ON A SMOOTH DUAL GAP FUNCTION FOR A CLASS OF QUASI-VARIATIONAL INEQUALITIES[1]

Nadja Harms, Tim Hoheisel, and Christian Kanzow

Preprint 318 October 2013

University of Würzburg
Institute of Mathematics
Emil-Fischer-Str. 30
97074 Würzburg
Germany

e-mail: nadja.harms@mathematik.uni-wuerzburg.de
hoheisel@mathematik.uni-wuerzburg.de
kanzow@mathematik.uni-wuerzburg.de

October 2, 2013

[1]This research was partially supported by the DFG (Deutsche Forschungsgemeinschaft) under grant KA 1296/18-1 as well as by a grant from the international doctorate program "Identification, Optimization, and Control with Applications in Modern Technologies" within the Elite-Network of Bavaria.

Abstract. A well-known technique for the solution of quasi-variational inequalities (QVIs) consists in the reformulation of QVIs as a constrained or unconstrained optimization problem by means of so-called gap functions. In contrast to standard variational inequalities, however, these gap functions turn out to be nonsmooth in general. Here it is shown that one can obtain an unconstrained optimization reformulation of a class of QVIs by using a continuously differentiable dual gap function. This extends an idea from Dietrich (*Journal of Mathematical Analysis and Applications 235, 1999, pp. 380–393*). Some numerical results illustrate the practical behavior of this dual gap function approach.

Key Words: Quasi-variational inequality, set-valued mapping, regularized gap function, DC optimization, conjugate function, dual gap function, PC^1 function, nonconvex duality.

1 Introduction

Given a function $F : \mathbb{R}^n \to \mathbb{R}^n$ and a set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ such that $S(x)$ is closed and convex (possibly empty) for any $x \in \mathbb{R}^n$, the finite-dimensional *quasi-variational inequality problem (QVI)* consists in finding a solution $x \in S(x)$ such that

$$F(x)^T(z - x) \geq 0 \qquad \forall z \in S(x). \tag{1}$$

This QVI was originally introduced in a series of papers by Bensoussan et al., see [3, 4, 5], where the authors consider an infinite-dimensional QVI arising from an application in impulse control problems. Several other applications from free boundary value problems can be found in the monograph [2]. Further applications are incorporated in the recent test problem collection [10] which also contains the corresponding references.

Although there exist plenty of papers dealing with several theoretical issues like the existence and uniqueness of solutions, numerical methods for the solution of QVIs are only starting to evolve. Most of the algorithmic papers deal with projection methods or fixed point iterations, see, e.g., [7, 23, 24, 25, 34, 35]. Essentially, the convergence theory for these methods considers only the case where the feasible set is given by $S(x) = c(x)+K$ for a suitable function $c : \mathbb{R}^n \to \mathbb{R}^n$ and a fixed closed and convex set $K \subseteq \mathbb{R}^n$. This class of problems is sometimes called the "moving set case". Other globally convergent methods, where also more general QVIs are treated, include the penalty-multiplier approach by Pang and Fukushima [26] and the potential-reduction interior-point method from Facchinei et al. [11]. Locally convergent Newton-type methods are presented by Outrata et al., see [27, 28, 29].

Here we follow the gap function idea which reformulates the QVI as a constrained or unconstrained optimization problem, see [1, 9, 14, 15, 16, 38] for more details. However, these gap functions are typically nonsmooth in the QVI-setting, except for the case where the feasible set is of the moving-set-type [9] or a suitable generalization of it [16]. In particular, this statement also holds for the regularized gap function that was originally introduced by Fukushima [13] in the context of standard variational inequalities. However, the paper by Dietrich [8] observed that this regularized gap function may be viewed as a difference of two convex functions and can therefore be used, by means of a suitable duality theory, to obtain a dual gap function which gives a *smooth* reformulation for a class of QVIs that are different from the moving set case. The aim of this paper is therefore to elaborate further on this approach. In particular, we get rid of the (implicit) assumption from [8] that the set $S(x)$ is always nonempty since, in many practical instances, this set is indeed empty for many x. Furthermore, we verify some stronger smoothness properties and present some numerical results obtained by the dual gap function approach.

The paper is organized as follows: Section 2 restates some definitions and standard results from convex and variational analysis. The dual gap function and its basic properties are then derived in Section 3. The piecewiese smoothness of this dual gap function is shown in Section 4 under a suitable assumption. Some promising numerical results are given in Section 5. We close with some final remarks in Section 6.

The notation used in this paper fairly standard. The symbol $\|\cdot\|$ always denotes the Euclidean norm. For a set $\Omega \subset \mathbb{R}^n$ and a matrix $D \in \mathbb{R}^{m\times n}$, we put $D\cdot\Omega := \{Dw \mid w \in \Omega\}$. Furthermore, if Ω is nonempty, closed, and convex, we denote by $P_\Omega(x)$ the Euclidean projection of a vector $x \in \mathbb{R}^n$ onto the set Ω. Also, we put $\mathbb{R}_> := \{x \in \mathbb{R} \mid x > 0\}$.

2 Preliminaries

In this section we review certain concepts from variational and convex analysis employed in the subsequent analysis. The notation and terminology is, in large parts, based on [33].

We first restate some definitions for set-valued mappings, see, e.g., [33, Chapter 5] for more details.

Definition 2.1 *Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. Then Φ is called*

(a) outer semicontinuous (osc) *at $\bar{x} \in \mathbb{R}^n$ if for all sequences $\{x^k\} \subset \mathbb{R}^n$ with $x^k \to \bar{x}$ and all sequences $z^k \to \bar{z}$ with $z^k \in \Phi(x^k)$ for all $k \in \mathbb{N}$ sufficiently large we have $\bar{z} \in \Phi(\bar{x})$;*

(b) outer semicontinuous (osc) *on $\mathbb{R}^n$ if it is osc at every $x \in \mathbb{R}^n$;*

(c) graph-convex *if its graph*

$$\operatorname{gph}\Phi = \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^m \mid z \in \Phi(x)\}$$

is a convex set.

The following properties of an osc and graph-convex set-valued mapping will be used in our subsequent analysis.

Lemma 2.2 *Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be an osc and graph-convex set-valued mapping. Then the following statements hold:*

(a) The sets $\Phi(x)$ are closed and convex (possibly empty).

(b) For all $x_1, x_2 \in \mathbb{R}^n$ with $\Phi(x_i) \neq \emptyset$ for $i = 1, 2$, and all $t \in [0, 1]$, we have

$$t\Phi(x_1) + (1-t)\Phi(x_2) \subset \Phi\big(tx_1 + (1-t)x_2\big),$$

in particular, the set on the right-hand side is nonempty.

(c) The set $\operatorname{gph}\Phi$ is closed and convex.

All statements are well known and easily verified; regarding assertion (b), we refer the reader, e.g., to [33, p. 155].

We next introduce some important concepts for extended real-valued functions, more precisely, for functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Handy tools for the analysis of such a function are its *epigraph*

$$\operatorname{epi} f := \{(x, \gamma) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \gamma\}$$

and its *domain*

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

Note that we call f *proper* if $\operatorname{dom} f \neq \emptyset$.

Definition 2.3 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper.*

(a) f is called lower semicontinuous *(lsc) if $\operatorname{epi} f$ is closed.*

(b) f is called convex *if $\operatorname{epi} f$ is convex.*

(c) f is called strongly convex *with modulus $c > 0$ if $f - \frac{c}{2}\|\cdot\|^2$ is convex.*

(d) If f is convex and $\bar{x} \in \mathbb{R}^n$ then the (possibly empty) set

$$\partial f(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid f(\bar{x}) + d^T(x - \bar{x}) \leq f(x) \quad \forall x \in \mathbb{R}^n \right\}$$

is called the subdifferential *of f at $\bar{x}$.*

(e) The conjugate *of f is the function $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left[x^T y - f(x) \right].$$

Note that for a proper and convex function f, the subdifferential $\partial f(\bar{x})$ is nonempty if x lies in the *(relative) interior* of $\operatorname{dom} f$, see [19, Theorem E 1.4.2].

Given a set $X \subset \mathbb{R}^n$, a very prominent extended real-valued function is the *indicator function* $\delta_X : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\delta_X(x) := \begin{cases} 0 & \text{if } x \in X, \\ +\infty & \text{if } x \notin X. \end{cases}$$

It is easily verified that δ_X is lsc if and only if X is closed, and convex if and only if X is convex.

The following result summarizes some well-known properties of the conjugate function.

Lemma 2.4 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Then the following statements hold:*

(a) The conjugate f^* of f is convex and lsc.

(b) The bi-conjugate function $f^{**} := (f^*)^*$ is convex and lsc.

(c) The inequality $f^{**}(x) \leq f(x)$ holds for all $x \in \mathbb{R}^n$.

(d) The equality $f^{**}(x) = f(x)$ holds for all $x \in \mathbb{R}^n$ if and only if f is a (convex and) lsc function.

(e) The Fenchel inequality $f(x) + f^*(y) \geq x^T y$ holds for all $x, y \in \mathbb{R}^n$.

(f) The equality $f(\bar{x}) + f^*(\bar{y}) = \bar{x}^T \bar{y}$ holds if and only if $\bar{y} \in \partial f(\bar{x})$.

All statements can entirely be found in [19, Chapter E], cf. [19, Thm. E 1.1.2, Thm. E 1.3.5, Cor. E 1.3.6, Eq. E 1.1.3, Thm. E 1.4.1].

Another useful observation on the conjugate function is restated in the following result which is a direct application of [33, Prop. 12.60].

Lemma 2.5 *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lsc convex function. Then f is strongly convex with modulus $c > 0$ if and only if f^* is differentiable with ∇f^* locally Lipschitz with modulus $\frac{1}{c}$.*

3 The Smooth Dual Gap Function

Let $\alpha > 0$ be a given parameter. Then the *regularized gap function* for QVIs is defined by

$$g_\alpha(x) := -\inf_{z \in S(x)} \left[F(x)^T (z - x) + \frac{\alpha}{2} \|z - x\|^2 \right] \tag{2}$$

and was introduced independently by Dietrich [9] and Taji [38], see also [16] for further details and [13] for its origin in the context of standard variational inequalities. Let

$$X := \{x \in \mathbb{R}^n \mid x \in S(x)\} \tag{3}$$

be the *fixed point set* of the set-valued mapping S which plays an important role in the context of QVIs and is often called the *feasible set* of the underlying QVI. Then the following basic properties of the regularized gap function are observed in [9, 38].

Lemma 3.1 *The following statements hold for the regularized gap function:*

(a) $g_\alpha(x) \geq 0$ for all $x \in X$.

(b) $g_\alpha(\bar{x}) = 0$ for some $\bar{x} \in X \iff \bar{x}$ is a solution of the QVI.

Hence the QVI (1) is equivalent to finding a solution of the constrained minimization problem

$$\min \ g_\alpha(x) \quad \text{s.t.} \quad x \in X \tag{4}$$

with zero optimal value. Note, however, that the objective function $g_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ is, in general, nonconvex and nondifferentiable, and takes the value $-\infty$ exactly for $x \notin M$, where

$$M := \operatorname{dom} S := \{x \in \mathbb{R}^n \mid S(x) \neq \emptyset\} \tag{5}$$

denotes the domain of the set-valued map S. Therefore, we can rewrite the constrained optimization problem (4) as the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \big[g_\alpha(x) + \delta_X(x)\big] \tag{6}$$

with convention $\eta + \infty = +\infty$ for all $\eta \in \mathbb{R} \cup \{\pm\infty\}$. Note that this convention makes sense in our case since the objective function from (6) should take the function value $+\infty$ outside of X, in particular, we would like to have $g_\alpha(x) + \delta_X(x) = +\infty$ also for all $x \notin X \cup M = M$.

Our next goal is to rewrite the objective function of (6) as a difference of two convex functions. To this end we have to make some assumptions on the class of QVIs that we are going to deal with.

Assumption 3.2

- *(a) The* feasible set X *of the QVI* (1) *defined in* (3) *is nonempty.*
- *(b) The function* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is given by* $F(x) = Ax + b$ *with* $A \in \mathbb{R}^{n \times n}$ *and* $b \in \mathbb{R}^n$.
- *(c) The set-valued mapping* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is graph-convex and osc on* $\mathbb{R}^n$.

Assumption 3.2 (a) does not limit the application of our theory since otherwise the QVI would not have a solution. Assumptions 3.2 (b) and (c), on the other hand, are more restrictive in the sense that we consider only (affine-)linear QVIs with suitable set-valued mappings S.

There are a couple of immediate consequences of Assumption 3.2 summarized in the following result.

Lemma 3.3 *Suppose that Assumption 3.2 holds. Then*

- *(a) The set* X *from* (3) *is nonempty, closed, and convex.*
- *(b) The set* M *from* (5) *is nonempty and convex.*

Proof. (a) The set X is nonempty by Assumption 3.2 (a). In order to show that X is also closed, let $\{x^k\} \subset X$ be an arbitrary sequence with $x^k \to \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$. Then $x^k \in S(x^k)$ for all $k \in \mathbb{N}$. Since the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc by Assumption 3.2 (c), it follows that $\bar{x} \in S(\bar{x})$. Hence $\bar{x} \in X$ so that X is a closed set.

We next show that X is also convex. To this end, let $x_1, x_2 \in X$ and $t \in [0,1]$ be arbitrarily given. Then $x_1 \in S(x_1)$ and $x_2 \in S(x_2)$. Using the assumed graph-convexity of S together with Lemma 2.2 (b), it follows that $tx_1 + (1-t)x_2 \in S\big(tx_1 + (1-t)x_2\big)$. This means that $tx_1 + (1-t)x_2 \in X$, i.e., X is a convex set.

(b) By Assumption 3.2 (a), there exists an element $x \in X$ which means that $x \in S(x)$, hence $x \in M$, so that M is nonempty.

Finally, we come to the convexity of M. Let $x_1, x_2 \in M$ and $t \in [0,1]$ be given. Then $S(x_1) \neq \emptyset$ and $S(x_2) \neq \emptyset$, hence there exist elements $z_1 \in S(x_1)$ and $z_2 \in S(x_2)$. Using Assumption 3.2 (c) together with Lemma 2.2 (b), this implies $tz_1 + (1-t)z_2 \in S\big(tx_1 + (1-t)x_2\big)$.

Consequently, the set on the right-hand side is nonempty, i.e., we have $tx_1 + (1-t)x_2 \in M$. $\square$

It is worth mentioning that even for an osc and graph-convex set-valued mapping S, its domain is not necessarily closed, as illustrated by the subsequent example.

Example 3.4 Let $S : \mathbb{R} \rightrightarrows \mathbb{R}$ be given by

$$S(x) := \begin{cases} \{y \in \mathbb{R} \mid y \geq \frac{1}{x}\} & \text{if } x > 0, \\ \emptyset & \text{if } x \leq 0. \end{cases}$$

Obviously, S is graph-convex. Also, S is osc, since, if $x_k \downarrow 0$, a sequence $\{z_k\}$ with $z_k \in S(x_k)$ is divergent, and all other cases are unproblematic. On the other hand, $M = \operatorname{dom} S = \mathbb{R}_>$ is not closed.

We next follow an observation by Dietrich [8] and reformulate the unconstrained objective function from problem (6) explicitly as a difference of two convex functions, i.e., we obtain a DC minimization problem, see [18] for a survey of DC programming. Having this DC formulation, it is pretty straightforward to obtain a reformulation as a difference of two strongly convex functions. Then we may invoke the duality theory by Toland [39] and Singer [36] in order to derive a smooth dual formulation of the original QVI.

The basic step to get a DC formulation is the following rearrangement of the regularized gap function:

$$\begin{aligned} g_\alpha(x) &= -\inf_{z \in S(x)} \left[-\frac{1}{2\alpha}\|F(x)\|^2 + \frac{\alpha}{2}\Big(\|z - x\|^2 + \frac{2}{\alpha}F(x)^T(z - x) + \frac{1}{\alpha^2}\|F(x)\|^2 \Big) \right] \\ &= \frac{1}{2\alpha}\|F(x)\|^2 - \frac{\alpha}{2} \inf_{z \in S(x)} \Big\| z - \big(x - \frac{1}{\alpha}F(x)\big) \Big\|^2 \qquad (7) \\ &= \frac{1}{2\alpha}\|F(x)\|^2 - \Phi_\alpha(x) \qquad (8) \end{aligned}$$

with the function $\Phi_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$\begin{aligned} \Phi_\alpha(x) &:= \frac{\alpha}{2} \inf_{z \in S(x)} \Big\| z - \big(x - \frac{1}{\alpha}F(x)\big) \Big\|^2 \qquad (9) \\ &= \begin{cases} \frac{\alpha}{2}\Big\| P_{S(x)}\big(x - \frac{1}{\alpha}F(x)\big) - \big(x - \frac{1}{\alpha}F(x)\big) \Big\|^2 & \text{if } x \in M, \\ +\infty & \text{if } x \notin M, \end{cases} \qquad (10) \end{aligned}$$

where we recall that, in view of Assumption 3.2 (c) and Lemma 2.2 (a), the set $S(x)$ is nonempty, closed, and convex, hence the projection $P_{S(x)}(y)$ of the point y onto this set is well-defined for all $x \in M$ with M being the set from (5).

Our next goal is to prove that Φ_α is lsc and convex. For these purposes, the following auxiliary result is pivotal.

Lemma 3.5 *Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be graph-convex and osc on $\mathbb{R}^n$. Then the function*

$$\Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, \quad \Psi(z, x) := \delta_{S(x)}(z)$$

is lsc and convex in (z, x).

Proof. First, we show that Ψ is lsc: To this end, let $\{(z^k, x^k, \gamma_k)\} \subset \operatorname{epi} \Psi$ such that $(z^k, x^k, \gamma_k) \to (\bar{z}, \bar{x}, \bar{\gamma})$. In particular, it holds that $\gamma_k \geq 0$, hence $\bar{\gamma} \geq 0$. On the other hand, we have $\delta_{S(x^k)}(z^k) \leq \gamma_k < \infty$, hence it necessarily follows from the definition of the indicator function that $\delta_{S(x^k)}(z^k) = 0$, i.e., $z^k \in S(x^k)$ holds for all $k \in \mathbb{N}$. Since S is osc, we therefore have $\bar{z} \in S(\bar{x})$ and thus, $\delta_{S(\bar{x})}(\bar{z}) = 0 \leq \bar{\gamma}$, hence $(\bar{z}, \bar{x}, \bar{\gamma}) \in \operatorname{epi} \Psi$. It follows that $\operatorname{epi} \Psi$ is closed, i.e., Ψ is lsc.

It remains to prove that Ψ is convex: For these purposes, let $(z, x, \gamma), (z', x', \gamma') \in \operatorname{epi} \Psi$ and $t \in (0, 1)$. Similar to the first part of the proof, it then follows that $z \in S(x)$ and $z' \in S(x')$. Consequently, we have $tz \in tS(x)$ and $(1-t)z' \in (1-t)S(x')$ and hence, due to the graph-convexity of S, we get $tz + (1-t)z' \in S(tx + (1-t)x')$, cf. Lemma 2.2 (b). Hence $\Psi(tz + (1-t)z', tx + (1-t)x') = 0 \leq t\gamma + (1-t)\gamma'$, and thus, $\operatorname{epi} \Psi$ is convex, i.e., Ψ is convex. □

Lemma 3.5 enables us to verify that the mapping Φ_α from (9) is lsc and convex.

Lemma 3.6 *Let Assumption 3.2 hold. Then the function Φ_α is lsc and convex on $\mathbb{R}^n$.*

Proof. In view of (9), we may rewrite Φ_α as

$$\Phi_\alpha(x) = \inf_{z \in \mathbb{R}^n} f(z, x),$$

where $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$f(z, x) := \frac{\alpha}{2} \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 + \delta_{S(x)}(z).$$

The first summand of f is convex as it is the composition of the convex map $\frac{\alpha}{2}\| \cdot \|^2$ and an affine function, see, e.g., [33, Ex. 2.20]. Moreover, the first summand is, in particular, continuous. The second summand is lsc and convex due to Lemma 3.5, hence f is lsc and convex (and proper, since $M \neq \emptyset$). Moreover, it holds that

$$\operatorname*{argmin}_z f(z, x) = \left\{ P_{S(x)} \left(x - \frac{1}{\alpha} F(x) \right) \right\} \quad \forall x \in M$$

is single-valued. Since $M \neq \emptyset$, the assertions therefore follow from [33, Cor. 3.32]. □

Note that Lemma 3.6 exploits the definition (9) of the mapping Φ_α in order to verify that it is both lsc and convex. Alternatively, one might try to use the representation (10) to rewrite Φ_α in the form

$$\Phi_\alpha(x) = \frac{\alpha}{2}\big\|P_{S(x)}\big(x - \frac{1}{\alpha}F(x)\big) - \big(x - \frac{1}{\alpha}F(x)\big)\big\|^2 + \delta_M(x).$$

This formulation can indeed be used to show convexity of Φ_α, but the verification of the lower semicontinuity is more difficult, especially since M is not necessarily closed, hence this formulation is, in general, not the sum of two lsc functions.

Moreover, we would like to point out that the proof of Lemma 3.6 exploits, for the first time, the assumption that $F(x) = Ax + b$ is an affine mapping, since it uses the fact that the composition of an outer convex function with an inner linear function remains convex. Similar situations will also arise in the subsequent analysis, and it is clear that there exist more general classes of functions F which have this property, but in order to avoid any technical conditions and to concentrate on the main ideas of our approach, Assumption 3.2 (b) takes F as a linear function.

In view of Lemma 3.6 and Assumption 3.2 (b), the representation (8) gives an explicit formulation of the regularized gap function as a DC optimization problem. In order to obtain better smoothness properties in a corresponding dual formulation, we add and substract a simple strongly convex quadratic term. This gives us the following DC decomposition of the unconstrained objective function from (6):

$$g_\alpha(x) + \delta_X(x) = f_\alpha(x) - h_\alpha(x)$$

with the two functions $f_\alpha, h_\alpha : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_\alpha(x) := \frac{\alpha}{2}\|x\|^2 + \frac{1}{2\alpha}\|F(x)\|^2 + \delta_X(x), \qquad h_\alpha(x) := \frac{\alpha}{2}\|x\|^2 + \Phi_\alpha(x). \tag{11}$$

We summarize the previous discussion in the following result.

Lemma 3.7 *Let Assumption 3.2 hold, and let f_α, h_α be defined as in* (11)*. Then:*

(a) The function f_α is lsc and convex on $\mathbb{R}^n$ as well as strongly convex on its domain $\operatorname{dom} f_\alpha = X$.

(b) The function h_α is lsc and convex on $\mathbb{R}^n$ as well as strongly convex on its domain $\operatorname{dom} h_\alpha = M$.

(c) $\bar{x}$ is a solution of the QVI if and only if it is a solution of the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} \,[f_\alpha(x) - h_\alpha(x)]$$

with optimal function value equal to zero.

We next want to apply the duality theory by Toland and Singer. This theory involves the conjugates of the two functions f_α and h_α. We therefore give explicit expressions for these two conjugate functions in the next two results.

Lemma 3.8 *Let Assumption 3.2 hold. Define*

$$Q_\alpha := \alpha\big(I + \frac{1}{\alpha^2}A^TA\big), \quad q_\alpha := \frac{1}{\alpha}A^Tb, \quad c_\alpha := \frac{1}{2\alpha}b^Tb, \quad \|x\|_{Q_\alpha} := \sqrt{x^TQ_\alpha x}. \tag{12}$$

Then the following statements hold for the conjugate f_α^* of f_α:

(a) f_α^* is given by

$$f_\alpha^*(y) = \frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha)\right\|_{Q_\alpha}^2 - \frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha) - x_\alpha^{f^*}(y)\right\|_{Q_\alpha}^2 - c_\alpha \tag{13}$$

where $x_\alpha^{f^*}(y)$ denotes the unique solution of the minimization problem

$$\min \frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha) - x\right\|_{Q_\alpha}^2 \quad s.t. \quad x \in X, \tag{14}$$

i.e., $x_\alpha^{f^*}(y)$ is the projection of the vector $Q_\alpha^{-1}(y-q_\alpha)$ onto the set X with respect to the Q_α-norm.

(b) f_α^* has the domain $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$.

(c) f_α^* is differentiable with locally Lipschitz gradient given by $\nabla f_\alpha^*(y) = x_\alpha^{f^*}(y)$.

Proof. Using Definition 2.3 (e) and the notation from (12), we obtain

$$\begin{aligned}
f_\alpha^*(y) &= \sup_{x\in\mathbb{R}^n}\Big[x^Ty - \frac{\alpha}{2}\|x\|^2 - \frac{1}{2\alpha}\|F(x)\|^2 - \delta_X(x)\Big] \\
&= \sup_{x\in\mathbb{R}^n}\Big[x^Ty - \frac{\alpha}{2}x^T\big(I + \frac{1}{\alpha^2}A^TA\big)x - \frac{1}{\alpha}b^TAx - \frac{1}{2\alpha}b^Tb - \delta_X(x)\Big] \\
&= \sup_{x\in\mathbb{R}^n}\Big[x^T(y-q_\alpha) - \frac{1}{2}\|x\|_{Q_\alpha}^2 - c_\alpha - \delta_X(x)\Big] \\
&= \sup_{x\in X}\Big[\frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha)\right\|_{Q_\alpha}^2 - c_\alpha - \frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha) - x\right\|_{Q_\alpha}^2\Big] \\
&= \frac{1}{2}\left\|Q_\alpha^{-1}(y-q_\alpha)\right\|_{Q_\alpha}^2 - c_\alpha - \frac{1}{2}\inf_{x\in X}\left\|Q_\alpha^{-1}(y-q_\alpha) - x\right\|_{Q_\alpha}^2.
\end{aligned} \tag{15}$$

Since the set X is nonempty, closed, and convex by Lemma 3.3 (a), and taking into account that the matrix Q_α is positive definite, the minimization problem (14) has a unique solution $x_\alpha^{f^*}(y)$ for all $y \in \mathbb{R}^n$. By definition, this solution is simply the projection of the vector $Q_\alpha^{-1}(y-q_\alpha)$ onto the set X with respect to the Q_α-norm and therefore known to be well-defined for all $y \in \mathbb{R}^n$, so that $\operatorname{dom} f_\alpha^* = \mathbb{R}^n$. This proves statements (a) and (b).

Part (c) can be derived as follows: Using the continuity of the projection operator, it follows that the mapping $y \mapsto x_\alpha^{f^*}(y)$ is continuous. Therefore, application of Danskin's Theorem (see, e.g., [6]) gives that f_α^* is continuously differentiable and directly yields $\nabla f_\alpha^*(y) = x_\alpha^{f^*}(y)$, cf. also the subsequent proof where a similar statement is carried out in some more detail. The fact that ∇f_α^* is even locally Lipschitz follows directly from Lemma 2.5. □

The following result computes the conjugate function of h_α and states some additional properties in the same spirit as in the previous result for the function f_α.

Lemma 3.9 *Let Assumption 3.2 hold. Then the following statements hold for the conjugate h_α^* of h_α:*

(a) $h_\alpha^*(y)$ is given by

$$h_\alpha^*(y) = \frac{1}{2\alpha}\|y\|^2 - \frac{\alpha}{2}\big\|x_\alpha^{h^*}(y) - \frac{1}{\alpha}y\big\|^2 - \frac{\alpha}{2}\big\|z_\alpha^{h^*}(y) - \big(x_\alpha^{h^*}(y) - \frac{1}{\alpha}F(x_\alpha^{h^*}(y))\big)\big\|^2 \quad (16)$$

where $\big(x_\alpha^{h^*}, z_\alpha^{h^*}\big)(y)$ is the unique solution of the minimization problem

$$\min \Big[\big\|x - \frac{1}{\alpha}y\big\|^2 + \big\|z - \big(x - \frac{1}{\alpha}F(x)\big)\big\|^2\Big] \quad s.t. \quad (x,z) \in \operatorname{gph} S.$$

(b) $h_\alpha^*(y)$ has the domain $\operatorname{dom} h_\alpha^* = \mathbb{R}^n$.

(c) $h_\alpha^*(y)$ is differentiable with locally Lipschitz gradient given by $\nabla h_\alpha^*(y) = x_\alpha^{h^*}(y)$.

Proof. Using Definition 2.3 (e), we have

$$\begin{aligned} h_\alpha^*(y) &= \sup_{x\in\mathbb{R}^n} \Big[x^T y - \frac{\alpha}{2}\|x\|^2 - \Phi_\alpha(x)\Big] \\ &= \sup_{x\in\mathbb{R}^n} \Big[\frac{1}{2\alpha}\|y\|^2 - \frac{\alpha}{2}\big(\|x\|^2 - \frac{2}{\alpha}x^T y + \frac{1}{\alpha^2}\|y\|^2\big) - \Phi_\alpha(x)\Big] \\ &= \frac{1}{2\alpha}\|y\|^2 - \inf_{x\in\mathbb{R}^n} \frac{\alpha}{2}\Big[\big\|x - \frac{1}{\alpha}y\big\|^2 + \inf_{z\in S(x)} \big\|z - \big(x - \frac{1}{\alpha}F(x)\big)\big\|^2\Big] \qquad (17) \\ &= \frac{1}{2\alpha}\|y\|^2 - \inf_{(x,z)\in\operatorname{gph} S} \frac{\alpha}{2}\Big[\big\|x - \frac{1}{\alpha}y\big\|^2 + \big\|z - \big(x - \frac{1}{\alpha}F(x)\big)\big\|^2\Big]. \qquad (18) \end{aligned}$$

Recall that, by Assumption 3.2 (a) and (c) and Lemma 2.2 (c), the set $\operatorname{gph} S$ is nonempty, closed, and convex. Furthermore, the mapping

$$\varphi_\alpha(x,y,z) := \big\|x - \frac{1}{\alpha}y\big\|^2 + \big\|z - \big(x - \frac{1}{\alpha}F(x)\big)\big\|^2$$

is strongly convex in (x,z) (uniformly in y) since

$$\nabla^2_{(x,z)(x,z)}\varphi(x,y,z) = 2\begin{pmatrix} 2I + \frac{1}{\alpha^2}A^TA - \frac{1}{\alpha}(A^T + A) & -I + \frac{1}{\alpha}A^T \\ -I + \frac{1}{\alpha}A & I \end{pmatrix} =: B_\alpha$$

is positive definite in (x, z) (uniformly in y) because we have

$$\begin{pmatrix} v^T & w^T \end{pmatrix} B_\alpha \begin{pmatrix} v \\ w \end{pmatrix} = 2\Big[\|v\|^2 + \big\|w - v + \frac{1}{\alpha} Av\big\|^2\Big] \geq 0 \quad \text{and}$$
$$\begin{pmatrix} v^T & w^T \end{pmatrix} B_\alpha \begin{pmatrix} v \\ w \end{pmatrix} = 0 \text{ if and only if } (v, w) = 0.$$

Hence, the infimum in (18) is uniquely attained for all $y \in \mathbb{R}^n$. We denote this unique solution by $\big(x_\alpha^{h^*}, z_\alpha^{h^*}\big)(y)$ and obtain (16) and $\operatorname{dom} h_\alpha^* = \mathbb{R}^n$. This proves statements (a) and (b).

Furthermore, the continuous differentiability of the conjugate convex function h_α^* follows from Lemma 2.5. Alternatively, we may invoke [20, Corollaries 8.1 and 9.1] to see that the mapping $y \mapsto \big(x_\alpha^{h^*}, z_\alpha^{h^*}\big)(y)$ is continuous, which together with Danskin's Theorem can be used to see that h_α^* is indeed continuously differentiable, with gradient given by

$$\nabla h_\alpha^*(y) = \frac{1}{\alpha} y - \frac{\alpha}{2} \nabla_y \varphi(x, y, z)\big|_{(x,z) = (x_\alpha^{h^*}, z_\alpha^{h^*})(y)} = \frac{1}{\alpha} y + x_\alpha^{h^*}(y) - \frac{1}{\alpha} y = x_\alpha^{h^*}(y).$$

The fact that ∇h_α^* is even locally Lipschitz is due to Lemma 2.5. This completes the proof. □

In order to illustrate the two previous and the subsequent results, we consider a simple example.

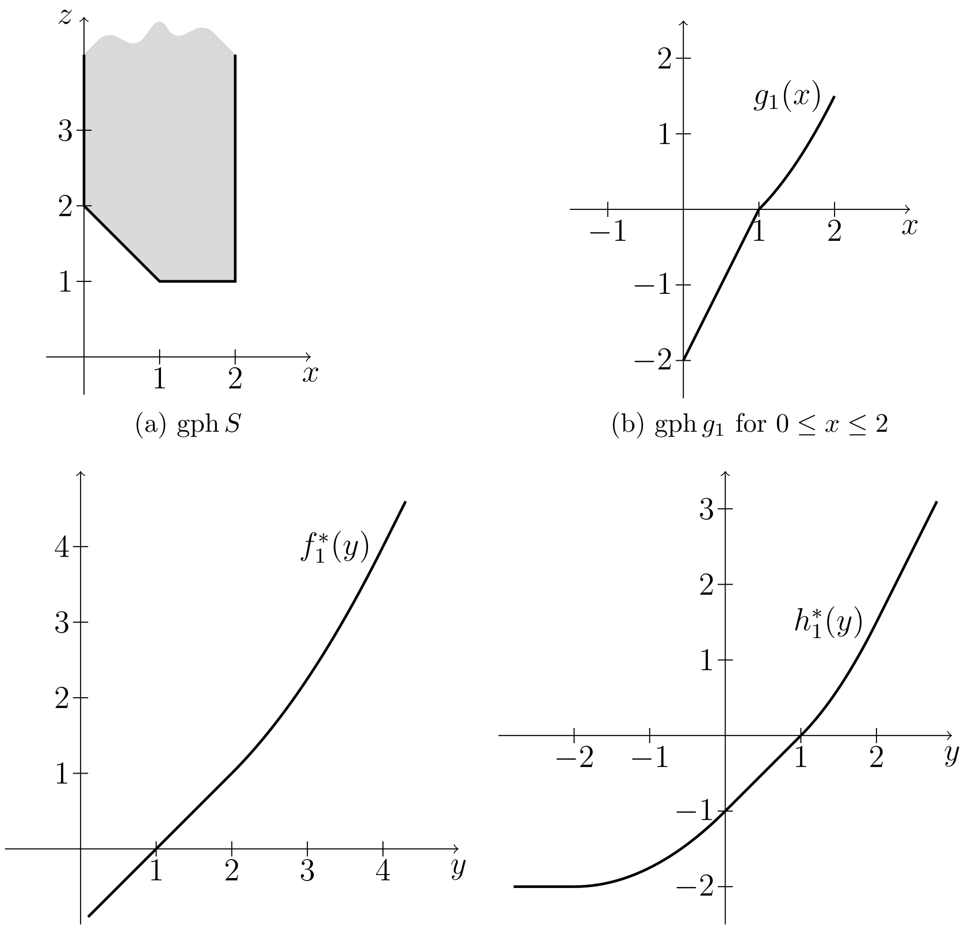
Example 3.10 Consider the QVI with $n = 1$, $F(x) = x$, and

$$S(x) = \begin{cases} [-x + 2, \infty) & \text{if } x \in [0, 1], \\ [1, \infty) & \text{if } x \in (1, 2], \\ \emptyset & \text{if } x \notin [0, 2]. \end{cases}$$

Note that $M = [0, 2]$ is the domain of S in this example, that $X = [1, 2]$ is the feasible set, and that all conditions from Assumption 3.2 are satisfied. Let $\alpha = 1$. Using (8), we may write the corresponding regularized gap function g_1 as

$$g_1(x) = \frac{1}{2} x^2 - \frac{1}{2} \inf_{z \in S(x)} z^2 = \begin{cases} 2x - 2 & \text{if } x \in [0, 1], \\ \frac{1}{2}(x^2 - 1) & \text{if } x \in (1, 2], \\ -\infty & \text{if } x \notin [0, 2]. \end{cases}$$

The graph of the set-valued mapping S and the graph of the function g_1 on $\operatorname{dom} g_1 = [0, 2]$ are illustrated in Figures 1a and 1b, respectively. We see that the function g_1 is zero only at $x = 1$, hence this point is the unique solution of the QVI, but g_1 has a 'kink' precisely at this solution point. On the other hand, for the functions $f_1(x) = x^2 + \delta_X(x)$



(a) gph S (b) gph g_1 for $0 \le x \le 2$

(c) gph f_1^* (d) gph h_1^*

Figure 1: Illustrations for Example 3.10

and $h_1(x) = \frac{1}{2}x^2 + \frac{1}{2} \inf\limits_{z \in S(x)} z^2$, we get the following conjugates, see Figures 1c and 1d:

$$f_1^*(y) = \begin{cases} y - 1 & \text{if } y < 2, \\ \frac{1}{4}y^2 & \text{if } y \in [2, 4], \\ 2y - 4 & \text{if } y > 4, \end{cases} \quad \text{and} \quad h_1^*(y) = \begin{cases} -2 & \text{if } y < -2, \\ \frac{1}{4}y^2 + y - 1 & \text{if } y \in [-2, 0], \\ y - 1 & \text{if } y \in (0, 1), \\ \frac{1}{2}(y^2 - 1) & \text{if } y \in [1, 2], \\ 2y - \frac{5}{2} & \text{if } y > 2. \end{cases}$$

Simple calculations show that both functions are continuously differentiable on $\mathbb{R}$ with gradients

$$\nabla f_1^*(y) = \begin{cases} 1 & \text{if } y < 2, \\ \frac{1}{2}y & \text{if } y \in [2,4], \\ 2 & \text{if } y > 4, \end{cases} \quad \text{and} \quad \nabla h_1^*(y) = \begin{cases} 0 & \text{if } y < -2, \\ \frac{1}{2}y + 1 & \text{if } y \in [-2,0], \\ 1 & \text{if } y \in (0,1), \\ y & \text{if } y \in [1,2], \\ 2 & \text{if } y > 2. \end{cases}$$

The same results follow from Lemma 3.8 and 3.9, respectively. ◊

We now apply Toland's and Singer's duality theory [40, Theorem 2.2] which states that

$$\inf_{x \in \mathbb{R}^n} [f(x) - h(x)] = \inf_{y \in \mathbb{R}^n} [h^*(y) - f^*(y)] \tag{19}$$

for all functions $f, h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with h convex and lower semicontinuous. Hence this duality fits perfectly within our framework and allows us to state the following main result of this section.

Theorem 3.11 *Let Assumption 3.2 hold, and define the dual gap function*

$$d_\alpha^* := h_\alpha^* - f_\alpha^*$$

with the functions f_α^* and h_α^* given by Lemmas 3.8 and 3.9, respectively. Then the following statements hold:

(a) The function d_α^* is continuously differentiable on $\mathbb{R}^n$.

(b) If $\bar{y}$ is a solution of the unconstrained minimization problem

$$\min \ d_\alpha^*(y), \quad y \in \mathbb{R}^n, \tag{20}$$

with $d_\alpha^*(\bar{y}) = 0$, then $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is a solution of the QVI.

(c) Conversely, if $\bar{x}$ is a solution of the QVI and $\bar{y} \in \partial h_\alpha(\bar{x})$, then $\bar{y}$ is a solution of (20) with $d_\alpha^*(\bar{y}) = 0$.

Proof. The result is essentially an application of the duality theory by Toland [39, 40] and Singer [36], but for the sake of completeness, we provide the details here.

(a) This follows immediately from the definition of the function d_α^* together with Lemmas 3.8 (c) and 3.9 (c).

(b) Let $\bar{y}$ be a solution of (20) with $d_\alpha^*(\bar{y}) = 0$. Then

$$0 = d_\alpha^*(\bar{y}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{21}$$

Moreover, the optimality of $\bar{y}$ and the convexity and continuous differentiability of f_α^* leads to

$$h_\alpha^*(y) - f_\alpha^*(y) \geq h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}) \quad \text{and} \quad f_\alpha^*(y) - f_\alpha^*(\bar{y}) \geq \nabla f_\alpha^*(\bar{y})^T(y - \bar{y})$$

for all $y \in \mathbb{R}^n$. Consequently, we have

$$h_\alpha^*(y) - h_\alpha^*(\bar{y}) \geq f_\alpha^*(y) - f_\alpha^*(\bar{y}) \geq \nabla f_\alpha^*(\bar{y})^T(y - \bar{y})$$

for all $y \in \mathbb{R}^n$. This shows that the vector $\bar{x} := \nabla f_\alpha^*(\bar{y})$ is an element of the subdifferentials $\partial f_\alpha^*(\bar{y})$ and $\partial h_\alpha^*(\bar{y})$. Since we also have $f_\alpha^{**} = f_\alpha$ and $h_\alpha^{**} = h_\alpha$ by Lemma 2.4 (d) and Lemma 3.7 (a), (b), we obtain from Lemma 2.4 (f) that

$$f_\alpha(\bar{x}) + f_\alpha^*(\bar{y}) = \bar{x}^T\bar{y} \quad \text{and} \quad h_\alpha(\bar{x}) + h_\alpha^*(\bar{y}) = \bar{x}^T\bar{y}. \tag{22}$$

Subtracting and rearranging these two equations shows

$$f_\alpha(\bar{x}) - h_\alpha(\bar{x}) = h_\alpha^*(\bar{y}) - f_\alpha^*(\bar{y}). \tag{23}$$

But the right-hand side is equal to zero in view of (21). Hence $\bar{x}$ is a minimum of the nonnegative function $f_\alpha - h_\alpha$ with function value equal to zero. Therefore Lemma 3.7 (c) implies that $\bar{x}$ is a solution of the QVI.

(c) The proof of this part is similar to the one of statement (b). Since $\bar{x}$ is a solution of the QVI, we have

$$0 = g_\alpha(\bar{x}) = \min_{x \in \mathbb{R}^n} \left[f_\alpha(x) - h_\alpha(x) \right] \tag{24}$$

in view of Lemma 3.7 (c). Hence

$$f_\alpha(x) - h_\alpha(x) \geq f_\alpha(\bar{x}) - h_\alpha(\bar{x}) \quad \forall x \in \mathbb{R}^n$$

and, using $\bar{y} \in \partial h_\alpha(\bar{x})$,

$$h_\alpha(x) - h_\alpha(\bar{x}) \geq \bar{y}^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n.$$

Combining these two inequalities yields

$$\bar{y}^T(\bar{x} - x) \geq h_\alpha(\bar{x}) - h_\alpha(x) \geq f_\alpha(\bar{x}) - f_\alpha(x)$$

which shows that the element $\bar{y}$ from the subdifferential $\partial h_\alpha(\bar{x})$ also belongs to the subdifferential $\partial f_\alpha(\bar{x})$. Using these two subdifferential relations, we obtain from Lemma 2.4 (f) that (22) holds which, in turn, implies that (23) is also true. But this time, the left-hand side of (23) is equal to zero. Consequently, the right-hand side is also equal to zero, meaning that $\bar{y}$ is a solution of the minimization problem (20) with $d_\alpha^*(\bar{y}) = 0$ because of (19). $\square$

To illustrate the results of Theorem 3.11, we return to Example 3.10.

Example 3.12 Consider once again the setting from Example 3.10. Calculating the difference of $h_1^* - f_1^*$, we obtain

$$h_1^*(y) - f_1^*(y) = \begin{cases} -y-1 & \text{if } y < -2, \\ \frac{1}{4}y^2 & \text{if } y \in [-2,0], \\ 0 & \text{if } y \in (0,1), \\ \frac{1}{2}y^2 - y + \frac{1}{2} & \text{if } y \in [1,2], \\ -\frac{1}{4}y^2 + 2y - \frac{5}{2} & \text{if } y \in (2,4], \\ \frac{3}{2} & \text{if } y > 4. \end{cases}$$

This function is illustrated in Figure 2. Due to the observations in Example 3.10, the

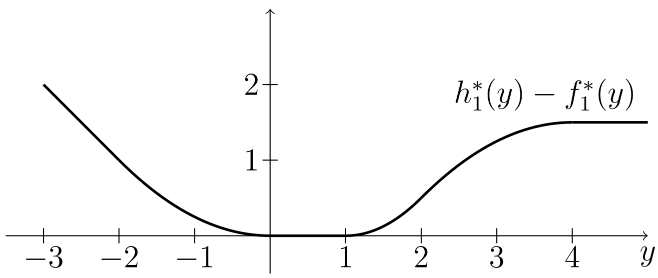


Figure 2: The graph of $h_1^* - f_1^*$

corresponding QVI has the unique solution $\bar{x} = 1$. Furthermore, it holds that $\partial h_1(1) = [0,1]$ since

$$h_1(x) = \frac{1}{2}x^2 + \frac{1}{2} \inf_{z \in S(x)} z^2 = \begin{cases} x^2 - 2x + 2 & \text{if } x \in [0,1], \\ \frac{1}{2}(x^2+1) & \text{if } x \in (1,2], \\ +\infty & \text{if } x \notin [0,2]. \end{cases}$$

In view of Theorem 3.11 (c), all $\bar{y} \in [0,1]$ solve the dual problem (20), and this statement is consistent with the graph of the dual problem shown in Figure 2. Furthermore, given any solution $\bar{y} \in [0,1]$ of (20), Theorem 3.11 (b) states that $\bar{x} = \nabla f_1^*(\bar{y})$ is a solution of the QVI. Since, in our case, we obtain $\nabla f_1^*(\bar{y}) = 1$ for all $\bar{y} \in [0,1]$, it follows that $\bar{x} = 1$ solves the QVI. This confirms a corresponding observation given in Example 3.10. $\Diamond$

Note that the dual gap function in the previous example has stationary points or local minima which are not solutions of the QVI. Since this example has relatively nice properties, this indicates that it might be difficult to obtain a result which says that, under suitable conditions, a stationary point is already a global minimum of the dual gap function. In fact, we were not able to derive such a result, but we have a partial result in this direction that is based on the following proposition.

Proposition 3.13 *Let Assumption 3.2 hold, let $d_\alpha^* = h_\alpha^* - f_\alpha^*$ be the dual gap function, and let $x_\alpha^{f^*}(y)$ and $x_\alpha^{h^*}(y), z_\alpha^{h^*}(y)$ denote the vectors defined in Lemma 3.8 and 3.9, respectively. Then the following statements are equivalent:*

(a) $d^*_\alpha(\bar{y}) = 0$.

(b) $x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}) = z^{h^*}_\alpha(\bar{y})$.

Proof. We first verify the simple implication (b) $\Longrightarrow$ (a). Hence assume that $x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}) = z^{h^*}_\alpha(\bar{y})$ holds. For simplicity of notation, let us denote this common vector by $\bar{x}$. Then, in particular, we have $\bar{x} \in X$, hence the definition of f^*_α yields

$$f^*_\alpha(\bar{y}) = \bar{x}^T\bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \frac{1}{2\alpha}\|F(\bar{x})\|^2,$$

whereas the definition of h^*_α implies

$$h^*_\alpha(\bar{y}) = \frac{1}{2\alpha}\|\bar{y}\|^2 - \frac{\alpha}{2}\left(\big\|\bar{x} - \frac{1}{\alpha}\bar{y}\big\|^2 + \big\|\frac{1}{\alpha}F(\bar{x})\big\|^2\right) = \bar{x}^T\bar{y} - \frac{\alpha}{2}\|\bar{x}\|^2 - \frac{1}{2\alpha}\|F(\bar{x})\|^2.$$

This immediately gives $d^*_\alpha(\bar{y}) = h^*_\alpha(\bar{y}) - f^*_\alpha(\bar{y}) = 0$.

Conversely, assume that $d^*_\alpha(\bar{y}) = 0$ holds. Then, in view of Theorem 3.11, $\bar{y}$ is a global minimum of the unconstrained optimization problem (20). Hence we have $\nabla d^*_\alpha(\bar{y}) = 0$. On the other hand, the definition of d^*_α together with Lemmas 3.8 and 3.9 yields

$$\nabla d^*_\alpha(\bar{y}) = \nabla h^*_\alpha(\bar{y}) - \nabla f^*_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}) - x^{f^*}_\alpha(\bar{y}).$$

Hence we obtain

$$x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}). \tag{25}$$

Furthermore, $d^*_\alpha(\bar{y}) = 0$ and Theorem 3.11 together imply that $\bar{x} := \nabla f^*_\alpha(\bar{y})$ is a solution of the QVI. Note that (25) and Lemma 3.8 yield

$$\bar{x} = x^{f^*}_\alpha(\bar{y}) = x^{h^*}_\alpha(\bar{y}). \tag{26}$$

The vector $\bar{x}$ being a solution of the QVI means that $\bar{x} \in X$ and $g_\alpha(\bar{x}) = 0$, where g_α denotes the regularized gap function, cf. Lemma 3.1. In view of (7), we may rewrite this regularized gap function as

$$\begin{aligned} g_\alpha(\bar{x}) &= \frac{1}{2\alpha}\|F(\bar{x})\|^2 - \frac{\alpha}{2}\inf_{z\in S(\bar{x})}\big\|z - (\bar{x} - \frac{1}{\alpha}F(\bar{x}))\big\|^2 \\ &= \frac{1}{2\alpha}\|F(\bar{x})\|^2 - \frac{\alpha}{2}\big\|z_\alpha(\bar{x}) - \big(\bar{x} - \frac{1}{\alpha}F(\bar{x})\big)\big\|^2 \end{aligned}$$

with the uniquely defined minimum

$$z_\alpha(\bar{x}) := \operatorname*{argmin}_{z\in S(\bar{x})}\big\|z - (\bar{x} - \frac{1}{\alpha}F(\bar{x}))\big\|^2.$$

According to Taji [38], $\bar{x}$ being a solution of the QVI is equivalent to $z_\alpha(\bar{x}) = \bar{x}$. However, in view of the representation (17) of the function $h^*_\alpha(\bar{y})$, it follows that $z_\alpha(\bar{x})$ is identical to $z^{h^*}_\alpha(\bar{y})$. Consequently, we also have $z^{h^*}_\alpha(\bar{y}) = \bar{x}$. Together with (26), this completes the proof. $\square$

Proposition 3.13 shows that $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y}) = z_\alpha^{h^*}(\bar{y}) =: \bar{x}$ implies $d_\alpha^*(\bar{y}) = 0$ and, therefore, that $\bar{x}$ is a solution of the QVI. This sufficient condition for a solution is partially satisfied at any stationary point of the dual gap function since, as noted in the previous proof, we always have $x_\alpha^{f^*}(\bar{y}) = x_\alpha^{h^*}(\bar{y})$ at a stationary point $\bar{y}$ of d_α^*. The missing part is therefore to verify that these two vectors are also equal to $z_\alpha^{h^*}(\bar{y})$ which seems to be the difficult part that is not satisfied in Example 3.10 for all $y \geq 4$.

Hence we have no complete answer for stationary points of the dual gap function d_α^* to be solutions of a QVI. On the other hand, since we know the optimal value of d_α^*, this disadvantage might not be that strong, since the function value itself tells us whether we are in a solution or not.

Theorem 3.11 gives, more or less, a one-to-one correspondence between the solutions of the QVI and the global minima of the dual gap function d_α^*. In fact, it shows that every solution of the optimization problem (20) yields a solution of the QVI, but the converse is not necessarily true, because statement (c) of Theorem 3.11 assumes (implicitly) that the subdifferential $\partial h_\alpha(\bar{x})$ is nonempty. As illustrated by the following counterexample, this subdifferential could be empty, and the infimum in the relation (19) is not necessarily attained.

Example 3.14 Consider the QVI with $n = 1$, $F(x) = x$, and

$$S(x) = \begin{cases} \left[1 - \sqrt{1-(x-2)^2}, 1 + \sqrt{1-(x-2)^2}\right] & \text{if } x \in [1,3], \\ \emptyset & \text{if } x \notin [1,3], \end{cases}$$

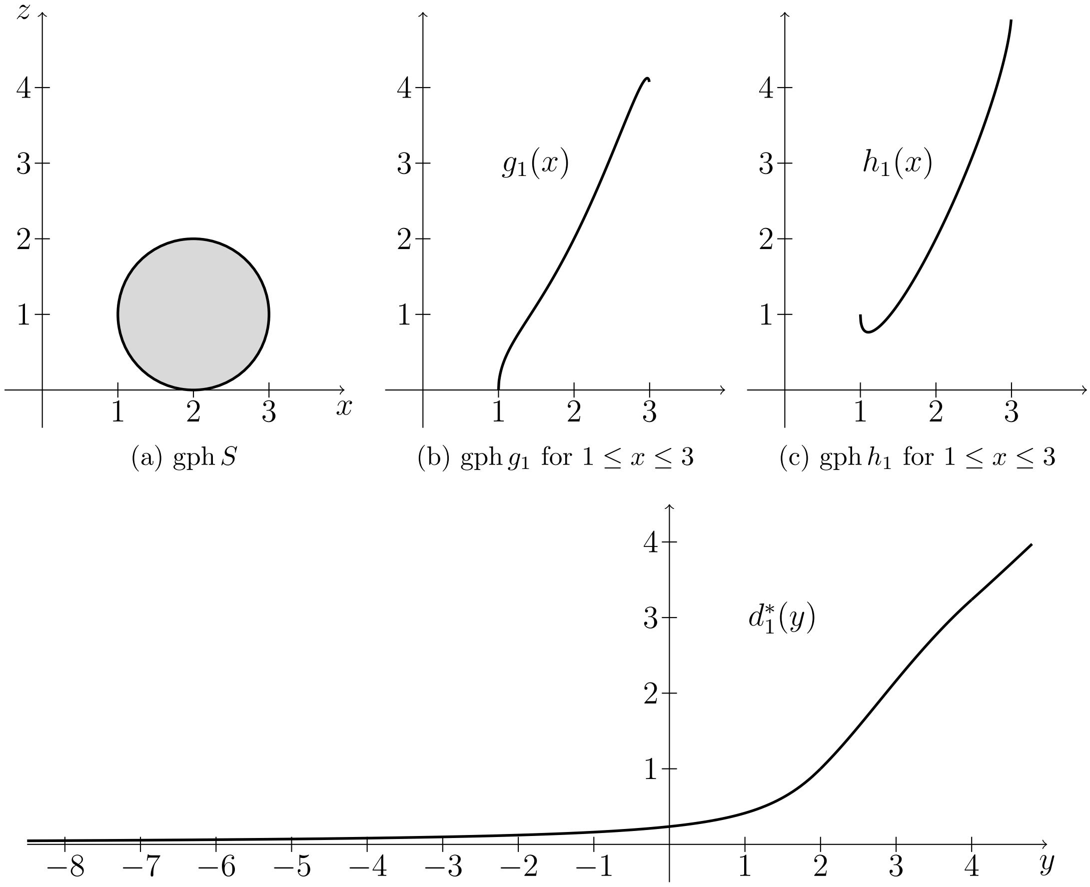
see Figure 3a. Note that $M = [1,3]$ is the domain of S in this example, that $X = [1,2]$ is the feasible set, and that all conditions from Assumption 3.2 are satisfied. Let $\alpha = 1$. The corresponding regularized gap function

$$g_1(x) = \frac{1}{2}x^2 - \frac{1}{2}\inf_{z \in S(x)} z^2 = \begin{cases} \frac{1}{2}x^2 - \frac{1}{2}\left(1 - \sqrt{1-(x-2)^2}\right)^2 & \text{if } x \in [1,3], \\ -\infty & \text{if } x \notin [1,3], \end{cases}$$

is illustrated in Figure 3b. For the functions $f_1(x) = x^2 + \delta_X(x)$ and $h_1(x) = \frac{1}{2}x^2 + \frac{1}{2}\inf_{z \in S(x)} z^2$ (see Figure 3c), we get the following conjugates:

$$f_1^*(y) = \begin{cases} y - 1 & \text{if } y < 2, \\ \frac{1}{4}y^2 & \text{if } y \in [2,4], \\ 2y - 4 & \text{if } y > 4, \end{cases} \quad \text{and} \quad h_1^*(y) = 2y + \sqrt{1+(y-2)^2} - 3.$$

We see that the function g_1 is zero only at $x = 1$, hence this point is the unique solution of the QVI. At this point, the slope of h_1 is infinite, and $\partial h_1(1) = \emptyset$. Hence, for this example, we cannot apply Theorem 3.11 to determine the solutions for the dual problem (20). We



(a) gph S (b) gph g_1 for $1 \le x \le 3$ (c) gph h_1 for $1 \le x \le 3$

(d) gph d_1^*

Figure 3: Illustrations for Example 3.14

further note that $d_1^*(y) = h_1^*(y) - f_1^*(y) > 0$ holds for all $y \in \mathbb{R}$ and $\lim_{y \to -\infty} d_1^*(y) = 0$, see Figure 3d. Therefore, zero is the infimum but not the minimum of the unconstrained minimization problem (20) which does not have a solution. $\Diamond$

4 PC^1 Property of the Dual Gap-Function

The dual gap function d_α^* turned out to be piecewise smooth in all previous examples. The aim of this section is therefore to show that this observation is true in a rather general setting. To this end, let us first recall the definition of a PC^1-mapping.

Definition 4.1 *A continuous function $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is called* PC^1 *near $\bar{x} \in D$ if there exists an open neighborhood $U \subset D$ of $\bar{x}$ and a finite family of continuously differentiable functions $f_i : U \to \mathbb{R}^m$ $(i = 1, \ldots, l)$ such that $f(x) \in \{f_1(x), \ldots, f_l(x)\}$ for all $x \in U$.*

Piecewise smooth functions arise naturally in the context of Euclidian projections onto convex sets. To this end, let us assume that we have a set $\Omega \subseteq \mathbb{R}^n$ described by

$$\Omega := \{x \mid c_i(x) \leq 0 \ (i = 1, \dots, m)\}, \tag{27}$$

with $c_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$ convex and twice continuously differentiable. The crucial constraint qualification about Ω in order to obtain a PC^1 property of the projection mapping is given in the next definition and goes back to [21].

Definition 4.2 *Let $c_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$ be continuously differentiable, and let Ω be defined by* (27)*. For $\bar{x} \in \Omega$ we put $I(\bar{x}) := \{i \mid c_i(\bar{x}) = 0\}$. Then we say that the* constant rank constraint qualification (CRCQ) *is satisfied at $\bar{x}$ for Ω if there exists a neighborhood U of $\bar{x}$ such that for all $K \subset I(\bar{x})$, the family of gradients $\{\nabla c_i(x) \mid i \in K\}$ has constant rank (depending on the set K) for all $x \in U$.*

Let Ω be the set from (27). Recall that the unique solution of the strongly convex minimization problem

$$\min_{w \in \Omega} \frac{1}{2} \|w - v\|^2$$

is called the *Euclidean projection* of a given vector $v \in \mathbb{R}^n$ onto the set Ω, denoted by $P_\Omega(v)$. The mapping $v \mapsto P_\Omega(v)$ is then called the *projection mapping*. It is well-known that this mapping is piecewise smooth under the CRCQ assumption. More precisely, the following result holds, see, e.g., [12, Thm. 4.5.2].

Theorem 4.3 *Let Ω be the set defined in* (27) *with twice continuously differentiable and convex functions c_i. Let $\bar{v} \in \mathbb{R}^n$ be given such that CRCQ holds at $\bar{w} := P_\Omega(\bar{v})$. Then the projection mapping P_Ω is a PC^1 function near $\bar{v}$.*

In order to apply this result to our case, recall that our two conjugate functions f_α^* and h_α^* also involve projections, but not with respect to the Euclidean norm. Instead, we are dealing with scaled projection problems of the form

$$\min_{w \in \Omega} \frac{1}{2} \|Dw - v\|^2, \tag{28}$$

where Ω denotes again the set from (27) and $D \in \mathbb{R}^{n \times n}$ is a nonsingular matrix. Then problem (28) is equivalent to the standard (Euclidean) projection problem

$$\min_{u \in D \cdot \Omega} \frac{1}{2} \|u - v\|^2$$

in the sense that the optimal values are equal with $\operatorname{argmin}_{u \in D \cdot \Omega} \frac{1}{2} \|u - v\|^2 = P_{D \cdot \Omega}(v)$ and $\operatorname{argmin}_{w \in \Omega} \frac{1}{2} \|Dw - v\|^2 = D^{-1} P_{D \cdot \Omega}(v)$. We are interested in the smoothness properties of the mapping $v \mapsto D^{-1} P_{D \cdot \Omega}(v)$.

To this end, we first state the following result which simply says that CRCQ still holds if the set is transformed in a simple way. The transformation is precisely the one that will be used in order to deal with projection-like problems as in (28).

Lemma 4.4 *Let $c_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$ be convex and continuously differentiable, $\Omega := \{x \mid c_i(x) \leq 0 \ (i = 1, \dots, m)\}$ and $\bar{v} \in \mathbb{R}^n$ such that CRCQ holds at $\bar{w} := D^{-1}P_{D\cdot\Omega}(\bar{v})$ for Ω. Then CRCQ holds at $\bar{u} := D\bar{w}$ for $D \cdot \Omega$.*

Proof. First, note that, setting $\tilde{c}_i(u) := c_i(D^{-1}u)$ $(i = 1, \dots, m)$ for all $u \in \mathbb{R}^n$, we have

$$D \cdot \Omega = \{u \mid \tilde{c}_i(u) \leq 0 \ (i = 1, \dots, m)\},$$

and thus,

$$I(\bar{u}) = \{i \mid \tilde{c}_i(\bar{u}) = 0\} = \{i \mid c_i(\bar{w}) = 0\} = I(\bar{w}).$$

By assumption, there exists a neighborhood W of $\bar{w}$ such that for all $K \subset I(\bar{w})$ the family of gradients $\{\nabla c_i(w) \mid i \in K\}$ has constant rank for all $w \in W$. Since D is nonsingular, the set $U := D \cdot W$ is a neighborhood of $\bar{u}$. Now, let $u, u' \in U$ and $K \subset I(\bar{u})$ be given, in particular, there exist $w, w' \in W$ such that $u = Dw$ and $u' = Dw'$. It holds that

$$\{\nabla \tilde{c}_i(u) \mid i \in K\} = D^{-1} \cdot \{\nabla c_i(w) \mid i \in K\}$$

and

$$\{\nabla \tilde{c}_i(u') \mid i \in K\} = D^{-1} \cdot \{\nabla c_i(w') \mid i \in K\},$$

and due to what was already argued above, both sets have the same rank, which concludes the proof. □

The previous result allows us to formulate the PC^1-property for the solution mapping of problems in the form (28).

Proposition 4.5 *Let the assumptions of Lemma 4.4 hold such that, in addition, c_i $(i = 1, \dots, l)$ is twice continuously differentiable. Then there exists a neighborhood V of $\bar{v}$ such that $v \mapsto D^{-1}P_{D\cdot\Omega}(v)$ is PC^1 on V.*

Proof. From Lemma 4.4 we infer that CRCQ holds at $\bar{u} := D\bar{w} = P_{D\cdot\Omega}(\bar{v})$ for $D \cdot \Omega$. Hence, from Theorem 4.3, we conclude that there exists a neighborhood U of $\bar{u}$ on which $v \mapsto P_{D\cdot\Omega}(v)$ is PC^1. Hence, the function $v \mapsto D^{-1}P_{D\cdot\Omega}(v)$ is PC^1 on $V := D \cdot U$ (recall that V is indeed a neighborhood of $\bar{v}$ due to the nonsingularity of the matrix D). □

We now want to apply the previous result in order to show that the gradient ∇d_α^* of the function d_α^* from Theorem 3.11 is PC^1.

For these purposes, we assume throughout that the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ takes the form

$$S(x) := \{z \in \mathbb{R}^n \mid s_i(z, x) \leq 0 \ (i = 1, \dots, m)\}, \tag{29}$$

where $s_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \dots, m)$ is twice continuously differentiable and convex in (z, x). Note that Assumption 3.2 (c) automatically holds in this case. Then we have

$$X = \{x \in \mathbb{R}^n \mid s_i(x, x) \leq 0 \ (i = 1, \dots, m)\}.$$

In order to verify the piecewise smoothness of the gradient of the dual gap function d_α^*, we show that both ∇h_α^* and ∇f_α^* are piecewise smooth. We begin with the mapping h_α^*.

Lemma 4.6 *Let Assumption 3.2 hold, and let $\bar{y} \in \mathbb{R}^n$ such that CRCQ holds at $(\bar{x}, \bar{z}) := D_h^{-1} P_{D_h \cdot \operatorname{gph} S}(\bar{y}, -b)$ for* $\operatorname{gph} S$, *where*

$$D_h := \begin{pmatrix} \alpha I & 0 \\ A - \alpha I & \alpha I \end{pmatrix}.$$

Then ∇h_α^* is PC^1 near $\bar{y}$.

Proof. It holds that, due to Lemma 3.9, for all $y \in \mathbb{R}^n$, we have $\nabla h_\alpha^*(y) = x_\alpha^{h^*}(y)$, where

$$\begin{aligned} \begin{pmatrix} x_\alpha^{h^*}(y) \\ z_\alpha^{h^*}(y) \end{pmatrix} &= \underset{(x,z)\in \operatorname{gph} S}{\operatorname{argmin}} \left\{ \left\| x - \frac{1}{\alpha} y \right\|^2 + \left\| z - \left(x - \frac{1}{\alpha} F(x) \right) \right\|^2 \right\} \\ &= \underset{(x,z)\in \operatorname{gph} S}{\operatorname{argmin}} \left\| \begin{pmatrix} \alpha I & 0 \\ A - \alpha I & \alpha I \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} - \begin{pmatrix} y \\ -b \end{pmatrix} \right\|^2 \\ &= D_h^{-1} P_{D_h \cdot \operatorname{gph} S}(y, -b). \end{aligned}$$

Hence, the assertion follows immediately from Proposition 4.5. □

Similar to the previous result, the next one proves that also the function ∇f_α^* is piecewise smooth under a suitable CRCQ assumption.

Lemma 4.7 *Let Assumption 3.2 hold, and let $\bar{y} \in \mathbb{R}^n$ such that CRCQ holds at $\bar{x} := D_f^{-1} P_{D_f \cdot X}(D_f^{-1}(\bar{y} - q_\alpha))$, where $D_f := Q_\alpha^{\frac{1}{2}}$ denotes the matrix square root of the matrix Q_α from* (12). *Then ∇f_α^* is PC^1 near $\bar{y}$.*

Proof. Due to Lemma 3.8, for all $y \in \mathbb{R}^n$, we have

$$\begin{aligned} \nabla f_\alpha^*(y) &= \underset{x\in X}{\operatorname{argmin}} \left\| Q_\alpha^{-1}(y - q_\alpha) - x \right\|_{Q_\alpha}^2 \\ &= \underset{x\in X}{\operatorname{argmin}} \left\| Q_\alpha^{-\frac{1}{2}}(y - q_\alpha) - Q_\alpha^{\frac{1}{2}} x \right\|^2 \\ &= D_f^{-1} P_{D_f \cdot X}(D_f^{-1}(y - q_\alpha)). \end{aligned}$$

Hence, the assertion follows immediately from Proposition 4.5. □

Summarizing the previous result, we obtain the following main result of this section.

Theorem 4.8 *Let Assumption 3.2 hold, and let $\bar{y} \in \mathbb{R}^n$ such that the assumptions of Lemmas 4.6 and 4.7 hold for $\bar{y}$. Then the gradient of the dual gap function ∇d_α^* is PC^1 near $\bar{y}$.*

Note that the two CRCQ conditions used in Lemmas 4.6 and 4.7 are independent of each other. A simple, but still important, case where the constant rank assumption holds, is the linear one. This yields the following consequence.

Corollary 4.9 *Let the functions s_i in* (29) *be (affine-)linear. Then the gradient of the dual gap function ∇d_α^* is a PC^1 mapping (in fact, it is piecewise (affine-)linear).*

Piecewise smooth functions are, in particular, semismooth in the sense of [30, 31], see, e.g., [12, Prop. 7.4.6]. In principle, this observation therefore allows the application of second-order Newton-type methods for the minimization of the dual gap function.

5 Numerical Results

In view of Theorem 3.11, a solution $\bar{y}$ of the the dual unconstrained minimization problem (20) implies a solution $\bar{x} = \nabla f_\alpha^*(\bar{y})$ of the corresponding QVI. In this section, we apply this theory to a class of examples from the QVILIB test problem collection [10] which satisfy Assumption 3.2.

For the solution of the unconstrained minimization problem (20), we use two different first-order methods: the spectral gradient (SG) method from [32] and a conjugate gradient (CG) method. The SG method is defined by

$$y^{k+1} := y^k - t_k \nabla d_\alpha^*(y^k)$$

with

$$t_0 := 1, \quad t_k := \frac{\|q^{k-1}\|^2}{(q^{k-1})^T r^{k-1}}, \quad q^{k-1} := y^k - y^{k-1}, \quad r^{k-1} := \nabla d_\alpha^*(y^k) - \nabla d_\alpha^*(y^{k-1})$$

if t_k satisfies the nonmonotone line search condition from [32]. In the CG method, we generate the search direction p^k for the iterates $y^{k+1} := y^k + t_k p^k$ using the Polak-Ribière updating scheme where

$$\begin{aligned} p^0 &:= -\nabla d_\alpha^*(y^0), \quad p^{k+1} := -\nabla d_\alpha^*(y^{k+1}) + \beta_k^{PR} p^k, \\ \beta_k^{PR} &:= \frac{\left(\nabla d_\alpha^*(y^{k+1}) - \nabla d_\alpha^*(y^k)\right)^T \nabla d_\alpha^*(y^{k+1})}{\left\|\nabla d_\alpha^*(y^k)\right\|^2}. \end{aligned}$$

Whenever $(p^k)^T \nabla d_\alpha^*(y^k) > 0$ holds, this CG algorithm has to be restarted with the negative gradient. For the examples in Table 2, however, this case never occured. Furthermore, we compute the step length t_k satisfying the strong Wolfe-Powell conditions whose implementation is based on the suggestion outlined in [22]. The computation of this stepsize uses, at each iteration k, the initial guess

$$t_k = \frac{-2 d_\alpha^*(y^k)}{(p^k)^T \nabla d_\alpha^*(y^k)}.$$

For both methods, the termination criteria are $\|\nabla d_\alpha^*(y_k)\| \leq 10^{-5}$ or $d_\alpha^*(y_k) \leq 10^{-6}$.

For the computation of the conjugate functions of f_α and h_α from Lemma 3.8 and 3.9, respectively, we use the TOMLAB/KNITRO solver with the active set Sequential Linear-Quadratic Programming (SLQP) optimizer by setting `Prob.KNITRO.options.ALG=3` and

Example	n	y^0	k	$\#d_\alpha$	d_α^*	$\|\nabla d_\alpha^*\|$
Scrim11	2400	$(0,\dots,0)$	32	33	1.8044e-08	1.9517e-05
Scrim11	2400	$(10,\dots,10)$	36	37	5.4686e-08	6.2638e-05
Scrim12	4800	$(0,\dots,0)$	32	33	2.7707e-08	3.3777e-05
Scrim12	4800	$(10,\dots,10)$	36	37	9.0629e-08	5.6006e-05
Scrim21	2400	$(0,\dots,0)$	32	33	1.9558e-08	1.8173e-05
Scrim21	2400	$(10,\dots,10)$	36	37	7.3633e-08	7.2570e-05
Scrim22	4800	$(0,\dots,0)$	32	33	1.7462e-08	2.9717e-05
Scrim22	4800	$(10,\dots,10)$	36	37	5.6927e-08	6.2776e-05

Table 1: Numerical results with the spectral gradient method

Example	n	y^0	k	$\#d_\alpha$	d_α^*	$\|\nabla d_\alpha^*\|$
Scrim11	2400	$(0,\dots,0)$	15	37	3.1869e-07	1.5863e-04
Scrim11	2400	$(10,\dots,10)$	20	47	7.6852e-07	3.5654e-04
Scrim12	4800	$(0,\dots,0)$	15	37	8.8592e-07	2.4624e-04
Scrim12	4800	$(10,\dots,10)$	23	50	9.3831e-07	1.5281e-04
Scrim21	2400	$(0,\dots,0)$	15	37	3.2617e-07	1.5942e-04
Scrim21	2400	$(10,\dots,10)$	20	47	7.5009e-07	3.5802e-04
Scrim22	4800	$(0,\dots,0)$	15	37	8.6840e-07	2.3947e-04
Scrim22	4800	$(10,\dots,10)$	23	50	8.8691e-07	1.5345e-04

Table 2: Numerical results with CG method with Polak-Ribière update of β

`Prob.KNITRO.options.FEASTOL=`10^{-10}, see the TOMLAB/KNITRO User's Guide on the web site `http://tomopt.com/tomlab/products/knitro/` for more information about the TOMLAB/KNITRO solver. Our implementation uses the regularization parameter $\alpha = 5$ for all test runs.

The class of test problems that we use here are named `Scrim*` in the test problem library QVILIB from [10]. This class corresponds to a large-scale transportation problem formulated as QVIs. Tables 1 and 2 contain the following data: The name of the example, the number of variables n, the starting point y^0, the number of iterations k, the cumulated number of dual gap function evaluations $\#d_\alpha^*$ needed until convergence, the final value of the dual gap function d_α^*, and the final value of the gradient norm $\|\nabla d_\alpha^*\|$.

In view of the large number of variables in each example, the evaluation of the dual gap function is more expensive than the computations in the outer iterations for both methods. In Tables 1 and 2, we observe that the total number of dual gap function evaluations in the SG method is less than in the CG method. Therefore, in spite of higher number of iterations, the total time until convergence in the SG method is less than in the CG method. Furthermore, we achieve the higher accuracy of the results with the SG method although the termination criteria for both methods are the same. In any case, both methods were able to find a solution for all instances of this class of QVIs, i.e., they never stopped at a local minimum of d_α^*.

6 Final Remarks

This paper shows that it is possible to reformulate a certain class of QVIs as an unconstrained and smooth optimization problem which, therefore, allows the application of some standard first-order software in order to solve the underlying QVI. In principle, since the objective function is continuously differentiable with a semismooth gradient (under suitable assumptions), the application of second-order methods is also possible. A natural candidate would be the semismooth Newton method from [30, 31], however, the computation of the corresponding generalized Jacobians (or Hessians, in our case) might be rather expensive. We therefore believe that another Newton-type method based on the idea of the computable generalized Jacobian from [37] (see also [17] for an application within the framework of generalized Nash equilibrium problems) might be the better choice. The corresponding details are left as part of our future research.

Acknowledgment. The authors would like to thank Oliver Stein for his comments on an earlier draft of this paper.

References

[1] D. Aussel, R. Correa, and M. Marechal: *Gap functions for quasivariational inequalities and generalized Nash equilibrium problems.* Journal of Optimization Theory and Applications 151, 2011, pp. 474–488.

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

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

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

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

[6] D.P. Bertsekas: *Nonlinear Programming.* Athena Scientific, Belmont, Massachusetts, second edition 1999.

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

[8] H. Dietrich: *A smooth dual gap function solution to a class of quasivariational inequalities.* Journal of Mathematical Analysis and Applications 235, 1999, pp. 380–393.

[9] H. Dietrich: *Optimal control problems for certain quasivariational inequalities.* Optimization 49, 2001, pp. 67–93.

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

[11] F. Facchinei, C. Kanzow, and S. Sagratella: *Solving quasi-variational inequalities via their KKT conditions.* Mathematical Programming, DOI:10.1007/s10107-013-0637-0.

[12] F. Facchinei and J.-S. Pang: *Finite Dimensional Variational Inequalitites and Complementarity Problems Volumes I and II.* Springer, New York, 2003.

[13] M. Fukushima: *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems.* Mathematical Programming 53, 1992, pp. 99–110.

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

[15] F. Giannessi: *Separation of sets and gap functions for quasi-variational inequalities.* In F. Giannessi and A. Maugeri (eds.): *Variational Inequality and Network Equilibrium Problems*, Plenum Press, New York, 1995, pp. 101–121.

[16] N. Harms, C. Kanzow, and O. Stein: *Smoothness properties of a regularized gap function for quasi-variational inequalities.* Optimization, to appear.

[17] A. von Heusinger, C. Kanzow, and M. Fukushima: *Newton's method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation.* Mathematical Programming 132, 2012, pp. 99–123.

[18] T. Horst and N.V. Thoai: *DC programming: Overview.* Journal of Optimization Theory and Applications 103, 1999, pp. 1–43.

[19] J.-B. Hiriart-Urruty and C. Lemaréchal: *Fundamentals of Convex Analysis.* Grundlehren Text Editions, Springer, Berlin, Heidelberg, 2001.

[20] W.W. Hogan: *Point-to-set maps in mathematical programming.* SIAM Review 15, 1973, pp. 591–603.

[21] R. Janin: *Directional derivative of the marginal function in nonlinear programming.* Mathematical Programming Study 21, 1984, pp. 110–126.

[22] J.J. Moré and D.C. Sorensen: *Newton's method.* In: G.H. Golub (ed): *Studies in Numerical Analysis.* The Mathematical Association of America, Washington, D.C., 1984, pp. 29–82.

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

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

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

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

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

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

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

[30] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[31] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[32] M. Raydan: *The Barzilai and Borwein gradient method for the large unconstrained minimization problem.* SIAM Journal on Optimization 7, 1997, pp. 26–33.

[33] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.

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

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

[36] I. Singer: *A Fenchel-Rockafellar type duality theorem for maximization.* Bulletin of the Australian Mathematical Society 20, 1979, pp. 193–198.

[37] D. Sun, M. Fukushima, and L. Qi: *A computable generalized Hessianof the D-gap function and Newton-type methods for variational inequality problems.* Complementarity and Variational Problems: State of the Art, M.C. Ferris and J.-S. Pang (eds.), SIAM, Philadelphia, 1997, pp. 452–472.

[38] K. Taji: *On gap functions for quasi-variational inequalities.* Abstract and Applied Analysis, 2008, Article ID 531361.

[39] J.F. Toland: *Duality in non-convex optimization.* Journal of Mathematical Analysis and Applications 66, 1978, pp. 399–415.

[40] J.F. Toland: *A duality principle for non-convex optimisation and the calculus of variations.* Archive for Rational Mechanics and Analysis 71, Number 1, 1979, pp. 41–61.