remain bounded and that every solution of (21) is a feasible point. Then the algorithm terminates finitely and produces a KKT point of the QVI.*

Proof. We assume that the method does not terminate finitely, i.e. we obtain sequences (x^k) and (ρ_k) with $\rho_k \to \infty$. Since (x^k) is bounded, we can choose a subset $K \subset \mathbb{N}$ such that $x^{k+1} \to_K \bar{x}$. By Theorem 6.3 and our assumptions, it follows that $\bar{x}$ is feasible. Using the notation from Lemma 6.4 and the fact that x^{k+1} is a zero of $L_{\rho_k}(x, \lambda(x))$, we now obtain a sequence of matrices M^k with $M^k s^k = 0$, where

$$M^k = \nabla_y g(x^{k+1}, x^{k+1})^T \nabla_y g(x^{k+1}, x^{k+1}) - G(x^{k+1})\operatorname{diag}(t^k).$$

But we have $t^k = \max\left\{-g(x^{k+1}, x^{k+1}), \lambda(x^{k+1})/\rho_k\right\} \to_K -g(\bar{x}, \bar{x})$ and, hence, $M^k \to_K M_0(\bar{x})$ with M_0 from Lemma 6.1. Since M_0 is regular, it follows that M^k is regular for sufficiently large $k \in K$. This implies $s^k = 0$ and the result follows from Lemma 6.4. □

Note that the above theorem uses two central assumptions: first, we require the sequence (x^k) to remain bounded. This is a rather standard condition in the context of similar exact penalty methods [9, 11]. Secondly, we require that every solution of the Feasibility QVI (which, in this case, takes on the form (21)) is a feasible point. This is a similar condition to those discussed in the context of the inexact augmented Lagrangian method, cf. Section 5. Furthermore, it is essentially equivalent to Assumption B from [11].

We close this section by noting that, for the special case of generalized Nash equilibrium problems, there already exist exact penalty methods, see [14, 18, 21]. In the terminology of optimization problems, however, these exact penalty methods correspond to the nonsmooth exact ℓ_1 penalty function, whereas our penalty approach is motivated by the differentiable exact penalty function from [11]. Consequently, even in the context of generalized Nash equilibrium problems, our approach has better smoothness properties than existing (exact) penalty schemes which implies that the resulting subproblems are usually easier to solve.

7 Numerical Results

The purpose of this section is to give detailed results illustrating the practical performance of both the augmented Lagrangian method (Algorithm 3.1) and the exact penalty method (Algorithm 6.2). To this end, we implemented both methods in MATLAB® and used the QVILIB library of test problems [16]. The QVIs in this library follow a simple structure: for each problem, the constraint set $K(x)$ is given by

$$K(x) = \{y \in \mathbb{R}^n \mid g^I(y) \leq 0 \text{ and } g^P(y, x) \leq 0\},$$

where g^I and g^P describe the *independent* and *parametrized* constraints, respectively. This structure lends itself to both partial and full penalization (recall that, for the exact penalty method, we always perform a full penalization). The resulting subproblems then become either standard VIs or, in the latter case, nonlinear equations. For the solution of these problems, we decided to employ a semismooth Levenberg-Marquardt type method together with the well-known Fischer-Burmeister complementarity function [15, 19], which allows us to transform a VI into a nonlinear equation. The implementation details of this method are rather standard and, hence, we do not explicitly report them here.

After the above discussion, we are left with four methods:

- the semismooth Levenberg-Marquardt method, applied directly to the KKT system of the QVI by use of the Fischer-Burmeister function. This method will be denoted by **Semi**.
- the augmented Lagrangian method (Algorithm 3.1) using partial penalization and the formula $u^k = \min\{\lambda^k, u^{\max}\}$. This method will be denoted by **ALMP**.
- the augmented Lagrangian method like above, but with full penalization, denoted **ALMF**.
- the exact penalty method (Algorithm 6.2), denoted **Exact**.

We now describe some implementation details for our methods. The starting points x^0 are given by the test problems themselves, and the subproblems occuring within the penalty methods are solved with a precision of 10^{-8}. Furthermore, we use the overall stopping criterion

$$\left\| \begin{pmatrix} F(x) + \nabla_y g(x,x)\lambda + \nabla_y h(x,x)\mu \\ \min\{-g(x,x),\lambda\} \\ \min\{-h(x,x),\mu\} \end{pmatrix} \right\|_\infty \leq \varepsilon := 10^{-4},$$

which can, of course, be written more tersely for the methods which do not use h. Finally, we use the initial Lagrange multipliers $(\lambda^0, \mu^0) = 0$ and the iteration parameters

$$\rho_0 = 1, \quad u^{\max} = 10^{10}, \quad \gamma = 5 \quad \text{and} \quad \tau = 0.9,$$

which are chosen to favour robustness over efficiency. Note that some algorithms, such as the exact penalty method, only use a subset of the above parameters.

The results are presented as follows. Each row represents a problem from the QVILIB library. The name of the problem is given in the first column, followed by the dimensions n, m and p. The final four columns list the iteration numbers for each of our four methods, where `-` denotes a failure. In view of the results, some remarks are in order:

1. The augmented Lagrangian methods (**ALMP** and **ALMF**) are able to solve most problems, the only exceptions being `Box1B` and `MovSet1B`. A quick analysis shows that the failure for these problems is due to the inability of the Newton method to solve the subproblems at certain iterations. This possibly could have been avoided with a different choice of sub-algorithm. However, a detailed discussion of such methods is outside the scope of this paper.

2. The Newton method has 5 failures, which shows that our implementation (although fairly simple) is quite robust (in particular, more robust than the standard semismooth Newton method investigated in [15]), but not as robust as the augmented Lagrangian methods. Note that the raw iteration numbers of this method are very hard to compare to those of the other methods, since a single step of the Newton method merely consists of one linear equation.

3. The exact penalty method is able to solve most of the smaller problems extremely quickly, usually requiring only 1 or 2 iterations. However, it exhibits failures for some of the larger problems. A quick analysis shows that, in particular, the `Kun*` and `Scrim*` examples do not satisfy LICQ, which makes the exact penalty method entirely unsuited for these problems.

4. We also tested all four algorithms with an overall accuracy of $\varepsilon = 10^{-8}$. For most problems, this did not cause any difficulties. The failure numbers in this setting are given by 8 (**Semi**), 4 (**ALMP** and **ALMF**), and 3 (**Exact**).

5. For some problem classes, the algorithms exhibit completely different behaviour. For instance, the `RHS*` examples turn out to be extremely easy for the augmented Lagrangian methods and quite hard for the semismooth Newton method.

Name	n	m	p	**Semi**	**ALMP**	**ALMF**	**Exact**
BiLin1A	5	3	10	27	23	22	2
BiLin1B	5	3	10	18	49	49	1
Box1A	5	10	0	5	38	38	1
Box1B	5	10	0	-	-	-	1
Box2A	500	1000	0	49	13	13	1
Box2B	500	1000	0	11	16	16	1
Box3A	500	1000	0	105	12	12	1
Box3B	500	1000	0	391	38	38	2
KunR11	2500	2500	0	245	12	12	$*$
KunR12	4900	4900	0	450	11	11	$*$
KunR21	2500	2500	0	8	1	1	$*$
KunR22	4900	4900	0	9	1	1	$*$
KunR31	2500	2500	0	112	29	29	$*$
KunR32	4900	4900	0	371	35	35	$*$
MovSet1A	5	1	0	8	36	36	1
MovSet1B	5	1	0	-	-	-	2
MovSet2A	5	1	0	9	42	42	1
MovSet2B	5	1	0	-	44	44	1
MovSet3A1	1000	1	0	53	3	3	1
MovSet3A2	2000	1	0	71	3	3	1
MovSet3B1	1000	1	0	59	3	3	4
OutKZ31	62	62	62	10	16	19	-
OutKZ41	82	82	82	9	22	10	-
OutZ40	2	2	4	5	1	1	1
OutZ41	2	2	4	5	1	1	1
OutZ42	4	4	4	9	6	6	1
OutZ43	4	4	0	6	8	8	1
OutZ44	4	4	0	6	7	7	1
RHS1A1	200	199	0	-	1	1	1
RHS1B1	200	199	0	774	1	1	1
RHS2A1	200	199	0	-	1	1	1
RHS2B1	200	199	0	527	1	1	1
Scrim21	2400	2400	2400	495	28	30	$*$
Scrim22	4800	4800	4800	512	28	30	$*$

Table 1: Numerical results of the four algorithms from Section 7. Note: $*$ denotes a problem where LICQ is violated.

6. The two problems `Box1B` and `MovSet1B` appear to be very hard for the first three methods, which agrees with the numerical results from [29]. Interestingly, though, the exact penalty method is able to solve these problems very easily, requiring only 1 and 2 iterations, respectively.

8 Final Remarks

We have revisited the augmented Lagrangian method for quasi-variational inequalities and described an elegant theoretical framework which explains its behaviour. This framework includes the known convergence properties from [29] and improves upon some of them, for instance, by always allowing the multipliers which arise from the subproblems to be negative. Another new feature of our analysis is a simple modification of the ALM which possesses an interesting exactness property and shows that the concept of exact penalty methods can be extended from classical optimization problems and VIs to QVIs without much additional work.

The numerical testing we have done on both methods indicates that they work quite well in practice. In particular, they are quite robust and achieve good accuracy; it should be mentioned, though, that the exact penalty method requires a quite strong regularity property to work properly.

Furthermore, we remark there are still many aspects which might lead to substantial numerical improvements. Aside from the fine-tuning of iteration parameters, we could consider some of the various improvements which the classical exact penalty method (for optimization problems) has seen throughout the last decades, cf. [12, 34] among others. Some possible extensions for the augmented Lagrangian method include second-order multiplier iterations or approaches such as the exponential method of multipliers [5].

Finally, one result of this paper that deserves a special mention is the *Feasibility QVI* which we introduced in Section 4. This is a generalization of a corresponding concept for classical optimization problems [6, 8] and generalized Nash equilibrium problems [30]. In the context of our work, it is the central building block of the whole theoretical framework, since it offers a simple and practical explanation of how the augmented Lagrangian method (and its exact penalty counterpart) achieve feasibility. Our analysis has also shown that an application of this concept to special classes of QVIs (cf. Section 5) yields a variety of strong convergence results. Hence, we hope that this concept will find further applications in the study of quasi-variational inequalities.

References

[1] C. Baiocchi and A. Capelo. *Variational and quasivariational inequalities: applications to free boundary problems.* A Wiley-Interscience publication. Wiley, 1984.

[2] A. Bensoussan. Points de Nash dans le cas de fonctionnelles quadratiques et jeux différentiels linéaires à N personnes. *SIAM J. Control*, 12:460–499, 1974.

[3] A. Bensoussan, M. Goursat, and J.-L. Lions. Contrôle impulsionnel et inéquations quasi-variationnelles stationnaires. *C. R. Acad. Sci. Paris Sér. A-B*, 276:A1279–A1284, 1973.

[4] P. Beremlijski, J. Haslinger, M. Kočvara, and J. Outrata. Shape optimization in contact problems with Coulomb friction. *SIAM J. Optim.*, 13(2):561–587, 2002.

[5] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press, Inc., New York-London, 1982.

[6] E. G. Birgin and J. M. Martínez. *Practical augmented Lagrangian methods for constrained optimization*, volume 10 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.

[7] M. C. Bliemer and P. H. Bovy. Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B: Methodological*, 37(6):501 – 519, 2003.

[8] J. V. Burke and S.-P. Han. A robust sequential quadratic programming method. *Math. Programming*, 43(3, (Ser. A)):277–303, 1989.

[9] T. A. de André and P. J. S. Silva. Exact penalties for variational inequalities with applications to nonlinear complementarity problems. *Comput. Optim. Appl.*, 47(3):401–429, 2010.

[10] M. De Luca and A. Maugeri. Discontinuous quasi-variational inequalities and applications to equilibrium problems. In *Nonsmooth optimization: methods and applications (Erice, 1991)*, pages 70–75. Gordon and Breach, Montreux, 1992.

[11] G. Di Pillo and L. Grippo. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. *SIAM J. Control Optim.*, 23(1):72–84, 1985.

[12] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. *SIAM J. Control Optim.*, 27(6):1333–1360, 1989.

[13] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Ann. Oper. Res.*, 175:177–211, 2010.

[14] F. Facchinei and C. Kanzow. Penalty methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.*, 20(5):2228–2253, 2010.

[15] F. Facchinei, C. Kanzow, S. Karl, and S. Sagratella. The semismooth Newton method for the solution of quasi-variational inequalities. *Comput. Optim. Appl.*, 62(1):85–109, 2015.

[16] F. Facchinei, C. Kanzow, and S. Sagratella. QVILIB: a library of quasi-variational inequality test problems. *Pac. J. Optim.*, 9(2):225–250, 2013.

[17] F. Facchinei, C. Kanzow, and S. Sagratella. Solving quasi-variational inequalities via their kkt conditions. *Mathematical Programming*, 144(1):369–412, 2014.

[18] F. Facchinei and L. Lampariello. Partial penalization for the solution of generalized Nash equilibrium problems. *J. Global Optim.*, 50(1):39–57, 2011.

[19] A. Fischer. A special Newton-type optimization method. *Optimization*, 24(3-4):269–284, 1992.

[20] A. Fischer, M. Herrich, and K. Schönefeld. Generalized nash equilibrium problems - recent advances and challenges. *Pesquisa Operacional*, 34:521 – 558, 12 2014.

[21] M. Fukushima. Restricted generalized Nash equilibria and controlled penalty algorithm. *Comput. Manag. Sci.*, 8(3):201–218, 2011.

[22] P. Hammerstein and R. Selten. Game theory and evolutionary biology. In *Handbook of game theory with economic applications, Vol. II*, volume 11 of *Handbooks in Econom.*, pages 929–993. North-Holland, Amsterdam, 1994.

[23] P. T. Harker. Generalized nash games and quasi-variational inequalities. *European Journal of Operational Research*, 54(1):81 – 94, 1991.

[24] 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(3-4):365–383, 1984.

[25] M. Hintermüller and C. N. Rautenberg. A sequential minimization technique for elliptic quasi-variational inequalities with gradient constraints. *SIAM J. Optim.*, 22(4):1224–1257, 2012.

[26] M. Hintermüller and C. N. Rautenberg. Parabolic quasi-variational inequalities with gradient-type constraints. *SIAM J. Optim.*, 23(4):2090–2123, 2013.

[27] T. Ichiishi. *Game theory for economic analysis.* Economic Theory, Econometrics, and Mathematical Economics. Academic Press, Inc., New York, 1983.

[28] W. Jing-Yuan and Y. Smeers. Spatial oligopolistic electricity models with cournot generators and regulated transmission prices. *Operations Research*, 47(1):102–112, 1999.

[29] C. Kanzow. On the multiplier-penalty-approach for quasi-variational inequalities. *Mathematical Programming*, pages 1–31, 2016.

[30] C. Kanzow and D. Steck. Augmented lagrangian methods for the solution of generalized nash equilibrium problems. *Technical Report*, 2016.

[31] I. Kharroubi, J. Ma, H. Pham, and J. Zhang. Backward SDEs with constrained jumps and quasi-variational inequalities. *Ann. Probab.*, 38(2):794–840, 2010.

[32] A. S. Kravchuk and P. J. Neittaanmäki. *Variational and quasi-variational inequalities in mechanics*, volume 147 of *Solid Mechanics and its Applications*. Springer, Dordrecht, 2007.

[33] M. Kunze and J. F. Rodrigues. An elliptic quasi-variational inequality with gradient constraints and some of its applications. *Math. Methods Appl. Sci.*, 23(10):897–908, 2000.

[34] S. Lucidi. New results on a continuously differentiable exact penalty function. *SIAM J. Optim.*, 2(4):558–574, 1992.

[35] U. Mosco. Implicit variational problems and quasi variational inequalities. In *Nonlinear operators and the calculus of variations (Summer School, Univ. Libre Bruxelles, Brussels, 1975)*, pages 83–156. Lecture Notes in Math., Vol. 543. Springer, Berlin, 1976.

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

[37] J. V. Outrata and M. Kočvara. On a class of quasi-variational inequalities. *Optimization Methods and Software*, 5(4):275–295, 1995.

[38] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Comput. Manag. Sci.*, 2(1):21–56, 2005.

[39] L. Scrimali. Quasi-variational inequalities in transportation networks. *Math. Models Methods Appl. Sci.*, 14(10):1541–1560, 2004.