Computational Optimization and Applications manuscript No.
(will be inserted by the editor)

A Class of Benders Decomposition Methods for Variational Inequalities

Juan Pablo Luna · Claudia Sagastizábal · Mikhail Solodov

Received: January 31, 2019 (revised September 16, 2019) / Accepted:

Abstract We develop new variants of Benders decomposition methods for variational inequality problems. The construction is done by applying the general class of Dantzig–Wolfe decomposition of [14] to an appropriately defined dual of the given variational inequality, and then passing back to the primal space. As compared to previous decomposition techniques of the Benders kind for variational inequalities, the following improvements are obtained. Instead of rather specific single-valued monotone mappings, the framework includes a rather broad class of multi-valued maximally monotone ones, and single-valued nonmonotone. Subproblems' solvability is guaranteed instead of assumed, and approximations of the subproblems' mapping are allowed (which may lead, in particular, to further decomposition of subproblems, which may otherwise be not possible). In addition, with a certain suitably chosen approximation, variational inequality subproblems become simple bound-constrained optimization problems, thus easier to solve.

Keywords Variational inequalities · Benders decomposition · Dantzig–Wolfe decomposition · stochastic Nash games

Mathematics Subject Classification (2010) 90C33 · 65K10 · 49J53

J. P. Luna
COPPE-UFRJ, Engenharia de Produção, Cidade Universitária, Rio de Janeiro - RJ, 21941-972, Brazil.
E-mail: jpluna@po.coppe.ufrj.br

C. Sagastizábal
IMECC - UNICAMP, Rua Sergio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil.
E-mail: sagastiz@unicamp.br

M. Solodov
IMPA – Instituto de Matemática Pura e Aplicada,
Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
E-mail: solodov@impa.br

1 Introduction

In applications, feasible sets of optimization or variational problems frequently have some special structure, amenable to decomposition, which opens the potential to handle larger problems' instances and to better computational performance. In this direction, two major classes of techniques are the Dantzig-Wolfe decomposition [8] and Benders decomposition [3]; see also [4, Chapter 11.1]. To outline the main ideas, it is sufficient to discuss the cases of two linear programs (LPs), structured as specified below. More precisely, consider

$$\begin{cases} \min_x \; f(x) \\ \quad \text{s.t. } H(x) \le 0 \\ \qquad\quad g(x) \le 0, \end{cases} \tag{1}$$

and

$$\begin{cases} \min_x \; f(x) + r(y) \\ \quad \text{s.t. } g(x) + h(y) \le 0, \end{cases} \tag{2}$$

where all functions are affine with appropriate dimensions. In the case of (1), suppose the function g has a special structure such that (1) is (much) easier to solve if the constraint $H(x) \le 0$ were to be removed. In the case of (2), suppose the problem becomes (much) easier to solve if we fix the value of the variable y. For example, those two cases occur when the Jacobian matrix g' defining the linear function g has the following block-separable structure:

$$g'(x) = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_j \end{bmatrix} \tag{3}$$

where A_i, $i = 1, \ldots, j$, are matrices of appropriate dimensions.

Problem (1) can be solved using Dantzig-Wolfe decomposition introduced in [8]. Without going into details, the Dantzig-Wolfe algorithm consists in iteratively solving two types of problems, called *master* program and *subproblem*. As a result of the master solution, an estimation of the Lagrange multiplier μ^k for the constraint $H(x) \le 0$, computed over the convex hull of past iterates, is obtained. The subproblem uses that estimate to solve the Lagrangian relaxation (with respect to the constraint $H(x) \le 0$) of the original problem (1), i.e.,

$$\begin{cases} \min_x \; f(x) + [\mu^k]^\top H(x) \\ \quad \text{s.t. } g(x) \le 0. \end{cases} \tag{4}$$

As all the functions in the above are linear and g has the separable structure according to (3), subproblem (4) decomposes into j smaller ones. Having solved the subproblem, the iterative process continues with a new master program, where the multiplier approximation is improved by incorporating in the convex hull of previous iterates the solution provided by the last subproblem (e.g., see (8) below in the setting of the Benders approach).

Concerning the second problem (2), its solution using Benders decomposition comes from distinguishing two kinds of variables, in a way that if we fix one of them (say, y), then the problem becomes much easier to solve. To motivate our approach for variational inequalities presented below, we recall that when solving the LP (2), Benders decomposition can be derived by applying the Dantzig-Wolfe decomposition algorithm to the dual of (2). Assume for simplicity (and without loss of generality), that $f(0) = 0$. Then the dual LP of (2) is given by

$$\begin{cases} \max_\mu \; r(0) + \langle \mu, g(0) + h(0) \rangle \\ \text{ s.t. } \;\; f'(0) + [g'(0)]^\top \mu = 0, \\ \qquad\;\; r'(0) + [h'(0)]^\top \mu = 0, \\ \qquad\;\; \mu \geq 0. \end{cases} \tag{5}$$

Note that if $g'(0) \equiv g'(\cdot)$ is block-separable like in (3), then the transposed Jacobian $[g'(0)]^\top$ inherits the decomposable structure. Hence, without the constraint $r'(0) + [h'(0)]^\top \mu = 0$, solving (5) would be easier. The situation is thus similar to problem (1), i.e., it is suitable for the Dantzig-Wolfe decomposition. In the considered dual setting, given at iteration k a Lagrange multiplier estimate y_M^k and using Lagrangian relaxation of the corresponding "difficult" constraint $r'(0) + [h'(0)]^\top \mu = 0$, the resulting Dantzig-Wolfe subproblem is

$$\begin{cases} \max_\mu \; r(y_M^k) + \langle \mu, g(0) + h(y_M^k) \rangle \\ \text{ s.t. } \;\; f'(0) + [g'(0)]^\top \mu = 0, \mu \geq 0. \end{cases} \tag{6}$$

With the structure at hand of g', the latter is easy to solve, to obtain a subproblem solution μ_S^{k+1} to be included in the master program. However, instead of solving directly the LP (6), we solve its dual:

$$\begin{cases} \min_x \; f(x) + r(y_M^k) \\ \text{ s.t. } \;\; g(x) + h(y_M^k) \leq 0, \end{cases} \tag{7}$$

obtaining an optimal Lagrange multiplier μ^{k+1} associated to the constraint. The latter is precisely the subproblem of the Benders decomposition scheme. Recall that in the setting of (2), for the fixed $y = y_M^k$, the subproblem (7) is easy to solve. Let a solution of (7) be denoted by x_S^{k+1}. At this stage, we have already computed Lagrange multipliers $\mu_S^0, \mu_S^1, \ldots, \mu_S^k$ associated to the previous subproblems. We can then perform the Dantzig-Wolfe master step, which consists in solving

$$\begin{cases} \max_\mu \; r(0) + \langle \mu, g(0) + h(0) \rangle \\ \text{ s.t. } \;\; r'(0) + [h'(0)]^\top \mu = 0, \\ \qquad\;\; \mu \in \operatorname{conv}\{\mu_S^0, \mu_S^1, \ldots, \mu_S^k\}, \end{cases} \tag{8}$$

where conv$D$ stands for the convex hull of the set D. Solving (8) gives a Lagrange multiplier y_M^k associated to the constraint $r'(0) + [h'(0)]^\top \mu = 0$, so that we can continue with the next subproblem in the Dantzig-Wolfe framework. However, again, instead of solving directly the LP (8), we solve its dual

$$\begin{cases} \min_{(y,t)} \ t + r(y) \\ \quad \text{s.t.} \quad \langle \mu_S^i, g(0) + h(y) \rangle \le t, \ i = 0, 1, \dots, k, \end{cases} \tag{9}$$

which can be rewritten, by using optimality conditions of (7), as

$$\begin{cases} \min_{(y,t)} \ t + r(y) \\ \quad \text{s.t.} \quad f(x_S^i) + \langle [h'(y_M^{i-1})]^\top \mu_S^i, y - y_M^{i-1} \rangle \le t, \ i = 1, 2, \dots, k. \end{cases} \tag{10}$$

In summary, applying to the dual (5) of the original LP (2) the Dantzig-Wolfe algorithm given by (6) and (8), is equivalent to solving iteratively (7) and (9). The latter is precisely the Benders decomposition method for LPs [3].

We emphasize that viewing Benders algorithm as an application of the Dantzig-Wolfe approach to the dual problem (as above), is not the only possible interpretation; see, e.g., [4, Chapter 11.1]. But this point of view has the advantage of possibly using those ideas for other problem classes. And this indeed would be our approach for deriving Benders decomposition for variational inequalities (VIs) in the sequel, by applying the Dantzig-Wolfe method for VIs of [14] to the appropriately defined dual of the original VI.

To define the class of variational problems in question, let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$, and let $H : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^q$ be continuous functions with convex components. Define the two convex sets $S_H = \{x \mid H(x) \le 0\}$ and $S_g = \{x \mid g(x) \le 0\}$. Consider the *variational inequality problem* $\mathrm{VI}(F, S_H \cap S_g)$ [10], which means to find

$$\begin{array}{c} \bar{x} \in S_H \cap S_g \text{ such that } \langle \bar{w}, x - \bar{x} \rangle \ge 0 \\ \text{for some } \bar{w} \in F(\bar{x}) \text{ and all } x \in S_H \cap S_g. \end{array} \tag{11}$$

As in the discussion of the Dantzig–Wolfe decomposition for the LP (1), we assume that the VI (11) would be much easier if the constraints $H(x) \le 0$ (the set S_H) are not present. A Dantzig–Wolfe method for this family of problems had been introduced for single-valued monotone VIs in [7, 11]. A more general framework, convergent under much weaker assumptions, is [14]. In particular, [14] is applicable to (single-valued) nonmonotone F and multi-valued (maximally monotone) F, allowing in addition a rich class of approximations of F and of the derivative of H, as well as inexact solution of subproblems.

The simplest form of Dantzig–Wolfe method for (11) comes from the following consideration. Under an appropriate constraint qualification [20], solving problem (11) is equivalent to finding $(\bar{x}, \bar{\mu})$ such that

$$\begin{cases} (\bar{x}, \bar{\mu}) \in S_H \times \mathbb{R}_+^q, \quad \bar{\mu} \perp H(\bar{x}), \\ \bar{x} \text{ solves } \mathrm{VI}(F(\cdot) + [H'(\cdot)]^\top \bar{\mu}, S_g), \end{cases} \tag{12}$$

where the notation $u \perp v$ means that $\langle u, v \rangle = 0$. In view of (12), having a current multiplier estimate μ_M^k for the $H(x) \le 0$ constraint, we may consider solving the subproblem

$$\mathrm{VI}(F(\cdot) + [H'(\cdot)]^\top \mu_M^k, S_g). \tag{13}$$

According to the discussion above, in the given context VI (13) is simpler than the original (11). And (13) is indeed one of the possible algorithmic options. However, it is not the only one. Algorithm 1 below solves subproblems with the structure in (13), but allows useful approximations of F and H', as well as proximal regularization of the subproblem operator (to induce solvability). We refer the reader to [14] for details on all the possible options. Here, we shall not discuss them in connection with Dantzig-Wolfe Algorithm 1; but some possibilities will be given for the Benders scheme developed in the sequel.

Algorithm 1 (Dantzig-Wolfe Decomposition for VIs)

1. Choose $x_S^0 \in S_g \cap S_H$, such that $H(x_S^0) < 0$ if H is not affine. Set $x_M^0 = x_S^0$. Choose $\mu_M^0 \in \mathbb{R}_+^q$ and $w_M^0 \in F(x_M^0)$. Set $k := 0$.
2. **Subproblem solution:** Choose an approximation $F^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of $F(\cdot)$, an approximation $H^k : \mathbb{R}^n \to \mathbb{R}^{q \times n}$ of $H'(\cdot)$, and a positive (semi)definite matrix $Q_k \in \mathbb{R}^{n \times n}$. Find x_S^{k+1}, an approximate solution of the problem
$$\mathrm{VI}(\hat{F}^k, S_g), \tag{14}$$
$$\hat{F}^k(x) = F^k(x) + [H^k(x)]^\top \mu_M^k + Q_k(x - x_M^k). \tag{15}$$
3. **Master program:** Choose finite set $X^{k+1} \subset S_g$ containing $\{x_S^0, \ldots, x_S^{k+1}\}$. Find a solution x_M^{k+1} of the problem
$$\mathrm{VI}(F, S_H \cap \operatorname{conv} X^{k+1}), \tag{16}$$
with the associated $w_M^{k+1} \in F(x_M^{k+1})$ and a Lagrange multiplier μ_M^{k+1} associated to the constraint $H(x) \leq 0$.
4. Set $k := k + 1$ and go to Step 2.

Note that although the master problem (16) formally involves the difficult constraint set S_H, by the change of variables, it is parametrized by the unit simplex associated to the convex hull of the set X^{k+1} (which again often gives a relatively easy problem). To alleviate potential computational burden, empirical approximations in the Dantzig-Wolfe master program were considered in [6], showing an improved performance for two energy market models.

In the subproblem step of the algorithm, the regularization matrix Q_k should be taken as zero if F (and then also F^k, for natural choices) is known to be strongly monotone. If strong monotonicity does not hold then Q_k should be positive definite, to guarantee that subproblems are solvable.

The rest of this paper is organized as follows. In Section 2, we discuss the type of VIs to which our Benders decomposition approach is applicable, cite some previous developments, and state our contributions. Section 3 is devoted to subproblems of Benders decomposition, in particular showing how appropriate approximations of the data can reduce them to simple optimization problems, and how further decomposition of them can be achieved in the case of stochastic Nash games. The master problem of Benders decomposition is stated in Section 4. The overall algorithm and its convergence analysis are presented in Section 5. The paper finishes with some concluding remarks in Section 6.

A few words on our terminology are in order. We say that the set-valued mapping $T : \mathbb{R}^l \rightrightarrows \mathbb{R}^l$ is monotone if $\langle u - v, x - y\rangle \geq 0$ for all $u \in T(x)$ and $v \in T(y)$ and all $x, y \in \mathbb{R}^l$. Such a mapping is further called maximally monotone if its graph is not properly contained in the graph of any other monotone mapping. T is $c-$strongly monotone if there exists $c > 0$ such that $\langle u - v, x - y\rangle \geq c\|x - y\|^2$ for all $u \in T(x)$ and $v \in T(y)$ and all $x, y \in \mathbb{R}^l$. The mapping $T : \mathbb{R}^l \rightrightarrows \mathbb{R}^l$ is outer semicontinuous if for any sequences $\{x^k\}$, $\{y^k\}$ such that $\{x^k\} \to \bar{x}$ and $\{y^k\} \to \bar{y}$ with $y^k \in T(x^k)$, it holds that $\bar{y} \in T(\bar{x})$. Recall that if T is either continuous or maximally monotone, then it is outer semicontinuous. We say that a family of set-valued mappings $\{T^k\}$ is equicontinuous on compact sets if for every compact set D and every $\epsilon > 0$ there exists $\delta > 0$ such that for any $x, y \in D$ with $\|x - y\| < \delta$, for every k it holds that $d_H(T^k(x), T^k(y)) < \epsilon$, where d_H is the Hausdorff distance between the sets, defined by

$$d_H(C, D) = \inf\{t > 0 : C \subset D + B(0, t) \text{ and } D \subset C + B(0, t)\},$$

$B(0, t)$ being the closed ball centered at the origin with radius t.

2 Primal-dual variational inequality framework for Benders decomposition

We start with specifying the VI structure (and its dual) suitable for Benders decomposition. Consider the given (primal) problem

$$\text{VI}(F_P, S_P) \tag{17}$$

with the following structure: $F_P : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ has the form

$$F_P(x, z) = F(x) \times G(z),$$

with $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ and

$$S_P = \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^m : Ax + Bz \leq d\},$$

where A and B are matrices of appropriate dimensions. We assume that after fixing the value of the variable z, the problem

$$\text{VI}\,(F, S_P(z)), \quad S_P(z) = \{x \in \mathbb{R}^n : Ax + Bz \leq d\},$$

is much easier to solve than (17). As in the LP case discussed in Section 1, this happens when A has block separable structure, something very common in various applications. For the eventual convergence analysis, we shall assume that F and G are outer semicontinuous mappings. The latter holds, in particular, if they are maximally monotone (possibly multi-valued) [5, Proposition 4.2.1], which is in particular the case (but not only) of subdifferentials of convex functions. For solvability of iterative problems involved in the construction, one of F and G should also either be surjective or its inverse has to be surjective. The latter is a technical assumption needed to ensure the maximal monotonicity of the mapping of the dual VI problem (defined in (19) further below). When F or G has domain or image bounded, then this technical assumption is automatic [5, Corollary 4.5.1].

The only other Benders type method for VIs that we are aware of is the one proposed in [12]; see [13] for an application. The differences between our development and [12], and our contributions, are summarized as follows. First, we consider VIs with a possibly multi-valued (maximally monotone) mapping, whereas in [12] the mapping is single-valued monotone and, even more importantly, it has a rather specific form, like $(F(x), u, v)$ with F continuous and invertible and u, v constant vectors. Clearly, our setting is much more general and covers far more applications. Moreover, as it will be seen below, this generalization does not complicate too much the iterations of the algorithm, i.e., the subproblems will still be computationally tractable. In fact, the iterative subproblems in our method related to the multi-valued part, that replaces the constant part in the previous work, are independent of the variable x and can be solved relatively easily. Second, the existence of solutions of primal subproblems in the previous work was an assumption. Here, we use regularization to ensure solvability, and thus dispense with such assumptions. Moreover, we allow approximations for the dual subproblem VI mapping, as in the Dantzig-Wolfe algorithm in [14], while previous work required the use of exact information. Apart from general importance of using approximations in many real-world applications, this is of special significance here, because it allows us to express the corresponding subproblem as a simpler minimization problem (instead of a VI), if an appropriate type of approximation is chosen.

As in the LP case discussed in Section 1, the VI Benders method is defined by applying the Dantzig-Wolfe technique to an appropriately defined dual formulation. Because (17) has a special structure, it is possible to define the dual problem by rearranging the corresponding rush-Kuhn-Tucker (KKT)-type conditions. It turns out that the resulting dual problem fits well the Dantzig-Wolfe decomposition scheme. Here, we have to mention that there exist a good number of other ways of attaching a dual VI problem to a given (primal) VI. This involves various degrees of abstraction, e.g., [1,2,9,16], among others. All these developments have their purpose and usefulness. However, they do not seem helpful for our task here, as they do not appear to connect to decomposition ideas.

We thus start with defining our own dual problem for VI (17) as follows:

$$\mathrm{VI}(F_D, S_D), \tag{18}$$

where $F_D : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is given by

$$F_D(w, \zeta, \mu) = F^{-1}(w) \times G^{-1}(\zeta) \times \{d\}, \tag{19}$$

and

$$S_D = \left\{ (w,\zeta,\mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p : \begin{array}{r} w + A^\top \mu = 0 \\ \zeta + B^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}.$$

In the above, for a set-valued mapping $T : \mathbb{R}^l \rightrightarrows \mathbb{R}^l$, we denote its inverse at $u \in \mathbb{R}^l$ by $T^{-1}(u) = \{v \in \mathbb{R}^l : T(v) = u\}$.

Again, if A has block-decomposable structure, the constraint $w + A^\top \mu = 0$ also has this property, which means that the problem would be easier to deal with if the constraint $\zeta + B^\top \mu = 0$ were to be removed. We then immediately recognize that (18) is amenable to the Dantzig-Wolfe decomposition for VIs in Algorithm 1.

But first, let us show that the original VI (17) and its dual given by (18) are equivalent in a certain sense.

Proposition 1 *For the data defined above, the following holds.*

1. *If the elements $(\bar{x}, \bar{z})$ together with $\bar{w} \in F(\bar{x})$ and $\bar{\zeta} \in G(\bar{z})$ solve the primal problem (17) with a Lagrange multiplier $\bar{\mu}$, then $(\bar{w}, \bar{\zeta}, \bar{\mu})$ together with $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{z} \in G^{-1}(\bar{\zeta})$ solve the dual problem (18) with multipliers $(-\bar{x}, -\bar{z}, -A\bar{x} - B\bar{z} + d)$.*
2. *If $(\bar{w}, \bar{\zeta}, \bar{\mu})$ together with $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{z} \in G^{-1}(\bar{\zeta})$ solve the dual problem (18) with multipliers $(-\bar{\alpha}, -\bar{\beta}, -\bar{\gamma})$, then*
   (a) $\bar{\alpha} = \bar{x}$, $\bar{\beta} = \bar{z}$ and $\bar{\gamma} = A\bar{x} + B\bar{z} - d$,
   (b) $(\bar{x}, \bar{z})$ together with $\bar{w} \in F(\bar{x})$ and $\bar{\zeta} \in G(\bar{z})$ solve the primal problem (17) with multiplier $\bar{\mu}$.

Proof Since the constraints defining S_P are linear, we have that the KKT type conditions hold for any solution of the primal problem (17):

$$\bar{w} + A^\top \bar{\mu} = 0, \tag{20a}$$

$$\bar{\zeta} + B^\top \bar{\mu} = 0, \tag{20b}$$

$$0 \leq \bar{\mu} \perp A\bar{x} + B\bar{z} - d \leq 0. \tag{20c}$$

This shows that $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_D$.

On the other hand, the following inclusion is immediate:

$$0 \in \begin{bmatrix} F^{-1}(\bar{w}) \\ G^{-1}(\bar{\zeta}) \\ d \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ A \end{bmatrix} (-\bar{x}) + \begin{bmatrix} 0 \\ I \\ B \end{bmatrix} (-\bar{z}) + \begin{bmatrix} 0 \\ 0 \\ -I \end{bmatrix} (d - A\bar{x} - B\bar{z}). \tag{21}$$

The latter, together with (20c), proves the first item.

The proof of the second item is analogous, reversing the roles of the primal and dual variables. $\square$

3 Benders subproblem

The full Benders decomposition algorithm for VIs is stated in Section 5 below; we give here an initial dual view. As in the Dantzig-Wolfe approach, there are two main blocks: the subproblem and the master program. In this section we describe the subproblems that the Dantzig-Wolfe VI scheme would solve if applied to the dual VI (defined above). The primal counterparts of these Dantzig-Wolfe subproblems define the Benders subproblems in Section 5.

3.1 Dantzig-Wolfe in the dual

At iteration k, we have $(w^k_M, \zeta^k_M, \mu^k_M) \in S_D$ with $x^k_M \in F^{-1}(w^k_M)$ and $z^k_M \in G^{-1}(\zeta^k_M)$ and a Lagrange multiplier estimate $(-z^k_M - \theta^k_M)$ associated to the constraint $\zeta + B^\top \mu = 0$. The multiplier estimate is expressed here as sum (or difference) of two terms for technical reasons concerning its relationship with the decision variables in the applications in Sections 3.2 and 3.3.

At the kth iteration, the Dantzig-Wolfe subproblem of Algorithm 1 applied to the dual VI (18) consists in solving

$$\mathrm{VI}(\hat{F}^k_D, S_{D_S}), \tag{22}$$

where the feasible set S_{D_S} is defined by

$$S_{D_S} = \left\{ (w, \zeta, \mu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p : \begin{array}{r} w + A^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}, \tag{23}$$

and the VI mapping $\hat{F}^k_D$ is defined using some approximations F_k^{-1} and G_k^{-1} of F^{-1} and G^{-1}, respectively, and some positive (semi)definite matrices P_k, Q_k and R_k. Specifically,

$$\hat{F}^k_D(w, \zeta, \mu) = \begin{bmatrix} F_k^{-1}(w) \\ G_k^{-1}(\zeta) \\ d \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ B \end{bmatrix} (-z^k_M - \theta^k_M) + \begin{bmatrix} Q_k(w - w^k_M) \\ R_k(\zeta - \zeta^k_M) \\ P_k(\mu - \mu^k_M) \end{bmatrix}. \tag{24}$$

The approximating functions F^k and G^k are chosen according to the requirements of the Dantzig-Wolfe scheme in [14, Subsection 2.1]. In particular,

$$x^k_M \in F_k^{-1}(w^k_M) \subset F^{-1}(w^k_M), \tag{25}$$

and

$$z^k_M \in G_k^{-1}(\zeta^k_M) \subset G^{-1}(\zeta^k_M). \tag{26}$$

Next, note the following two important structural features. In the feasible set (23) the variable ζ is unconstrained, and in the VI mapping (24) the entry corresponding to this variable is independent of the other variables. It then follows that the subproblem (22) equivalently splits into the following two *independent* steps:

1. Find $\zeta \in \mathbb{R}^m$ such that

$$0 \in G_k^{-1}(\zeta) - z_M^k - \theta_M^k + R_k(\zeta - \zeta_M^k). \tag{27}$$

2. Solve

$$\mathrm{VI}(\hat{F}_{D2}^k, S_{D_{S2}}), \tag{28}$$

where

$$\hat{F}_{D2}^k(w,\mu) = \begin{bmatrix} F_k^{-1}(w) + Q_k(w - w_M^k) \\ d - B(z_M^k + \theta_M^k) + P_k(\mu - \mu_M^k) \end{bmatrix}, \tag{29}$$

and

$$S_{D_{S2}} = \left\{ (w,\mu) \in \mathbb{R}^n \times \mathbb{R}^p : \begin{array}{r} w + A^\top \mu = 0 \\ \mu \geq 0 \end{array} \right\}. \tag{30}$$

At least in full generality, the two problems above have the disadvantage in that they are defined in terms of the inverse operators that in practice could be difficult to deal with. In fact, in most applications those inverses would not be known explicitly. Fortunately, in the dual framework that we defined, it is possible to solve (27) and (28) via their duals, not involving the inverse operators. The following propositions show those dual relations.

Proposition 2 *If z_S^{k+1} together with $\zeta_S^{k+1} \in G_k(z_S^{k+1})$ solve the problem*

$$0 \in z - z_M^k - \theta_M^k + R_k(G_k(z) - \zeta_M^k), \tag{31}$$

then ζ_S^{k+1} solves (27) with $z_S^{k+1} \in G_k^{-1}(\zeta_S^{k+1})$.

Also, the existence of solutions of (27) implies the existence of solutions of (31). In particular, solutions exist if G_k is maximal monotone and R_k is positive definite.

Proof The first assertion is obtained by direct inspection. The existence of solutions is by [18, Theorem 5]. □

Proposition 3 *Let (x_S^{k+1}, u_S^{k+1}) be any solution of*

$$VI(\hat{F}_P^k, S_P(z_M^k, \theta_M^k)), \tag{32}$$

where

$$\hat{F}_P^k(x,\mu) = \begin{pmatrix} F_k(x) \\ (AQ_kA^\top + P_k)\mu \end{pmatrix}, \tag{33}$$

and

$$S_P(z_M^k, \theta_M^k) = \{(x,\mu) : Ax + B(z_M^k + \theta_M^k) \leq d + (AQ_kA^\top + P_k)(\mu - \mu_M^k)\}. \tag{34}$$

Let $w_S^{k+1} \in F_k(x_S^{k+1})$ and the Lagrange multiplier μ_S^{k+1} verify the KKT conditions for (32) *at its solution (x_S^{k+1}, u_S^{k+1}).*

Then the following holds.

1. *We have that $(AQ_kA^\top + P_k)(\mu_S^{k+1} - u_S^{k+1}) = 0$; and, when P_k is symmetric positive define, then $\mu_S^{k+1} = u_S^{k+1}$.*

2. *The element* (x_S^{k+1}, μ_S^{k+1}) *solves* $VI(\hat{F}_P^k, S_P(z_M^k, \theta_M^k))$ *with* $w_S^{k+1} \in F^k(x_S^{k+1})$ *and multiplier* μ_S^{k+1}.
3. *The element* (w_S^{k+1}, μ_S^{k+1}) *solves* (28) *with* $x_S^{k+1} \in F_k^{-1}(w_S^{k+1})$ *and multipliers* $(-x_S^{k+1} - Q_k(w_S^{k+1} - w_M^k), d + (AQ_kA^\top + P_k)(\mu_S^{k+1} - \mu_M^k) - Ax_S^{k+1} - B(z_M^k + \theta_M^k))$.

Also, the existence of solutions of (28) *implies the existence of solutions of (32). In particular, solutions exist if* F_k *is maximal monotone, with the following holding:* $\{-A^\top \mu : \mu \geq 0\} \cap int(F_k(\mathbb{R}^n)) \neq \emptyset$ *and* P_k, Q_k *are positive definite.*

Proof Writing the KKT conditions corresponding to (32), which hold by the linearity of constraints in this problem, we have that

$$0 = w_S^{k+1} + A^\top \mu_S^{k+1}, \tag{35a}$$

$$0 = (AQ_kA^\top + P_k)u_S^{k+1} - (AQ_kA^\top + P_k)^\top \mu_S^{k+1}, \tag{35b}$$

$$0 \leq \mu_S^{k+1} \perp Ax_S^{k+1} + B(z_M^k + \theta_M^k) - d - (AQ_kA^\top + P_k)(u_S^{k+1} - \mu_M^k) \leq 0. \tag{35c}$$

Then (35b) implies the first item of the proposition, which together with the system above shows also the second item. The third item is obtained in a way similar to Proposition 1; we omit the details.

The existence assertion follows by applying [18, Theorem 5] to (28). □

In Dantzig-Wolfe decomposition for VIs, the role of using approximations is (at least) two-fold. First, when F is not monotone, using its appropriate approximation (for example, the value of F at the current point) would induce monotonicity in the subproblem, which is a desirable feature both theoretically and computationally. Second, when using F itself yields a subproblem which does not decompose further, an appropriate approximation may induce this further decomposition. Again, we refer to [14] for details and discussions.

In the Benders approach above, we face some difficulties. Note that the function F_k^{-1} is defined as an approximation of F^{-1} around w_M^k. In practice, it may be difficult to obtain enough information about F^{-1}, and so choosing F_k may not be so easy. In the case when F is monotone, we can choose F_k as F itself. However, F_k would not have further decomposable structure, if F does not have it. Also, there exist important classes of problems (e.g., Nash games), where the corresponding operator F is typically nonmonotone. In Sections 3.2 and 3.3, we develop suitable constructions for stochastic Nash games with risk aversion and with smoothed AVaR risk measures, respectively.

The option that is always available is to take

$$F_k^{-1}(w) = x_M^k, \text{ for all } w.$$

Observe that in this case, the subproblem (28) is equivalent to

$$\text{VI}(\Phi_k(\mu), \mathbb{R}_+^p), \tag{36}$$

where

$$\Phi_k(\mu) = d - Ax_M^k - B(z_M^k + \theta_M^k) + (AQ_kA^\top + P_k)(\mu - \mu_M^k).$$

The important issue is that, in this case, (36) is the optimality condition for the following *strongly convex* (if P_k and Q_k are symmetric positive definite) *quadratic programming problem* with simple bounds:

$$\begin{array}{ll} \min & \frac{1}{2}\langle (AQ_kA^\top + P_k)(\mu - \mu_M^k), \mu - \mu_M^k\rangle + \langle d - Ax_M^k - B(z_M^k + \theta_M^k), \mu\rangle \\ \text{s.t.} & \mu \geq 0. \end{array}$$

This approach is thus quite appealing, as there is a wealth of powerful software to solve such simple problems. Also in this case, $w_S^{k+1} = -A^\top \mu_S^{k+1}$.

Observe that we can always choose the matrices Q_k and P_k in a convenient way, preserving any structure of the problem data. For example, as diagonal matrices, or block diagonal. This is particularly important when the matrix A has structure which we would like to exploit. Suitable choices maintain the decomposability in the matrix $AQ_kA^\top + P_k$, thus allowing (36) to split into smaller problems. Also, the set (34) splits according to the given pattern, which makes solving (32) much easier. Of course, in (33) the mapping $F_k(x)$ may not be decomposable for some choices (the entry corresponding to μ is clearly decomposable). However, even in that case we can still use special methods in order to take advantage of the structure of the feasible set (e.g., the parallel variable distribution coupled with sequential quadratic programming [19], if we are in the optimization setting). Of course, it is desirable to have F_k decomposable. In [14] various decomposable monotone operators F_{Pk} that approximate F around x_M^k are stated, even when F is nondecomposable or nonmonotone. Taking such approximations, we can then define F_k as

$$F_k(x) = F_{Pk}(x) + U_k(x - x_M^k), \tag{37}$$

where U_k is a positive definite matrix chosen in such a way that F_k is still decomposable and strongly monotone. Now, since F_k is injective we can ensure that F_k^{-1} approximates F^{-1} in the sense of (25). This way of defining F_k is especially useful when we are dealing with nontrivial functions, as is the case of risk-measures considered in the sequel.

3.2 Decomposition Scheme for Stochastic Nash Games

In this subsection, we describe in detail a decomposition scheme for Stochastic Nash Games (see, e.g., [15]), resulting from the Benders approach above. Assume we have N players and W scenarios. Each player $i \in \{1, \ldots, N\}$ solves

$$\begin{array}{c} \min_{(q^i, z^i)} \; I^i(z^i, z^{-i}) + R^i\Big(\big(f_\omega^i(q_\omega^i, q_\omega^{-i})\big)_{w=1}^W\Big) \\ \text{s.t. } A_\omega q_\omega + B_\omega z \leq d_\omega, \quad \text{for } \omega = 1, \ldots, W, \end{array} \tag{38}$$

where (q^i, z^i) denote the decision variables of player i, while (q^{-i}, z^{-i}) are the decision variables of the other players, and I^i and f_ω^i are real-valued functions.

In this kind of settings, z model "here and now" variables that must be decided before any future event takes place, while q^i_ω are "wait and see" variables which depend on future uncertain events (scenario ω). That is why each player uses a risk measure R^i to hedge against uncertain events. Note that the scenario-dependent variables are coupled by the risk measure and the constraints.

To compute a variational equilibrium of this game we need to solve

$$VI\Big((F(q), G(z)), \{(q,z) : Aq + Bz \leq d)\}\Big)$$

where $q = (q_1, \dots, q_W)$,

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_W \end{bmatrix}, \quad B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_W \end{bmatrix}, \quad d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_W \end{bmatrix}$$

and for $\omega = 1, \dots, W$ we have $q_\omega = (q^1_\omega, \dots, q^N_\omega)$,

$$G(z) = \big(\partial_{z^i} I^i(z)\big)_{i=1}^N$$

and

$$F(q) = \left(\left(\partial_{q^i_\omega}\left[R^i\Big(\big(f^i_l(q_l)\big)_{l=1}^W\Big)\right]\right)_{i=1}^N\right)_{\omega=1}^W.$$

The notation $\partial\phi$ stands for the Convex Analysis subdifferential.

The matrix A naturally decomposes along scenarios, the risk measure R^i, on the other hand, couples inter-scenarios variables. For this reason, it is important to construct approximations for F around a point q_M that are decomposable along scenarios (and monotone). In order to do so, first note that

$$\partial_{q^i_\omega}\left[R^i\Big(\big(f^i_l(q_l)\big)_{l=1}^W\Big)\right] = \partial_\omega R^i\Big(\big(f^i_l(q_l)\big)_{l=1}^W\Big)\partial_{q^i_\omega} f^i_\omega(q_\omega).$$

For each scenario ω, the corresponding VI operator $F_\omega(q_\omega)$ giving a variational equilibrium is

$$F_\omega(q_\omega) = \Big(\partial_{q^i_\omega} f^i_\omega(q_\omega)\Big)_{i=1}^N.$$

The *decomposable monotone* approximation of $F_\omega(q_\omega)$ around $q_{M\,\omega}$

$$\hat{F}_\omega(q_\omega) = \Big(\hat{F}^i_\omega(q^i_\omega)\Big)_{i=1}^N \tag{39}$$

can be defined in various ways, the simplest is the constant approximation

$$\hat{F}^i_\omega(q^i_\omega) = F^i_\omega(q^i_{M\,\omega}).$$

More sophisticated options can be obtained proceeding in two steps, as follows.

1. First, a monotone approximation $\mathbb{F}_\omega$ is built. This is a tricky part, since building a good monotone approximation depends on the nature of the problem at hand, to which it must be tailored accordingly. But general approaches are also possible. For example:
   - $\mathbb{F}_\omega(q_\omega) = F_\omega(q_{M\,\omega})$
   - $\mathbb{F}_\omega(q_\omega) = F_\omega(q_{M\,\omega}) + M_\omega(q_\omega - q_{M\,\omega})$, where M_ω is some positive semidefinite matrix.
2. Next, the decomposable monotone operator $\hat{F}_\omega$ is defined by

$$\hat{F}^i_\omega(q^i_\omega) = \mathbb{F}_\omega(q^i_\omega, q^{-i}_{M\,\omega}), \quad \text{for } i = 1, 2, \dots, N.$$

Using the constructions above, we build the monotone approximation

$$\widetilde{F}(q) = \left(\left(\widetilde{F}^i_\omega(q^i_\omega)\right)_{i=1}^N\right)_{w=1}^W$$

of F by setting

$$\widetilde{F}^i_\omega(q^i_\omega) = \begin{cases} \partial_\omega R^i\Big(\big(f^i_l(q_{M\,l})\big)_{l=1}^W\Big)\hat{F}^i_\omega(q^i_\omega) & \text{if} \quad \partial_\omega R^i\Big(\big(f^i_l(q_{M\,l})\big)_{l=1}^W\Big) \geq 0, \\ \partial_\omega R^i\Big(\big(f^i_l(q_{M\,l})\big)_{l=1}^W\Big)\hat{F}^i_\omega(q_{M\,\omega}^{\;i}) & \text{otherwise.} \end{cases}$$

Taking block-diagonal matrices Q_k and P_k,

$$Q_k = \begin{bmatrix} Q_{k\,1} & & & \\ & Q_{k\,2} & & \\ & & \ddots & \\ & & & Q_{k\,W} \end{bmatrix} \quad \text{and} \quad P_k = \begin{bmatrix} P_{k\,1} & & & \\ & P_{k\,2} & & \\ & & \ddots & \\ & & & P_{k\,W} \end{bmatrix},$$

the structure is preserved in the product below

$$AQ_kA^\top + P_k = \begin{bmatrix} A_1Q_{k\,1}{A_1}^\top + P_{k\,1} & & & \\ & A_2Q_{k\,2}{A_2}^\top + P_{k\,2} & & \\ & & \ddots & \\ & & & A_WQ_{k\,W}{A_W}^\top + P_{k\,W} \end{bmatrix}.$$

As a result, solving the Benders subproblem (32) is equivalent to solving W smaller problems

$$VI(\hat{F}^k_{P_\omega}, S_{P_\omega}(z^k_M, \theta^k_M)), \tag{40}$$

$\omega = 1, \dots, W$, where

$$\hat{F}^k_{P_\omega}(q_\omega, \mu_\omega) = \begin{pmatrix} \left(\widetilde{F}^i_\omega(q^i_\omega)\right)_{i=1}^N \\ (A_\omega Q_k A^\top_\omega + P_{k\omega})\mu_\omega \end{pmatrix}, \tag{41}$$

and

$$S_{P_\omega}(z^k_M, \theta^k_M) = \left\{ (q_\omega, \mu_\omega) : \begin{array}{c} A_\omega q_\omega + B_\omega(z^k_M + \theta^k_M) \leq d_\omega + \\ (A_\omega Q_{k\omega} A^\top_\omega + P_{k\omega})(\mu_\omega - {\mu_M}^k_\omega) \end{array} \right\}. \tag{42}$$

3.3 Nash Game with Explicit Smoothed AVaR.

One special case associated to the stochastic Nash equilibrium is when the risk measure R^i in (38) is a smoothing of the average value-at-risk (AVaR), see [15]. Specifically, let

$$R^i(u) = \min_{v^i} \left\{ v^i + \frac{1}{1-\varepsilon_i} \mathbb{E}\left[\sigma^i_{\tau^i}(u_\omega - v^i)\right] \right\},$$

where $\sigma^i_\tau(\cdot)$ is an appropriate smoothing of the plus-function $\max\{0,\cdot\}$. In this case we can write (38) as

$$\begin{array}{ll} \min_{(q^i,z^i,v^i)} I^i(z^i,z^{-i}) + v^i + \frac{1}{1-\varepsilon_i}\mathbb{E}\left[\sigma^i_{\tau^i}\Big(\big(f^i_\omega(q^i_\omega, q^{-i}_\omega)\big)_{\omega=1}^W - v^i\Big)\right] \\ \qquad \text{s.t. } A_\omega q_\omega + B_\omega z \le d_\omega, \quad \text{for } \omega = 1,\ldots,W, \end{array} \tag{43}$$

and the associated VI is

$$VI\Big((F(q,v), G(z)), \{(q,v,z) : [A, 0]\begin{bmatrix} q \\ v \end{bmatrix} + Bz \le d)\}\Big),$$

where $v = (v^1, v^2, \ldots, v^N) \in \mathbb{R}^N$ and

$$\begin{aligned} F(q,v) = &\Bigg(\Bigg(\Bigg(\partial_{q^i_\omega}\left[\frac{1}{1-\varepsilon_i}\mathbb{E}\left[\sigma^i_{\tau^i}\Big(f^i_l(q_l) - v^i\Big)\right]\right]\Bigg)_{i=1}^N\Bigg)_{\omega=1}^W, \\ &\Bigg(\partial_{v^i}\left[v^i + \frac{1}{1-\varepsilon_i}\mathbb{E}\left[\sigma^i_{\tau^i}\Big(f^i_\omega(q_\omega) - v^i\Big)\right]\right]\Bigg)_{i=1}^N\Bigg) \\ = &\Bigg(\Bigg(\Bigg(\frac{P_\omega}{1-\varepsilon_i}(\sigma^i_{\tau^i})'\Big(f^i_\omega(q_\omega) - v^i\Big)\partial_{q^i_\omega} f^i_\omega(q_\omega)\Bigg)_{i=1}^N\Bigg)_{\omega=1}^W, \\ &\Bigg(1 - \frac{1}{1-\varepsilon_i}\mathbb{E}\left[(\sigma^i_{\tau^i})'\Big(f^i_\omega(q_\omega) - v^i\Big)\right]\Bigg)_{i=1}^N\Bigg). \end{aligned} \tag{44}$$

Observe that variables v^i are coupling the scenario-dependent variables q_ω. To overcome this difficulty, we can take the monotone approximation (39) of $F_\omega(q_\omega) = \Big(\partial_{q^i_\omega} f^i_\omega(q_\omega)\Big)_{i=1}^N$. Using the approximation $\hat{F}_\omega(q_\omega) = \Big(\hat{F}^i_\omega(q^i_\omega)\Big)_{i=1}^N$ defined in the previous section, we construct the monotone approximation $\widetilde{F}(q,v) = (((\widetilde{F}^i_\omega(q^i_\omega))_{\omega=1}^W)_{i=1}^N, (\widetilde{F}^i_v(v^i))_{i=1}^N)$ of $F(q,v)$ by

$$\widetilde{F}^i_\omega(q^i_\omega) = \begin{cases} \dfrac{P_\omega}{1-\varepsilon^i}(\sigma^i_{\tau^i})'\big(f^i_\omega(q_{M\,\omega}) - v^i_M\big)\hat{F}^i_\omega(q^i_\omega) & \text{if } (\sigma^i_{\tau^i})'\big(f^i_\omega(q_{M\,\omega}) - v^i_M\big) \ge 0 \\ \dfrac{P_\omega}{1-\varepsilon^i}(\sigma^i_{\tau^i})'\big(f^i_\omega(q_{M\,\omega}) - v^i_M\big)\hat{F}^i_\omega(q_{M}{}^i_\omega) & \text{otherwise,} \end{cases} \tag{45}$$

and

$$\widetilde{F}_v^i(v^i) = 1 - \frac{1}{1-\varepsilon_i}\mathbb{E}\left[(\sigma^i_{\tau^i})'\Big(f^i_\omega(q_{M\omega}) - v^i\Big)\right].$$

We are interested in the structure of the Benders subproblem (32) in this particular case. The corresponding feasible set (34) is given by elements (q, v, μ) that satisfy

$$[A,0]\begin{bmatrix} q \\ v \end{bmatrix} + B(z_M^k + \theta_M^k) \le d + \left([A,0]\begin{bmatrix} Q_k & Q_k^{12} \\ Q_k^{21} & Q_k^{22}\end{bmatrix}\begin{bmatrix} A^\top \\ 0 \end{bmatrix} + P_k\right)(\mu - {\mu_M}^k)\},$$

where the matrices Q_k^{ij} are chosen so that $\begin{bmatrix} Q_k & Q_k^{12} \\ Q_k^{21} & Q_k^{22}\end{bmatrix}$ is positive definite. Next, performing the matrix products that define this feasible set, we obtain that it actually has the following form:

$$\{(q, v, \mu) : Aq + B(z_M^k + \theta_M^k) \le d + (AQ_kA^\top + P_k)(\mu - {\mu_M}^k)\}.$$

In particular, we see that the variable v is free, and this set has decomposable structure along scenarios. Even more, since the approximating operator $\widetilde{F}(q, v)$ is decomposable by scenarios, and since the variable v is free, it turns out that solving the subproblem (32) is equivalent to solving the smaller problems (40) – (42), where VI operators are defined by (37) and (45). The output of these subproblems gives the values of q_S, μ_S that will be required by the master problem. Note also that formally, for solving "in full" the Benders subproblem (32), since the variables v^i are free, we need also to solve the equation

$$\widetilde{F}_v^i(v^i) + U(v^i - v_M^i) = 0.$$

Even though of easy solution ($U > 0$), in practice the equation is not solved, because the master problem does not require that information.

4 Benders master program

As already mentioned, the full Benders decomposition for VIs is stated in Section 5 further below. In Section 3, we discussed the Benders subproblems block. In this section, we deal with the second block; in particular, we relate Benders master problem for the primal VI to the Dantzig-Wolfe master program applied to the dual VI.

At iteration $k \ge 0$, given $X^{k+1} = \{(w_S^i, \zeta_S^i, \mu_S^i)\}_{i=0}^{k+1}$, the kth master problem consists in solving

$$\mathrm{VI}(F_D, S_{D_M}), \quad S_{D_M} = \{(w, \zeta, \mu) \in \operatorname{conv} X^{k+1} : 0 = \zeta + B^\top \mu\}. \qquad (46)$$

Under the technical assumptions stated in the beginning of Section 2, it holds that F_D is maximally monotone. Then, since S_{D_M} is compact, the master problem (46) is solvable [18, Theorem 5].

Let us remind that the points $(w_S^i, \zeta_S^i, \mu_S^i)$, for $i \ge 1$, are computed by solving the subproblems described in Section 3 above, whereas $(w_S^0, \zeta_S^0, \mu_S^0)$ is a feasible point of (18) chosen at the beginning of the algorithm (as prescribed by the Dantzig-Wolfe framework in [14]).

Using the matrices

$$W_{k+1} = [w_S^0 | w_S^1 | \cdots | w_S^{k+1}],$$
$$Z_{k+1} = [\zeta_S^0 | \zeta_S^1 | \cdots | \zeta_S^{k+1}],$$
$$M_{k+1} = [\mu_S^0 | \mu_S^1 | \cdots | \mu_S^{k+1}],$$

problem (46) can be reformulated as

$$\mathrm{VI}(F_{D_\triangle}, S_{D_\triangle}), \tag{47}$$

where

$$F_{D_\triangle}(w, \zeta, \alpha) = F^{-1}(w) \times G^{-1}(\zeta) \times \{M_{k+1}^\top d\},$$

and $S_{D_\triangle}$ is the set of points $(w, \zeta, \alpha) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{k+2}$ that satisfy the following constraints:

$$w - W_{k+1}\alpha = 0, \tag{48a}$$
$$\zeta + B^\top M_{k+1}\alpha = 0, \tag{48b}$$
$$\zeta - Z_{k+1}\alpha = 0, \tag{48c}$$
$$1 - \mathbf{1}^\top \alpha = 0, \tag{48d}$$
$$\alpha \geq 0, \tag{48e}$$

where $\mathbf{1}$ is the vector of ones of the appropriate dimension.

It is worth to note that since all points $(w_S^i, \zeta_S^i, \mu_S^i)$ are feasible for (22), we have that $W_{k+1} = -A^\top M_{k+1}$.

For the same reasons as above (i.e., the involvement of the likely difficult inverse functions), instead of solving directly (47), we shall solve instead its dual, described next.

Proposition 4 *Defining*

$$F_P(x, z, \beta, \theta) = F(x) \times G(z) \times \{-1\} \times \{0\}, \tag{49}$$

and

$$S_{P_{k+1}} = \{(x, z, \beta, \theta) : M_{k+1}^\top [Ax + Bz - d] \leq -\mathbf{1}\beta - [Z_{k+1}^\top + M_{k+1}^\top B]\theta\}, \tag{50}$$

we have that if $(x_M^{k+1}, z_M^{k+1}, \beta_M^{k+1}, \theta_M^{k+1})$ *solves*

$$VI(F_P, S_{P_{k+1}}), \tag{51}$$

with some $w_M^{k+1} \in F(x_M^{k+1})$, $\zeta_M^{k+1} \in G(z_M^{k+1})$ *and some Lagrange multiplier* α_M^{k+1}, *then* $(w_M^{k+1}, \zeta_M^{k+1}, \alpha_M^{k+1})$ *solves* $VI(F_{D_\triangle}, S_{D_\triangle})$ *with* $x_M^{k+1} \in F^{-1}(w_M^{k+1})$, $z_M^{k+1} \in G^{-1}(\zeta_M^{k+1})$ *and Lagrange multipliers* $-x_M^{k+1}$, $-z_M^{k+1} - \theta_M^{k+1}$, θ_M^{k+1}, β_M^{k+1} *and* $M_{k+1}^\top[d - Ax_M^{k+1} - Bz_M^{k+1}] - [Z_{k+1}^\top + M_{k+1}^\top B]\theta_M^{k+1} - \mathbf{1}\beta_M^{k+1}$.

Also, $VI(F_P, S_{P_{k+1}})$ *has solutions if, and only if,* $VI(F_{D_\triangle}, S_{D_\triangle})$ *has solutions.*

Proof The proof is analogous to that of Proposition 1. □

5 Benders decomposition algorithm for VIs and its convergence analysis

We now state formally the algorithm and then analyze its convergence properties. The Benders decomposition for VI (17) follows the following pattern.

Algorithm 2 (Benders Decomposition for VIs)

1. Choose a feasible dual point $(w_M^0, \zeta_M^0, \mu_M^0) \in S_D$. Choose $x_M^0 \in F(w_M^0)$, $z_M^0 \in G(\zeta_M^0)$ and $\theta_M^0 \in \mathbb{R}^m$. Set $w_S^0 = \mu_M^0$, $\mu_M^0 = \mu_S^0$, $\zeta_M^0 = \zeta_S^0$ and $k := 0$.
2. **Subproblem solution:** Choose mappings approximations $F^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $G^k : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ of $F(\cdot)$ and $G(\cdot)$, and symmetric positive (semi)definite matrices $Q_k \in \mathbb{R}^{n\times n}$, $R_k \in \mathbb{R}^{m\times m}$ and $P_k \in \mathbb{R}^{p\times p}$. Find the primal-dual points (x_S^{k+1}, z_S^{k+1}) and $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})$ by solving the problems (27) or (31), and (32) or (36), according to the approximation functions chosen.
3. **Master program:** Find $(x_M^{k+1}, z_M^{k+1}, \theta_M^{k+1})$ and $(w_M^{k+1}, \zeta_M^{k+1}, \mu_M^{k+1})$ by solving (51), with a Lagrange multiplier α_M^{k+1} associated to the constraint in (50). Compute $\mu_M^{k+1} = M_{k+1}\alpha_M^{k+1}$.
4. Set $k := k + 1$ and go to Step 2.

We proceed to analyze convergence properties of Algorithm 2. We start with the associated convergence gap quantity. For the Dantzig-Wolfe method, the *gap of convergence* Δ_k is defined by [14, Eq (19)]. This quantity drives convergence of the algorithm, closing the gap as it tends to zero. In the Benders setting, this would be clear, for example, from Proposition 5 below (which shows that as Δ_k tends to zero, so is the distance between the master problems and the subproblems solutions); see also the subsequent Theorem 1 for the eventual consequences of this fact.

In the case under consideration, Δ_k defined by [14, Eq (19)] has the form

$$\begin{aligned}\Delta_k &= \Big\langle (x_M^k, -\theta_M^k, d - Bz_M^k - B\theta_M^k), (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k) \Big\rangle \\ &= \langle x_M^k, w_S^{k+1} - w_M^k\rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k\rangle \\ &\quad + \langle d - Bz_M^k - B\theta_M^k, \mu_S^{k+1} - \mu_M^k\rangle,\end{aligned} \tag{52}$$

where we have used the fact that

$$(x_M^k, z_M^k, d) + (0, -z_M^k - \theta_M^k, -Bz_M^k - B\theta_M^k) \in \hat{F}_D^k(w_M^k, \zeta_M^k, \mu_M^k),$$

which follows from (24).

Since $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) \in S_{D_S}$ and $(w_M^k, \zeta_M^k, \mu_M^k) \in S_D$, we further obtain that

$$\begin{aligned}\Delta_k &= \langle Ax_M^k, \mu_M^k - \mu_S^{k+1}\rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k\rangle + \langle d - Bz_M^k - B\theta_M^k, \mu_S^{k+1} - \mu_M^k\rangle \\ &= \langle Ax_M^k + Bz_M^k - d, \mu_M^k - \mu_S^{k+1}\rangle + \langle -\theta_M^k, \zeta_S^{k+1} - \zeta_M^k\rangle - \langle B\theta_M^k, \mu_S^{k+1} - \mu_M^k\rangle \\ &= \langle d - Ax_M^k - Bz_M^k, \mu_S^{k+1} - \mu_M^k\rangle - \langle \theta_M^k, \zeta_S^{k+1} + B^\top \mu_S^{k+1}\rangle.\end{aligned}$$

Proposition 5 *If for each k the function $\hat{F}_D^k$ is chosen c_k-strongly monotone, then*

$$c_k\|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 + \Delta_k \leq 0. \tag{53}$$

In particular, $\Delta_k \leq 0$ for every k.

Proof Using the c_k-strong monotonicity of $\hat{F}_D^k$, since

$$(x_M^k, -\theta_M^k, d - Bz_M^k - B\theta_M^k) \in \hat{F}_D^k(w_M^k, \zeta_M^k, \mu_M^k)$$

and since $(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})$ solves $\mathrm{VI}(\hat{F}_D^k, S_{D_S})$ with some

$$y_S^{k+1} \in \hat{F}_D^k(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}),$$

we have that

$$\begin{aligned} &c_k\|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 \\ &\leq \langle y_S^{k+1} - (x_M^k, -\theta_M^k, d - Bz_M^k - B\theta_M^k), (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\rangle. \end{aligned}$$

We then further obtain

$$\begin{aligned} &c_k\|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\|^2 \\ &\leq \langle y_S^{k+1}, (w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\rangle - \Delta_k. \end{aligned}$$

Using the latter relation and $(w_M^k, \zeta_M^k, \mu_M^k) \in S_{D_S}$, (53) follows. □

We are now in position to state convergence properties of Benders decomposition of VIs with the given structure. Theorem 1 below assumes that F and G are outer semicontinuous (the definition is given in the end of Section 1), which is the only property used in the proof, if the existence of solutions/iterations is a given. Outer semicontinuity is in fact automatic from the initial assumptions stated in the beginning of Section 2, which also guarantee the existence of solutions of all the (sub)problems along the iterations. We also assume the equicontinuity (the definition is given in the end of Section 1) of the approximating families $\{F_k^{-1}\}$ and $\{G_k^{-1}\}$. Again, there is a number of ways to ensure the latter. Constant approximations is one option. If first-order (Newtonian) approximations are used for single-valued smooth data, choosing bounded $\{H_{F_k}^{-1}\}$ does the job. Finally, if the exact information F and G is employed (no approximations) in the single-valued case, the continuity of those functions is sufficient.

Theorem 1 *Suppose that F and G are outer semicontinuous (which holds, in particular, if they are maximally monotone). For the iterative sequences generated by Algorithm 2, the following holds.*

1. *If the sequences $\{\mu_S^{k+1}\}$, $\{z_S^{k+1}\}$, $\{z_M^k\}$ and $\{\theta_M^k\}$ are bounded and if the family of matrices $\{R_k\}$ is uniformly positive definite, then the sequences $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ and $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ are bounded.*

2. *If the sequences* $\{x_M^k\}$, $\{\mu_S^{k+1}\}$, $\{z_S^{k+1}\}$, $\{z_M^k\}$ *and* $\{\theta_M^k\}$ *are bounded, the family of matrices* $\{R_k\}$ *is uniformly positive definite, and the approximations* $\{\hat{F}_D^k\}$ *are chosen monotone, then*

$$\lim_{k\to\infty} \Delta_k = 0.$$

In particular, if the elements of $\{\hat{F}_D^k\}$ *are chosen uniformly strongly monotone, then*

$$\lim_{k\to\infty} \|(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1}) - (w_M^k, \zeta_M^k, \mu_M^k)\| = 0. \tag{54}$$

3. *Suppose that condition* (54) *holds; and the sequences* $\{P_k\}$,$\{Q_k\}$,$\{R_k\}$, $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ *and* $\{(x_S^{k+1}, z_S^{k+1})\}$ *are bounded. Then, if the approximations* $\{F_k^{-1}\}$ *and* $\{G_k^{-1}\}$ *are equicontinuous on compact sets, every cluster point of* $\{(x_S^{k+1}, z_S^{k+1})\}$ *is a solution of VI* (17).

Proof 1. Let $\{\mu_S^{k+1}\}$ be bounded. Since $\mu_M^k \in \text{conv}(\{\mu_S^j\})$, it follows that the sequence $\{\mu_M^k\}$ is bounded. Furthermore, as

$$w_S^{k+1} = -A^\top \mu_S^{k+1}, \quad w_M^k = -A^\top \mu_M^k, \quad \zeta_M^k = -B^\top \mu_M^k,$$

it follows that the sequences $\{w_S^{k+1}\}$ and $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ are bounded. Now boundedness of $\{\zeta_S^{k+1}\}$ follows from

$$z_S^{k+1} - z_M^k - \theta_M^k + R_k(\zeta_S^{k+1} - \zeta_M^k) = 0$$

and the uniform positive definite property of the family $\{R_k\}$.

2. By the first item, we have that the sequences $\{(w_S^{k+1}, \zeta_S^{k+1}, \mu_S^{k+1})\}$ and $\{(w_M^k, \zeta_M^k, \mu_M^k)\}$ are bounded.
Using (53) with $c_k = 0$ (i.e., monotonicity instead of strong monotonicity), we have that $\Delta_k \le 0$. Hence,

$$\bar{\Delta} = \liminf_{k\to\infty} \Delta_k \le \limsup_{k\to\infty} \Delta_k \le 0.$$

We take a subsequence $\{\Delta_{k_j}\}$ such that $\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta}$. Without loss of generality, we can assume convergence of the corresponding subsequences, in particular: $\{(w_M^{k_j}, \zeta_M^{k_j}, \mu_M^{k_j})\} \to (\bar{w}, \bar{\zeta}, \bar{\mu})$, $\{(x_M^{k_j}, z_M^{k_j}, \theta_M^{k_j})\} \to (\bar{x}, \bar{z}, \bar{\theta})$, $\{(w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1})\} \to (\hat{w}, \hat{\zeta}, \hat{\mu})$. Then from (52), we have that

$$\lim_{j\to\infty} \Delta_{k_j} = \bar{\Delta} = \Big\langle (\bar{x}, -\bar{\theta}, d - B(\bar{z} + \bar{\theta})), (\hat{w}, \hat{\zeta}, \hat{\mu}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \Big\rangle.$$

Consider the problem that results from $\text{VI}(F_D, S_{D_M})$ given by (46), after relaxing the constraint $\zeta + B^\top \mu = 0$ using the multiplier $(-z_M^{k_i} - \theta_M^{k_i})$. Fix any index j. Since $(w_M^{k_i}, \zeta_M^{k_i}, \mu_M^{k_i})$ is a solution in (46), and recalling the definition (19) of F_D, we have that

$$\left\langle \begin{bmatrix} x_M^{k_i} \\ z_M^{k_i} \\ d \end{bmatrix} + \begin{bmatrix} 0 \\ I \\ B \end{bmatrix} (-z_M^{k_i} - \theta_M^{k_i}), \begin{bmatrix} w \\ \zeta \\ \mu \end{bmatrix} - \begin{bmatrix} w_M^{k_i} \\ \zeta_M^{k_i} \\ \mu_M^{k_i} \end{bmatrix} \right\rangle \ge 0,$$

for any $(w, \zeta, \mu) \in S_{D_M}$. Note further that for every $i > j$, we have that $(w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) \in S_{D_M}$. Therefore,

$$\left\langle (x_M^{k_i}, -\theta_M^{k_i}, d - B(z_M^{k_i} + \theta_M^{k_i})), (w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) - (w_M^{k_i}, \zeta_M^{k_i}, \mu_M^{k_i}) \right\rangle \geq 0,$$

and passing onto the limit as $i \to \infty$, we obtain that

$$\left\langle (\bar{x}, -\bar{\theta}, d - B(\bar{z} + \bar{\theta})), (w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \right\rangle \geq 0.$$

Passing onto the limit again, now as $j \to \infty$, we obtain that

$$\bar{\Delta} = \left\langle (\bar{x}, -\bar{\theta}, d - B(\bar{z} + \bar{\theta})), (\hat{w}, \hat{\zeta}, \hat{\mu}) - (\bar{w}, \bar{\zeta}, \bar{\mu}) \right\rangle \geq 0,$$

which shows that $\lim_{k\to\infty} \Delta_k = 0$. Finally, (53) implies the last assertion of this item.

3. Suppose that $(\bar{x}, \bar{z})$ is an accumulation point of $\{(x_S^{k+1}, z_S^{k+1})\}$ and that the subsequence $\{(x_S^{k_j+1}, z_S^{k_j+1})\}$ converges to it as $j \to \infty$. Since

$$z_S^{k_j+1} - z_M^{k_j} - \theta_M^{k_j} + R_{k_j}(\zeta_S^{k_j+1} - \zeta_M^{k_j}) = 0,$$

using the stated hypotheses we conclude that

$$\lim_{j\to\infty} (z_M^{k_j} + \theta_M^{k_j}) = \bar{z}.$$

On the other hand, taking into account (54) and passing onto further subsequences if necessary, we can assume that

$$\lim_{j\to\infty} w_S^{k_j+1} = \lim_{j\to\infty} w_M^{k_j} = \bar{w},$$
$$\lim_{j\to\infty} \zeta_S^{k_j+1} = \lim_{j\to\infty} \zeta_M^{k_j} = \bar{\zeta},$$
$$\lim_{j\to\infty} \mu_S^{k_j+1} = \lim_{j\to\infty} \mu_M^{k_j} = \bar{\mu}.$$

Since the families of approximations $\{F_k^{-1}\}$ and $\{G_k^{-1}\}$ are equicontinuous on compact sets, there exists, for each t, some k_{j_t} such that for every k it holds that

$$d_H\left(F_k^{-1}(w_S^{k_{j_t}+1}), F_k^{-1}(w_M^{k_{j_t}})\right) < \frac{1}{t}.$$

In particular, there exists $\hat{x}_M^{k_{j_t}} \in F_k^{-1}(w_M^{k_{j_t}}) \subset F^{-1}(w_M^{k_{j_t}})$ such that

$$\|x_S^{k_{j_t}+1} - \hat{x}_M^{k_{j_t}}\| < \frac{1}{t}.$$

Hence,

$$\lim_{t\to\infty} \hat{x}_M^{k_{j_t}} = \bar{x}.$$

Then, since F is outer semicontinuous, we conclude that $\bar{x} \in F^{-1}(\bar{w})$. In a similar way we can conclude that $\bar{z} \in G^{-1}(\bar{\zeta})$. Also, note that since

$(w_S^{k_j+1}, \zeta_S^{k_j+1}, \mu_S^{k_j+1}) \in S_{D_S}$, we have that $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_{D_S}$, and since $(w_M^{k_j}, \zeta_M^{k_j}, \mu_M^{k_j}) \in S_{D_M}$, also $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_{D_M}$. Thus $(\bar{w}, \bar{\zeta}, \bar{\mu}) \in S_D$.
Finally, since $(w_S^{k_j+1}, \mu_S^{k_j+1})$ solves (28) with $x_S^{k_j+1} \in F_{k_j}^{-1}(w_S^{k_j+1})$, we have that for every $(w, \mu) \in S_{D_{S2}}$ the following inequality holds:

$$\begin{aligned}\langle x_S^{k_j+1} + Q_{k_j}(w_S^{k_j+1} - w_M^{k_j}), w - w_S^{k_j+1}\rangle \\ + \langle d - B(z_M^{k_j} + \theta_M^{k_j}) + P_{k_j}(\mu_S^{k_j+1} - \mu_M^{k_j}), \mu - \mu_S^{k_j+1}\rangle \geq 0. \quad (55)\end{aligned}$$

Thus, passing onto the limit as $j \to \infty$, we obtain that

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d - B\bar{z}, \mu - \bar{\mu}\rangle \geq 0,$$

that is,

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d, \mu - \bar{\mu}\rangle + \langle \bar{z}, -\bar{\zeta} - B^\top \mu\rangle \geq 0.$$

So, for each $(w, \zeta, \mu) \in S_D$ we have that

$$\langle \bar{x}, w - \bar{w}\rangle + \langle d, \mu - \bar{\mu}\rangle + \langle \bar{z}, \zeta - \bar{\zeta}\rangle \geq 0.$$

Since also $\bar{x} \in F^{-1}(\bar{w})$ and $\bar{z} \in G^{-1}(\bar{\zeta})$, we conclude that $(\bar{w}, \bar{\zeta}, \bar{\mu})$ is a solution of VI(F_D, S_D). The latter implies, by Proposition 1, that $(\bar{x}, \bar{z})$ is a solution of (17).

□

6 Concluding remarks

Whereas Dantzig-Wolfe decomposition deals with coupling constraints (a setting that fits naturally VIs arising from generalized Nash equilibrium problems, for example), the Benders decomposition algorithm is intended for VIs where we recognize the existence of "coupling variables" (denoted by z in this work). These coupling variables have the property that whenever they are fixed to some values, the resulting problem on the remaining variables (denoted by q in this work), becomes much easier to solve. This frequently occurs when the variables q have a decomposable structure that allows the problem to be split in several smaller ones. This is the case in a wide range of diverse applications; for example: power management, investment planning, capacity expansion, etc. (see [17]). To deal with such decomposable structures, we derived a fairly general Benders decomposition scheme for VIs by applying Dantzig-Wolfe decomposition to a suitably defined dual VI, and then passing back to the primal space. Among the specific contributions of our development is the ability to handle multi-valued mappings (previously only rather specific single-valued case was considered), the possibility to further decompose the subproblems using appropriate approximations of the VI data, as well as the possibility to reduce the subproblems to simple optimization problems. Finally, such developments also open the possibility of combining Benders decomposition (say, along scenarios) with Dantzig-Wolfe decomposition (say, along the players in the game). As an example, we considered the case of the Stochastic Nash Equilibrium Problem. There, when using the Benders decomposition scheme, at each iteration we fix the variable z to z_M^k and solve the subproblem (32). This problem can be split in smaller problems, along scenarios. Each of these smaller problems constitutes a generalized Nash equilibrium problem where the uncertainty has realized. If those smaller problems are still too hard to solve because of their size, they can be decomposed further, this time along the agents, via the Dantzig-Wolfe algorithm.

Acknowledgments

The authors thank the two referees for their useful comments.

All the three authors are supported by the FAPERJ Grant 203.052/2016 and by FAPERJ Grant E-26/210.908/2016 (PRONEX–Optimization). The second author's research is also supported by CNPq Grant 303905/2015-8, by Gaspard Monge Visiting Professor Program, and by EPSRC grant number EP/ R014604/1 (Isaac Newton Institute for Mathematical Sciences, programme Mathematics for Energy Systems). The third author is also supported by CNPq Grant 303724/2015-3, and by Russian Foundation for Basic Research Grant 19-51-12003 NNIOa.

References

1. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**(1), 1–24 (1996)
2. Auslender, A., Teboulle, M.: Lagrangian duality and related multiplier methods for variational inequality problems. SIAM J. Optim. **10**(4), 1097–1115 (2000).
3. Benders, J.F.: Partitioning procedures for solving mixed-variables programming problems. Numerische mathematik **4**(1), 238–252 (1962)
4. Bonnans, J., Gilbert, J., Lemaréchal, C., Sagastizábal, C.: Numerical Optimization: Theoretical and Practical Aspects. Springer, Berlin, Germany (2006). Second Edition
5. Burachik, R.S., Iusem, A.N.: Set-valued mappings and enlargements of monotone operators, *Springer Optimization and Its Applications*, vol. 8. Springer, New York (2008)
6. Çelebi, E., Fuller, J.D.: Master problem approximations in Dantzig-Wolfe decomposition of variational inequality problems with applications to two energy market models. Comput. Oper. Res. **40**(11), 2724–2739 (2013).
7. Chung, W., Fuller, J.D.: Subproblem approximation in Dantzig-Wolfe decomposition of variational inequality models with an application to a multicommodity economic equilibrium model. Operations Research **58**(5), 1318–1327 (2010)
8. Dantzig, G.B., Wolfe, P.: The decomposition algorithm for linear programs. Econometrica: Journal of the Econometric Society **29**(4), pp. 767–778 (1961)
9. Eckstein, J., Ferris, M.C.: Smooth methods of multipliers for complementarity problems. Math. Program. **86**(1, Ser. A), 65–90 (1999).
10. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Springer Science & Business Media (2007)
11. Fuller, J.D., Chung, W.: Dantzig-Wolfe decomposition of variational inequalities. Computational Economics **25**(4), 303–326 (2005)
12. Fuller, J.D., Chung, W.: Benders decomposition for a class of variational inequalities. European J. Oper. Res. **185**(1), 76–91 (2008).

13. Gabriel, S.A., Fuller, J.D.: A Benders Decomposition Method for Solving Stochastic Complementarity Problems with an Application in Energy. Computational Economics **35**(4), 301–329 (2010)
14. Luna, J.P., Sagastizábal, C., Solodov, M.: A class of Dantzig-Wolfe type decomposition methods for variational inequality problems. Mathematical Programming **143**(1-2), 177–209 (2014).
15. Luna, J.P., Sagastizábal, C., Solodov, M.: An approximation scheme for a class of risk-averse stochastic equilibrium problems. Mathematical Programming **157**(2), 451–481 (2016).
16. Mosco, U.: Dual variational inequalities. J. Math. Anal. Appl. **40**, 202–206 (1972)
17. Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W.: The Benders decomposition algorithm: A literature review. European Journal of Operational Research **259**(3), 801–817 (2017)
18. Rockafellar, R.T.: On the maximality of sums of nonlinear monotone operators. Trans. Amer. Math. Soc. **149**, 75–88 (1970)
19. Sagastizábal, C.A., Solodov, M.V.: Parallel variable distribution for constrained optimization. Comput. Optim. Appl. **22**(1), 111–131 (2002).
20. Solodov, M.V.: Constraint qualifications. In: J.J. Cochran, et al. (eds.) Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons, Inc. (2010)