

The Study of Coordinate-Wise Decomposition Descent Method for Optimization Problems

Salahuddin*

Department of Mathematics, Jazan University, Jazan-45142, Kingdom of Saudi Arabia

*Corresponding author: smohammad@jazanu.edu.sa, dr_salah12@yahoo.com

Abstract. The aim of this paper is to consider a general non-stationary optimization problem whose objective function need not be smooth in general and only approximation sequences are known instead of exact values of the functions. We apply a two-step technique where approximate solutions of a sequence of a generalized mixed variational inequality problem (GMVIP) are inserted in the iterative method of a selective coordinate-wise decomposition descent method. Its convergence is achieved under coercivity-type assumptions.

1. Introduction

This paper presents a systematic approach to non-stationary optimization problems for a class of non-stationary variational inequalities. The theoretical results are delivered in the general framework of abstract inequalities in an abstract space. The main feature of such general non-stationary optimizations lies in the fact that they are governed by a target function that cannot be smooth in general and only approximation sequences are known instead of exact values of the functions.

The variational inequalities appear in a variety of mathematical, physical and mechanical problems, *for example*, the unilateral contact problems in nonlinear elasticity, the problems describing the adhesive and friction effects, the nonconvex semi permeability problems, the nonlinear optimization problems, the masonry structures, and the delamination problems in multi-layered compositions. Variational inequalities have been introduced by G. Stampacchia [1] in 1968 as the variational formulation of important classes of unilateral, boundary-valued problems and inequality problems in mathematical

Received: Apr. 2, 2023.

2020 *Mathematics Subject Classification.* 49J40, 47H09, 47J20, 54H25.

Key words and phrases. Non-stationary optimization problems; auxiliary properties; decomposition descent method; coercivity assumptions.

mechanics. The notion of optimization problem is a generalization of variational inequality for a case where the function involved is nonconvex, nonsmoothed and nonstationary. They cover optimization problems for partial differential equations with nonmonotone, possibly multivalued and nonconvex nonlinearities. In the last few years, many kinds of optimization problems and variational inequalities have been investigated see, [2] and the study of optimization and variational inequalities has emerged today as a new and interesting branch of mathematics.

Various models in applied sciences can be conveniently formulated as optimization problems involving certain constraints/parameters. These constraints/parameters are either known or unknown but they often characterize some physical properties of the underlying model. In this context, the direct problem consists in solving the optimization problem. In recent years, the field of optimization and variational inequalities emerged as one of the most vibrant and developing branches of applied and industrial mathematics because of their wide applications, see [3, 4]. Rather recently, the decomposition approach was suggested for variational inequalities with binding constraints in [5]. Within this approach, the initial problem is treated as a two-level one using a share allocation procedure, leading to a set-valued variational inequality as a master problem. Further, a decomposable dual regularization (penalty) method that deals with a single-valued approximation of the master problem for each fixed share allocation was suggested. However, the optimization problem of identifying constraints/parameters in optimization and variational inequalities is still an untreated topic in the literature, which is the motivation of the present work.

The general optimization problem consists of finding the minimal value of some goal function p over the corresponding feasible set \mathcal{X} . For brevity, we write this problem as

$$\min_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}), \quad (1.1)$$

its solution set is denoted by \mathcal{X}^* and the optimal value of the function by p^* , defined by

$$p^* = \inf_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}).$$

We denote \mathbb{R}^s by the real s -dimensional Euclidean space, all elements of such spaces being column vectors represented by boldface, e.g. \mathbf{x} . For any vectors \mathbf{x} and \mathbf{y} of \mathbb{R}^s , we denote by $\langle \mathbf{x}, \mathbf{y} \rangle$ their scalar product, i.e.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^s x_i y_i,$$

and by $\|\mathbf{x}\|$ the Euclidean norm of \mathbf{x} , i.e.

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

We define for brevity $\mathbf{M} = \{1, \dots, n\}$, $|\mathcal{A}|$ will denote the cardinality of a finite set \mathcal{A} . As usual, \mathbb{R} will denote the set of real numbers, $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

Let us consider a partition of the N -dimensional space

$$\mathbb{R}^N = \mathbb{R}^{N_1} \times \dots \times \mathbb{R}^{N_n}, \tag{1.2}$$

i.e.

$$\mathcal{N} = \bigcup_{i=1}^n \mathcal{N}_i,$$

where $\mathcal{N} = \{1, \dots, N\}$, $N = |\mathcal{N}|$, $N_i = |\mathcal{N}_i|$, and

$$\mathcal{N}_i \cap \mathcal{N}_j = \emptyset, \text{ if } i \neq j.$$

This means that any point $\mathbf{x} = (x_1, \dots, x_N)^\top \in \mathbb{R}^N$ is represented by $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ where $\mathbf{x}_i = (x_j)_{j \in \mathcal{N}_i} \in \mathbb{R}^{N_i}$ for $i \in \mathbf{M}$. The simplest case, where $n_i = 1$ for all $i \in \mathbf{M}$ and $n = N$ corresponds to the scalar coordinate partition.

Assume that $f_1, f_2 : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ are two functions, which are continuous on \mathcal{X} and $f_1 + f_2 = f : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ is a continuous function on \mathcal{X} . The partially decomposable optimization problems played a significant role in various data type applications; see, e.g. [6–9]. In these problems, the cost function and feasible set are specialized as follows:

$$p(\mathbf{x}) = f(\mathbf{x}) + h(\mathbf{x}), \tag{1.3}$$

$$h(\mathbf{x}) = \sum_{i=1}^n h_i(\mathbf{x}_i), \tag{1.4}$$

$$\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n = \prod_{i=1}^n \mathcal{X}_i, \tag{1.5}$$

where $h_i : \mathbb{R}^{N_i} \rightarrow \bar{\mathbb{R}}$ is a convex function, and \mathcal{X}_i is a convex set in \mathbb{R}^{N_i} for $i = 1, \dots, n$. Note that the function $f : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ is not supposed to be convex in general. That is, we have to solve a non-convex non-differentiable optimization problem, which appears too difficult to solve with usual subgradient-type methods. One can develop efficient coordinate-wise decent decomposition methods for finding stationary points problem (1.1), (1.3)-(1.5) for a smooth f ; see, e.g. [7, 10–12]. Then the stationary points can be defined as a solution to the following generalized mixed variational inequality problems (GMVIP): Find a point $\mathbf{x}^* \in \mathcal{X}$ such that

$$\langle f'(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle + \sum_{i=1}^n [h_i(\mathbf{y}_i) - h_i(\mathbf{x}_i^*)] \geq 0, \forall \mathbf{y}_i \in \mathcal{X}_i \text{ for } i = 1, \dots, n \tag{1.6}$$

where $f'(\mathbf{x}^*) = f'_1(\mathbf{x}^*) + f'_2(\mathbf{x}^*)$ and $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top$.

In this paper, we intend to suggest a coordinate-wise decomposition descent method for the following problem: Find a point $\mathbf{x}^* \in \mathcal{X}$ such that

$$\exists \mathbf{g}^* \in \mathbf{G}(\mathbf{x}^*), \mathbf{q}^* \in \mathbf{Q}(\mathbf{x}^*), \langle \mathbf{g}^* + \mathbf{q}^*, \mathbf{y} - \mathbf{x}^* \rangle + \sum_{i=1}^n [h_i(\mathbf{y}_i) - h_i(\mathbf{x}_i^*)] \geq 0, \forall \mathbf{y}_i \in \mathcal{X}_i, \tag{1.7}$$

for $i = 1, \dots, n$, where $\mathbf{G}, \mathbf{Q} : \mathcal{X} \rightarrow \Pi(\mathbb{R}^N)$ are point to set mappings whose values are considered as generalized gradient sets of the functions f_1, f_2 respectively, i.e. $f; cf$ (1.6). Here $\Pi(A)$ denotes the family of all nonempty subsets of a set A . For instance, if f is a locally Lipschitz function, we can set \mathbf{G} and \mathbf{Q} to be its Clarke subdifferential mapping. If f is a convex function, simply set $\mathbf{G}(\mathbf{x})$ and $\mathbf{Q}(\mathbf{x})$ to be the usual subdifferential

$$\partial f(\mathbf{x}) = \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})$$

of f at \mathbf{x} , then each solution of (1.7) exactly solves (1.1), (1.3)-(1.5). Next, we suppose that only sequences of approximations are known instead of the exact values of \mathbf{G} , \mathbf{Q} and h .

2. Preliminaries

Let us consider a partially partitionable optimization problem of the form

$$\min_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x}) = \{\mu(\mathbf{x}) + \eta(\mathbf{x})\}, \quad (2.1)$$

where the function $\mu : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ is smooth on \mathcal{X} , but not necessary convex. This problem will serve as an approximation of the basic problem (1.1), (1.3)-(1.5). We use the same partition (1.2) of the space \mathbb{R}^N and fix the assumption on the feasible set.

(A1) It holds that (1.5) where \mathcal{X}_i are non-empty, convex and closed sets in \mathbb{R}^{N_i} for $i = 1, \dots, n$.

We suppose that

$$\eta(\mathbf{x}) = \sum_{i=1}^n \eta_i(\mathbf{x}_i), \quad (2.2)$$

where $\eta_i : \mathbb{R}^{N_i} \rightarrow \bar{\mathbb{R}}$ is convex and has the non-empty subdifferential $\partial \eta_i(\mathbf{x}_i)$ at each point $\mathbf{x}_i \in \mathcal{X}_i$, for $i \in \mathbf{M}$. Then each function η_i is lower semi-continuous, hence the function η is also lower semi-continuous, and

$$\partial \eta(\mathbf{x}) = \partial \eta_1(\mathbf{x}_1) \times \dots \times \partial \eta_n(\mathbf{x}_n).$$

Therefore, problems (2.1) and (2.2) are rewritten as

$$\min_{\mathbf{x} \in \mathcal{X}_1 \times \dots \times \mathcal{X}_n} \varphi(\mathbf{x}) = \left\{ \mu(\mathbf{x}) + \sum_{i=1}^n \eta_i(\mathbf{x}_i) \right\}. \quad (2.3)$$

Set $\mathbf{g}(\mathbf{x}) + \mathbf{q}(\mathbf{x}) = \mu'_1(\mathbf{x}) + \mu'_2(\mathbf{x}) = \mu'(\mathbf{x})$, then

$$\mathbf{g}(\mathbf{x}) + \mathbf{q}(\mathbf{x}) = (\mathbf{g}_1(\mathbf{x}) + \mathbf{q}_1(\mathbf{x}), \dots, \mathbf{g}_n(\mathbf{x}) + \mathbf{q}_n(\mathbf{x}))^\top,$$

where

$$\mathbf{g}_i(\mathbf{x}) + \mathbf{q}_i(\mathbf{x}) = \left\{ \frac{\partial \mu_1(\mathbf{x})}{\partial x_j} + \frac{\partial \mu_2(\mathbf{x})}{\partial x_j} \right\}_{j \in \mathcal{N}_i} = \left\{ \frac{\partial \mu(\mathbf{x})}{\partial x_j} \right\}_{j \in \mathcal{N}_i} \in \mathbb{R}^{N_i}, i = 1, \dots, n.$$

Given a point $\mathbf{x} \in \mathcal{X}$, we say that a vector \mathbf{d} is feasible for \mathbf{x} if $\mathbf{x} + \alpha \mathbf{d} \in \mathcal{X}$ for some $\alpha > 0$. From the above assumptions, it follows that the function φ is directionally differentiable at each point $\mathbf{x} \in \mathcal{X}$, that is, its directional derivative concerning any feasible vector \mathbf{d} is defined by the formula:

$$\varphi'(\mathbf{x}; \mathbf{d}) = \langle g(\mathbf{x}) + q(\mathbf{x}), \mathbf{d} \rangle + \eta'(\mathbf{x}; \mathbf{d}), \text{ with } \eta'(\mathbf{x}; \mathbf{d}) = \sum_{i=1}^n \max_{\mathbf{b}_i \in \partial \eta_i(x_i)} \langle \mathbf{b}_i, \mathbf{d}_i \rangle;$$

see, e.g. [14].

Recall a function $f : \mathbb{R}^s \rightarrow \bar{\mathbb{R}}$ is said to be coercive on a set $D \subset \mathbb{R}^s$ if $\{f(\mathbf{u}^k)\} \rightarrow +\infty$ for any sequence $\{\mathbf{u}^k\} \subset D, \|\mathbf{u}^k\| \rightarrow \infty$. Suppose that the function $\varphi : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$ is coercive on \mathcal{X} , then problem (2.1) and (2.2) (or (2.3)) has a solution.

Lemma 2.1. [15]

(a) Each solution of problem (2.3) is a solution of the following GMVIP: Find a point $\mathbf{x}^* \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ such that

$$\sum_{i=1}^n \langle \mathbf{g}_i(\mathbf{x}^*) + \mathbf{q}_i(\mathbf{x}^*), \mathbf{y}_i - \mathbf{x}_i^* \rangle + \sum_{i=1}^n [\eta_i(\mathbf{y}_i) - \eta_i(\mathbf{x}_i^*)] \geq 0, \forall \mathbf{y} \in \mathcal{X}. \tag{2.4}$$

(b) If μ is convex, then each solution of GMVIP (2.4) solves problem (2.3).

Now, we denote by $\tilde{\mathcal{X}}^0$ the solution set of GMVIP (2.4) and call it the set of stationary points of problem (2.3); cf(1.6).

Fix $\alpha > 0$. For each point $\mathbf{x} \in \mathcal{X}$ we can define $\mathbf{y}(\mathbf{x}) = (\mathbf{y}_1(\mathbf{x}), \dots, \mathbf{y}_n(\mathbf{x}))^\top \in \mathcal{X}$ such that

$$\sum_{i=1}^n \langle \mathbf{g}_i(\mathbf{x}) + \mathbf{q}_i(\mathbf{x}) + \alpha(\mathbf{y}_i(\mathbf{x}) - \mathbf{x}_i), \mathbf{y}_i - \mathbf{y}_i(\mathbf{x}) \rangle + \sum_{i=1}^n [\eta_i(\mathbf{y}_i) - \eta_i(\mathbf{y}_i(\mathbf{x}))] \geq 0, \forall \mathbf{y} \in \mathcal{X}. \tag{2.5}$$

This GMVIP gives a necessary and sufficient optimality condition for the optimization problem:

$$\min_{\mathbf{y} \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n} \sum_{i=1}^n \Phi_i(\mathbf{x}, \mathbf{y}_i), \tag{2.6}$$

$$\Phi_i(\mathbf{x}, \mathbf{y}_i) = \langle \mathbf{g}_i(\mathbf{x}) + \mathbf{q}_i(\mathbf{x}), \mathbf{y}_i \rangle + 0.5\alpha \|\mathbf{x}_i - \mathbf{y}_i\|^2 + \eta_i(\mathbf{y}_i), \text{ for } i = 1, \dots, n. \tag{2.7}$$

From the above assumptions each $\Phi_i(\mathbf{x}, \cdot)$ is strongly convex then problem (2.6) and (2.7) (or (2.5)) has the unique solution $\mathbf{y}(\mathbf{x})$, thus defining the single valued mapping $\mathbf{x} \mapsto \mathbf{y}(\mathbf{x})$. Observe that all the components of $\mathbf{y}(\mathbf{x})$ can be found independently, i.e. (2.6) and (2.7) is equivalent to n independent optimization problems of form

$$\min_{\mathbf{y}_i \in \mathcal{X}_i} \Phi_i(\mathbf{x}, \mathbf{y}_i), \tag{2.8}$$

for $i = 1, \dots, n$ and $\mathbf{y}_i(\mathbf{x})$ just solves (2.8).

Lemma 2.2. [15]

- (a) $\mathbf{x} = \mathbf{y}(\mathbf{x}) \Leftrightarrow \mathbf{x} \in \tilde{\mathcal{X}}^0$;
- (b) The mapping $\mathbf{x} \mapsto \mathbf{y}(\mathbf{x})$ is continuous on \mathcal{X} .

Set

$$\Delta(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}(\mathbf{x})\|,$$

then

$$\Delta^2(\mathbf{x}) = \sum_{i=1}^n \Delta_i^2(\mathbf{x})$$

where $\Delta_i(\mathbf{x}) = \|\mathbf{x}_i - \mathbf{y}_i(\mathbf{x})\|$.

From Lemma 2.2, we see that the value $\Delta(\mathbf{x})$ can serve as an accuracy measure for GMVIP (2.4).

Lemma 2.3. [15] *Take any point $\mathbf{x} \in \mathcal{X}$ and index $i \in M$. If*

$$\mathbf{d}_s = \begin{cases} \mathbf{y}_i(\mathbf{x}) - \mathbf{x}_i & \text{if } s = i, \\ \mathbf{0} & \text{if } s \neq i, \end{cases}$$

then

$$\varphi'(\mathbf{x}; \mathbf{d}) \leq -\alpha \|\mathbf{y}_i(\mathbf{x}) - \mathbf{x}_i\|^2.$$

Denote by \mathbb{Z}_+ the set of non-negative integers. From [15], we describe the basic algorithm for GMVIP (2.4) as follows

Algorithm 2.1. (Decomposition Descent Schemes (DDS.))

Input: A point $\mathbf{x}^0 \in \mathcal{X}$.

Output: A point \mathbf{z} .

Parameters: Numbers $\alpha > 0, \delta > 0, \beta \in (0, \alpha), \theta \in (0, 1)$.

At the κ -th iteration, $\kappa = 0, 1, \dots$, we have a point $\mathbf{x}^\kappa \in \mathcal{X}$.

Step 1: Choose an index $i \in \mathbf{M}$ such that $\Delta_i(\mathbf{x}^\kappa) \geq \delta$, set $i_\kappa = i$,

$$\mathbf{d}_s^\kappa = \begin{cases} \mathbf{y}_s(\mathbf{x}^\kappa) - \mathbf{x}_s^\kappa & \text{if } s = i_\kappa, \\ \mathbf{0} & \text{if } s \neq i_\kappa, \end{cases}$$

and go to Step 3.

Otherwise (i.e., when $\Delta_s(\mathbf{x}^\kappa) < \delta$ for all $s \in \mathbf{M}$) go to Step 2.

Step 2: Set $\mathbf{z} = \mathbf{x}^\kappa$ and stop.

Step 3: Determine m as the smallest number in \mathbb{Z}_+ such that

$$\varphi(\mathbf{x}^\kappa + \theta^m \mathbf{d}^\kappa) \leq \varphi(\mathbf{x}^\kappa) - \beta \theta^m \Delta_i^2(\mathbf{x}^\kappa), \quad (2.9)$$

set $\lambda_\kappa = \theta^m$, $\mathbf{x}^{\kappa+1} = \mathbf{x}^\kappa + \lambda_\kappa \mathbf{d}^\kappa$, and $\kappa = \kappa + 1$. The iteration is complete.

Lemma 2.4. *The line search procedure in Step 3 is always finite.*

Proof. We assume that, if the line search procedure is infinite, then

$$\theta^{-m}(\varphi(\mathbf{x}^\kappa + \theta^m \mathbf{d}^\kappa) - \varphi(\mathbf{x}^\kappa)) > -\beta \Delta_i^2(\mathbf{x}^\kappa), \text{ for } m \rightarrow \infty,$$

hence, by taking the limit, we get

$$\varphi'(\mathbf{x}^\kappa; \mathbf{d}^\kappa) \geq -\beta \Delta_i^2(\mathbf{x}^\kappa),$$

but Lemma 2.3 gives

$$\varphi'(\mathbf{x}^\kappa; \mathbf{d}^\kappa) \leq -\alpha \Delta_i^2(\mathbf{x}^\kappa),$$

hence $\alpha \leq \beta$, a contradiction. □

Proposition 2.1. *The number of iterations in Algorithm DDS is finite.*

Proof. By construction, we have

$$-\infty < \varphi^* \leq \varphi(\mathbf{x}^\kappa)$$

and

$$\varphi(\mathbf{x}^{\kappa+1}) \leq \varphi(\mathbf{x}^\kappa) - \beta \delta^2 \lambda_\kappa,$$

hence the sequence $\{\mathbf{x}^\kappa\}$ is bounded and has limits points, besides,

$$\lim_{\kappa \rightarrow \infty} \lambda_\kappa = 0.$$

Assume that the sequence $\{\mathbf{x}^\kappa\}$ is infinite. Since the set \mathbf{M} is finite, there is an index $i_\kappa = i$, which is repeated infinitely. Take the corresponding subsequence $\{\kappa_s\}$, then, without loss of generality, we can assume that the subsequence $\{\mathbf{x}^{\kappa_s}\}$ converges to a point $\bar{\mathbf{x}}$, besides,

$$\Delta_{i_{\kappa_s}}(\mathbf{x}^{\kappa_s}) = \|\mathbf{d}_i^{\kappa_s}\|,$$

and we have

$$\left(\frac{\lambda_{\kappa_s}}{\theta}\right)^{-1} \left(\varphi\left(\mathbf{x}^{\kappa_s} + \left(\frac{\lambda_{\kappa_s}}{\theta}\right) \mathbf{d}^{\kappa_s}\right) - \varphi(\mathbf{x}^{\kappa_s})\right) > -\beta \|\mathbf{d}_i^{\kappa_s}\|^2.$$

Using the mean value theorem (see e.g. [14]), we obtain

$$\langle \mathbf{g}_i^{\kappa_s} + \mathbf{q}_i^{\kappa_s} + \mathbf{t}_i^{\kappa_s}, \mathbf{d}_i^{\kappa_s} \rangle = \langle \mathbf{g}^{\kappa_s} + \mathbf{q}^{\kappa_s} + \mathbf{t}^{\kappa_s}, \mathbf{d}^{\kappa_s} \rangle > -\beta \|\mathbf{d}_i^{\kappa_s}\|^2,$$

for some $\mathbf{g}^{\kappa_s} + \mathbf{q}^{\kappa_s} = \mu'(\mathbf{x}^{\kappa_s} + \left(\frac{\lambda_{\kappa_s}}{\theta}\right) \xi_{\kappa_s} \mathbf{d}^{\kappa_s})$, $\mathbf{t}^{\kappa_s} \in \partial\eta(\mathbf{x}^{\kappa_s} + \left(\frac{\lambda_{\kappa_s}}{\theta}\right) \xi_{\kappa_s} \mathbf{d}^{\kappa_s})$, $\xi_{\kappa_s} \in (0, 1)$. By taking the limit $s \rightarrow \infty$, we have

$$\langle \mu'(\bar{\mathbf{x}}) + \bar{\mathbf{t}}, \bar{\mathbf{d}} \rangle = \langle \mathbf{g}_i(\bar{\mathbf{x}}) + \mathbf{q}_i(\bar{\mathbf{x}}), \bar{\mathbf{d}}_i \rangle \geq -\beta \|\bar{\mathbf{d}}_i\|^2,$$

for some $\bar{\mathbf{t}} \in \partial\eta(\bar{\mathbf{x}})$, where

$$\bar{\mathbf{d}}_s = \begin{cases} \mathbf{y}_i(\bar{\mathbf{x}}) - \bar{\mathbf{x}}_i & \text{if } s = i, \\ \mathbf{0} & \text{if } s \neq i. \end{cases}$$

On the other hand, using Lemma 2.3 gives

$$\langle \mu'(\bar{\mathbf{x}}) + \bar{\mathbf{t}}, \bar{\mathbf{d}} \rangle \leq \varphi'(\bar{\mathbf{x}}, \bar{\mathbf{d}}) \leq -\alpha \|\bar{\mathbf{d}}_i\|^2,$$

besides, by construction, we have

$$\|\mathbf{d}_i^{\kappa_s}\| \geq \delta,$$

hence

$$\|\bar{\mathbf{d}}_i\| \geq \delta > 0$$

and

$$\alpha \leq \beta,$$

which is a contradiction. \square

3. Main Results

We now intend to describe a general iterative method for the non-stationary (or limit) generalized mixed variational inequality problems (GMVIP) (1.7). First, we introduce the approximation assumptions.

(A2) There exists a sequence of continuous mappings $G_\ell, Q_\ell : \mathcal{X} \rightarrow \mathbb{R}^N$, which are the gradients of functions $f_{1,\ell} + f_{2,\ell} = f_\ell : \mathbb{R}^N \rightarrow \bar{\mathbb{R}}$, $\ell = 1, \dots$, such that the relations $\{\mathbf{y}^\ell\} \rightarrow \bar{\mathbf{y}}$ and $\mathbf{y}^\ell \in \mathcal{X}$ imply $\{\mathbf{G}_\ell(\mathbf{y}^\ell)\} \rightarrow \bar{\mathbf{g}} \in \mathbf{G}(\bar{\mathbf{y}})$, $\{\mathbf{Q}_\ell(\mathbf{y}^\ell)\} \rightarrow \bar{\mathbf{q}} \in \mathbf{Q}(\bar{\mathbf{y}})$.

(A3) For each $i = 1, \dots, n$, there exists a sequence of convex functions $h_{\ell,i} : \mathbb{R}^{N_i} \rightarrow \bar{\mathbb{R}}$, such that each of them is subdifferentiable on \mathcal{X}_i and that the relations $\{\mathbf{u}^\ell\} \rightarrow \bar{\mathbf{u}}$ and $\mathbf{u}^\ell \in \mathcal{X}_i$ imply $\{h_{\ell,i}(\mathbf{u}^\ell)\} \rightarrow h_i(\bar{\mathbf{u}})$.

Under condition **(A2)**, the limit set valued mappings \mathbf{G} and \mathbf{Q} at any point is approximated by a sequence of gradients $\{\mathbf{G}_\ell\}$ and $\{\mathbf{Q}_\ell\}$. In fact, if \mathbf{G}, \mathbf{Q} are the Clarke subdifferential of a locally Lipschitz functions $f_1 + f_2 = f$, it can be always approximated by a sequence of gradients within condition **(A2)**; see [16, 17]. Also, observe that if there is a subsequence $\mathbf{y}^{\ell_s} \in \mathcal{X}$ with $\{\mathbf{y}^{\ell_s}\} \rightarrow \bar{\mathbf{y}}$, then **(A2)** implies $\{\mathbf{G}_{\ell_s}(\mathbf{y}^{\ell_s})\} \rightarrow \bar{\mathbf{g}} \in \mathbf{G}(\bar{\mathbf{y}})$, $\{\mathbf{Q}_{\ell_s}(\mathbf{y}^{\ell_s})\} \rightarrow \bar{\mathbf{q}} \in \mathbf{Q}(\bar{\mathbf{y}})$ and the same is true for **(A3)**. At the same time, the non-differentiability of the functions $f_1 + f_2 = f$ or h is not obligatory, the main property is the existence of the approximation sequences indicated in **(A2)** and **(A3)**.

We replace GMVIP (1.7) with a sequence of GMVIPs: Find a point $\bar{\mathbf{z}}^\ell \in \mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$ such that

$$\sum_{i=1}^n \langle \mathbf{G}_{\ell,i}(\bar{\mathbf{z}}^\ell) + \mathbf{Q}_{\ell,i}(\bar{\mathbf{z}}^\ell), \mathbf{y}_i - \bar{\mathbf{z}}_i^\ell \rangle + \sum_{i=1}^n [h_{\ell,i}(\mathbf{y}_i) - h_{\ell,i}(\bar{\mathbf{z}}_i^\ell)] \geq 0, \quad \forall \mathbf{y}_i \in \mathcal{X}_i, \text{ for } i = 1, \dots, n; \quad (3.1)$$

where we use the partition of \mathbf{G}_ℓ and \mathbf{Q}_ℓ which corresponds to that of the space \mathbb{R}^N , i.e.

$$\mathbf{G}_\ell(\mathbf{x}) + \mathbf{Q}_\ell(\mathbf{x}) = (\mathbf{G}_{\ell,1}(\mathbf{x}) + \mathbf{Q}_{\ell,1}(\mathbf{x}), \dots, \mathbf{G}_{\ell,n}(\mathbf{x}) + \mathbf{Q}_{\ell,n}(\mathbf{x}))^\top,$$

where $\mathbf{G}_{\ell,i}(\mathbf{x}) \in \mathbb{R}^{N_i}$ and $\mathbf{Q}_{\ell,i}(\mathbf{x}) \in \mathbb{R}^{N_i}$, $i = 1, \dots, n$.

Similarly, we set

$$h_\ell(\mathbf{x}) = \sum_{i=1}^n h_{\ell,i}(\mathbf{x}_i).$$

Since the feasible set \mathcal{X} may be unbounded, we introduce coercivity conditions.

- (C1)** For each fixed $\ell = 1, 2, \dots$, the function $f_\ell(\mathbf{x}) + h_\ell(\mathbf{x})$ is coercive on the set \mathcal{X} , that is $\{f_\ell(\mathbf{w}^\kappa) + h_\ell(\mathbf{w}^\kappa)\} \rightarrow +\infty$ if $\{\mathbf{w}^\kappa\} \subset \mathcal{X}$, $\|\mathbf{w}^\kappa\| \rightarrow \infty$ as $\kappa \rightarrow \infty$.
- (C2)** There exists a number $\sigma > 0$ and a point $\bar{\mathbf{v}} \in \mathcal{X}$ such that for any sequences $\{\mathbf{u}^\ell\}$ and $\{\mathbf{d}^\ell\}$ satisfying the conditions:

$$\mathbf{u}^\ell \in \mathcal{X}, \{\|\mathbf{u}^\ell\|\} \rightarrow +\infty, \{\mathbf{d}^\ell\} \rightarrow \mathbf{0};$$

it holds that

$$\liminf_{\ell \rightarrow \infty} \{ \langle \mathbf{G}_\ell(\mathbf{u}^\ell) + \mathbf{Q}_\ell(\mathbf{u}^\ell) + \tau \mathbf{d}^\ell, \bar{\mathbf{v}} - \mathbf{u}^\ell - \mathbf{d}^\ell \rangle + [h_\ell(\bar{\mathbf{v}}) - h_\ell(\mathbf{u}^\ell - \mathbf{d}^\ell)] \} \leq -\sigma, \text{ if } \tau > 0.$$

Clearly, **(C1)** gives a custom coercivity condition for each function $f_\ell(\mathbf{x}) + h_\ell(\mathbf{x})$, which provides existence of solutions of each particular problem (3.1). Obvious, **(C1)** holds if \mathcal{X} is bounded. At the same time, **(C2)** gives a similar coercivity condition for the whole sequence of these problems approximation the limit GMVIP (1.7). It also holds if \mathcal{X} is bounded. In the unbounded case **(C2)** is weaker than the following coercivity condition:

$$\|\bar{\mathbf{v}} - \mathbf{u}^\ell - \mathbf{d}^\ell\|^{-1} \{ \langle \mathbf{G}_\ell(\mathbf{u}^\ell) + \mathbf{Q}_\ell(\mathbf{u}^\ell) + \tau \mathbf{d}^\ell, \bar{\mathbf{v}} - \mathbf{u}^\ell - \mathbf{d}^\ell \rangle + [h_\ell(\bar{\mathbf{v}}) - h_\ell(\mathbf{u}^\ell - \mathbf{d}^\ell)] \} \rightarrow -\infty \text{ as } \ell \rightarrow \infty.$$

Therefore, we conclude that the conditions **C1** and **C2** are not restrictive. The whole decomposition descent method for the non-stationary GMVIP (1.7) has a two-level iteration scheme where each stage of the upper level invokes Algorithm (DDS) with different parameters.

Decomposition Non-Stationary Method (DNS) Choose a point $\mathbf{z}^0 \in \mathcal{X}$ and a sequence $\{\delta_\ell\} \rightarrow +0$. At the ℓ -th stage, $\ell = 1, 2, \dots$, we have a point $\mathbf{z}^{\ell-1} \in \mathcal{X}$ and a number δ_ℓ .

Set

$$\mu(\mathbf{x}) = f_\ell(\mathbf{x}), \quad \eta(\mathbf{x}) = h_\ell(\mathbf{x}),$$

apply Algorithm (DDS) with $\mathbf{x}^0 = \mathbf{z}^{\ell-1}$, $\delta = \delta_\ell$ and obtain a $\mathbf{z}^\ell = \mathbf{z}$ as its output.

Now, we establish the main convergence result.

Theorem 3.1. Assume that the assumptions **(A1)-(A3)** and **(C1)-(C2)** hold, besides, $\{\delta_\ell\} \rightarrow +0$. Then

- (i) problem (3.1) has a solution;
- (ii) the number of iterations at each stage of Method (DNS) is finite;
- (iii) the sequence $\{\mathbf{z}^\ell\}$ generated by Method (DNS) has limit points and all these limit points are solutions of GMVIP (1.7);
- (iv) if f is convex, then all the limit points of $\{\mathbf{z}^\ell\}$ belong to \mathcal{X}^* .

Proof. We see that **(C1)** implies that each problem (3.1) has a solution. Since the cost function

$$\mu(\mathbf{x}) = f_\ell(\mathbf{x}) + h_\ell(\mathbf{x})$$

is coercive, hence the set

$$\mathcal{X}_\ell(\mathbf{x}^0) = \{ \mathbf{y} \in \mathcal{X} \mid \mu(\mathbf{y}) \leq \mu(\mathbf{x}^0) \}$$

is bounded. It follows that the optimization problem

$$\min_{\mathbf{x} \in \mathcal{X}} \mu(\mathbf{x})$$

has a solution and so is GMVIP (3.1). Hence, assertion (i) is true. Next, from Proposition 2.1, the assertion (ii) is also true.

By (ii), the sequence $\{\mathbf{z}^\ell\}$ is well-defined and (2.5) implies

$$\langle \mathbf{G}_\ell(\mathbf{z}^\ell) + \mathbf{Q}_\ell(\mathbf{z}^\ell) + \alpha(\mathbf{y}^\ell(\mathbf{z}^\ell)) - \mathbf{z}^\ell, \mathbf{y} - \mathbf{y}^\ell(\mathbf{z}^\ell) \rangle + [h_\ell(\mathbf{y}) - h_\ell(\mathbf{y}^\ell(\mathbf{z}^\ell))] \geq 0, \forall \mathbf{y} \in \mathcal{X}. \quad (3.2)$$

Besides, the stopping rule in Algorithm (DDS) gives

$$\alpha \delta_\ell \sqrt{n} \geq \alpha \|\mathbf{z}^\ell - \mathbf{y}^\ell(\mathbf{z}^\ell)\|. \quad (3.3)$$

Now, we proceed to show that $\{\mathbf{z}^\ell\}$ is bounded. Conversely, assume that $\{\|\mathbf{z}^\ell\|\} \rightarrow +\infty$. Applying (3.2) with $\mathbf{y} = \bar{\mathbf{v}}$, we have

$$0 \leq \langle \mathbf{g}^\ell + \mathbf{q}^\ell + \mathbf{d}^\ell, \bar{\mathbf{v}} - \tilde{\mathbf{z}}^\ell \rangle + [h_\ell(\bar{\mathbf{v}}) - h_\ell(\tilde{\mathbf{z}}^\ell)].$$

Here and below, for brevity we set $\mathbf{g}^\ell = \mathbf{G}_\ell(\mathbf{z}^\ell)$, $\mathbf{q}^\ell = \mathbf{Q}_\ell(\mathbf{z}^\ell)$, $\tilde{\mathbf{z}}^\ell = \mathbf{y}^\ell(\mathbf{z}^\ell)$, and $\mathbf{d}^\ell = \alpha(\mathbf{y}(\mathbf{z}^\ell) - \mathbf{z}^\ell)$.

Take a subsequence $\{\ell_s\}$ such that

$$\begin{aligned} & \lim_{s \rightarrow \infty} \{ \langle \mathbf{g}^{\ell_s} + \mathbf{q}^{\ell_s} + \mathbf{d}^{\ell_s}, \bar{\mathbf{v}} - \tilde{\mathbf{z}}^{\ell_s} \rangle + [h_{\ell_s}(\bar{\mathbf{v}}) - h_{\ell_s}(\tilde{\mathbf{z}}^{\ell_s})] \} \\ & = \liminf_{\ell \rightarrow \infty} \{ \langle \mathbf{g}^\ell + \mathbf{q}^\ell + \mathbf{d}^\ell, \bar{\mathbf{v}} - \tilde{\mathbf{z}}^\ell \rangle + [h_\ell(\bar{\mathbf{v}}) - h_\ell(\tilde{\mathbf{z}}^\ell)] \}, \end{aligned}$$

then, from (C2), we have

$$0 \leq \lim_{s \rightarrow \infty} \{ \langle \mathbf{g}^{\ell_s} + \mathbf{q}^{\ell_s} + \mathbf{d}^{\ell_s}, \bar{\mathbf{v}} - \tilde{\mathbf{z}}^{\ell_s} \rangle + [h_{\ell_s}(\bar{\mathbf{v}}) - h_{\ell_s}(\tilde{\mathbf{z}}^{\ell_s})] \} \leq -\sigma < 0,$$

a contradiction. Therefore, the sequence $\{\mathbf{z}^\ell\}$ is bounded and has limit points. Let $\bar{\mathbf{z}}$ be an arbitrary limit point for $\{\mathbf{z}^\ell\}$, i.e.

$$\bar{\mathbf{z}} = \lim_{s \rightarrow \infty} \mathbf{z}^{\ell_s}.$$

Since $\mathbf{z}^\ell \in \mathcal{X}$, we have $\bar{\mathbf{z}} \in \mathcal{X}$. It follows from (A2) that

$$\lim_{s \rightarrow \infty} \mathbf{g}^{\ell_s} = \bar{\mathbf{g}} \in \mathbf{G}(\bar{\mathbf{z}})$$

and

$$\lim_{s \rightarrow \infty} \mathbf{q}^{\ell_s} = \bar{\mathbf{q}} \in \mathbf{Q}(\bar{\mathbf{z}}).$$

Fix an arbitrary point $\mathbf{y} \in \mathcal{X}$, then, using (3.2) and (3.3) and (A3), we have

$$\begin{aligned} \langle \bar{\mathbf{g}} + \bar{\mathbf{q}}, \mathbf{y} - \bar{\mathbf{z}} \rangle + [h(\mathbf{y}) - h(\bar{\mathbf{z}})] &= \lim_{s \rightarrow \infty} \{ \langle \mathbf{g}^{\ell_s} + \mathbf{q}^{\ell_s}, \mathbf{y} - \mathbf{z}^{\ell_s} \rangle + [h_{\ell_s}(\mathbf{y}) - h_{\ell_s}(\mathbf{z}^{\ell_s})] \} \\ &= \lim_{s \rightarrow \infty} \{ \langle \mathbf{g}^{\ell_s} + \mathbf{q}^{\ell_s} + \mathbf{d}^{\ell_s}, \mathbf{y} - \tilde{\mathbf{z}}^{\ell_s} \rangle + [h_{\ell_s}(\mathbf{y}) - h_{\ell_s}(\tilde{\mathbf{z}}^{\ell_s})] \} \geq 0. \end{aligned}$$

Therefore, $\bar{\mathbf{z}}$ solves GMVIP (1.7) and assertion (iii) holds.

Next, if f is convex, then so is p (a goal function defined on a feasible set \mathcal{X}) and each limit point of

$\{\mathbf{z}^\ell\}$ belongs to \mathcal{X}^* , which gives assertion (iv), and GMVIP (1.7) has a solution. □

4. Examples

We can take the exact one-dimensional minimization rule instead of the current Armijo line-search (2.9) in Algorithm (DDS), then the assertions of Theorem 3.1 remain true. Next, if the function μ each function f_ℓ) is convex, we can replace (2.9) with the following:

$$\langle \mathbf{g}_i(\mathbf{x}^\kappa + \theta^m \mathbf{d}^\kappa) + \mathbf{q}_i(\mathbf{x}^\kappa + \theta^m \mathbf{d}^\kappa) + \mathbf{d}_i^\kappa \rangle + \theta^{-m} \{ \eta_i(\mathbf{x}_i^\kappa + \theta^m \mathbf{d}_i^\kappa) - \eta_i(\mathbf{x}_i^\kappa) \} \leq -\beta \alpha^{-1} \Delta_i^2(\mathbf{x}^\kappa).$$

Moreover, if the gradient of the function μ is Lipschitz continuous, we have an explicit lower bound for the step size and utilize the fixed step-size version of the Algorithm (DDS), which leads to further reduction of computational expenses.

Now, we give only two instances to illustrate possible applications.

The first instance is the linear inverse problem that arises very often in signal and image processing, see [18]. The problem consists of solving a linear system of equations

$$\mathbf{Ax} = \mathbf{b},$$

where A is a $m \times n$ matrix, b is a vector in \mathbb{R}^m , whose exact values may be unknown or admit some noise perturbations. If $A^\top A$ is ill-conditioned, the custom approach based on the least square minimization problem

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2$$

may give very inexact approximations. To enhance its properties, one can utilize a family of regularized problems of the form

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|^2 + \varepsilon h(\mathbf{x}), \tag{4.1}$$

where $h(\mathbf{x}) = \|\mathbf{x}\|^2$ or

$$h(\mathbf{x}) = \|\mathbf{x}\|_1 \triangleq \sum_{i=1}^n |x_i|,$$

$\varepsilon > 0$ is a parameter. Note that the non-smooth regularization term yields additionally sparse solutions with rather small number of non-zero components; see, e.g. [6, 19].

The second instance is the basic machine learning problem, which is called the linear support vector machine. It consists in finding the optimal partition of the feature space \mathbb{R}^n by using some given training sequence \mathbf{x}^i , $i = 1, \dots, \ell$ where each point \mathbf{x}^i has a binary label $y_i \in \{-1, +1\}$ indicating the class. We have to find a separating hyperplane. Usually, its parameters are found in the solution of the optimization problem

$$\min_{\mathbf{w} \in \mathbb{R}^n} \left(\frac{1}{\rho} \right) \|\mathbf{w}\|_\rho^p + C \sum_{i=1}^{\ell} L(\langle \mathbf{w}, \mathbf{x}^i \rangle; y_i), \tag{4.2}$$

where L is a loss function and $C > 0$ is a penalty parameter. The usual choice is

$$L(z; y) = \max\{0; 1 - yz\}$$

and p is either 1 or 2. We observe that the data of the observation points \mathbf{x}^i can be inexact or even non-stationary.

Next, taking $p = 2$, we can rewrite this problem as

$$\min_{\mathbf{w}, \xi} 0.5 \|\mathbf{w}\|^2 + C \sum_{i=1}^{\ell} \xi_i,$$

subject to

$$1 - y_i \langle \mathbf{w}, \mathbf{x}^i \rangle \leq \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, \ell.$$

Its dual has the quadratic programming format:

$$\max_{0 \leq \alpha_i \leq C, i=1, \dots, \ell} \sum_{i=1}^{\ell} \alpha_i - 0.5 \sum_{s=1}^{\ell} \sum_{t=1}^{\ell} (\alpha_s y_s) (\alpha_t y_t) \langle \mathbf{x}^s, \mathbf{x}^t \rangle. \quad (4.3)$$

We see that all these problems fall into format (1.1), (1.3)-(1.5) has a solution.

5. Conclusions

We discussed a new class of selective coordinate-wise descent splitting methods for non-stationary decomposable composite optimization problems and proved the convergence of the problems involving the non-smooth set-valued functions where all coordinate variations together change the tolerance parameters corresponding to the sequence of GMVIP.

References

- [1] G. Stampacchia, Variational Inequalities, in: Proc. NATO Advanced Study Inst., Theory and Application of Monotone Operators, Venice, 1968, pp. 101–192.
- [2] M. Sofonea, W. Han, S. Migórski, Numerical Analysis of History-Dependent Variational–hemivariational Inequalities With Applications to Contact Problems, *Eur. J. Appl. Math.* 26 (2015), 427–452. <https://doi.org/10.1017/s095679251500011x>.
- [3] M. Hintermüller, Inverse Coefficient Problems for Variational Inequalities: Optimality Conditions and Numerical Realization, *ESAIM: M2AN.* 35 (2001), 129–152. <https://doi.org/10.1051/m2an:2001109>.
- [4] Salahuddin, A Coordinate Wise Variational Method with Tolerance Functions, *J. Appl. Nonlinear Dyn.* 9 (2020), 541–549. <https://doi.org/10.5890/jand.2020.12.002>.
- [5] I.V. Konnov, Right-Hand Side Decomposition for Variational Inequalities, *J. Optim. Theory Appl.* 160 (2013), 221–238. <https://doi.org/10.1007/s10957-013-0370-0>.
- [6] V. Cevher, S. Becker, M. Schmidt, Convex Optimization for Big Data: Scalable, Randomized, and Parallel Algorithms for Big Data Analytics, *IEEE Signal Process. Mag.* 31 (2014), 32–43. <https://doi.org/10.1109/msp.2014.2329397>.
- [7] F. Facchinei, G. Scutari, S. Sagratella, Parallel Selective Algorithms for Nonconvex Big Data Optimization, *IEEE Trans. Signal Process.* 63 (2015), 1874–1889. <https://doi.org/10.1109/tsp.2015.2399858>.

- [8] I. Konnov, Decomposition Descent Method for Limit Optimization Problems, in: R. Battiti, D.E. Kvasov, Y.D. Sergeyev (Eds.), *Learning and Intelligent Optimization*, Springer International Publishing, Cham, 2017: pp. 166–179. https://doi.org/10.1007/978-3-319-69404-7_12.
- [9] I.V. Konnov, Selective Bi-coordinate Method for Limit Non-Smooth Resource Allocation Type Problems, *Set-Valued Var. Anal.* 27 (2017), 191–211. <https://doi.org/10.1007/s11228-017-0447-2>.
- [10] P. Tseng, S. Yun, A Coordinate Gradient Descent Method for Nonsmooth Separable Minimization, *Math. Program.* 117 (2009), 387–423. <https://doi.org/10.1007/s10107-007-0170-0>.
- [11] I.V. Konnov, Salahuddin, Two-Level Iterative Method for Non-Stationary Mixed Variational Inequalities, *Russian Math.* 61 (2017), 44–53. <https://doi.org/10.3103/s1066369x17100061>.
- [12] Salahuddin, Iterative Method for Non-Stationary Mixed Variational Inequalities, *Discontin. Nonlinear. Complex.* 9 (2020), 647–655. <https://doi.org/10.5890/dnc.2020.12.015>.
- [13] G. Salmon, V.H. Nguyen, J.J. Strodiot, Coupling the Auxiliary Problem Principle and Epiconvergence Theory to Solve General Variational Inequalities, *J. Optim. Theory Appl.* 104 (2000), 629–657. <https://doi.org/10.1023/a:1004693710334>.
- [14] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [15] I.V. Konnov, Sequential Threshold Control in Descent Splitting Methods for Decomposable Optimization Problems, *Optim. Meth. Softw.* 30 (2015), 1238–1254. <https://doi.org/10.1080/10556788.2015.1030015>.
- [16] Y.M. Ermoliev, V.I. Norkin, R.J.-B. Wets, The Minimization of Semicontinuous Functions: Mollifier Subgradients, *SIAM J. Control Optim.* 33 (1995), 149–167. <https://doi.org/10.1137/s0363012992238369>.
- [17] M.O. Czarnecki, L. Rifford, Approximation and Regularization of Lipschitz Functions: Convergence of the Gradients, *Trans. Amer. Math. Soc.* 358 (2006), 4467–4520.
- [18] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.
- [19] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, *J. R. Stat. Soc.: Ser. B (Methodol.)* 58 (1996), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.