First-Order Sensitivity of Linearly Constrained Strongly Monotone Composite Variational Inequalities*

Jong-Shi Pang† Defeng Sun‡

December 23, 2008

Abstract

Extending the characterization of the directional derivatives of a strongly regular solution to a variational inequality under parameter perturbations, this paper derives a first-order sensitivity result for a monotone-plus affine variational inequality (AVI) having non-isolated solutions. The result employs a "directional AVI" whose solutions provide the directional derivatives, if they exist, of some invariant constants of the original AVI when its data are perturbed. Solvability of the directional AVI is characterized in terms of the direction of data perturbation. An application of this result to the support-vector machine quadratic program is discussed. An extension to a linearly constrained variational inequality defined by a strongly monotone composite mapping is also addressed.

1 Introduction

Starting with the seminal work of S.M. Robinson on differentiable inequality systems [21] in the early 1970's, modern sensitivity analysis of nonlinear programs (NLPs) and variational inequalities (VIs) under parameter perturbations is a well researched topic in continuous optimization. The reference [2] provides an excellent comprehensive account of this subject, including the treatment of infinite-dimensional problems. Offering a different approach, Chapter 5 of [10] treats the finite-dimensional variational inequality and complementarity problems in detail. These analyses are generally of 2 kinds: local analysis of an isolated solution, and global continuity properties of the solution sets. Under the strong regularity property introduced in [22], the former local analysis includes the directional differentiability of the solution and the characterization of such a derivative. In contrast, relying on an upper Lipschitzian property of polyhedral multifunctions discovered by Robinson [23], the sensitivity analysis of the solution set of a VI is restricted mainly to affine problems.

The present paper is motivated by several sources. Foremost is a question raised by the above background; namely, is it possible to derive a "first-order" sensitivity result for certain constants of an affine variational inequality (AVI) with non-unique solutions, and more generally, to a nonlinear variational inequality (VI)? To give an example to explain this question, consider a convex quadratic program (QP). It is known, by a basic result due to Mangasarian [18], that the gradient of the objective function is a constant on the set of optimal solutions. Thus by a result due originally to M.S. Gowda [10, Exercise 5.6.14], it follows that the said gradient is a single-valued, piecewise linear, thus Lipschitz

*This work was supported in part by the Office of Naval Research under grant N00014-06-1-0014 and the Air Force Office of Sponsored Research under grant FA9550-08-1-0061.

†Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Illinois, USA. Email: jspang@illinois.edu.

‡Department of Mathematics and Risk Management Institute, National University of Singapore, Republic of Singapore. The authors research was supported by the Academic Research Fund under Grant R-146-000-104-112 and the Risk Management Institute under Grants R-703-000-004-720 and R-703-000-004-646. Email: matsundf@nus.edu.sg.

continuous, function of the right-hand side of the constraints of the program. Such Lipschitz continuity is a zeroth-order sensitivity property. A first-order sensitivity property would be the differentiability of the optimal gradient as a function of the data. By known theory of piecewise linear functions, cf. the habilitation thesis of Scholtes [25] and [10, Lemma 4.6.1], it follows that the directional derivatives of the optimal gradient exist. Nevertheless, the characterization and computation of such derivatives remain open. In particular, it is not known whether a directional derivative can be derived from the optimal solutions of a related quadratic program. Part of the objective of this paper is to provide an affirmative answer to the latter question in the broader context of a "monotone-plus" AVI and its generalization to a linearly constrained "strongly monotone composite" VI. A recent paper [24] has established that the solution set of such an AVI is a Lipschitz multifunction of the right-hand side of the constraints defining the polyhedron of the VI whose defining function is not perturbed. A related paper [15] establishes a local sensitivity property of an isolated solution of a linearly constrained VI under a coherent orientation assumption on the solution but without the co-coercivity assumption, which is a global property on the defining function.

Another important motivation for the present work occurs in the context of a specific convex quadratic program that provides the optimization formulation for a support vector machine (SVM). Such a QP depends on 2 crucial parameters whose choice has a significant effect on the solution of the problem, and more importantly, on the generalization power of the SVM. Several recent works in the machine learning community [6, 7, 8] have suggested the choice of these parameters via the minimization of an "out-of-sample" error by a gradient descent method; in general, one has to be careful with the application of such a gradient-based method because the out-of-sample error is only a directionally differentiable function of the model parameters. Most recently, this idea of out-of-sample error minimization is significantly expanded in several papers [1, 13, 14], in which a bilevel programming approach to cross-validated support-vector regression and classification is proposed. The sensitivity results obtained in this paper provide a rigorous exposition of the directional differentiability of the SVM solution as a function of the model parameters.

2 Monotone-Plus AVIs

For a given n-vector q, an $n \times n$ matrix M (not necessarily symmetric), and a polyhedron

$$K \triangleq \{ x \in \mathbb{R}^n : Ax \leq b \},$$

where A is an $m \times n$ matrix and b is an m-vector, consider the affine variational inequality, which we denote AVI $(q, b; M, A)$, of finding a vector $x \in K$ such that

$$(y - x)^T (q + Mx) \geq 0, \quad \forall y \in K.$$

Throughout this section, the pair of matrices (M, A) is fixed but the pair of vectors (q, b) is subject to perturbations; moreover, we make the blanket assumption that the set K is nonempty and the matrix M is *positive semidefinite-plus*; the latter means that M is positive semidefinite and satisfies the implication:

$$x^T Mx = 0 \Rightarrow Mx = 0.$$

Such a monotone-plus AVI includes the case of the convex quadratic program:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & q^T x + \frac{1}{2} x^T Mx \\ \text{subject to} & Ax \leq b, \end{array} \tag{1}$$

where M is symmetric positive semidefinite. By a basic result of Luo and Tseng [17], it is known that M is positive semidefinite-plus if and only if $M = EGE^T$ for some $n \times \ell$ matrix E and some $\ell \times \ell$ positive definite (not necessarily symmetric) matrix G; see also [10, Exercise 2.9.4]. This characterization of a positive semidefinite-plus matrix serves as the basis for a subsequent generalization to a certain class of linearly constrained variational inequalities defined by a (nonlinear) "strongly monotone composite" map; see Section 4.

A special case of the monotone-plus AVI is worth-mentioning, namely when $M = EGE^T$ with G positive definite and $q = E^T r$ for some vector r. In this case, the monotone AVI $(q, b; M, A)$ in the variable x is equivalent to the *strongly monotone* AVI of finding a vector $y \in \widehat{K} \triangleq \{y = E^T x \text{ for some } x \in K\}$ such that $(y' - y)^T(r + Gy) \geq 0$ for all $y' \in \widehat{K}$. In this case, the first-order sensitivity analysis of the unique solution y to the latter AVI as a function of r for fixed $(b; E, G, A)$ is well understood. This analysis pertains to the perturbation of q in the range of E^T with b fixed. While the simultaneous perturbation of (r, b) can be similarly analyzed using the positive definiteness of G, the general case where q does not belong to the range of E^T is not covered by existing results; the treatment of this case, which is not trivial, is the goal of this section.

Let $\mathcal{R}(M, A)$ denotes the set of pairs (q, b) for which the AVI $(q, b; M, A)$ has a solution. It is well known that $\mathcal{R}(M, A)$ consists of all pairs (q, b) for which there exists $(x, \lambda) \in \mathbb{R}^{n+m}$ such that

$$\begin{aligned} 0 &= q + Mx + A^T\lambda \\ 0 &\leq \lambda, \quad b - Ax \geq 0. \end{aligned}$$

Thus $\mathcal{R}(M, A) = \begin{bmatrix} -M & -A^T \\ A & 0 \end{bmatrix} (\mathbb{R}^n \times \mathbb{R}^m_+) + (\{0\} \times \mathbb{R}^m_+)$, showing that $\mathcal{R}(M, A)$ is a convex polyhedral cone. For each $(q, b) \in \mathcal{R}(M, A)$, let $\text{SOL}(q, b; M, A)$ denote the nonempty set of solutions to the AVI $(q, b; M, A)$. While $\text{SOL}(q, b; M, A)$ is in general a convex polyhedron with multiple elements, it follows from the positive semidefinite-plus property of M that $M\text{SOL}(q, b; M, A)$ is a singleton [10, Corollary 2.3.10]. (In the context of the QP (1), this property follows from a result of Mangasarian [18] which asserts that the gradient of the objective function of a convex program is a constant on the set of optimal solutions of the program.) Moreover, since $\text{SOL}(\bullet, \bullet; M, A) : \mathbb{R}^{n+m} \to \mathbb{R}^n$ is a polyhedral multifunction with a convex domain (equal to $\mathcal{R}(M, A)$), it follows that $q + M\text{SOL}(q, b; M, A)$, which we denote $w(q, b)$ for simplicity, is a piecewise linear, thus Lipschitz continuous, function of (q, b) on $\mathcal{R}(M, A)$. (In the case of the QP (1), $w(q, b) = \nabla_x \theta(\text{SOL}(q, b; M, A); q)$.)

Let (q, b) be a given pair in $\mathcal{R}(M, A)$ and (q', b') be a pair such that $(q(\tau), b(\tau)) \triangleq (q, b) + \tau(q', b')$ remains in $\mathcal{R}(M, A)$ for all $\tau > 0$ sufficiently small. We wish to show that the limit

$$\lim_{\tau \downarrow 0} \frac{w(q(\tau), b(\tau)) - w(q, b)}{\tau} \tag{2}$$

exists and is equal to $q' + M\widehat{x}$, where $\widehat{x}$ is any solution to a certain AVI that is determined by the tuple (q, b, q', b'). In principle, the existence of the limit (2) is known from the theory of piecewise linear functions [25]. Nevertheless the proof of this result in the latter thesis is rather involved and sheds no light on what the "directional AVI" should be. The general sensitivity theory in [10, Section 5.4] provides some insight but is not applicable because of the non-isolatedness of the elements in $\text{SOL}(q, b; M, A)$. Besides identifying what the directional AVI should be in this case of non-unique solutions, our analysis gives a direct existence proof of the limit (2) that is specific to the AVI $(q, b; M, A)$.

There are two complications in the analysis: one is due to the multiplicity of vectors in the solution set $\text{SOL}(q, b; M, A)$; and two is the non-uniqueness of the associated multipliers λ that satisfy the following Karush-Kuhn-Tucker (KKT) conditions of (1):

$$\begin{aligned} 0 &= q + Mx + A^T\lambda \\ 0 &\leq b - Ax \perp \lambda \geq 0. \end{aligned} \tag{3}$$

For each $x \in \mathrm{SOL}(q,b;M,A)$, we let $\Lambda(q,b;x)$ be the (polyhedral) set of multipliers λ for which (x,λ) satisfies the KKT system (3); $\Lambda(q,b;x)$ has the following algebraic representation:

$$\Lambda(q,b;x) \triangleq \left\{ \begin{array}{ll} \lambda \geq 0 : & 0 = w(q,b) + A^T\lambda \\ & 0 = \lambda_i \quad \forall\, i \notin \mathcal{A}(x) \end{array} \right\},$$

where $\mathcal{A}(x) \triangleq \{i : (\, b - Ax\,)_i = 0\}$ is the index set of active constraints at x. It turns out that, due to positive semidefinitess of M, the sets $\Lambda(q,b;x)$ are all equal to a common set that can be defined in terms of one single index set:

$$\alpha(q,b) \triangleq \{\, i : \exists \text{ a pair } (x,\lambda) \text{ satisfying (3) such that } \lambda_i > 0\,\},$$

which depends only on the tuple $(q,b;M,A)$ and not on the solutions of the AVI. Indices in $\alpha(q,b)$ correspond to the *maximally complementary* solutions of the KKT system (3) as a mixed monotone linear complementarity problem [11, 12]; this index set can be determined by an interior-point method together with a strongly polynomial rounding procedure as described in the latter reference. Supplementing [10, Proposition 2.3.12], the following lemma shows that $\Lambda(q,b;x)$ is equal to the set

$$\Lambda(q,b) \triangleq \left\{ \begin{array}{ll} \lambda \geq 0 : & 0 = w(q,b) + A^T\lambda \\ & 0 = \lambda_i \quad \forall\, i \notin \alpha(q,b) \end{array} \right\}$$

for all solutions x of the AVI $(q,b;M,A)$.

Lemma 1. For every $x \in \mathrm{SOL}(q,b;M,A)$, $\Lambda(q,b;x) = \Lambda(q,b)$. Moreover, $\Lambda(q,b)$ is further equal to the optimal solution set of the linear program:

$$\begin{array}{ll} \underset{\lambda \geq 0}{\text{minimize}} & b^T\lambda \\ \text{subject to} & w(q,b) + A^T\lambda = 0. \end{array}$$

Proof. The inclusion $\Lambda(q,b;x) \subseteq \Lambda(q,b)$ is immediate by the definition of the index set $\alpha(q,b)$. Conversely, if $\lambda \in \Lambda(q,b)$, then for every $x \in \mathrm{SOL}(q,b;M,A)$, it suffices to verify that (x,λ) satisfies the complementary slackness condition: for every index i, either $\lambda_i = 0$ or $(b - Ax)_i = 0$. This is clear if $i \notin \alpha(q,b)$ by the definition of λ as an element of $\Lambda(q,b)$. If $i \in \alpha(q,b)$, then there exists (x',λ') satisfying (3) such that $\lambda'_i > 0$. By the *cross complementarity* property of the KKT system (3), it follows that $(b - Ax)_i = 0$, as desired. The second assertion of the lemma is proved in [10, Proposition 2.3.12]. □

For an arbitrary vector b' and a given solution $x \in \mathrm{SOL}(q,b;M,A)$, consider the linear program (LP):

$$\begin{array}{ll} \underset{x' \in \mathbb{R}^n}{\text{minimum}} & (\,x'\,)^T w(q,b) \\ \text{subject to} & (\,b' - Ax'\,)_i \geq 0, \quad \forall\, i \in \mathcal{A}(x), \end{array} \tag{4}$$

whose dual is (with $A_{i\bullet}$ denoting the i-th row of A):

$$\begin{array}{lll} \underset{\lambda_i \,:\, i \in \mathcal{A}(x)}{\text{minimize}} & \displaystyle\sum_{i \in \mathcal{A}(x)} b'_i \lambda_i & \\ \text{subject to} & w(q,b) + \displaystyle\sum_{i \in \mathcal{A}(x)} (\,A_{i\bullet}\,)^T \lambda_i = 0 & \\ \text{and} & \lambda_i \geq 0, \quad \forall\, i \in \mathcal{A}(x). & \end{array} \tag{5}$$

By Lemma 4, (5) is equivalent to

$$\theta_{\min}(q,b;b') \triangleq \underset{\lambda \in \Lambda(q,b)}{\text{minimum}} \quad (\,b'\,)^T\lambda, \tag{6}$$

where the dependence on the solution x disappears. By the algebraic representation of the set $\Lambda(q,b)$, (6) is equivalent to

$$\begin{array}{lll} \underset{\lambda_i \,:\, i \in \alpha(q,b)}{\text{minimize}} & \displaystyle\sum_{i\in\alpha(q,b)} b_i'\lambda_i & \\ \text{subject to} & w(q,b) + \displaystyle\sum_{i\in\alpha(q,b)} (\, A_{i\bullet}\,)^T\lambda_i \,=\, 0 & \\ \text{and} & \lambda_i \,\geq\, 0, \quad \forall\, i \,\in\, \alpha(q,b), & \end{array}$$

whose dual is

$$\begin{array}{ll} \underset{x' \in \mathbb{R}^n}{\text{minimize}} & (\,x'\,)^T w(q,b) \\ \text{subject to} & (\,b' - Ax'\,)_i \,\geq\, 0, \quad \forall\, i \,\in\, \alpha(q,b). \end{array} \tag{7}$$

It is important to point out that while (5) and (6) are equivalent, (4) and (7), which are defined in the same (primal) space of the variable x, are not. This is illustrated by the following simple example.

Example 2. Consider the trivial univariate quadratic program:

$$\begin{array}{ll} \text{minimize} & \tfrac{1}{2}\,x^2 \\ \text{subject to} & x \,\geq\, 0, \end{array}$$

which has a unique optimal solution $x = 0$ and a unique pair $(x,\lambda) = (0,0)$ satisfying the KKT conditions:

$$\begin{array}{l} 0 \,=\, x - \lambda \\ 0 \,\leq\, x \,\perp\, \lambda \,\geq\, 0. \end{array}$$

We have $w(q,b) = 0$. The index set $\alpha(q,b)$ is empty in this case; thus (7) is the *unconstrained* LP: minimize $0 \cdot x$ whose optimal solution is the entire real line. On the other hand, (4) is the *constrained* LP: minimize $0 \cdot x$ subject to $b' - x \geq 0$ whose optimal solution set is the interval $(\,-\infty, b'\,]$. □

The non-equivalence of (4) and (7) is in terms of their optimal solution sets. It is clear, however, that if either one of these two LPs are solvable, then so is the other; moreover, their optimal objective values are equal. In turn, this happens if and only if (6) has an optimal solution. Since $\Lambda(q,b)$ is nonempty for $(q,b) \in \mathcal{R}(M,A)$, it follows that (6) has an optimal solution if and only if it is feasible, or equivalently, if and only if $(\,b'\,)^T\lambda$ is bounded below on $\Lambda(q,b)$. In this case, (4), (5), and (7) all have optimal solutions with the same optimal objective value, denoted $\theta_{\min}(q,b;b')$, as (6). For such a b', we let $D(q,b;b';x)$ denote the set of optimal solutions of the LP (4). Thus, for any $x \in \mathrm{SOL}(q,b;M,A)$, $D(q,b;b';x)$ is nonempty if and only if $(\,b'\,)^T\lambda$ is bounded below on $\Lambda(q,b)$. A noteworthy point is that the latter condition is independent of the solution x.

An element $x' \in D(q,b;b';x)$ is characterized by the existence of a multiplier $\lambda \in \Lambda(q,b)$ such that

$$\begin{array}{l} 0 \,=\, w(q,b) + \displaystyle\sum_{i\in\mathcal{A}(x)} (\, A_{i\bullet}\,)^T\lambda_i \\ 0 \,\leq\, (\,b' - Ax'\,)_i \,\perp\, \lambda_i \,\geq\, 0, \quad \forall\, i \,\in\, \mathcal{A}(x); \end{array}$$

or alternatively, as a vector x' satisfying

$$\begin{array}{l} (\,x'\,)^T w(q,b) \,\leq\, \theta_{\min}(q,b;b') \\ (\,b' - Ax'\,)_i \,\geq\, 0, \quad \forall\, i \,\in\, \mathcal{A}(x). \end{array} \tag{8}$$

For a given b' such that (6) has a finite optimal solution, for an arbitrary q', and for any x in $\mathrm{SOL}(q,b;M,A)$, consider the AVI of finding a vector $x' \in D(q,b;b';x)$ such that

$$(\,y' - x'\,)^T(\,q' + Mx'\,) \,\geq\, 0, \quad \forall\, y' \,\in\, D(q,b;b';x). \tag{9}$$

A vector x' is a solution of the latter AVI, which we call a *directional AVI* and denote by $\text{AVI}'(q,b;q',b';x)$, if and only if scalars η and μ_i for $i \in \mathcal{A}(x)$ exist such that

$$\begin{aligned} 0 \,&=\, q' + Mx' + \eta\, w(q,b) + \sum_{i \in \mathcal{A}(x)} (\, A_{i\bullet} \,)^T \mu_i \\ 0 \,&\leq\, (\, b' - Ax' \,)_i \,\perp\, \mu_i \,\geq\, 0, \quad i \,\in\, \mathcal{A}(x) \\ 0 \,&\leq\, \eta \,\perp\, \theta_{\min}(q,b;b') - (\, x' \,)^T w(q,b) \,\geq\, 0. \end{aligned} \tag{10}$$

We let $\text{SOL}'(q,b;q',b';x)$ denote the (possibly empty) solution set of the directional AVI defined by the tuple $(q,b;q',b';x)$. Note that any solution x' of this directional AVI must satisfy $(\, x' \,)^T w(q,b) = \theta_{\min}(q,b;b')$ as an equation. Moreover, $q' + Mx'$ is a constant on $\text{SOL}'(q,b;q',b';x)$. By Lemma 1 applied to this AVI, it follows that while $\text{SOL}'(q,b;q',b';x)$ may contain multiple elements, there must exist a scalar $\rho_x > 0$ and multipliers μ_i for $i \in \mathcal{A}(x)$ and η_x of (10), all dependent on the tuple $(q,b;q',b';x)$ only, and not on the solution x', such that

$$\eta_x \,\leq\, \rho_x \, [\, \| \, q' + M\text{SOL}'(q,b;q',b';x) \, \| + \| \, b' \, \| \,]$$

The following result shows that the vector $q' + M\text{SOL}'(q,b;q',b';x)$ depends on the tuple $(q,b;q',b')$ only and is independent of the solution $x \in \text{SOL}(q,b;M,A)$.

Lemma 3. Suppose that $\text{SOL}'(q,b;q',b';x)$ is nonempty for some $x \in \text{SOL}(q,b;M,A)$. Then a scalar $\tau_x > 0$ exists such that

$$q' + M\text{SOL}'(q,b;q',b';x) \,=\, \tau^{-1} \, (\, w(q(\tau),b(\tau)) - w(q,b) \,)$$

for all $\tau \in (0, \tau_x)$. Consequently, if $\widehat{x} \in \text{SOL}(q,b;M,A)$ is such that $\text{SOL}'(q,b;q',b';\widehat{x}\,)$ is nonempty, then

$$q' + M\text{SOL}'(q,b;q',b';x) \,=\, q' + M\text{SOL}'(q,b;q',b';\widehat{x}\,).$$

Proof. To prove the first assertion of the lemma, fix an arbitrary solution $x' \in \text{SOL}'(q,b;q',b';x)$. Let τ_x be a positive scalar such that

$$\tau_x \,<\, \frac{1}{\rho_x \, [\, \| \, q' + M\text{SOL}'(q,b;q',b';x) \, \| + \| \, b' \, \| \,]}$$

and $b + \tau b' - A(x + \tau x') \geq 0$ for all $\tau \in (0, \tau_x\,)$. Let $\tau \in (0, \tau_x\,)$ be arbitrary; let

$$\tau_0 \,\triangleq\, \frac{\tau}{1 - \tau\,\eta_x} \,>\, 0, \quad \text{which yields} \quad \tau \,=\, \frac{\tau_0}{1 + \tau_0\,\eta_x}.$$

Let $\lambda \in \Lambda(q,b)$ be an optimal solution of (5). By complementary slackness, we have

$$0 \,\leq\, \lambda_i \,\perp\, (\, b' - Ax' \,)_i \,\geq\, 0, \quad \forall\, i \,\in\, \mathcal{A}(x).$$

Multiplying

$$0 \,=\, q' + Mx' + \eta_x\, w(q,b) + \sum_{i \in \mathcal{A}(x)} (\, A_{i\bullet} \,)^T \mu_i$$

by τ_0, adding to

$$0 \,=\, q + Mx + \sum_{i \in \mathcal{A}(x)} (\, A_{i\bullet} \,)^T \lambda_i,$$

and dividing by $(1+\tau_0\eta_x)$, we obtain:

$$\begin{aligned} 0 &= q + Mx + \frac{\tau_0}{1+\tau_0\,\eta_x}\,(\,q' + Mx'\,) + \sum_{i\in\mathcal{A}(x)}(\,A_{i\bullet}\,)^T\,\widehat{\lambda}_i, \quad \text{where} \quad \widehat{\lambda}_i \triangleq \frac{\lambda_i + \tau_0\,\mu_i}{1+\tau_0\,\eta_x} \\ &= (\,q + \tau\,q'\,) + M(\,x + \tau\,x'\,) + \sum_{i\in\mathcal{A}(x)}(\,A_{i\bullet}\,)^T\,\widehat{\lambda}_i, \end{aligned}$$

Since

$$0 \le \widehat{\lambda}_i \;\perp\; [\,b + \tau b' - A(x+\tau x')\,]_i \ge 0, \quad \forall\, i \in \mathcal{A}(x)$$
$$\text{and} \;\; [\,b + \tau b' - A(x+\tau x')\,]_i \ge 0, \quad \forall\, i \notin \mathcal{A}(x),$$

it follows that $x + \tau x' \in \mathrm{SOL}(q(\tau), b(\tau); M, A)$; thus $q(\tau) + M(x+\tau x') = w(q(\tau), b(\tau))$, which yields

$$q' + M\mathrm{SOL}'(q, b; q', b'; x) \;=\; q' + Mx' \;=\; \tau^{-1}\,(\,w(q(\tau), b(\tau)) - w(q,b)\,).$$

This completes the proof of the first assertion of the lemma. The second assertion follows readily from the first by taking a common scalar $\tau < \min(\,\tau_x, \tau_{\widehat{x}}\,)$. $\square$

Based on the above lemma, we have the following main result for the limit (2).

Theorem 4. Let M be positive semidefinite-plus. Let (q, b, q', b') be such that $(q, b) + \tau(q', b')$ remains in $\mathcal{R}(M, A)$ for all $\tau \ge 0$ sufficiently small. The following three statements are valid.

(a) A solution $x \in \mathrm{SOL}(q, b; M, A)$ exists such that $\mathrm{SOL}'(q, b; q', b'; x)$ is nonempty.

(b) The limit (2), which we denote $w'(q, b; q', b')$, exists and is equal to $q' + M\mathrm{SOL}'(q, b; q', b', \widehat{x}\,)$ for all $\widehat{x} \in \mathrm{SOL}(q, b; M, A)$ such that $\mathrm{SOL}'(q, b; q', b', \widehat{x}\,)$ is nonempty.

(c) For all $(\widehat{q}, \widehat{b}\,) \in \mathcal{R}(M, A)$ sufficiently close to (q, b), we have

$$w(\widehat{q}, \widehat{b}\,) \;=\; w(q, b) + w'(q, b; \widehat{q} - q, \widehat{b} - b). \tag{11}$$

Proof. For each $\tau > 0$ sufficiently small, let (x^τ, λ^τ) satisfy the KKT conditions:

$$\begin{aligned} 0 &= q + \tau\,q' + Mx^\tau + A^T\lambda^\tau \\ 0 &\le b + \tau\,b' - Ax^\tau \;\perp\; \lambda^\tau \ge 0. \end{aligned}$$

Let $\{\tau_k\}$ be an arbitrary sequence of positive scalars converging to zero. Without loss of generality, by working with a subsequence of these scalars if necessary, we may assume that an index subset γ of $\{1, \cdots, m\}$ exists such that for all k,

$$\begin{aligned} 0 &= q + \tau_k\,q' + Mx^k + A^T\lambda^k \\ \left[\,b + \tau_k\,b' - Ax^k\,\right]_\gamma &= 0 \le \lambda^k_\gamma \\ \left[\,b + \tau_k\,b' - Ax^k\,\right]_{\bar{\gamma}} &\ge 0 = \lambda^k_{\bar{\gamma}}, \end{aligned}$$

where $x^k \triangleq x^{\tau_k}$ and $\lambda^k \triangleq \lambda^{\tau_k}$. Consequently, there exists a pair $(\bar{x}, \bar{\lambda})$ such that

$$\begin{aligned} 0 &= q + M\bar{x} + A^T\bar{\lambda} \\ [\,b - A\bar{x}\,]_\gamma &= 0 \le \bar{\lambda}_\gamma \\ [\,b - A\bar{x}\,]_{\bar{\gamma}} &\ge 0 = \bar{\lambda}_{\bar{\gamma}}. \end{aligned}$$

Thus $\bar{x} \in \mathrm{SOL}(q, b; M, A)$ and $\bar{\lambda} \in \Lambda(\bar{x}; q, b) = \Lambda(q, b)$. Note that we did not claim that $(\bar{x}, \bar{\lambda})$ is the limit of the sequence $\{(x^k, \lambda^k)\}$; indeed the latter sequence is not necessarily bounded. We clearly have $\gamma \subseteq \mathcal{A}(\bar{x})$ and

$$\left[b' - A \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right]_\gamma = 0;$$

furthermore,

$$\left[b' - A \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right]_i \geq 0, \quad \forall\, i \in \mathcal{A}(\bar{x}) \setminus \gamma.$$

Consequently,

$$0 = q + M\bar{x} + \sum_{i \in \mathcal{A}(\bar{x})} (A_{i\bullet})^T \bar{\lambda}_i$$

$$\left[b' - A \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right]_i = 0 \leq \bar{\lambda}_i, \quad \forall\, i \in \gamma$$

$$\left[b' - A \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right]_i \geq 0 = \bar{\lambda}_i, \quad \forall\, i \in \mathcal{A}(\bar{x}) \setminus \gamma,$$

which implies that $\tau_k^{-1}(x^k - \bar{x}) \in D(q, b; b'; \bar{x})$ for all k. We have

$$\begin{aligned} 0 &= q' + M \left(\frac{x^k - \bar{x}}{\tau_k} \right) + \sum_{i \in \gamma} (A_{i\bullet})^T \left(\frac{\lambda_i^k - \bar{\lambda}_i}{\tau_k} \right) \\ &= q' + M \left(\frac{x^k - \bar{x}}{\tau_k} \right) + \sum_{i \in \gamma} (A_{i\bullet})^T \left(\frac{\lambda_i^k - \bar{\lambda}_i}{\tau_k} \right) + \tau_k^{-1} \left[q + M\bar{x} + A^T\bar{\lambda} \right] \\ &= q' + M \left(\frac{x^k - \bar{x}}{\tau_k} \right) + \tau_k^{-1} w(q, b) + \sum_{i \in \gamma} (A_{i\bullet})^T \left(\frac{\lambda_i^k}{\tau_k} \right). \end{aligned}$$

Let $x' \in D(q, b; b'; \bar{x})$ be arbitrary. Then

$$(x')^T w(q, b) = \left[\frac{x^k - \bar{x}}{\tau_k} \right]^T w(q, b) \quad \text{and} \quad (b' - Ax')_\gamma \geq 0 = \left[b' - A \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right]_\gamma.$$

Hence,

$$\left[x' - \frac{x^k - \bar{x}}{\tau_k} \right]^T \left[q' + M \left(\frac{x^k - \bar{x}}{\tau_k} \right) \right] \geq 0,$$

establishing that $\tau_k^{-1}(x^k - \bar{x}) \in \mathrm{SOL}'(q, b; q', b'; \bar{x})$. Thus part (a) of the theorem holds.

With $(q^k, b^k) \triangleq (q, b) + \tau_k(q', b')$, we have, for all k,

$$\begin{aligned} w(q^k, b^k) &= q^k + Mx^k = q + M\bar{x} + \tau_k \left[q' + \tau_k^{-1} M(x^k - \bar{x}) \right] \\ &= w(q, b) + \tau_k \left[q' + M\mathrm{SOL}'(q, b; q', b'; \bar{x}) \right]. \end{aligned}$$

This establishes part (b) with $\widehat{x} = \bar{x}$. For another $\widehat{x}$, part (b) follows from Lemma 3. To show that (11) holds for all $(\widehat{q}, \widehat{b}) \in \mathcal{R}(M, A)$ sufficiently close to (q, b), we assume by way of contradiction that a sequence $\{(\widehat{q}^{\,k}, \widehat{b}^{\,k})\} \subset \mathcal{R}(M, A)$ converging to (q, b) exists such that for each k,

$$w(\widehat{q}^{\,k}, \widehat{b}^{\,k}) \neq w(q, b) + w'(q, b; \widehat{q}^{\,k} - q, \widehat{b}^{\,k} - b).$$

Proceeding as in the above proof using the pair $(\widehat{q}^{\,k}, \widehat{b}^{\,k})$ instead of the pair $(q + \tau_k q', b + \tau_k b')$, we may deduce the existence of $x^k \in \mathrm{SOL}(\widehat{q}^{\,k}, \widehat{b}^{\,k}; M, A)$ for each k and a vector $\bar{x} \in \mathrm{SOL}(q, b; M, A)$ such that $x^k - \bar{x}$ belongs to $D(q, b; \widehat{b}^k - b; \bar{x})$ for each k; moreover there exists a multiplier $\lambda^k \in \Lambda(x^k; q^k, b^k)$ and an index set $\gamma \subseteq \mathcal{A}(\bar{x})$ such that for all k,

$$\widehat{q}^{\,k} - q + M(\, x^k - \bar{x}\,) + w(q, b) + \sum_{i \in \gamma} (\, A_{i\bullet}\,)^T \lambda_i^k \; = \; 0$$

$$(\,\widehat{b}^k - b - Ax'\,)_\gamma \; \geq \; 0 \; = \; \left[\widehat{b}^k - b - A(\, x^k - \bar{x}\,)\right]_\gamma, \quad \forall\, x' \in \, D(q, b; \widehat{b}^k - b; \bar{x}).$$

This shows that $x^k - \bar{x} \in \mathrm{SOL}'(q, b; \widehat{q}^{\,k} - q, \widehat{b}^{\,k} - b; \bar{x})$; hence $w(\widehat{q}^{\,k}, \widehat{b}^{\,k}) = w(q, b) + w'(q, b; \widehat{q}^{\,k} - q, \widehat{b}^{\,k} - b)$, which is a contradiction. $\square$

We remark that if M is written as $M = EGE^T$ for some positive definite matrix G, then $E^T\mathrm{SOL}(q, b)$ is also a single-valued piecewise linear function on $\mathcal{R}(M, A)$; moreover, for all $(\widehat{q}, \widehat{b}\,) \in \mathcal{R}(M, A)$ sufficiently close to (q, b),

$$E^T\mathrm{SOL}(\widehat{q}, \widehat{b}; M, A\,) \; = \; E^T\mathrm{SOL}(q, b; M, A) + E^T\mathrm{SOL}'(q, b; \widehat{q} - q, \widehat{b} - b).$$

This equality follows from the last part of the proof of Theorem 4 and provides the basis for a subsequent generalization to a class of nonlinear VIs.

2.1 Choice of solutions

As it stands, Theorem 4 establishes the well-definedness and characterization of the directional derivative $w'(q, b; q', b')$ under the reasonable condition that $(q, b) + \tau(q', b')$ remains in $\mathcal{R}(M, A)$ for all $\tau \geq 0$ sufficiently small. The theorem does not provide a constructive way to calculate such a derivative. Ideally, one would want to be able to deduce that $\mathrm{SOL}'(q, b; q', b', x)$ is nonempty for *all* $x \in \mathrm{SOL}(q, b; M, A)$, which would imply that one could use *any* solution of the original AVI $(q, b; M, A)$ for the calculation of $w'(q, b; q', b')$. In what follows, we establish a necessary and sufficient condition on the pair (q', b') in order for this to be true. For this purpose, we define the index set:

$$\gamma(q, b) \; \triangleq \; \{\, i \,:\, (\, b - Ax\,)_i \; = \; 0 \text{ for all } x \, \in \, \mathrm{SOL}(q, b; M, A)\,\}.$$

Note that $\alpha(q, b) \subseteq \gamma(q, b) \subseteq \mathcal{A}(x)$ for all $x \in \mathrm{SOL}(q, b; M, A)$. Unlike $\alpha(q, b)$, the index set $\gamma(q, b)$ is defined by the (primal) variable x only. An important property of the set $\gamma(q, b)$ is that there must exist $\bar{x} \in \mathrm{SOL}(q, b; M, A)$ such that $(b - A\bar{x})_i > 0$ for *all* $i \notin \gamma(q, b)$, i.e., such that $\gamma(q, b) = \mathcal{A}(\bar{x})$. This is because for each $i \notin \gamma(q, b)$, a solution $x^i \in \mathrm{SOL}(q, b; M, A)$ exists such that $(b - Ax^i)_i > 0$; we can simply let $\bar{x}$ be the average of these solutions x^i.

Theorem 5. Let M be positive semidefinite-plus and $(q, b) \in \mathcal{R}(M, A)$. A necessary and sufficient condition for $\mathrm{SOL}'(q, b; q', b', x)$ to be nonempty for all $x \in \mathrm{SOL}(q, b; M, A)$ is that $(b')^T\lambda$ is bounded below on $\Lambda(q, b)$ and the following implication holds:

$$\left.\begin{array}{rcl} My & = & 0 \\ q^T y & \leq & 0 \\ A_{i\bullet} y & \leq & 0, \quad \forall\, i \,\in\, \gamma(q, b) \end{array}\right\} \Rightarrow \; (\, q'\,)^T y \, \geq \, 0. \tag{12}$$

Moreover, if this condition holds, then $w'(q, b; q', b')$ is well-defined and equal to $\mathrm{SOL}'(q, b; q', b', x)$ for all $x \in \mathrm{SOL}(q, b; M, A)$.

Proof. Sufficiency. Suppose that $(b')^T\lambda$ is bounded below on $\Lambda(q,b)$ and (12) holds. The set $D(q,b;b')$ is then not empty. Let $x \in \mathrm{SOL}(q,b;M,A)$ be arbitrary. Since the KKT system of AVI $(q,b;q',b',x)$ is a monotone mixed linear complementarity problem, this AVI is solvable if and only if its KKT system is feasible, i.e., if and only if there exists a tuple (x',η,μ_i) such that

$$\begin{aligned} 0 &= q' + Mx' + \eta\, w(q,b) + \sum_{i\in\mathcal{A}(x)} (\,A_{i\bullet}\,)^T\mu_i \\ 0 &\le (\,b' - Ax'\,)_i, \quad \mu_i \ge 0, \quad i \in \mathcal{A}(x) \\ 0 &\le \eta, \quad \theta_{\min}(q,b;b') - (\,x'\,)^T w(q,b) \ge 0. \end{aligned}$$

In turn, the latter is true if and only if the implication below holds:

$$\left.\begin{aligned} M^Ty - \sum_{i\in\mathcal{A}(x)} (\,A_{i\bullet}\,)^T\xi_i - \sigma\, w(q,b) &= 0 \\ y^T w(q,b) \le 0, \quad \sigma &\ge 0 \\ A_{i\bullet}y \le 0, \quad \xi_i \ge 0, \quad \forall\, i &\in \mathcal{A}(x) \end{aligned}\right\} \Rightarrow (\,q'\,)^Ty + \sum_{i\in\mathcal{A}(x)} b'_i\xi_i + \sigma\,\theta_{\min}(q,b;b') \ge 0. \tag{13}$$

Let (y,ξ_i,σ) satisfy the left-hand conditions. Multiplying the first equation by y^T, we deduce $y^TM^Ty = 0$, which implies, by the positive semidefinite-plus property of M, $My = M^Ty = 0$. Thus $y^Tw(q,b) = y^Tq$. Since $\gamma(q,b) \subseteq \mathcal{A}(x)$, it follows that y satisfies the left-hand conditions in the implication (12). Thus, $(q')^Ty \ge 0$. Moreover, we have

$$\begin{aligned} \sum_{i\in\mathcal{A}(x)} (\,A_{i\bullet}\,)^T\xi_i + \sigma\, w(q,b) &= 0 \\ \xi_i \ge 0, \quad \forall\, i \in \mathcal{A}(x), \quad \text{and} \quad \sigma &\ge 0. \end{aligned}$$

Since the set $D(q,b;b')$ is nonempty, by the inequality representation (8) of elements of this set, it follows that

$$\sum_{i\in\mathcal{A}(x)} b'_i\xi_i + \sigma\,\theta_{\min}(q,b;b') \ge 0.$$

Thus the implication (13) holds, completing the sufficiency proof of the theorem.

Necessity. Conversely, suppose that $\mathrm{SOL}'(q,b;q',b',x)$ is not empty for all $x \in \mathrm{SOL}(q,b;M,A)$. It suffices to prove the implication (12). The set $\mathrm{SOL}'(q,b;q',b',\bar{x}) \neq \emptyset$ for the solution $\bar{x} \in \mathrm{SOL}(q,b;M,A)$ such that $\mathcal{A}(\bar{x}) = \gamma(q,b)$. Let y be a vector satisfying the left-hand conditions in (12). Then, with $x \triangleq \bar{x}$, $\xi_i = \sigma = 0$, the left-hand conditions in (13) are satisfied. Hence by the latter implication, which is equivalent to the nonemptiness of $\mathrm{SOL}'(q,b;q',b',\bar{x})$, we deduce $(q')^Ty \ge 0$ as desired.

To complete the proof of the theorem, it remains to show that for all $\tau > 0$ sufficiently small, $(q+\tau q', b+\tau b') \in \mathcal{R}(M,A)$. Once this is shown, the last assertion of the theorem follows from Theorem 4. By the above proof, the $\mathrm{AVI}'(q,b;q',b';\bar{x})$ has a solution for some $\bar{x} \in \mathrm{SOL}(q,b;M,A)$. By Lemma 3, it follows that the AVI $(q+\tau q', b+\tau b'; M,A)$ has a solution for all $\tau > 0$ as desired. □

Theorem 5 gives a necessary and sufficient condition for the $\mathrm{AVI}'(q,b;q',b',x)$ to be solvable for all $x \in \mathrm{SOL}(q,b;M,A)$. For completeness, we present a necessary and sufficient condition for the $\mathrm{AVI}'(q,b;q',b',x)$ to be solvable for some $x \in \mathrm{SOL}(q,b;M,A)$. This result is simply a combination of Lemma 3 and Theorem 4; we omit its proof.

Theorem 6. Let M be positive semidefinite-plus and $(q,b) \in \mathcal{R}(M,A)$. A necessary and sufficient condition for $\mathrm{SOL}'(q,b;q',b',x) \neq \emptyset$ for some $x \in \mathrm{SOL}(q,b;M,A)$ is that $(q+\tau q', b+\tau b') \in \mathcal{R}(M,A)$ for all $\tau > 0$ sufficiently small. □

Theorem 6 still does not provide a constructive way to identify a desired $x \in \text{SOL}(q, b; M, A)$ such that $\text{SOL}'(q, b; q', b', x) \neq \emptyset$. In what follows, we briefly describe one situation where such an x can be identified by parametric principal pivoting [9, Section 4.5] and the directional derivative $w'(q, b; q', b')$ readily computed. We begin with the parametric KKT system with $\tau \geq 0$ as the parameter:

$$\begin{aligned} 0 &= q + \tau\, q' + Mx + A^T\lambda \\ 0 \leq \lambda \perp b &+ \tau\, b' - Ax \geq 0, \end{aligned}$$

which is a parametric mixed LCP. Apply the subspace-removal technique described in [10, Exercise 1.8.10], which involves only linear algebraic operations, to reduce the above mixed LCP to an equivalent parametric LCP:

$$\begin{aligned} 0 \leq \widehat{x} &\perp \widehat{q} + \tau\,\widehat{q}\,' + \widehat{M}\,\widehat{x} + \widehat{A}^T\widehat{\lambda} \geq 0 \\ 0 \leq \widehat{\lambda} &\perp \widehat{b} + \tau\,\widehat{b}' - \widehat{A}\,\widehat{x} \geq 0, \end{aligned} \tag{14}$$

where the matrix $\widehat{M}$ remains positive semidefinite-plus. [The cited exercise does not assert the plus-property of the reduced system; but this is not difficult to show based on the description of $\widehat{M}$ contained therein.] Originally due to [4, 5] and justified under the assumption that M is nonsingular on the null space of A, this process removes the equation $0 = q + \tau q' + Mx + A^T\lambda$, solving out the free variable x, and pivots the complementarity condition $0 \leq \lambda \perp b + \tau b' - Ax \geq 0$ into the above form (14). From a solution of (14), we can recover the variable x by a simple backward substitution.

For simplicity, write (14) as

$$0 \leq u \perp r + \tau\, s + Ru \geq 0,$$

where R is a positive semidefinite matrix. Apply parametric principal pivoting, which in principle requires a nondegeneracy assumption, to the latter parametric LCP, obtaining a scalar $\bar{\tau} > 0$ and an index set γ with complement $\bar{\gamma}$ such that for all $\tau \in [0, \bar{\tau}]$

$$\begin{aligned} u_\gamma &\geq 0 = [\, r + \tau\, s + Ru\,]_\gamma \\ u_{\bar{\gamma}} &= 0 \leq [\, r + \tau\, s + Ru\,]_{\bar{\gamma}}. \end{aligned}$$

From the latter system, we may recover a solution $x(\tau)$ to the AVI $(q(\tau), b(\tau); M, A)$ for all $\tau \in [0, \bar{\tau}]$. By the proof of Theorem 4, $\text{SOL}'(q, b; q', b', x(0)) \neq \emptyset$; moreover, the desired directional derivative $w'(q, b; q', b')$ is given by $q' + \tau^{-1}M(x(\tau) - x(0))$.

3 An Application: Support Vector Machine

Consider the following convex piecewise quadratic minimization problem derived from the support vector machine (SVM) model in data mining: given data $\left\{ (x^i, y_i)_{i=1}^m \right\} \subset \mathbb{R}^{n+1}$ and parameters $C > 0$ and $\varepsilon \geq 0$,

$$\underset{(w,b)\in\mathbb{R}^{n+1}}{\text{minimize}} \quad C \sum_{i=1}^m \max\left\{ \left| w^T x^i + b - y_i \right| - \varepsilon,\, 0 \right\} + \tfrac{1}{2}\, w^T w. \tag{15}$$

In what follows, we call w the support vector and b the bias. By introducing the auxiliary variable: $e_i \triangleq \max\left\{ \left| w^T x^i + b - y_i \right| - \varepsilon,\, 0 \right\}$, it is easily seen that (15) is equivalent to a convex (but not strictly convex) quadratic program in the variables $(w, e, b) \in \mathbb{R}^{n+m+1}$:

$$\begin{aligned} \theta_{\text{opt}}(C, \varepsilon) \triangleq \underset{(w,e,b)\in\mathbb{R}^{n+m+1}}{\text{minimize}} & \quad C \sum_{i=1}^m e_i + \tfrac{1}{2}\, w^T w \\ \text{subject to} & \quad \left\{ \begin{array}{ll} e_i \geq & w^T x^i + b - y_i - \varepsilon \\ e_i \geq & -w^T x^i - b + y_i - \varepsilon \\ e_i \geq 0 & \end{array} \right\} \quad i = 1, \cdots, m. \end{aligned} \tag{16}$$

For every pair (C, ε), the above QP has a unique optimal w-solution, which we denote $w(C, \varepsilon)$; nevertheless, no assertion is made about the uniqueness of the optimal bias b.

For a given pair (C, ε), the KKT conditions of (15) are:

$$
\begin{aligned}
w \;&=\; \sum_{i=1}^{m} (\, \lambda_i^- - \lambda_i^+ \,)\, x^i \\
0 \;&=\; \sum_{i=1}^{m} (\, \lambda_i^- - \lambda_i^+ \,) \\
&\left\{ \begin{array}{l}
0 \,\leq\, \lambda_i^+ \,\perp\, e_i - w^T x^i - b + y_i + \varepsilon \,\geq\, 0 \\
0 \,\leq\, \lambda_i^- \,\perp\, e_i + w^T x^i + b - y_i + \varepsilon \,\geq\, 0 \\
0 \,\leq\, e_i \,\perp\, C - \lambda_i^+ - \lambda_i^- \,\geq\, 0
\end{array} \right\} \quad i \,=\, 1, \cdots, m.
\end{aligned}
$$

Upon eliminating w using the first equation, we obtain a mixed LCP in the variables $\left(b,\, \{\lambda_i^{\pm}, e_i\}_{i=1}^m\right)$ (where the bias b is the only free variable) with parameters (C, ε):

$$
\left\{ \begin{array}{l}
0 \,=\, \displaystyle\sum_{i=1}^{m} (\, \lambda_i^+ - \lambda_i^- \,) \\
0 \,\leq\, \lambda_i^+ \,\perp\, e_i + \displaystyle\sum_{j=1}^{m} \left[\, (\, x^j \,)^T x^i \,\right] (\, \lambda_j^+ - \lambda_j^- \,) - b + y_i + \varepsilon \,\geq\, 0 \\
0 \,\leq\, \lambda_i^- \,\perp\, e_i - \displaystyle\sum_{j=1}^{m} \left[\, (\, x^j \,)^T x^i \,\right] (\, \lambda_j^+ - \lambda_j^- \,) + b - y_i + \varepsilon \,\geq\, 0 \\
0 \,\leq\, e_i \,\perp\, C - \lambda_i^+ - \lambda_i^- \,\geq\, 0
\end{array} \right\} \quad i \,=\, 1, \cdots, m. \tag{17}
$$

The optimal solution $w(C, \varepsilon)$ is a piecewise linear function of $(C, \varepsilon) \geq 0$. In what follows, we use the general AVI theory developed in the last section to characterize the directional derivative

$$
w'(C, \varepsilon; C', \varepsilon') \;=\; \lim_{\tau \downarrow 0} \frac{w(C + \tau C', \varepsilon + \tau \varepsilon') - w(C, \varepsilon)}{\tau},
$$

at a pair $(C, \varepsilon) > 0$ and along a direction (C', ε'). Since $\varepsilon > 0$, any solution to the above KKT conditions must satisfy $\min(\lambda_i^+, \lambda_i^-) = 0$. Unlike the nominal pair (C, ε), which is positive, the direction pair $(\, C', \varepsilon' \,)$ is not necessarily so.

Following the general theory, we define several index sets in terms of the multipliers:

$$
\begin{aligned}
\alpha^+(C, \varepsilon) \;&\triangleq\; \left\{\, i \,:\, \exists \left(b,\, \left(\lambda_i^{\pm}, e_i\right)_{i=1}^m\right) \text{ satisfying (17) such that } \lambda_i^+ \,>\, 0 \,\right\} \\
\alpha^-(C, \varepsilon) \;&\triangleq\; \left\{\, i \,:\, \exists \left(b,\, \left(\lambda_i^{\pm}, e_i\right)_{i=1}^m\right) \text{ satisfying (17) such that } \lambda_i^- \,>\, 0 \,\right\} \\
\alpha^e(C, \varepsilon) \;&\triangleq\; \left\{\, i \,:\, \exists \left(b,\, \left(\lambda_i^{\pm}, e_i\right)_{i=1}^m\right) \text{ satisfying (17) such that } C - \lambda_i^+ - \lambda_i^- \,>\, 0 \,\right\}.
\end{aligned}
$$

Note that $\alpha^e(C, \varepsilon) \cup \alpha^+(C, \varepsilon) \cup \alpha^-(C, \varepsilon) = \{1, \cdots, m\}$ because if $i \notin \alpha^e(C, \varepsilon)$, then $C = \lambda_i^+ + \lambda_i^- > 0$ for all solutions $\left(b,\, \left(\lambda_i^{\pm}, e_i\right)_{i=1}^m\right)$ of (17); so either λ_i^+ or λ_i^- must be positive because $C > 0$. In terms of the above defined index sets, the set of optimal multipliers of (17) is

$$\Lambda(C,\varepsilon) \triangleq \left\{ \begin{array}{rclr} \lambda^{\pm} \geq 0 \;:\; w(C,\varepsilon) & = & \displaystyle\sum_{i=1}^{m} (\,\lambda_i^- - \lambda_i^+\,)\,x^i & \\ 0 & = & \displaystyle\sum_{i=1}^{m} (\,\lambda_i^- - \lambda_i^+\,) & \\ 0 & = & \lambda_i^+, & i \notin \alpha^+(C,\varepsilon) \\ 0 & = & \lambda_i^-, & i \notin \alpha^-(C,\varepsilon) \\ C & \geq & \lambda_i^+ + \lambda_i^-, & i \in \alpha^e(C,\varepsilon) \\ C & = & \lambda_i^+ + \lambda_i^-, & i \notin \alpha^e(C,\varepsilon) \end{array} \right\},$$

which is a bounded polyhedron. We can similarly define several primal index sets:

$$\begin{array}{rcl} \gamma^+(C,\varepsilon) & \triangleq & \left\{\, i \,:\, w(C,\varepsilon)^T x^i + b - y_i \,\geq\, \varepsilon \text{ for all optimal bias } b \,\right\} \\ \gamma^-(C,\varepsilon) & \triangleq & \left\{\, i \,:\, w(C,\varepsilon)^T x^i + b - y_i \,\leq\, -\varepsilon \text{ for all optimal bias } b \,\right\} \\ \gamma^e(C,\varepsilon) & \triangleq & \left\{\, i \,:\, |\, w(C,\varepsilon)^T x^i + b - y_i \,| \,\leq\, \varepsilon \text{ for all optimal bias } b \,\right\}, \end{array}$$

and the following bias-dependent index sets: for an optimal b,

$$\begin{array}{rcl} \alpha^+(C,\varepsilon;b) & \triangleq & \left\{\, i \,:\, w(C,\varepsilon)^T x^i + b - y_i \,\geq\, \varepsilon \,\right\} \\ \alpha^-(C,\varepsilon;b) & \triangleq & \left\{\, i \,:\, w(C,\varepsilon)^T x^i + b - y_i \,\leq\, -\varepsilon \,\right\} \\ \alpha^e(C,\varepsilon;b) & \triangleq & \left\{\, i \,:\, |\, w(C,\varepsilon)^T x^i + b - y_i \,| \,\leq\, \varepsilon \,\right\}. \end{array}$$

In terms of the latter index sets, we can define the optimal bias dependent multiplier sets:

$$\Lambda(C,\varepsilon;b) \triangleq \left\{ \begin{array}{rclr} \lambda^{\pm} \geq 0 \;:\; w(C,\varepsilon) & = & \displaystyle\sum_{i=1}^{m} (\,\lambda_i^- - \lambda_i^+\,)\,x^i & \\ 0 & = & \displaystyle\sum_{i=1}^{m} (\,\lambda_i^- - \lambda_i^+\,) & \\ 0 & = & \lambda_i^+, & i \notin \alpha^+(C,\varepsilon;b) \\ 0 & = & \lambda_i^-, & i \notin \alpha^-(C,\varepsilon;b) \\ C & \geq & \lambda_i^+ + \lambda_i^-, & i \in \alpha^e(C,\varepsilon;b) \\ C & = & \lambda_i^+ + \lambda_i^-, & i \notin \alpha^e(C,\varepsilon;b) \end{array} \right\}.$$

Note that $\alpha^e(C,\varepsilon;b) \cup \alpha^+(C,\varepsilon;b) \cup \alpha^-(C,\varepsilon;b) = \{1, \cdots, m\}$. The linear program (4) in this context is

$$\begin{array}{ll} \underset{(w',e',b') \in \mathbb{R}^{n+m+1}}{\text{minimize}} & C \displaystyle\sum_{i=1}^{m} e_i' + (\,w'\,)^T w(C,\varepsilon) \\ \text{subject to} & e_i' \geq (\,w'\,)^T x^i + b' - \varepsilon', \quad i \in \alpha^+(C,\varepsilon;b) \\ & e_i' \geq -(\,w'\,)^T x^i - b' - \varepsilon', \quad i \in \alpha^-(C,\varepsilon;b) \\ & e_i' \geq 0, \quad i \in \alpha^e(C,\varepsilon;b), \end{array} \tag{18}$$

which is clearly feasible, and whose dual is

$$\underset{\lambda\in\Lambda(C,\varepsilon;b)}{\text{minimize}} \quad \varepsilon' \left[\sum_{i\in\alpha^+(C,\varepsilon;b)} \lambda_i^+ + \sum_{i\in\alpha^-(C,\varepsilon;b)} \lambda_i^- \right]$$

$$= \begin{cases} \varepsilon' \underset{\lambda\in\Lambda(C,\varepsilon;b)}{\text{minimize}} \left[\displaystyle\sum_{i\in\alpha^+(C,\varepsilon;b)} \lambda_i^+ + \sum_{i\in\alpha^-(C,\varepsilon;b)} \lambda_i^- \right] & \text{if } \varepsilon' \geq 0 \\ \\ \varepsilon' \underset{\lambda\in\Lambda(C,\varepsilon;b)}{\text{maximize}} \left[\displaystyle\sum_{i\in\alpha^+(C,\varepsilon;b)} \lambda_i^+ + \sum_{i\in\alpha^-(C,\varepsilon;b)} \lambda_i^- \right] & \text{if } \varepsilon' \leq 0. \end{cases}$$

Let $D(C,\varepsilon;\varepsilon';b)$ denote the optimal solution set of (18), which must be nonempty because $\Lambda(C,\varepsilon;b) = \Lambda(C;\varepsilon)$ is bounded. For any nonzero ε', this set is equal to the optimal solution set of a constant LP scaled by ε'; specifically, for any $\varepsilon' \neq 0$, $D(C,\varepsilon;\varepsilon';b) = |\,\varepsilon'\,|\, D^{\pm}(C,\varepsilon;b)$, where the $\pm$ sign is determined by the sign of ε' and

$$\begin{aligned} D^+(C,\varepsilon;b) \quad \triangleq \quad & \underset{(w',e',b')\in\mathbb{R}^{n+m+1}}{\text{argmin}} \quad C\sum_{i=1}^m e_i' + (\,w'\,)^T w(C,\varepsilon) \\ & \text{subject to} \quad e_i' \geq \ \ (\,w'\,)^T x^i + b' - 1, \quad i \in \alpha^+(C,\varepsilon;b) \\ & \qquad\qquad\quad\ e_i' \geq -(\,w'\,)^T x^i - b' - 1, \quad i \in \alpha^-(C,\varepsilon;b) \\ & \qquad\qquad\quad\ e_i' \geq \ \ 0, \quad i \in \alpha^e(C,\varepsilon;b), \end{aligned}$$

and

$$\begin{aligned} D^-(C,\varepsilon;b) \quad \triangleq \quad & \underset{(w',e',b')\in\mathbb{R}^{n+m+1}}{\text{argmin}} \quad C\sum_{i=1}^m e_i' + (\,w'\,)^T w(C,\varepsilon) \\ & \text{subject to} \quad e_i' \geq \ \ (\,w'\,)^T x^i + b' + 1, \quad i \in \alpha^+(C,\varepsilon;b) \\ & \qquad\qquad\quad\ e_i' \geq -(\,w'\,)^T x^i - b' + 1, \quad i \in \alpha^-(C,\varepsilon;b) \\ & \qquad\qquad\quad\ e_i' \geq \ \ 0, \quad i \in \alpha^e(C,\varepsilon;b). \end{aligned}$$

By Theorem 4, for all $(\widehat{C},\widehat{\varepsilon}\,)$ sufficiently close to (C,ε), we have

$$w(\widehat{C},\widehat{\varepsilon}\,) \;=\; w(C,\varepsilon) + w'(C,\varepsilon;\widehat{C}-C,\widehat{\varepsilon}-\varepsilon),$$

where $w'(C,\varepsilon;\widehat{C}-C,\widehat{\varepsilon}-\varepsilon)$ is the unique w' in the optimal solution of the quadratic program:

$$\underset{(w',e',b')\in D(C,\varepsilon;\widehat{\varepsilon}-\varepsilon;b)}{\text{minimize}} \quad (\,\widehat{C}-C\,)\sum_{i=1}^m e_i' + \tfrac{1}{2}\,(\,w'\,)^T w',$$

for any optimal bias b for which the above quadratic program is solvable. In turn, the latter directional quadratic program is equivalent to, for $\widehat{\varepsilon} \neq \varepsilon$,

$$\begin{aligned} & \underset{w',e',b'}{\text{minimize}} \quad \frac{\widehat{C}-C}{|\,\widehat{\varepsilon}-\varepsilon\,|}\sum_{i=1}^m e_i' + \tfrac{1}{2}\,(\,w'\,)^T w' \\ & \text{subject to} \quad (\,w',e',b'\,) \in D^{\pm}(C,\varepsilon;b). \end{aligned}$$

Consequently, by solving 2 parametric QPs:

$$\begin{array}{ll} \underset{w',e',b'}{\text{minimize}} & \tau \sum_{i=1}^{m} e_i' + \frac{1}{2} (w')^T w' \\ \text{subject to} & (w', e', b') \in D^{\pm}(C, \varepsilon; b) \end{array}$$

for all nonzero values of τ, we can determine the directional derivatives $w'(C,\varepsilon;\widehat{C}-C,\widehat{\varepsilon}-\varepsilon)$ for all pairs $(\widehat{C},\widehat{\varepsilon})$ with $\widehat{\varepsilon} \neq \varepsilon$.

We close this section by pointing out that the uniqueness of the optimal bias b has been studied in [3] wherein necessary and sufficient conditions are derived for such uniqueness to hold. Specifically, define the four index sets corresponding to an optimal bias b:

$$\begin{array}{llll} \mathcal{N}_1 \triangleq \{ i : y_i - w(C,\varepsilon)^T x^i - b > \varepsilon \}, & \mathcal{N}_3 \triangleq \{ i : y_i - w(C,\varepsilon)^T x^i - b = -\varepsilon \} \\ \mathcal{N}_2 \triangleq \{ i : y_i - w(C,\varepsilon)^T x^i - b = \varepsilon \}, & \mathcal{N}_4 \triangleq \{ i : y_i - w(C,\varepsilon)^T x^i - b < -\varepsilon \}. \end{array}$$

The cited references shows that the optimal b in (15) is unique if and only if $|\mathcal{N}_1 \cup \mathcal{N}_2| \neq |\mathcal{N}_4|$ and $|\mathcal{N}_3 \cup \mathcal{N}_4| \neq |\mathcal{N}_1|$.

4 Nonlinear Extensions

We consider a nonlinear extension of the AVI $(q,b;M,A)$ by replacing the monotone-plus affine map $x \mapsto q + Mx$ by a strongly monotone composite map $x \mapsto q + EG(E^T x)$ [10, Definition 2.3.9(d)], where E is an arbitrary $n \times \ell$ matrix and $G : \Omega \subseteq \mathbb{R}^\ell \to \mathbb{R}^\ell$ is a continuously differentiable, strongly monotone, and Lipschitz continuous function defined on an open convex set Ω containing the range of E^T. Given the tuple $(q,b;E,G,A)$, we formally define the variational inequality, denoted VI $(q,b;E,G,A)$, as the problem of finding a vector $x \in K \triangleq \{x : Ax \leq b\}$ such that

$$(y - x)^T [q + EG(E^T x)] \geq 0, \quad \forall y \in K.$$

To ensure that the solution set of this VI, which we denote $\text{SOL}(q,b;E,G,A)$, is nonempty, we assume throughout this section that $K_\infty \cap \ker(E^T) = \{0\}$, where K_∞ is the recession cone of K and $\ker(E^T)$ denotes the kernel of the matrix E^T. Under this assumption, it follows that $\text{SOL}(q,b;E,G,A)$ is a nonempty compact polyhedron for every $q \in \mathbb{R}^n$ and every b for which the set K is nonempty—the nonemptiness of $\text{SOL}(q,b;E,G,A)$ is by [19, Proposition 6.3], the boundedness is by [10, Theorem 2.3.16], and the polyhedrality is by [10, Corollary 2.3.13]; indeed, the last corollary implies that $E^T\text{SOL}(q,b;E,G,A)$ is a singleton, which we denote $v(q,b)$, and

$$\text{SOL}(q,b;E,G,A) = (E^T)^{-1}(v(q,b)) \cap \operatorname{argmin} \{ x^T [q + EG(v(q,b)] : x \in K \}.$$

Our goal in this section is to derive a first-order sensitivity property of the function $v(q,b)$. It is known [10, Corollary 6.2.11] that for fixed (q,b), there exist positive constants c and δ such that

$$\| q - q' \| \leq \delta \Rightarrow \| v(q,b) - v(q',b) \| \leq c \| q - q' \|.$$

This result pertains to small perturbations of q with b unchanged. Unlike the affine case, even the joint continuity of $v(q,b)$ in its arguments is not known in the literature. The analysis herein aims at establishing the locally Lipschitz continuity and the directional differentiability of this solution function in (q,b) jointly and at characterizing the directional derivatives. Our first step is to derive some bounds for the solutions and multipliers of the VI $(q,b;E,G,A)$. For this purpose, let $\varsigma > 0$ be a strong monotonicity modulus of G on Ω, i.e.,

$$(y - y')^T (G(y) - G(y')) \geq \varsigma \| y - y' \|^2, \quad \forall y, y' \in \Omega;$$

also let $L_G > 0$ be a Lipschitz modulus of G on Ω, i.e.,

$$\| G(y) - G(y') \| \leq L_G \| y - y' \|, \quad \forall\, y, y' \in \Omega.$$

The KKT conditions of the VI $(q, b; E, G, A)$ is:

$$\begin{aligned} 0 &= q + EG(E^T x) + A^T \lambda \\ 0 &\leq \lambda \perp b - Ax \geq 0. \end{aligned} \tag{19}$$

As before, define the index set

$$\alpha(q,b) \triangleq \{ i : \exists \text{ a solution } (x, \lambda) \text{ of (19) such that } \lambda_i > 0 \}.$$

The set of KKT multipliers is equal to the polyhedron

$$\Lambda(q,b) = \left\{ \begin{array}{rl} \lambda \geq 0 : & 0 = q + EG(v(q,b)) + A^T \lambda \\ & 0 = \lambda_i, \quad i \notin \alpha(q,b) \end{array} \right\}.$$

Consequently, there exists a constant $\rho > 0$, dependent on A only, such that for all (q, b) for which $\Lambda(q, b) \neq \emptyset$, a $\lambda \in \Lambda(q, b)$ exists such that

$$\| \lambda \| \leq \rho \| q + EG(v(q,b)) \| \leq \rho \, [\| q \| + L (\| EG(0) \| + \| E \| \| x \|)]$$

for any $x \in \mathrm{SOL}(q, b; E, G, A)$. The next result derives a bound on the solutions in $\mathrm{SOL}(q, b; E, G, A)$; its proof is similar to that of [19, Proposition 6.3]. We let $\mathcal{P}(A)$ denote the set of all vectors b for which the polyhedron $\{x : Ax \leq b\}$ is nonempty.

Proposition 7. Suppose $K_\infty \cap \ker(E^T) = \{0\}$. Let G be strongly monotone and Lipschitz continuous on an open convex set containing the range of E^T. A constant $\xi > 0$ exists such that for all (q, b) with $b \in \mathcal{P}(A)$,

$$\sup\{ \| x \| : x \in \mathrm{SOL}(q, b; E, G, A) \} \leq \xi \, (\| q \| + \| b \|).$$

Proof. From the KKT conditions (19), it follows that for all $x \in \mathrm{SOL}(q, b; E, G, A)$,

$$\begin{aligned} 0 &= x^T q + (E^T x)^T G(E^T x) + \lambda^T b \\ &\geq x^T q + (E^T x)^T G(0) + \varsigma \| E^T x \|^2 + \lambda^T b. \end{aligned}$$

Hence,

$$\varsigma \| E^T x \|^2 \leq \| x \| \| EG(0) \| + \| b \| \rho \, [\| q \| + L_G (\| EG(0) \| + \| E \| \| x \|)] .$$

Suppose no such constant $\kappa > 0$ exists. Sequences $\{(q^k, b^k)\}$ and $\{x^k\}$ with $x^k \in \mathrm{SOL}(q^k, b^k; E, G, A)$ for every k exist such that

$$\| x^k \| > k (\| q^k \| + \| b^k \|).$$

Hence, the sequences $\{q^k/\|x^k\|\}$ and $\{b^k/\|x^k\|\}$ both converge to zero. Without loss of generality, we may assume that the normalized sequence $\{x^k/\|x^k\|\}$ converge to a vector x^∞ of unit length, which must necessarily belong to K_∞. Dividing by $\|x^k\|^2$ in the inequality,

$$\varsigma \| E^T x^k \|^2 \leq \| x^k \| \| EG(0) \| + \| b^k \| \rho \left[\| q^k \| + L_G (\| EG(0) \| + \| E \| \| x^k \|) \right],$$

we deduce that $E^T x^\infty = 0$. This contradicts the assumption that $K_\infty \cap \ker(E^T) = \{0\}$. □

The next result establishes the Lipschitz continuity of $v(q, b)$ at all pairs (q, b) with $b \in \mathcal{P}(A)$. We write $JG(y)$ for the Jacobian matrix of G at the vector y provided that G is differentiable at y.

Theorem 8. Suppose $K_\infty \cap \ker(E^T) = \{0\}$. Let G be continuously differentiable, strongly monotone, and Lipschitz continuous on an open convex set containing the range of E^T. Let (q, b) with $b \in \mathcal{P}(A)$ be arbitrary. A neighborhood $\mathcal{N}$ of (q, b) and a scalar $L_v > 0$ exist such that for all $(q^i, b^i) \in \mathcal{N}$ with $b^i \in \mathcal{P}(A)$ for $i = 1, 2$,

$$\| v(q^1, b^1) - v(q^2, b^2) \| \leq L_v [\, \| q^1 - q^2 \| + \| b^1 - b^2 \| \,].$$

Proof. We first show that if $\{(q^k, b^k)\}$ is a sequence converging to (q, b) such that $\text{SOL}(q^k, b^k; E, G, A) \neq \emptyset$ for all k, then the sequence $\{v(q^k, b^k)\}$ converges to $v(q, b)$. For each k, let $x^k \in \text{SOL}(q^k, b^k; E, G, A)$ be such that $E^T x^k = v(q^k, b^k) \triangleq y^k$ and let λ^k be a KKT multiplier of the VI $(q^k, b^k; E, G, A)$; thus we have

$$\begin{aligned} 0 &= q^k + EG(y^k) + A^T \lambda^k \\ 0 &\leq \lambda^k \perp b^k - Ax^k \geq 0 \\ 0 &= y^k - E^T x^k. \end{aligned}$$

Since the sequence $\{y^k\}$ is bounded, let $\{y^k : k \in \kappa\}$ be any subsequence converging to some vector y^∞. It follows that $(x^\infty, \lambda^\infty)$ exists satisfying

$$\begin{aligned} 0 &= q + EG(y^\infty) + A^T \lambda^\infty \\ 0 &\leq \lambda^\infty \perp b - Ax^\infty \geq 0 \\ 0 &= y^\infty - E^T x^\infty. \end{aligned}$$

Hence $x^\infty \in \text{SOL}(q, b; E, G, A)$ and $y^\infty = v(q, b)$. Consequently, there is a unique accumulation point of the sequence $\{y^k\}$, which is $v(q, b)$. Thus, $\{v(q^k, b^k)\}$ converges to $v(q, b)$ as desired.

Suppose that no such constant L_v exists. Two sequences $\{(q^{1,k}, b^{1,k})\}$ and $\{(q^{2,k}, b^{2,k})\}$, both converging to (q, b), exist such that for every k, $b^k \in \mathcal{P}(A)$ and

$$\| v(q^{1,k}, b^{1,k}) - v(q^{2,k}, b^{2,k}) \| > k [\, \| q^{1,k} - q^{2,k} \| + \| b^{1,k} - b^{2,k} \| \,].$$

Both sequences $\{v(q^{1,k}, b^{1,k})\}$ and $\{v(q^{2,k}, b^{2,k})\}$ converge to $v(q, b)$. Since

$$\begin{aligned} EG(v(q^{i,k}, b^{i,k})) &= EJG(v(q, b))v(q^{i,k}, b^{i,k}) + e^{i,k}, \quad i = 1, 2 \\ &= EJG(v(q, b))E^T x^{i,k} + e^{i,k} \end{aligned}$$

for any solution $x^{i,k} \in \text{SOL}(q^{i,k}, b^{i,k}; E, G, A)$, where $e^{i,k} \triangleq EG(v(q^{i,k}, b^{i,k})) - EJG(v(q, b))v(q^{i,k}, b^{i,k})$, it follows that any such solution $x^{i,k}$ is a solution of the AVI $(q^{i,k} + e^{i,k}, b^{i,k}; EJG(v(q, b))E^T, A)$. Since the matrix $JG(v(q, b))$ is positive definite, by the strong monotonicity of G, a constant $\eta > 0$, dependent on the pair $(EJG(v(q, b))E^T, A)$ only, exists such that

$$\begin{aligned} &\| v(q^{1,k}, b^{1,k}) - v(q^{2,k}, b^{2,k}) \| \\ &\quad = \| E^T x^{1,k} - E^T x^{2,k} \| \\ &\quad \leq \eta \left[\| q^{1,k} - q^{2,k} \| + \| e^{1,k} - e^{2,k} \| + \| b^{1,k} - b^{2,k} \| \right] \\ &\quad \leq \eta \| E \| \left[\| G(v(q^{1,k}, b^{1,k})) - G(v(q^{2,k}, b^{2,k})) - JG(v(q, b))(\, v(q^{1,k}, b^{1,k}) - v(q^{2,k}, b^{2,k}) \,) \| \right] \\ &\qquad + \eta \left[\| q^{1,k} - q^{2,k} \| + \| b^{1,k} - b^{2,k} \| \right]. \end{aligned}$$

By the continuous differentiability of G near $v(q, b)$, it follows that for every $\varepsilon > 0$, a positive integer $\bar{k}$ exists such that for all $k > \bar{k}$,

$$\begin{aligned} &\| G(v(q^{1,k}, b^{1,k})) - G(v(q^{2,k}, b^{2,k})) - JG(v(q, b))(\, v(q^{1,k}, b^{1,k}) - v(q^{2,k}, b^{2,k}) \,) \| \\ &\quad \leq \varepsilon \| v(q^{1,k}, b^{1,k})) - v(q^{2,k}, b^{2,k}) \|. \end{aligned}$$

Hence, for $\varepsilon > 0$ sufficiently small,

$$\| v(q^{1,k}, b^{1,k}) - v(q^{2,k}, b^{2,k}) \| \leq \frac{\eta}{1 - \eta \| E \| \varepsilon} \left[\| q^{1,k} - q^{2,k} \| + \| b^{1,k} - b^{2,k} \| \right],$$

which is a contradiction. □

We now use the equation (11) to establish the directional differentiability of the function $v(q, b)$. For any pair $(\widehat{q}, \widehat{b})$ and any solution $\widehat{x} \in \mathrm{SOL}(\widehat{q}, \widehat{b}; E, G, A)$, the above proof shows that $\widehat{x}$ is a solution of the AVI $(\widehat{h}, \widehat{b}; M, A)$, where $M \triangleq EJG(v(q,b))E^T$ and $\widehat{h} \triangleq \widehat{q} + EG(v(\widehat{q}, \widehat{b})) - EJG(v(q,b))v(\widehat{q}, \widehat{b})$. Consider the nominal AVI $(h, b; M, A)$, where $h \triangleq q + EG(v(q,b)) - EJG(v(q,b))v(q,b)$. It is not difficult to see that $\mathrm{SOL}(h, b; M, A) = \mathrm{SOL}(q, b; E, G, A)$. Indeed, since both h and M are independent of the solutions in $\mathrm{SOL}(q, b; E, G, A)$, we have $\mathrm{SOL}(q, b; E, G, A) \subseteq \mathrm{SOL}(h, b; M, A)$. Conversely, if x^i for $i = 1, 2$ are any two solutions of the AVI $(h, b; M, A)$, then $Mx^1 = Mx^2$; hence $E^T x^i = v(q, b)$ and $w(h, b) \triangleq h + Mx^i = q + EG(E^T x^i)$ for $i = 1, 2$. This establishes that $\mathrm{SOL}(h, b; M, A) = \mathrm{SOL}(q, b; E, G, A)$ and $w(h, b) = q + EG(E^T \mathrm{SOL}(q, b; E, G, A))$. Moreover, the AVI $(h, b; M, A)$ and the VI $(q, b; E, G, A)$ have the same set of multipliers $\Lambda(q, b)$. Note that if $(\widehat{q}, \widehat{b})$ is sufficiently close to (q, b), then so is $\widehat{h}$ to h, by the continuity of the function $v(\bullet, \bullet)$. For each pair (h', b') and each solution $x \in \mathrm{SOL}(h, b; M, A)$, we may define the directional $\mathrm{AVI}'(h, b; h', b'; x)$ as in Section 2. By the remark made at the end of Section 2, it follows that for all $(\widehat{q}, \widehat{b})$ sufficiently close to (q, b) with $\widehat{b} \in \mathcal{P}(A)$, we have

$$v(\widehat{q}, \widehat{b}) = v(q, b) + E^T \mathrm{SOL}'(h, b; \widehat{h} - h; \widehat{b} - b; x)$$

for any solution $x \in \mathrm{SOL}(h, b; M, A)$ such that $\mathrm{SOL}'(h, b; \widehat{h} - h; \widehat{b} - b; x)$ is nonempty.

By Theorem 5, the $\mathrm{AVI}'(h, b; h', b'; x)$ has a solution for all $x \in \mathrm{SOL}(h, b; M, A) = \mathrm{SOL}(q, b; E, G, A)$ if and only if $(b')^T \lambda$ is bounded below on $\Lambda(q, b)$ and the implication below holds:

$$\left. \begin{array}{rcl} EJG(v(q,b))E^T y &=& 0 \\ h^T y &\leq& 0 \\ A_{i\bullet} y &\leq& 0, \quad \forall\, i \in \gamma(y, b) \end{array} \right\} \Rightarrow (h')^T y \geq 0,$$

where

$$\gamma(q, b) \equiv \{ i : (b - Ax)_i = 0 \text{ for all } x \in \mathrm{SOL}(q, b; E, G, A) \}.$$

With

$$h \triangleq q + EG(v(q,b)) - EJG(v(q,b))v(q,b)$$

and

$$h' \triangleq q(\tau) + EG(v(q(\tau), b(\tau))) - EJG(v(q,b))v(q(\tau), b(\tau)),$$

where $(q(\tau), b(\tau)) \triangleq (q, b) + \tau(q', b')$ for $\tau > 0$, the above implication reduces to

$$\left. \begin{array}{rcl} E^T y &=& 0 \\ q^T y &\leq& 0 \\ A_{i\bullet} y &\leq& 0, \quad \forall\, i \in \gamma(q, b) \end{array} \right\} \Rightarrow (q')^T y \geq 0, \tag{20}$$

which is independent of τ. The following theorem, which, together with Theorem 8, offers a complete first-order sensitivity analysis of the strongly monotone composite VI $(q, b; E, G, A)$.

Theorem 9. Suppose $K_\infty \cap \ker(E^T) = \{0\}$. Let G be continuously differentiable, strongly monotone, and Lipschitz continuous on an open convex set containing the range of E^T. For every tuple (q, b, q', b') such that $b \in \mathcal{P}(A)$, $(b')^T \lambda$ is bounded below on $\Lambda(q, b)$, and the implication (20) holds, the limit

$$\lim_{\tau \downarrow 0} \frac{v(q(\tau), b(\tau)) - v(q, b)}{\tau}$$

exists and equals $E^T \mathrm{SOL}'(h, b; q', b'; x)$ for any solution $x \in \mathrm{SOL}(q, b; E, G, A)$.

Proof. Let $\{\tau_k\}$ be a sequence of positive scalars for which the limit

$$\lim_{k\to\infty} \frac{v(q(\tau_k), b(\tau_k)) - v(q,b)}{\tau_k} \triangleq v^\infty$$

exists. With $h(\tau) \triangleq q + \tau\, q' + EG(v(q(\tau), h(\tau))) - EJG(v(q,b))v(q(\tau), b(\tau))$, we have

$$h(\tau) - h \,=\, \tau\, q' + E\,[\, G(v(q(\tau), h(\tau))) - G(v(q,b)) - JG(v(q,b))(\, v(q(\tau), b(\tau)) - v(q,b)\,)\,]\,.$$

Thus,

$$\lim_{\tau\downarrow 0} \frac{h(\tau) - h}{\tau} \,=\, q'.$$

Hence, for any $x \in \mathrm{SOL}(h, b; M, A) = \mathrm{SOL}(q, b; E, G, A)$,

$$\begin{aligned} v^\infty &= \lim_{k\to\infty} \frac{v(q(\tau_k), b(\tau_k)) - v(q,b)}{\tau_k} \\ &= \lim_{k\to\infty} E^T \mathrm{SOL}'(h, b; h(\tau_k) - h; b(\tau_k) - b; x) \,=\, E^T \mathrm{SOL}'(h, b; q', b'; x). \end{aligned}$$

Since $E^T\mathrm{SOL}'(h, b; q', b'; x)$ does not depend on the sequence $\{\tau_k\}$, the desired conclusion of the theorem follows. □

A parametric extension

We discuss a generalization of Theorem 9 to a (parametric) VI of finding a vector $x \in K \triangleq \{x : Ax \le b\}$ such that

$$(\, y - x\,)^T [\, q + EH(E^T x; p)\,] \,\ge\, 0, \quad \forall\, y \,\in\, K,$$

where $H : \mathbb{R}^{\ell+k} \to \mathbb{R}^\ell$ is a mapping of two arguments $(v, p) \in \mathbb{R}^{\ell+k}$, with $v = E^T x$, x being the primary variable of the VI and p being the parameter varying near a base value p^*. We continue to assume that $K_\infty \cap \ker E^T = \{0\}$. Restricted to a local analysis near a tuple (p^*, q, b), we assume that H is continuously differentiable in an open convex set containing (range of $E^T \times \{p^*\}$) and $H(\bullet; p)$ is strongly monotone for all p near p^*. It follows that for all such p and for all $b \in \mathcal{P}(A)$, the VI $(q, b; E, H(\bullet; p), A)$ has a nonempty compact polyhedral solution set with $E^T\mathrm{SOL}(q, b; E, H(\bullet; p), A)$ being a singleton, which we denote $v(p; q, b)$. We claim that this solution function $v(p; q, b)$ has the same first-order sensitivity properties as (the previous) $v(q, b)$ for all p sufficiently near p^*.

For any tuple $(\widehat{p}; \widehat{q}, \widehat{b}\,)$ and any solution $\widehat{x}$ of the VI $(\widehat{q}, \widehat{b}; E, H(\bullet, \widehat{p}\,), A)$, $\widehat{x}$ is a solution of the AVI $(\widehat{q}{+}\widehat{e}, \widehat{b}; EJ_vH(v(p^*; q, b); p^*)E^T, A)$, where $J_vH(v(p^*; q, b); p^*)$ denotes the Jacobian matrix of $H(\bullet; p^*)$ at $v(p^*; q, b)$ and $\widehat{e} \triangleq EH(v(\widehat{p}, \widehat{q}, \widehat{b}\,); \widehat{p}\,) - EJ_vH(v(p^*; q, b); p^*)v(\widehat{p}; \widehat{q}, \widehat{b}\,)$. By the continuous differentiability of H, it follows that for all (v, p) and (v', p') sufficiently near $(v(p^*; q, b); p^*)$,

$$\begin{aligned} &\|\, H(v; p) - H(v', p') - J_vH(v(p^*; q, b); p^*)(\, v - v'\,)\,\| \\ &\quad \le \,\|\, H(v; p) - H(v', p') - J_vH(v(p^*; q, b); p^*)(\, v - v'\,) - J_pH(v(p^*; q, b), p^*)(\, p - p'\,)\,\| \\ &\qquad\qquad + \|\, J_pH(v(p^*; q, b), p^*)(\, p - p'\,)\,\| \\ &\quad = \, o(\|\, v - v'\,\|) + O(\|\, p - p'\,\|\,). \end{aligned}$$

From this inequality, and applying the same (non-parametric) derivations as before, we can establish that for any given tuple $(p^*; q, b)$ with $b \in \mathcal{P}(A)$ satisfying the above stipulations, the solution function $v(\widehat{p}; \widehat{q}, \widehat{b}\,)$ is Lipschitz continuous in a neighborhood of $(p^*; q, b)$. Moreover, based on the nominal AVI $(h, b; M, A)$, where $M \triangleq EJ_vH(v(p^*; q, b)E^T$ and $h \triangleq q + EH(v(p^*, q, b); p^*) - EJ_vH(v(p^*; q, b); p^*)v(p^*; q, b)$, we can carry out the same directional analysis as in the previous case. The details are omitted.

Closing remarks. This paper has focused on the monotone-plus AVI and the linearly constrained, strongly monotone composite VI. It is natural to ask what if the constraints are nonlinear but remain convex and satisfy appropriate constraint qualifications. This question remains unanswered at this time. One can ask a similar question for the class of VIs in positive semidefinite matrices. In other words, can one develop a first-order sensitivity theory for a strongly monotone composite VI in positive semidefinite matrices without the strong regularity assumption, or for a semidefinite program without the strong second-order sufficiency condition, thereby extending the results in the papers [20, 26]?

Acknowledgement. The authors gratefully acknowledge the discussion with Dr. Weihong Yang of Fudan University on a previous draft of the paper during his visit to the University of Illinois at Urbana-Champaign.

References

[1] K. Bennett, X. Ji, J. Hu, G. Kunapuli, and J.S. Pang. Model selection via bilevel programming, *Proceedings of the International Joint Conference on Neural Networks (IJCNN'06) Vancouver, B.C. Canada, July 16–21, 2006* pp. 1922–1929.

[2] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer New York (2000).

[3] C.J.C. Burges and D.J. Crisp. Uniqueness of the SVM solution. *Advances in Neural Information Processing Systems* 13 (1999) 223–229.

[4] M. Cao and M.C. Ferris. Lineality removal for copositive-plus normal maps. *Communications on Applied Nonlinear Analysis* 2 (1995) 110.

[5] M. Cao and M.C. Ferris. A pivotal method for affine variational inequalities. *Mathematics of Operations Research* 21 (1996) 44-64.

[6] M.W. Chang and C.J. Lin. Leave-one-out bounds for support vector regression model selection. *Neural Computation* 17 (2005) 1188–1222.

[7] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee. Choosing multiple parameters with support vector machines. *Machine Learning* 46 (2002) 131–159.

[8] K.M. Chung, W.C. Kao, C.L. Sun, L.L. Wang, and C.J. Lin. Radius margin bounds for support vector machines with the RBF kernel. *Neural Computation* 15 (2003) 2643–2681.

[9] R.W. Cottle, J.S. Pang, R.E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston (1992).

[10] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Volumes I and II, Springer New York (2003).

[11] O. Güler and Y. Ye. Convergence behavior of interior-point algorithms. *Mathematical Programming* 60 (1993) 215–228.

[12] T. Illés, J.M. Peng, C. Roos, and T. Terlaky. A strongly polynomially rounding procedure yielding a maximally complementary solution for $P_*(\kappa)$ linear complementarity problems. *SIAM Journal on Optimization* 11 (2000) 320–340.

[13] G. Kunapuli, K. Bennett, J. Hu, and J.S. Pang. Classification model selection via bilevel programming, *Optimization Methods and Software*, in print.

[14] G. Kunapuli, K. Bennett, J. Hu, and J.S. Pang Bilevel model selection for support vector machines, *Centre de Recherches Mathématiques CRM Proceedings and Lecture Notes*. [P. Pardalos and P. Hansen, editors. *Data Mining and Mathematical Programming*] 45 (2008) pp. 129–158.

[15] S. Lu and S.M. Robinson. Variational inequalities over perturbed polyhedral convex sets. *Mathematics of Operations Research* (2008) DOI: 10.1287/moor.1070.0297.

[16] Z.Q. Luo, J.S. Pang, and D. Ralph. *Mathematical Programs With Equilibrium Constraints*. Cambridge University Press, Cambridge, England (1996).

[17] Z.Q. Luo and P. Tseng. A decomposition property for a class of square matrices. *Applied Mathematics Letters* 4 (1991) 67–69.

[18] O.L. Mangasarian. A simple characterization of solution sets of convex programs. *Operations Research Letters* 7 (1988) 21–26.

[19] J.S. Pang and D.E. Stewart. Differential variational inequalities. *Mathematical Programming* 113 (2008) 345–424.

[20] J.S. Pang, D.F. Sun, and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Mathematics of Operations Research* 28 (2003) 39–63.

[21] S.M. Robinson. Stability theory for systems of inequalities. II. Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis* 13 (1976) 497-513.

[22] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research* 5 (1980) 43–62.

[23] S.M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study* 14 (1981) 206–214.

[24] S.M. Robinson. Solution continuity in monotone affine variational inequalities. *SIAM Journal on Optimization* 18 (2008) 1046–1127.

[25] S. Scholtes. *Introduction to Piecewise Differentiable Equations*. Habilitation thesis, University of Karlsruhe, Karlsruhe, Germany (1994).

[26] D.F. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research* 31 (2006) 761–776.