Journal of Global Optimization **17:** 201–224, 2000.

Auxiliary Problem Principle and Proximal Point Methods

A. KAPLAN and R. TICHATSCHKE
Department of Mathematics, University of Trier, Trier, Germany

(Received for publication May 2000)

Abstract. An extension of the auxiliary problem principle to variational inequalities with non-symmetric multi-valued operators in Hilbert spaces is studied. This extension concerns the case that the operator is split into the sum of a single-valued operator $\mathcal{F}$, possessing a kind of pseudo Dunn property, and a maximal monotone operator $\mathcal{Q}$. The current auxiliary problem is constructed by fixing $\mathcal{F}$ at the previous iterate, whereas $\mathcal{Q}$ (or its single-valued approximation $\mathcal{Q}^k$) is considered at a variable point. Using auxiliary operators of the form $\mathcal{L}^k + \chi_k \nabla h$, with $\chi_k > 0$, the standard for the auxiliary problem principle assumption of the strong convexity of the function h can be weakened exploiting mutual properties of $\mathcal{Q}$ and h. Convergence of the general scheme is analyzed and some applications are sketched briefly.

Key words: Auxiliary problem principle; Convex and nonconvex optimization; Ill-posed problems; Proximal point methods; Regularization; Variational inequalities

AMS subject classification: 90C48, 90C25, 90C26, 49J40, 65K05.

1. Introduction

The *auxiliary problem principle* (APP), originallly introduced by Cohen [1, 2] as a general framework to analyze optimization algorithms of gradient and subgradient types as well as decomposition algorithms, was extended later to different numerical methods for solving variational inequalities. In the majority of the papers dedicated to such extensions, the $(k+1)$-th auxiliary problem is constructed by applying the operator of the variational inequality (here called main operator) to the k-th iterate. The idea to take this operator (or some additive part of it) at a variable point leads to a scheme which appears to be a generalization of proximal-like methods, too.

In order to illustrate this, let us consider the problem

$$\text{find} \quad x \in X: \quad 0 \in \Psi(x)\,, \tag{1.1}$$

with Ψ a given multi-valued maximal monotone operator in a Hilbert space X. The APP is taken in the form

$$\text{find} \quad x^{k+1} \in X: \quad 0 \in \frac{1}{\epsilon_k}[\Omega(x^{k+1}) - \Omega(x^k)] + \Psi(x^k)\,, \tag{1.2}$$

with $\Omega : X \to X$ an auxiliary operator, $\epsilon_k > 0$, and this iterative scheme can be modified as follows:

$$\text{find} \quad x^{k+1} \in X: \quad 0 \in \frac{1}{\epsilon_k}[\Omega(x^{k+1}) - \Omega(x^k)] + \Psi(x^{k+1}), \tag{1.3}$$

or

$$\text{find} \quad x^{k+1} \in X: \quad 0 \in \frac{1}{\epsilon}[\Omega(x^{k+1}) - \Omega(x^k)] + \Psi_1(x^k) + \Psi_2(x^{k+1}), \tag{1.4}$$

where $\epsilon > 0$ and Ψ is decomposed into a sum of a single-valued operator Ψ_1 and a monotone operator Ψ_2.

Then, in case $\Omega = \mathscr{I}$ ($\mathscr{I}$ – identity operator), the following well-known methods arise:

- inclusion (1.2) leads to

  $$x^{k+1} \in x^k - \epsilon_k \Psi(x^k),$$

  which is an analogon of the subgradient method;
- on using (1.3) we obtain

  $$x^{k+1} = (\mathscr{I} + \epsilon_k \Psi)^{-1}(x^k),$$

  the proximal point method;
- (1.4) produces

  $$x^{k+1} = (\mathscr{I} + \epsilon \Psi_2)^{-1}(1 - \epsilon \Psi_1)(x^k), \tag{1.5}$$

  the splitting algorithm, suggested by Lions and Mercier [13] (see also Gabay [6]) and Passty [19].

Obviously, the latter algorithm can be represented as

$$z^k = x^k - \epsilon \Psi_1(x^k), \qquad x^{k+1} = (\mathscr{I} + \epsilon \Psi_2)^{-1}(z^k),$$

where x^{k+1} is calculated from z^k by means of the proximal mapping. Tseng [26, 27] has used method (1.5) (with a variable ϵ) as a basic process to investigate convergence of several known but also new splitting methods for solving variational inequalities with separability properties as well as for related convex optimization problems with linear constraints and for linear complementarity problems.

In the present paper the APP is studied for variational inequalities of the type

$$(P) \quad \text{find } x^* \in K: \ \langle \mathscr{F}(x^*) + \mathscr{Q}(x^*), x - x^* \rangle \geq 0 \ \ \forall x \in K,$$

with K a convex closed subset of a Hilbert space $(X, \|\cdot\|)$, $\mathscr{F}$ a single-valued operator from X into the dual space X' and $\mathscr{Q} : X \to 2^{X'}$ a maximal monotone (in general, multi-valued) operator; $\langle \cdot, \cdot \rangle$ denotes the duality pairing between X and X'.

The here suggested auxiliary problems have the form:

$$
\begin{aligned}
&(P^k) \quad \text{find } x^{k+1} \in K: \\
&\quad \langle \mathscr{F}(x^k) + \mathscr{Q}^k(x^{k+1}) + \mathscr{L}^k(x^{k+1}) - \mathscr{L}^k(x^k) \\
&\quad + \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geq 0 \quad \forall x \in K ,
\end{aligned}
$$

where $h: X \to \mathbb{R}$ is a convex Gâteaux-differentiable functional, $\mathscr{Q}^k : X \to X'$ is a monotone operator approximating $\mathscr{Q}$, $\mathscr{L}^k : X \to X'$ is a monotone operator and χ_k is a positive scalar. In this case, $\mathscr{L}^k + \chi_k \nabla h$ corresponds to the customary notion of an auxiliary operator. In the sequel, we refer to this scheme as the *proximal auxiliary problem method* (PAP-method).

For the exact conditions which are supposed to be valid for both problems (P) and (P^k) see the Assumptions 1 and 2 below.

Applying for the sources of the PAP-method, we begin with a version of the APP for convex optimization problems, where a linearization of the Gâteaux-differentiable term J_1 of an objective functional $J = J_1 + J_2$ is used. This was already studied in the mentioned paper [1]. Taking this partial linearization, the term $\langle \nabla J_1(x^k), \cdot - x^k \rangle + J_2(\cdot)$ is inserted into the objective functional of the $(k+1)$-th auxiliary problem. Such an approach is of special interest for constructing decomposition methods. In fact, if the problem $\min\{J_2(x) : x \in K\}$ splits up into independent subproblems, then the mentioned linearization permits to provide the same splitting in the framework of the APP for the original problem $\min\{J(x) : x \in K\}$, i.e., the corresponding auxiliary problems can be split up, too.

The general scheme oriented to decomposition methods for variational inequalities of the form

$$
\text{find} \quad x^* \in X: \quad \langle \Psi(x^*), x - x^* \rangle + f(x) - f(x^*) \geq 0 , \quad \forall x \in X , \tag{1.6}
$$

with Ψ a single-valued monotone operator and f a convex, lower semi-continuous (lsc) and additive w.r.t. a Cartesian factorization of the space X functional, has been developed in the paper of Makler-Scheimberg et al. [14]. Here the APP is combined with the approximation of the functional f on the basis of the concept of the Mosco-convergence. An extension of this scheme (without accentuating decomposition methods), described by Salmon et al. [24], is connected with a relaxation of the monotonicity condition for Ψ and with the use of a wider class of auxiliary operators (as distinct from [14], these operators may be non-symmetric). The auxiliary problems in [24] can be written as

$$
\begin{aligned}
&\text{find} \quad x^{k+1} \in K: \\
&\quad \langle \Psi(x^k) + \mathscr{L}^k(x^{k+1}) - \mathscr{L}^k(x^k) + \chi_k(\nabla h^k(x^{k+1}) - \nabla h^k(x^k)), x - x^{k+1} \rangle \\
&\quad + f^k(x) - f^k(x^{k+1}) \geq 0 \quad \forall x \in K ,
\end{aligned}
$$

and the conditions concerning $\mathscr{L}^k$, h^k and f^k made there permit to cover a lot of earlier versions of the APP and special algorithms.

Our PAP-method may be considered as a perturbed version of the method studied by Zhu and Marcotte [28] for the case where $X = \mathbb{R}^n$ and $\mathscr{F} + \mathscr{Q}$ is a continuous operator: the auxiliary problems in [28] correspond to (P^k) with $\mathcal{Q}^k = \mathcal{Q}$, $\mathcal{L}^k = \mathcal{L}$, $\chi_k = \chi$, under stronger assumptions w.r.t. h and $\mathcal{F} - \mathcal{L}$.

The paper of Renaud and Cohen [21] should also be mentioned, where for Problem (1.1) the APP is studied in the form

$$\text{find} \quad x^{k+1} \in X: \quad 0 \in \frac{1}{\epsilon}[\Xi(x^{k+1}) - \Xi(x^k)] + \Psi(x^{k+1}), \tag{1.7}$$

with a single-valued (in general, non-symmetric) auxiliary operator $\Xi = \Xi_1 + \epsilon\Xi_2$. Here, Ξ_1 is supposed to be the gradient of a strongly convex functional, and observing the following relation between Ψ and Ξ_2:

$$\exists \gamma_0 > 0: \quad \langle \psi(x) - \psi(y), x - y \rangle \geqslant \gamma_0 \|\Xi_2(x) - \Xi_2(y)\|^2,$$
$$\forall \psi(x) \in \Psi(x), \psi(y) \in \Psi(y), \quad \forall x, y \in \operatorname{dom} \Psi,$$

the operator Ξ_2 may be hemicontinuous only. Note that the auxiliary problem (1.4) is equivalent to (1.7) setting $\Xi = \Omega - \epsilon\Psi_1$.

In [21] the general convergence results for method (1.7) have been also adapted to prove convergence of a new algorithm for solving saddle-point problems with convex–concave functions on the product of convex sets. Depending on the decomposition of the related maximal monotone operator, the Arrow–Hurwicz algorithm and the proximal point method can be obtained as particular cases.

The PAP-method studied in the present paper has the following peculiarities:

- as distinct from [24, 14], the operator $\mathcal{Q}$ is supposed to be not necessarily the subdifferential of a convex functional (note that the variational inequality (1.6) corresponds to Problem (P) with $\mathcal{Q} = \partial f$);
- as distinct from [21], the main operator $\mathcal{F} + \mathcal{Q}$ is not necessarily monotone, an approximation of $\mathcal{Q}$ is included, and the auxiliary operator may vary after each step.

Besides, we weaken the standard (for the APP) assumption on strong convexity of the auxiliary function h in the Problems (P^k): h is supposed to be convex and the operators $\mathcal{Q}^k + \nabla h$ have to be strongly monotone with a common modulus for all k.

Note that the conditions joining the main and auxiliary operators are not completely comparable in the schemes suggested here and in [14, 21, 24].

The paper is organized as follows: In Section 2 we start with the full description of the problem under consideration and discuss some assumptions. The conditions w.r.t. the successive approximation of the problem and the convergence analysis of the PAP-method are described in Section 3, and Section 4 contains some applications.

2. Proximal auxiliary problem method

We consider the variational inequality (P) under the following basic assumptions.

Assumption 1

(i) *$K \subset X$ is a non-empty, convex set;*

(ii) *$\mathscr{Q} : X \to 2^{X'}$ is a maximal monotone operator, $D(\mathscr{Q}) \cap K$ is a convex set, and*

$$\mathscr{Q}_K : y \to \begin{cases} \mathscr{Q}(y) & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}$$

is locally hemi-bounded at each point of $D(\mathscr{Q}) \cap K$;

(iii) *the operator $\mathscr{Q} + \mathscr{N}_K$ is maximal monotone, where*

$$\mathscr{N}_K : y \to \begin{cases} \{z \in X' : \langle z, y - x\rangle \geqslant 0 \ \ \forall x \in K\} & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}$$

is the normality operator for K;

(iv) *$\mathscr{F} : X \to X'$ is a single-valued compact on K operator;*

(v) *given a family $\{\mathscr{L}_y\}$, $\mathscr{L}_y : X \to X'$, of monotone on $D(\mathscr{Q}) \cap K$ operators parameterized by $y \in K$; if*

$$\langle \mathscr{F}(x) + q(x), y - x\rangle \geqslant 0$$

holds for some $x, y \in D(\mathscr{Q}) \cap K$ and some $q(x) \in \mathscr{Q}(x)$, then

$$\begin{aligned} &\langle \mathscr{F}(y) - \mathscr{L}_y(y) + \mathscr{L}_y(x) + q(x), y - x\rangle \\ &\quad \geqslant \gamma \|\mathscr{F}(y) - \mathscr{L}_y(y) - \mathscr{F}(x) + \mathscr{L}_y(x)\|^2_{X'}, \end{aligned}$$

is valid, where $\gamma > 0$ is independent of x, y;

(vi) *Problem (P) is solvable.*

Referring in the sequel to the separate conditions described in Assumptions 1 and 2 (below), we write (1-i), (1-ii), . . . and (2-i), (2-ii), . . . , respectively.

Let us discuss some mentioned notations and conditions.

- By definition, an element $x^* \in K$ is a solution of Problem (P) if the inequality

  $$\langle F(x^*) + q^*(x^*), x - x^*\rangle \geqslant 0 \quad \forall x \in K$$

  is valid for some $q^*(x^*) \in \mathscr{Q}(x^*)$. In the sequel, the notation $q^*(\cdot)$ taken to diverse elements of the solution set X^* has the same meaning.
- Local hemi-boundedness of an operator $\mathscr{M}$ at a point $x^0 \in D(\mathscr{M})$ means: for each $x \in D(\mathscr{M})$, $x \neq x^0$, there exists a number $t_0(x^0, x) > 0$ such that $x^0 + t(x - x^0) \in D(\mathscr{M})$ holds for $0 \leqslant t \leqslant t_0(x^0, x)$ and the set

  $$\bigcup_{0 < t \leqslant t_0(x^0,x)} \mathscr{M}(x^0 + t(x - x^0)) \text{ is bounded in } X'.$$

  Here we use a weakened notion of local hemi-boundedness: the standard notion supposes boundedness of $\cup_{0 \leqslant t \leqslant t_0(x^0,x)} \mathscr{M}(x^0 + t(x - x^0))$. The simple example

$\mathcal{M} = \mathcal{N}_C$, where $C = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leqslant 1\}$, shows that this relaxation may be very essential.

- With $\mathcal{Q}$ a maximal monotone operator and K a convex closed set, the operator $\mathcal{Q} + \mathcal{N}_K$ is maximal monotone if, for instance, int$D(\mathcal{Q}) \cap K \neq \emptyset$ or $\mathcal{Q}$ is locally bounded[1] at some $x \in K \cap clD(\mathcal{Q})$ (see [22]).
- Assumption (1-v) is not so strange as it seems at the first glance. For instance, (1-v) is certainly fulfilled if for each $y \in K$ the operator $\mathcal{F} - \mathcal{L}_y$ possesses the Dunn property (is co-coercive):

$$\exists \gamma > 0: \quad \langle \mathcal{F}(x) - \mathcal{L}_y(x) - \mathcal{F}(v) + \mathcal{L}_y(v), x - v \rangle$$
$$\geqslant \gamma \|\mathcal{F}(x) - \mathcal{L}_y(x) - \mathcal{F}(v) + \mathcal{L}_y(v)\|_{X'}^2, \; \forall x, v \in K ,$$

which is a rather standard hypothesis for the APP. On the other hand, the simple example

$$X = \mathbb{R}^1, \;\; \mathcal{F} : x \to x^2, \;\; \mathcal{Q} : x \to x + 4, \;\; K = [-1, 1] \text{ and } \mathcal{L}_y \equiv 0$$

illustrates the situation that (1-v) is fulfilled, although the operator $\mathcal{F} - \mathcal{L}_y$ does not possess the Dunn property and even not the following weaker pseudo Dunn property:

$$\exists \gamma_1 > 0: \quad \langle \mathcal{F}(v) - \mathcal{L}_y(v), v - x \rangle \geqslant \gamma_1 \|\mathcal{F}(v) - \mathcal{L}_y(v) - \mathcal{F}(x) + \mathcal{L}_y(x)\|_{X'}^2$$

holds, whenever

$$\langle \mathcal{F}(x) - \mathcal{L}_y(x), v - x \rangle \geqslant 0 \text{ for some } x, v \in K .$$

- If the operator $\mathcal{F}$ is monotone and hemicontinuous, Assumption (1-iv) can be weakened: instead of the compactness of $\mathcal{F}$ it suffices that $\mathcal{F}$ is bounded on K (i.e., $\mathcal{F}$ carries bounded subsets of K into bounded subsets of X'). This change causes only minor modifications of the proofs of Lemma 3 and Theorem 1 below.

Now, the method suggested reads as follows:

Proximal auxiliary problem method (PAP-method). *Starting with $x^0 \in K$, the sequence $\{x^k\}$ is defined by solving successively the auxiliary problems (P^k), $k = 0, 1, \ldots,$ where $\mathcal{L}^k = \mathcal{L}_y|_{y=x^k}$.*

3. Convergence analysis

We study the convergence of the PAP-method, using Assumption 1 and the following conditions w.r.t. the data of the auxiliary problems (P^k).

[1] Local boundedness of $\mathcal{Q}$ at x means that $\mathcal{Q}$ carries some neighborhood x into a bounded set.

Assumption 2

(i) *The operators* $\mathscr{L}_y$, $y \in K$, *are monotone and Lipschitz continuous on* K, *with a common Lipschitz constant* $l_{\mathscr{L}}$;

(ii) *for a given monotone operator* $\mathscr{G} : X \to X'$ *the inequality*

$$h(y) - h(x) - \langle \nabla h(x), y - x \rangle \geqslant \langle \mathscr{G}(y) - \mathscr{G}(x), y - x \rangle \quad \forall x, y \in K$$

is satisfied;

(iii) *the mapping* ∇h *is Lipschitz continuous on* K, *with a Lipschitz constant* l_h;

(iv) $\mathscr{Q}^k$ *are single-valued operators and*

$$\langle \mathscr{Q}^k(x) - \mathscr{Q}^k(y), x - y \rangle \geqslant \langle \mathscr{B}(x - y), x - y \rangle \quad \forall x, y \in K \cap D(\mathscr{Q}^k),$$

where $\mathscr{B} : X \to X'$ *is a given linear continuous and monotone operator with the symmetry property* $\langle \mathscr{B}x, y \rangle = \langle \mathscr{B}y, x \rangle$;

(v) *with given constants* $\tilde{\chi} > 0$, $m > 0$, *the inequality*

$$\frac{1}{2} \tilde{\chi} \langle \mathscr{B}(x - y), x - y \rangle + \langle \mathscr{G}(x) - \mathscr{G}(y), x - y \rangle \geqslant m \|x - y\|^2$$

is valid for all $x, y \in K$;

(vi) *for the regularization parameters it holds* $0 < \underline{\chi} \leqslant \chi_k \leqslant \chi_{k+1} \leqslant \overline{\chi} < \infty$ $\forall k$;

(vii) *for all* k *and* $y \in K$, *the operators* $\mathscr{Q}^k + \mathscr{L}_y + \mathscr{N}_K + \chi_k \nabla h$ *are maximal monotone*;

(viii) *for each* $w \in D(\mathscr{Q}) \cap K$, *there exists a sequence* $\{w^k\}$, $w^k \in D(\mathscr{Q}^k) \cap K$, *such that*

$$\lim_{k \to \infty} \|w^k - w\| = 0, \qquad \lim_{k \to \infty} \|\mathscr{Q}^k(w^k) - q(w)\|_{X'} = 0,$$

with $q(w) \in \mathscr{Q}(w)$ (*in general*, $q(w)$ *depends on* $\{w^k\}$);

(ix) *for some solution* x^* *of Problem* (P) *there exist a constant* $\alpha > 1$ *and a sequence* $\{w^k\}$, $w^k \in D(\mathscr{Q}^k) \cap K$, *such that*

$$\lim_{k \to \infty} k^\alpha \|w^k - x^*\| = 0, \qquad \lim_{k \to \infty} k^\alpha \|\mathscr{Q}^k(w^k) - q^*(x^*)\|_{X'} = 0.$$

Under (1-i), (2-i)–(2-iii), a condition guaranteeing that (2-vii) is valid is that each operator $\mathscr{Q}^k$ is maximal monotone and locally bounded at some $x \in clD(\mathscr{Q}^k) \cap K$ (this follows from the Theorems 1 and 3 in [22]). The assumptions (2-viii), (2-ix), concerning the successive approximation of the operator $\mathscr{Q}$, are closely related to Mosco's conditions [15] for the approximation of variational inequalities by using the Browder–Tichonov regularization.

We start the convergence analysis with some preliminary results. As usual, the symbol $\rightharpoonup$ denotes weak convergence.

LEMMA 1. *Let* $C \subset X$ *be a convex closed set, the operators* $\mathscr{A}_0 : X \to 2^{X'}$, $\mathscr{A}_0 + \mathscr{N}_C$ *be maximal monotone and* $D(\mathscr{A}_0) \cap C$ *be a convex set. Moreover, assume that the operator*

$$\mathscr{A}_C : v \to \begin{cases} \mathscr{A}_0(v) & \textit{if } v \in C \\ \emptyset & \textit{otherwise} \end{cases}$$

is locally hemi-bounded at each point $v \in D(\mathscr{A}_0) \cap C$ *and that, for some* $u \in D(\mathscr{A}_0) \cap C$ *and each* $v \in D(\mathscr{A}_0) \cap C$, *there exists* $\eta(v) \in \mathscr{A}_0(v)$ *satisfying*

$$\langle \eta(v), v - u \rangle \geqslant 0 . \tag{3.1}$$

Then, with some $\eta \in \mathscr{A}_0(u)$, *the inequality*

$$\langle \eta, v - u \rangle \geqslant 0 \tag{3.2}$$

holds for all $v \in C$.

Proof. In view of the maximal monotonicity of $\mathscr{A}_0$ and $\mathscr{A}_0 + \mathscr{N}_C$, the operators $\mathscr{A} : v \to \mathscr{A}_0(v) + \mathscr{I}(v - u)$ and $\mathscr{A}_1 = \mathscr{A} + \mathscr{N}_C$ (with $\mathscr{I} : X \to X'$ a canonical isometry) are also maximal monotone. Moreover, they are strongly monotone. Therefore, there exists $w \in D(\mathscr{A}_0) \cap C$ such that $0 \in \mathscr{A}(w) + \mathscr{N}_C(w)$, and due to the definition of the normality operator, this yields

$$\langle \eta(w), v - w \rangle \geqslant 0 \quad \forall v \in C , \tag{3.3}$$

with some $\eta(w) \in \mathscr{A}(w)$.

If $w = u$, then, of course, $\eta(w) \in \mathscr{A}_0(w)$, hence, the conclusion of the lemma is valid. Otherwise, we use the relation

$$\langle \overline{\eta}(v), v - u \rangle \geqslant 0 \quad \forall v \in D(\mathscr{A}_0) \cap C , \tag{3.4}$$

which follows from (3.1) *taking* $\overline{\eta}(v) = \eta(v) + \mathscr{I}(v - u) \in \mathscr{A}(v)$.

Let $w_\lambda = u + \lambda(w - u)$ for $\lambda \in (0, 1]$. Obviously, $w_\lambda \in D(\mathscr{A}_0) \cap C$, and according to (3.4) there exists $\overline{\eta}(w_\lambda) \in \mathscr{A}(w_\lambda)$ ensuring

$$\langle \overline{\eta}(w_\lambda), w - u \rangle \geqslant 0 .$$

Because the operator $\mathscr{A}_C$ is locally hemi-bounded at u, the set $\{\overline{\eta}(w_\lambda) : \lambda \in (0, \lambda_0]\}$ is bounded in V' for a sufficiently small $\lambda_0 > 0$. Hence, if λ tends to 0 in an appropriate manner, the corresponding sequence $\{\overline{\eta}(w_\lambda)\}$ converges weakly in V' to some $\overline{\eta}$. Taking into account that $\lim_{\lambda \to 0} \|w_\lambda - u\| = 0$ and that $\mathscr{A}$ is maximal monotone, one can conclude that $\overline{\eta} \in \mathscr{A}(u)$ and

$$0 \leqslant \lim \langle \overline{\eta}(w_\lambda), w - u \rangle = \langle \overline{\eta}, w - u \rangle .$$

Combining this inequality and inequality (3.3) given with $v = u$, we obtain

$$\langle \overline{\eta} - \eta(w), u - w \rangle \leqslant 0 ,$$

but that contradicts the strong monotonicity of $\mathscr{A}$. □

REMARK 1. The Assumptions (2-ii), (2-iv)–(2-vi) provide strong monotonicity of the operator $\mathscr{Q}^k + \chi_k \nabla h$, and together with (1-i), (2-i), this yields strong monotonicity of $\mathscr{Q}^k + \mathscr{L}^k + \mathscr{N}_K + \chi_k \nabla h$. Moreover, according to (2-vii), the operator

$\mathscr{Q}^k + \mathscr{L}^k + \mathscr{N}_K + \chi_k \nabla h$ is maximal monotone. Hence, for each k, Problem (P^k) is uniquely solvable.

With $x^* \in X^*$, $q^*(x^*)$ as in (2-ix), define

$$\Gamma^k(x^*, x) = \tilde{\chi}\langle \mathscr{B}(x - x^*), x - x^*\rangle + h(x^*) - h(x) - \langle \nabla h(x), x^* - x\rangle + \frac{1}{\chi_k}\langle \mathscr{F}(x^*) + q^*(x^*), x - x^*\rangle . \tag{3.5}$$

For $x \in K$, under (2-ii), (2-iv)–(2-vi) it holds

$$\Gamma^k(x^*, x) \geqslant m\|x^* - x\|^2 \quad \text{and} \quad \Gamma^{k+1}(x^*, x) \leqslant \Gamma^k(x^*, x) . \tag{3.6}$$

The sequence $\{\Gamma^k\}$ plays the role of a Ljapunov function in the further analysis.

LEMMA 2. *Let the Assumptions (1-i), (1-v), (1-vi) and (2-i)–(2-vii), (2-ix) be fulfilled and*

$$\frac{1}{4\gamma m} < \underline{\chi} , \quad 2\tilde{\chi}\tilde{\chi} < 1. \tag{3.7}$$

Then the sequence $\{x^k\}$, *generated by the PAP-method is bounded, for the iterates it holds* $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$ *and the sequence* $\{\Gamma^k(x^*, x^k)\}$ *converges.*

This statement can be proved by modifying the proof of Theorem 2.1 in [24], and due to the rather technical character of the modification, the proof of Lemma 2 is given in the Appendix.

LEMMA 3. *Let the Assumptions (1-i)–(1-iv), 1-vi) and (2-i), (2-iii), (2-iv), (2-viii) be fulfilled, and* $0 < \chi_k \leqslant \bar{\chi}$ *holds for all k. Moreover, let the sequence* $\{x^k\}$ *generated by the PAP-method be bounded, and* $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$. *Then each weak limit point of* $\{x^k\}$ *is a solution of Problem* (P).

Proof. Let $\bar{x}$ be an arbitrary weak limit point of $\{x^k\}$ and let $\{x^k\}_{k\in\mathfrak{K}}$ converge weakly to $\bar{x}$. Since $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$, one gets $x^{k+1} \rightharpoonup \bar{x}$ if $k \in \mathfrak{K}$, $k \to \infty$.

According to (2-viii), for each $y \in D(\mathscr{Q}) \cap K$ one can choose a sequence $\{y^k\}$, $y^k \in D(\mathscr{Q}^k) \cap K$ such that $\lim_{k\to\infty}\|y^k - y\| = 0$, and

$$\lim_{k\to\infty}\|\mathscr{Q}^k(y^k) - q(y)\|_{X'} = 0 \tag{3.8}$$

is valid with some $q(y) \in \mathscr{Q}(y)$. By definition of x^{k+1}, we obtain

$$\langle \mathscr{F}(x^k) + \mathscr{Q}^k(x^{k+1}) + \mathscr{L}^k(x^{k+1}) - \mathscr{L}^k(x^k) + \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), y^k - x^{k+1}\rangle \geqslant 0 ,$$

and the monotonicity of $\mathscr{Q}^k$ (see (2-iv)) leads to

$$\langle \mathscr{F}(x^k) + \mathscr{Q}^k(y^k) + \mathscr{L}^k(x^{k+1}) - \mathscr{L}^k(x^k) \\ + \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), y^k - x^{k+1}\rangle \geqslant 0 , \tag{3.9}$$

Passing to the limit for $k \in \mathfrak{K}$ in (3.9), in view of

$$\lim_{k\to\infty}\|y^k - y\| = 0 , \quad x^{k+1} \rightharpoonup \bar{x} \ (k \in \mathfrak{K}), \lim_{k\to\infty}\|x^{k+1} - x^k\| = 0 ,$$

as well as $0 < \chi_k \leqslant \bar{\chi}$, inequality (3.8) and the conditions (1-iv), (2-i), (2-iii), we obtain

$$\langle \mathscr{F}(\bar{x}) + q(y), y - \bar{x}\rangle \geqslant 0 .$$

Moreover, due to (1-ii), (1-iii), the operators $\mathscr{Q}_0 : y \to \mathscr{F}(\bar{x}) + \mathscr{Q}(y)$ and $\mathscr{Q}_0 + \mathscr{N}_K$ are maximal monotone, and the operator

$$\mathscr{Q}_K : y \to \begin{cases} \mathscr{Q}_0(y) & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}$$

is locally hemi-bounded at each point of K. Thus, applying Lemma 1 with $u = \bar{x}$, $C = K$ and $\mathscr{A}_0 = \mathscr{Q}_0$, we obtain

$$\langle \mathscr{F}(\bar{x}) + q(\bar{x}), y - \bar{x}\rangle \geqslant 0 \quad \forall y \in K ,$$

where $q(\bar{x}) \in \mathscr{Q}(\bar{x})$. Hence, $\bar{x}$ is a solution of (P). □

REMARK 2. If we suppose instead of (1-iv) that the operator $\mathscr{F}$ is monotone and hemicontinuous on X and that $\mathscr{F}$ is bounded on K, then from (3.9) and

$$\langle \mathscr{F}(x^k), y^k - x^{k+1}\rangle \leqslant \langle \mathscr{F}(y), y - x^k\rangle + \langle \mathscr{F}(x^k), y^k - y\rangle + \langle \mathscr{F}(x^k), x^k - x^{k+1}\rangle$$

one gets

$$\langle \mathscr{F}(y) + q(y), y - \bar{x}\rangle \geqslant 0 \quad \forall y \in D(\mathscr{Q}) \cap K ,$$

with $q(y)$ as in (3.8).

The operators $\mathscr{F} + \mathscr{Q}$ and $\mathscr{F} + \mathscr{Q} + \mathscr{N}_K$ are maximal monotone in this case and the operator

$$y \to \begin{cases} \mathscr{F}(y) + \mathscr{Q}(y) & \text{if } y \in K \\ \emptyset & \text{otherwise} \end{cases}$$

is locally hemi-bounded at each point of K. Thus, Lemma 1 can be applied, proving that $\bar{x}$ solves Problem (P).

THEOREM 1. *Let the Assumptions 1 and 2 and condition* (3.7) *be fulfilled. Then the following conclusions are true*:

(i) *Problem* (P^k) *is uniquely solvable for each k, the sequence* $\{x^k\}$ *generated by the PAP-method is bounded, and each weak limit point of* $\{x^k\}$ *is a solution of Problem* (P);

(ii) *if, in addition, (2-ix) with some* $\alpha > 1$ *is valid for each* $x \in X^*$ *and*

$$z^k \rightharpoonup z \text{ in } X,\ z^k \in K \ \Rightarrow\ \nabla h(z^k) \rightharpoonup \nabla h(z) \text{ in } X', \tag{3.10}$$

then the whole sequence $\{x^k\}$ *converges weakly to a solution* x^* *of Problem* (P);

(iii) *if, moreover,*

$$\lim_{k\to\infty} \langle \mathscr{G}(x^k) - \mathscr{G}(x^*), x^k - x^* \rangle = 0 \tag{3.11}$$

holds with x^* *as in* (ii), *then* $\{x^k\}$ *converges strongly to* x^*.

Proof. Conclusion (i) follows immediately from the Lemmata 2, 3 and Remark 1. To prove (ii), suppose that $\{x^k\}_{k\in\mathfrak{R}_1}$, $\{x^k\}_{k\in\mathfrak{R}_2}$ are two subsequences converging weakly to $\bar{x}$, $\tilde{x}$, respectively. Then, according to Lemma 3, $\bar{x}$, $\tilde{x}$ belong to X^*, and because (2-ix) is valid for each $x \in X^*$, Lemma 2 ensures that the sequences $\{\Gamma^k(\bar{x}, x^k)\}_{k\in\mathbb{N}}$, $\{\Gamma^k(\tilde{x}, x^k)\}_{k\in\mathbb{N}}$ are convergent.

By definition of Γ^k and with regard to $\bar{x} \in X^*$, the symmetry of the operator $\mathscr{B}$ and (2-ii), (2-v), we obtain for $x \in K$

$$\begin{aligned}
&\Gamma^k(\bar{x}, x) - \Gamma^k(\tilde{x}, x) \\
&\quad = (h(\bar{x}) - h(\tilde{x}) - \langle \nabla h(\tilde{x}), \bar{x} - \tilde{x}\rangle) + \langle \nabla h(\tilde{x}) - \nabla h(x), \bar{x} - \tilde{x}\rangle \\
&\qquad + \frac{1}{\chi_k} \langle \mathscr{F}(\bar{x}) + q^*(\bar{x}), \tilde{x} - \bar{x}\rangle \\
&\qquad + \frac{1}{\chi_k} \langle \mathscr{F}(\bar{x}) + q^*(\bar{x}) - \mathscr{F}(\tilde{x}) - q^*(\tilde{x}), x - \tilde{x}\rangle \\
&\qquad + \tilde{\chi}\langle \mathscr{B}(\bar{x} - \tilde{x}), \bar{x} - \tilde{x}\rangle + 2\tilde{\chi}\langle \mathscr{B}(\bar{x} - \tilde{x}), \tilde{x} - x\rangle \\
&\quad \geqslant m\|\bar{x} - \tilde{x}\|^2 + \langle \nabla h(\tilde{x}) - \nabla h(x), \bar{x} - \tilde{x}\rangle \\
&\qquad + \frac{1}{\chi_k} \langle \mathscr{F}(\bar{x}) + q^*(\bar{x}) - \mathscr{F}(\tilde{x}) - q^*(\tilde{x}), x - \tilde{x}\rangle \\
&\qquad + 2\tilde{\chi}\langle \mathscr{B}(\bar{x} - \tilde{x}), \tilde{x} - x\rangle\,.
\end{aligned} \tag{3.12}$$

Inserting $x = x^k$ in (3.12) and passing to the limit for $k \in \mathfrak{R}_2$, one can conclude from (3.10) and (3.12) that

$$\bar{\gamma} - \tilde{\gamma} \geqslant m\|\bar{x} - \tilde{x}\|^2\,,$$

where $\bar{\gamma} = \lim_{k\to\infty} \Gamma^k(\bar{x}, x^k)$, $\tilde{\gamma} = \lim_{k\to\infty} \Gamma^k(\tilde{x}, x^k)$. Obviously, in the same way the 'symmetric' inequality

$$\tilde{\gamma} - \bar{\gamma} \geqslant m\|\bar{x} - \tilde{x}\|^2$$

can be concluded, and therefore $\bar{x} = \tilde{x}$ is valid, proving the uniqueness of the weak limit point for $\{x^k\}$.

Denoting this limit point by x^*, now we assume additionally that relation (3.11) is fulfilled. With $\{w^k\}$ chosen according to (2-ix), from (2-iv) one gets

$$\begin{aligned}
&\langle \mathcal{B}(x^{k+1}-x^*), x^{k+1}-x^*\rangle \\
&\quad = \langle \mathcal{B}(x^{k+1}-w^k), x^{k+1}-w^k\rangle - \langle \mathcal{B}(x^*-w^k), x^{k+1}-x^*\rangle \\
&\qquad - \langle \mathcal{B}(x^{k+1}-w^k), x^*-w^k\rangle \\
&\quad \leqslant \langle \mathcal{Q}^k(x^{k+1}) - \mathcal{Q}^k(w^k), x^{k+1}-w^k\rangle - \langle \mathcal{B}(x^{k+1}-x^*), x^*-w^k\rangle \\
&\qquad - \langle \mathcal{B}(x^{k+1}-w^k), x^*-w^k\rangle\,. \qquad (3.13)
\end{aligned}$$

To estimate the term $\langle \mathcal{Q}^k(x^{k+1}), x^{k+1}-w^k\rangle$, we use Problem (P^k). Together with (3.13) this gives

$$\begin{aligned}
&\langle \mathcal{B}(x^{k+1}-x^*), x^{k+1}-x^*\rangle \leqslant \langle \mathcal{Q}^k(w^k) - q^*(x^*), w^k - x^{k+1}\rangle \\
&\quad + \langle q^*(x^*), w^k - x^*\rangle + \langle q^*(x^*), x^* - x^{k+1}\rangle \\
&\quad + \chi_k\langle \nabla h(x^{k+1}) - \nabla h(x^k), w^k - x^{k+1}\rangle \\
&\quad + \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), w^k - x^{k+1}\rangle \\
&\quad + \langle \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k), w^k - x^{k+1}\rangle + \langle \mathcal{F}(x^*), w^k - x^*\rangle \\
&\quad + \langle \mathcal{F}(x^*), x^* - x^{k+1}\rangle + \langle \mathcal{B}x^* + \mathcal{B}w^k - 2\mathcal{B}x^{k+1}, x^* - w^k\rangle \qquad (3.14)
\end{aligned}$$

Now, using (2-ix), (1-iv), (2-iii), (2-i), (2-iv), and taking into account the boundedness of $\{x^k\}$ and $x^k \rightharpoonup x^*$, $\|x^k - x^{k+1}\| \to 0$, the relation

$$\lim_{k\to\infty} \langle \mathcal{B}(x^{k+1}-x^*), x^{k+1}-x^*\rangle = 0$$

can be deduced from (3.14). Together with (3.11) and (2-v) this ensures conclusion (iii). □

Replacing Assumption (1-iv) as mentioned in Remark 2, the modification in the proof of Theorem 1 is connected only with the estimation of the term $\langle \mathcal{F}(x^k) - \mathcal{F}(x^*),\ w^k - x^{k+1}\rangle$. Due to the monotonicity of $\mathcal{F}$, we obtain

$$\begin{aligned}
&\langle \mathcal{F}(x^k) - \mathcal{F}(x^*), w^k - x^{k+1}\rangle \\
&\quad \leqslant \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), w^k - x^* + x^k - x^{k+1}\rangle + \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), x^* - x^k\rangle \\
&\quad \leqslant \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), w^k - x^* + x^k - x^{k+1}\rangle
\end{aligned}$$

and the boundedness of $\mathcal{F}$ together with (2-ix) and $\|x^k - x^{k+1}\| \to 0$ ensure

$$\overline{\lim_{k\to\infty}}\, \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), w^k - x^{k+1}\rangle \leqslant 0\,.$$

REMARK 3. Obviously, the conditions (3.7), used in Lemma 2 and also in Theorem 1, are compatible if and only if $2\gamma m \geqslant \tilde{\chi}$. But, they are certainly compatible, for instance, if the regularizing functional h is strongly convex. Then, (2-ii) can be satisfied with a strongly monotone operator $\mathcal{G}$, and assuming m is the modulus of the strong monotonicity, in (2-v) an arbitrary small $\tilde{\chi}$ is appropriate. In case $\mathcal{F}$ is monotone and we deal with proximal-like methods, which correspond formally to the PAP-method by setting $\mathcal{Q}^k := \mathcal{Q}^k + \mathcal{F}$, $\mathcal{F} := \mathbf{0}$, $\mathcal{L}^k := \mathbf{0}$, condition (1-v) is valid for arbitrary large γ, and hence, $2m\gamma \geqslant \tilde{\chi}$ can be fulfilled, too. This kind of proximal-like methods with weak regularization and regularization on a subspace have been developed in [9–11] for solving problems in elasticity theory and optimal control, and in [8] for abstract variational inequalities with non-symmetric, multi-valued and monotone operators.

In general, one should choose the constants $\tilde{\chi}$ and m such that (2-v) is satisfied and the ratio $\tilde{\chi}/m$ is as small as possible.

If Assumptions 1 and (2-i), (2-viii), (2-ix) (for all $x \in X^*$) are fulfilled, the operators $\mathcal{Q}^k + \mathcal{L}_y + \mathcal{N}_K$ are maximal monotone and $\mathcal{Q}^k$ are strongly monotone on K with a common modulus m_1, and if $2\gamma m_1 \geqslant 1$, then the iterates of the PAP-method applied with $h \equiv 0$ converge strongly to a solution of Problem (P).

To apply this analysis in the case when the auxiliary problems do not include an approximation of the operator $\mathcal{Q}$ (see, for instance [4, 21]), one has to modify Problem (P^k) and Assumption 2 as follows:

- the point $x^{k+1} \in K$ is defined such that

$$\begin{aligned} &\langle \mathcal{F}(x^k) + q^{k+1} + \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k) \\ &\quad + \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle \geqslant 0 \quad \forall x \in K \end{aligned} \tag{3.15}$$

  is valid with some $q^{k+1} \in \mathcal{Q}(x^{k+1})$;
- in (2-iv) the operator $\mathcal{Q}^k$ is replaced by $\mathcal{Q}$, and the inequality

$$\langle q(x) - q(y), x - y \rangle \geqslant \langle \mathcal{B}(x - y), x - y \rangle$$

  has to be fulfilled for all $x, y \in D(\mathcal{Q}) \cap K$ and all $q(x) \in \mathcal{Q}(x)$, $q(y) \in \mathcal{Q}(y)$;
- in (2-vii) $\mathcal{Q}^k$ is replaced by $\mathcal{Q}$;
- the conditions (2-viii) and (2-ix) are skipped.

THEOREM 2. *Lemmata 2, 3, and Theorem 1 remain true under the modifications described above*.

The proof of Lemma 2 remains true in this case if we take $w^k = x^*$ and replace $\mathcal{Q}^k(w^k)$ by $q^*(x^*)$ and $\mathcal{Q}^k(x^{k+1})$ by q^{k+1}.

In the proof of Lemma 3 one has only to take $y^k \equiv y$ and to substitute an arbitrary $q(y) \in \mathcal{Q}(y)$ for $\mathcal{Q}^k(y^k)$. The proof of Theorem 1 remains true completely.

4. Applications

We start with some observations concerning the application of the PAP-method to different types of variational inequalities.

For the variational inequality

$(\tilde{P})$ find $x^* \in X$: $\langle \mathcal{F}(x^*) + \mathcal{Q}(x^*), x - x^* \rangle + f(x) - f(x^*) \geqslant 0, \quad \forall x \in X$,

where $f : X \to \overline{\mathbb{R}} \equiv \mathbb{R} \cup \{+\infty\}$ is a convex lsc functional and the operators $\mathcal{Q} + \partial f$ (instead of $\mathcal{Q}$) and $\mathcal{F}$ satisfy Assumption 1 with $K = dom\ f$, the studied scheme can be directly applied if we construct the auxiliary problems (P^k) being caused from the equivalent formulation for $(\tilde{P})$:

$$\text{find } x^* \in \text{dom } f\text{: } \langle \mathcal{F}(x^*) + \mathcal{Q}(x^*) + \partial f(x^*), x - x^* \rangle \geqslant 0 \quad \forall x \in \text{dom } f\ .$$

However, one can deal with more convenient requirements for the approximation of f (see [24]) if the auxiliary problems have the form

$(\tilde{P}^k)$ find $x^{k+1} \in X$:

$$\begin{aligned}&\langle \mathcal{F}(x^k) + \mathcal{Q}^k(x^{k+1}) + \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k)\\ &+ \chi_k(\nabla h(x^{k+1}) - \nabla h(x^k)), x - x^{k+1} \rangle\\ &+ f^k(x) - f^k(x^{k+1}) \geqslant 0 \quad \forall x \in X\ ,\end{aligned}$$

where $f^k : X \to \overline{\mathbb{R}}$ is a convex lsc functional. In the case, compiling the convergence analysis in [24] and in this paper, Theorem 1 can be proved under the following modifications concerning the Assumptions 1 and 2:
everywhere K is replaced by $dom\ f$ and $\mathcal{N}_K$ by ∂f;

(1-v)′: for $y \in \text{dom } f$, $\mathcal{L}_y$ is monotone on $D(\mathcal{Q}) \cap \text{dom } f$, and

$$\begin{aligned}&\langle \mathcal{F}(y) - \mathcal{L}_y(y) + \mathcal{L}_y(x) + q(x), y - x \rangle + f(y) - f(x)\\ &\quad \geqslant \gamma \| \mathcal{F}(y) - \mathcal{L}_y(y) - \mathcal{F}(x) + \mathcal{L}_y(x) \|^2_{X'} \quad (\gamma > 0 - \text{const.})\end{aligned}$$

holds true whenever

$$\langle \mathcal{F}(x) + q(x), y - x \rangle + f(y) - f(x) \geqslant 0 \quad \text{with some } q(x) \in \mathcal{Q}(x);$$

(2-iv)′: $\langle \mathcal{Q}^k(x) - \mathcal{Q}^k(y), x - y \rangle + f^k(x) - f^k(y) - \langle g^k(y), x - y \rangle \geqslant \langle \mathcal{B}(x - y), x - y \rangle$, $\forall x, y \in D(\mathcal{Q}^k) \cap D(\partial f^k)$, $\forall g^k(y) \in \partial f^k(y)$, with $\mathcal{B}$ as in (2-iv);

(2-viii)′: $f^k \geqslant f$ and for each $w \in D(\mathcal{Q}) \cap \text{dom } f$ there exists a sequence $\{w^k\}$, $w^k \in D(\mathcal{Q}^k) \cap \text{dom } f^k$, such that

$$\lim_{k \to \infty} \|w^k - w\| = 0 \ \lim_{k \to \infty} f^k(w^k) = f(w), \ \lim_{k \to \infty} \|\mathcal{Q}^k(w^k) - q(w)\|_{X'} = 0,$$

with $q(w) \in \mathcal{Q}(w)$;

(2-ix)′: for some solution x^* of problem $(\tilde{P})$ there exist a constant $\alpha > 1$ and a sequence $\{w^k\}$, $w^k \in D(\mathcal{Q}^k) \cap \text{dom } f^k$, such that

$$\lim_{k \to \infty} k^\alpha \|w^k - x^*\| = 0, \ \lim_{k \to \infty} k^\alpha \|\mathcal{Q}^k(w^k) - q^*(x^*)\|_{X'} = 0$$

and $\lim_{k \to \infty} k^\alpha \max\{f^k(w^k) - f(x^*), 0\} = 0$.

REMARK 4. Of course, the conditions (2-viii), (2-ix) and especially (2-viii)′, (2-ix)′ are not intended for a wide class of monotone operators, because real possibilities to satisfy these conditions are connected with individual properties of $\mathcal{Q}$ or mutual properties of $\mathcal{Q}$ and f. The general way based on the Moreau–Yosida regularization is very expensive (however, the use of this regularization in a special algorithm of the APP in [5] seems to be promising). A couple of problems in mathematical physics (the problem of linear elasticity with friction, Bingham's problem, etc.) takes the form of a variational inequality $(\tilde{P})$ with a single-valued operator $\mathcal{Q}$, and a uniform approximation of f by a sequence of differentiable functionals is possible (see [7, 8, 16, 17]. In this case, there are no serious difficulties in satisfying (2-viii)′, (2-ix)′.

The extension of Theorems 1 and 2 to the methods developed for inclusion (1.1) can be carried out using the relationship

$$0 \in \Psi(x) \iff \langle \Psi(x), y - x \rangle \geqslant 0 \quad \forall y \in K\,,$$

where K is an arbitrary convex closed set containing $D(\Psi)$.

For instance, method (1.5) applied to inclusion (1.1) can be rewritten in the form

$$\text{find } x^{k+1} \in X: \quad 0 \in \frac{1}{\epsilon}[x^{k+1} - x^k] + \Psi_1(x^k) + \Psi_2(x^{k+1})\,, \tag{4.1}$$

and if Ψ_2 is a maximal monotone operator (that is usually supposed), then the operator

$$\Psi^k : x \to \frac{1}{\epsilon}(x - x^k) + \Psi_1(x^k) + \Psi_2(x)\,,$$

is maximal monotone, too, and $D(\Psi^k) = D(\Psi_2)$. Hence, (4.1) is equivalent to

$$\text{find } x^{k+1} \in K: \\ \left\langle \Psi_1(x^k) + \frac{1}{\epsilon}(x^{k+1} - x^k) + \Psi_2(x^{k+1}), x - x^{k+1} \right\rangle \geqslant 0 \quad \forall x \in K\,, \tag{4.2}$$

with $K \supset D(\Psi_2)$ an arbitrary convex closed set, and (4.2) is a partial case of Problem (3.15) with

$$\mathcal{F} = \Psi_1,\ \mathcal{Q} = \Psi_2,\ \mathcal{L}^k \equiv \mathbf{0},\ \chi_k = \frac{1}{\epsilon},\ h : x \to \frac{1}{2}\|x\|^2\,.$$

Analogously, taking $\Psi = \Psi_1 + \Psi_2$, the auxiliary problem (1.7) may be rewritten in the form (3.15) with

$$\mathcal{F} = \Psi_1,\ q^{k+1} \in \Psi_2(x^{k+1}),\ h \text{ with } \nabla h = \Xi_1,\ \mathcal{L}^k = \Xi_2 + \Psi_1,\ \chi_k = \frac{1}{\epsilon},$$

and $K \supset D(\Psi_2)$ a convex closed set. Of course, the decomposition $\Psi = \Psi_1 + \Psi_2$ has to be performed maintaining the conditions for $\mathcal{F}$ and $\mathcal{Q}$ in Assumption 1.

4.1. DECOMPOSITION

Due to the splitting of the main operator in (P) into a sum $\mathscr{F}+\mathscr{Q}$, certain decomposition properties are already inherent in the auxiliary problem (P^k) as well as in (3.15), where x^{k+1} is calculated with $\mathscr{F}$ fixed at the point x^k. Such a splitting can be caused by several reasons, in particular:

- usually, applications of the APP in its traditional way assume that the operator in the variational inequality is single-valued, whereas applying the proximal methods the operator is supposed to be monotone, but not necessarily single-valued (concerning relaxations of these conditions see [2] Sect. 3, [12, 25]). The class of problems admitting the use of the PAP-method is wider, including variational inequalities with an operator $\mathscr{F}+\mathscr{Q}$, whose 'geometrical' properties are defined by Assumptions (1-ii), (1-v);
- for some variational inequalities, the problems (P^k) or (3.15) can be solved easier than the auxiliary problems arising in proximal methods. This fact was the motivation for Lions–Mercier's method (1.5) for Problem (1.1), although the conditions assumed for the operator Ψ in [13, 6] permit a straightforward use of the proximal point method, too.

It is of special interest when, under an appropriate splitting of the operator, the application of the PAP-method leads to a decomposition of the auxiliary problem in the space X.

Let X be a Cartesian product of the Hilbert spaces X_1 and X_2 with their duals X_1' and X_2', respectively. In the sequel, subscript i, $i=1,2$, indicates that a point or a set belongs to X_i or that an operator acts from X_i into X_i'.

Assume that $K=K_1\times K_2$ and the operator $\mathscr{Q}$ possesses the following separability property:

$$\langle \mathscr{Q}(x), y\rangle = \langle \mathscr{Q}_1(x_1), y_1\rangle + \langle \mathscr{Q}_2(x_2), y_2\rangle$$

for $x=(x_1,x_2)$, $y=(y_1,y_2)$. In this case, it is natural that the approximating operators $\mathscr{Q}^k$ (if an approximation of $\mathscr{Q}$ is needed) have the same separability property. Then, choosing the auxiliary operators $\mathscr{L}^k: x\to(\mathscr{L}_1^k(x_1), \mathscr{L}_2^k(x_2))$ and the function $h(x)=h_1(x_1)+h_2(x_2)$, Problem (P^k) can be decomposed into the pair of variational inequalities ($i=1,2$):

$$\begin{aligned} &\text{find } x_i^{k+1}\in K_i:\\ &\quad\langle \mathscr{F}_i(x_i^k) + \mathscr{Q}_i^k(x_i^{k+1}) + \mathscr{L}_i^k(x_i^{k+1}) - \mathscr{L}_i^k(x_i^k)\\ &\quad + \chi_k(\nabla h_i(x_i^{k+1}) - \nabla h_i(x_i^k)), x_i - x_i^{k+1}\rangle \geqslant 0 \quad \forall x_i\in K_i\,, \end{aligned}$$

where $\mathscr{F}: X_i\to X_i'$ is the composition of $\mathscr{F}$ and the canonical projection onto X_i'.

A related method for Problem $(\tilde{P})$ with $\mathscr{Q}\equiv\mathbf{0}$ and $\mathscr{F}$ a monotone operator was investigated in [14].

For the variational inequality (P) with $X = \mathbb{R}^n$, $\mathcal{Q} = \mathbf{0}$, $K = K_1 \times K_2$ and $\mathcal{F}$ a monotone affine operator of a special structure, Tseng [26] developed a decomposition algorithm where at each step two minimization problems with separable, strongly convex, quadratic functions and feasible sets K_1 and K_2, respectively, have to be solved one after the other. This algorithm was derived (by means of an appropriately chosen matrix $\mathcal{D}$) from the *asymmetric projection method*, which can be written as:

$$\text{find } x^{k+1} \in K: \\ \langle \mathcal{F}(x^k) + \mathcal{D}(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0 \quad \forall x \in K, \tag{4.3}$$

with $\mathcal{D}$ an $n \times n$ positive definite (non-symmetric) matrix. Obviously, the auxiliary variational inequality (4.3) can be considered as a partial case of Problem (P^k) with $\mathcal{Q}^k \equiv \mathbf{0}$, $\chi_k \equiv \chi$, $\mathcal{L}^k = \mathcal{D} - \chi \mathcal{I}$ and $h(x) = (\chi/2)\|x\|^2$. It is worth noting that the auxiliary operator corresponding to Tseng's decomposition algorithm is not separable, in general.

4.2. LINEAR APPROXIMATION METHODS

These methods have been studied extensively, mainly for finite-dimensional problems (P) with $\mathcal{Q} \equiv \mathbf{0}$, and we refer to [18] and the bibliography therein. The operator $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$ is approximated at the point x^k by $\mathcal{F}(x^k) + \mathcal{A}(x^k)(x - x^k)$, where $\mathcal{A}(x^k)$ is an $n \times n$-matrix, and $x^{k+1} \in K$ is defined by solving the variational inequality

$$\langle \mathcal{F}(x^k) + \mathcal{A}(x^k)(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0 \quad \forall x \in K. \tag{4.4}$$

Depending on the choice of $\mathcal{A}(x^k)$ well-known methods can be obtained:

- the settings $\mathcal{A}(x^k) = \nabla\mathcal{F}(x^k)$, $\mathcal{A}(x^k) = \frac{1}{2}[\nabla\mathcal{F}(x^k) + \nabla\mathcal{F}(x^k)^T]$ or $\mathcal{A}(x^k)$ is an approximation of $\nabla\mathcal{F}(x^k)$ correspond to the Newton, symmetrized Newton or quasi Newton methods, respectively;
- the cases $\mathcal{A}(x^k) = \mathcal{L}(x^k) + (1/\omega)\mathcal{D}(x^k)$ or $\mathcal{A}(x^k) = \mathcal{U}(x^k) + (1/\omega)\mathcal{D}(x^k)$ ($\mathcal{L}(x^k)$, $\mathcal{U}(x^k)$ and $\mathcal{D}(x^k)$ are strictly lower triangular, strictly upper triangular and diagonal parts of $\nabla\mathcal{F}(x^k)$, respectively, and $0 < \omega < 2$), correspond to SOR-methods;
- in case $\mathcal{A}(x^k) = \mathcal{M}$, where $\mathcal{M}$ is a symmetric positive-definite matrix, a projection method is designed, etc.

If there exists a symmetric positive definite matrix $\mathcal{S}$ such that $\mathcal{A}(x) - \mathcal{S}$ is positive semi-definite for all $x \in K$, then Problem (4.4) can be transformed into the auxiliary problem (P^k) of the PAP-method with $\mathcal{Q}^k \equiv \mathbf{0}$, $\mathcal{L}^k : x \to (\mathcal{A}(x^k) - \mathcal{S})x$, $h(x) = \frac{1}{2}x^T\mathcal{S}x$ and $\chi_k \equiv 1$.

For the general variational inequality (P) with $X = \mathbb{R}^n$, the methods considered can be extended using the same $\mathcal{L}^k$, h and χ_k and an appropriate approximation $\mathcal{Q}^k$

of $\mathscr{Q}$ as in Assumption 2. Moreover, taking into account the mutual properties of $\mathscr{Q}^k$ and h, the requirements for h may be weakened according to (2-ii), (2-iv), (2-v).

Of course, in both cases the choice of $\mathscr{S}$ has to satisfy Assumption (1-v).

4.3. ON THE USE OF WEAK REGULARIZATION

Weak regularization in proximal methods deals with a regularizing functional, being not strongly convex on $(X, \|\cdot\|)$, but strongly convex on X endowed by a weaker norm than $\|\cdot\|$.

Taking into account the convergence results for the classical proximal point method (see [23]), an application of weak regularization is justified if the mutual properties of the operator of the variational inequality and the regularizing functional provide weak convergence of the iterates in the space $(X, \|\cdot\|)$. We illustrate this approach using the well-known model of linear elasticity with friction (see [3], Sect. 3 and [17], Sect. 5). Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with a Lipschitz-continuous boundary Γ, $g \in L_\infty(\Gamma)$ be a given non-negative function ($\text{mes}_\Gamma \text{ supp } g > 0$), $a_{klpm} \in L_\infty(\Omega)$ $(k, l, p, m = 1, 2)$ be given functions with symmetry property

$$a_{klpm} = a_{lkpm} = a_{pmkl} . \tag{4.5}$$

Moreover, it is supposed that there exists a positive constant c_0 such that

$$a_{klpm}(x)\sigma_{kl}\sigma_{pm} \geqslant c_0 \sigma_{kl}\sigma_{kl} \text{ a.e. on } \Omega \tag{4.6}$$

holds for all symmetric matrices $[\sigma_{kl}]_{k,l=1,2}$ (here and in the sequel we follow Einstein's summation convention, i.e., the summation is performed over terms with repeating indices).

Let $V = [H^1(\Omega)]^2$ and V' be the dual of V. We consider the variational inequality

$$\text{find } u \in V: \quad \langle \mathscr{A}u - l, v - u \rangle + j(v) - j(u) \geqslant 0 \quad \forall v \in V , \tag{4.7}$$

where $l \in V'$ is a given linear functional,

$$\langle \mathscr{A}u, v \rangle = \int_\Omega a_{klpm} \epsilon_{kl}(u) \epsilon_{pm}(v) \, d\Omega ,$$

$$j(u) = \int_\Gamma g|u_t| \, d\Gamma ,$$

with $\epsilon_{kl}(u) = \frac{1}{2}(\partial u_k / \partial x_l + \partial u_l / \partial x_k)$ (the component of the strain tensor) and u_t the tangential (w.r.t. Γ) component of u.

The linear operator $\mathscr{A}$ is not strongly (and even not strictly) monotone, its kernel has the structure

$$\{z = (z_1, z_2) : z_1 = a_1 - bx_2, z_2 = a_2 + bx_1\}$$

with arbitrary $a_1, a_2, b \in \mathbb{R}$. However, the second Korn inequality (see [17]) and (4.5), (4.6) ensure that there exists a positive constant c_1 such that

$$\langle \mathscr{A}u, u\rangle + \|u\|^2_{[L_2(\Omega)]^2} \geqslant c_1\|u\|^2_V ,$$

and hence, taking $h(u) = \|u\|^2_{[L_2(\Omega)]^2}$, the operator $\mathscr{A}u + \nabla h(u)$ (with $\nabla h : V \to V'$) is strongly monotone in V, although h is not strongly convex w.r.t. H^1-norm, but only w.r.t. L_2-norm.

The proximal method considered in [10] couples weak regularization with successive discretization of the original problems (by means of the finite element method on a sequence of subspaces $\{V_k\}$) and a successive approximation of the non-differential functional j. This reads as follows:

$$\begin{aligned}&\text{find } u^{k+1} \in V:\\ &\quad \langle \mathscr{A}u^{k+1} - l + \chi_k(\nabla h(u^{k+1}) - \nabla h(u^k)), v - u^{k+1}\rangle\\ &\quad + j^k(v) - j^k(u^{k+1}) \geqslant 0 \quad \forall v \in V ,\end{aligned}$$

where

$$j^k(u) = \begin{cases} \int_\Gamma g\sqrt{u_t^2 + r_k}\, \mathrm{d}\Gamma & \text{if } u \in V_{k+1}\\ +\infty & \text{otherwise}\end{cases}$$

and $r_k > 0$, $\lim_{k\to\infty} r_k = 0$.

Obviously, this method is a special case of the PAP-method for a variational inequality in the form $(\tilde{P})$ with auxiliary problems $(\tilde{P}^k)$, in which

$$\mathscr{F} \equiv \mathbf{0},\ \ \mathscr{Q}^k \equiv \mathscr{Q} : u \to \mathscr{A}u - l \text{ and } f^k = j^k .$$

One can show that, if a solution of (4.7) belongs to $[H^2(\Omega)]^2$, then a suitable choice of $\{V_k\}$ and $\{r_k\}$ permits to satisfy the modified Assumptions 1 and 2 at the beginning of this section, moreover, (2-ix) can be guaranteed for any solution of (4.7).

Appendix

Proof of Lemma 2. In the sequel, we make use of the following inequalities, which are valid for arbitrary $a, b, x \in X$, $p \in X'$ and $\epsilon > 0$:

$$\langle p, a\rangle \leqslant \frac{1}{2\epsilon}\|p\|^2_{X'} + \frac{\epsilon}{2}\|a\|^2 , \tag{A.1}$$

$$\begin{aligned}&\langle \mathscr{B}(a-b), a-b\rangle\\ &\quad \leqslant (1+\epsilon)\langle \mathscr{B}(a-x), a-x\rangle + \frac{1+\epsilon}{\epsilon}\langle \mathscr{B}(b-x), b-x\rangle .\end{aligned} \tag{A.2}$$

In order to estimate $\Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k)$, where Γ^k is defined by (3.5), we obtain from (3.6) that

$$\Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \leqslant \Gamma^k(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) .$$

The right-hand side of this inequality can be split up as follows:

$$\Gamma^k(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \equiv s_1 + s_2 + s_3 + s_4 ,$$

with

$$s_1 = h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle ,$$

$$s_2 = \langle \nabla h(x^k) - \nabla h(x^{k+1}), w^k - x^{k+1} \rangle + \frac{1}{\chi_k} \{ \mathcal{F}(x^*) + q^*(x^*), x^{k+1} - x^k \rangle ,$$

$$s_3 = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - w^k \rangle$$

and

$$s_4 = \tilde{\chi} \langle \mathcal{B}(x^{k+1} - x^*), x^{k+1} - x^* \rangle - \tilde{\chi} \langle \mathcal{B}(x^k - x^*), x^k - x^* \rangle . \tag{A.3}$$

We suppose that the sequence $\{w^k\}$ satisfies (2-ix). In view of (2-ii), one gets

$$s_1 \leqslant -\langle \mathcal{G}(x^{k+1}) - \mathcal{G}(x^k), x^{k+1} - x^k \rangle , \tag{A.4}$$

and (2-iii) and (A.1) yield

$$s_3 \leqslant \frac{\tau}{2} \|x^{k+1} - x^k\|^2 + \frac{1}{2\tau} l_h^2 \|x^* - w^k\|^2 , \tag{A.5}$$

with an arbitrary $\tau > 0$.

Setting $x = w^k$ in Problem (P^k), we obtain

$$s_2 \leqslant \frac{1}{\chi_k} \langle \mathcal{F}(x^k) + \mathcal{Q}^k(x^{k+1}) + \mathcal{L}^k(x^{k+1}) - (\mathcal{L}^k(x^k), w^k - x^{k+1} \rangle + \frac{1}{\chi_k} \langle \mathcal{F}(x^*) + q^*(x^*), x^{k+1} - x^k \rangle ,$$

and due to (2-iv),

$$\begin{aligned} \chi_k s_2 \leqslant & \langle \mathcal{F}(x^k) + \mathcal{Q}^k(w^k) + \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k), w^k - x^{k+1} \rangle \\ & - \langle \mathcal{B}(w^k - x^{k+1}), w^k - x^{k+1} \rangle + \langle \mathcal{F}(x^*) + q^*(x^*), x^{k+1} - x^k \rangle \\ = & -\langle \mathcal{B}(w^k - x^{k+1}), w^k - x^{k+1} \rangle + \langle \mathcal{F}(x^k) + q^*(x^*), x^* - x^k \rangle \\ & + \langle \mathcal{Q}^k(w^k) - q^*(x^*), x^* - x^{k+1} \rangle + \langle \mathcal{F}(x^k) + q^*(x^*), x^k - x^{k+1} \rangle \\ & + \langle \mathcal{Q}^k(w^k) - q^*(x^*), w^k - x^* \rangle + \langle q^*(x^*), w^k - x^* \rangle \\ & + \langle \mathcal{F}(x^k), w^k - x^* \rangle + \langle \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k), w^k - x^{k+1} \rangle \\ & + \langle \mathcal{F}(x^*) + q^*(x^*), x^{k+1} - x^k \rangle . \end{aligned}$$

But, by definition of x^*, $q^*(x^*)$,

$$\langle \mathcal{F}(x^*) + q^*(x^*), x^k - x^* \rangle \geqslant 0 ,$$

and with (1-v) one can continue:

$$\begin{aligned}
\chi_k s_2 \leqslant & -\langle \mathcal{B}(w^k - x^{k+1}), w^k - x^{k+1} \rangle \\
& - \gamma \| \mathcal{F}(x^k) - \mathcal{L}^k(x^k) - \mathcal{F}(x^*) + \mathcal{L}^k(x^*) \|^2_{X'} \\
& + \langle \mathcal{Q}^k(w^k) - q^*(x^*), x^* - x^{k+1} \rangle + \langle \mathcal{Q}^k(w^k) - q^*(x^*), w^k - x^* \rangle \\
& + \langle q^*(x^*), w^k - x^* \rangle + \langle \mathcal{F}(x^k), w^k - x^* \rangle \\
& + \langle \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^k), w^k - x^{k+1} \rangle + \langle \mathcal{L}^k(x^*) - \mathcal{L}^k(x^k), x^k - x^* \rangle \\
& + \langle \mathcal{F}(x^k) - \mathcal{F}(x^*), x^k - x^{k+1} \rangle \\
= & - \gamma \| \mathcal{F}(x^k) - \mathcal{L}^k(x^k) - \mathcal{F}(x^*) + \mathcal{L}^k(x^*) \|^2_{X'} \\
& - \langle \mathcal{B}(w^k - x^{k+1}), w^k - x^{k+1} \rangle \\
& + \langle \mathcal{F}(x^k) - \mathcal{L}^k(x^k) - \mathcal{F}(x^*) + \mathcal{L}^k(x^*), x^k - x^{k+1} \rangle \\
& + \langle \mathcal{F}(x^k) - \mathcal{L}^k(x^k) - \mathcal{F}(x^*) + \mathcal{L}^k(x^*), w^k - x^* \rangle \\
& + \langle \mathcal{L}^k(x^*) - \mathcal{L}^k(x^{k+1}), x^{k+1} - w^k \rangle \\
& + \langle \mathcal{F}(x^*) + q^*(x^*), w^k - x^* \rangle + \langle \mathcal{Q}^k(w^k) - q^*(x^*), x^* - x^{k+1} \rangle \\
& + \langle \mathcal{Q}^k(w^k) - q^*(x^*), w^k - x^* \rangle . \qquad \text{(A.6)}
\end{aligned}$$

Due to the monotonicity of $\mathcal{L}^k$,

$$\langle \mathcal{L}^k(x^*) - \mathcal{L}^k(x^{k+1}), x^{k+1} - w^k \rangle \leqslant \langle \mathcal{L}^k(x^{k+1}\rangle - \mathcal{L}^k(x^*), w^k - x^* \rangle$$

is valid, and using the inequalities (A.1), (A.2) together with (2-i) in order to estimate the right-hand side of (A.6) as well as the term $\langle \mathcal{L}^k(x^{k+1}) - \mathcal{L}^k(x^*), w^k - x^* \rangle$, we obtain

$$\begin{aligned}
\chi_k s_2 \leqslant & \frac{\mu + \eta - 2\gamma}{2} \| \mathcal{F}(x^k) - \mathcal{L}^k(x^k) - \mathcal{F}(x^*) + \mathcal{L}^k(x^*) \|^2_{X'} \\
& + \frac{1}{2\mu} \|x^k - x^{k+1}\|^2 + \frac{1}{2\eta} \|w^k - x^*\|^2 \\
& - \frac{1}{1+\epsilon} \langle \mathcal{B}(x^{k+1} - x^*), x^{k+1} - x^* \rangle + \frac{1}{\epsilon} \langle \mathcal{B}(w^k - x^*), w^k - x^* \rangle \\
& + \| \mathcal{F}(x^*) + q^*(x^*) \|_{X'} \|w^k - x^*\| + \frac{l_{\mathcal{L}}}{\theta_k} \|w^k - x^*\|^2 \\
& + l_{\mathcal{L}} \theta_k \|x^{k+1} - x^*\|^2 + \left(\frac{1}{2 l_{\mathcal{L}} \theta_k} + \frac{\theta_k}{2 l_{\mathcal{L}}} \right) \| \mathcal{Q}^k(w^k) - q^*(x^*) \|^2_{X'} \qquad \text{(A.7)}
\end{aligned}$$

with arbitrary positive μ, η, θ_k and ϵ.

Choosing

$$\epsilon = \frac{1}{2\tilde{\chi}\bar{\chi}} - 1, \qquad \mu \in \left(\frac{1}{2\underline{\chi} m}, 2\gamma \right), \qquad \tau = m - \frac{1}{2\underline{\chi}\mu}, \qquad \text{(A.8)}$$

one can conclude from (2-v), (2-vi), (3.7) and (A.2) that

$$\left(\tilde{\chi} - \frac{1}{\chi_k}\frac{1}{1+\epsilon}\right)\langle \mathscr{B}(x^{k+1}-x^*), x^{k+1}-x^*\rangle - \tilde{\chi}\langle \mathscr{B}(x^k - x^*), x^k - x^*\rangle$$
$$- \langle \mathscr{G}(x^{k+1}) - \mathscr{G}(x^k), x^{k+1}-x^k\rangle + \left(\frac{1}{2\chi_k \mu} + \frac{\tau}{2}\right)\|x^{k+1}-x^k\|^2$$
$$\leqslant -\tilde{\chi}[\langle \mathscr{B}(x^{k+1}-x^*), x^{k+1}-x^*\rangle + \langle \mathscr{B}(x^k - x^*), x^k - x^*\rangle]$$
$$- \langle \mathscr{G}(x^{k+1}) - \mathscr{G}(x^k), x^{k+1}-x^k\rangle + \left(\frac{1}{2\chi_k \mu} + \frac{m}{2} - \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1}-x^k\|^2$$
$$\leqslant -\frac{\tilde{\chi}}{2}\langle \mathscr{B}(x^{K+1}-x^k), x^{k+1}-x^k\rangle - \langle \mathscr{G}(x^{k+1}) - \mathscr{G}(x^k), x^{k+1}-x^k\rangle$$
$$+\left(\frac{m}{2} + \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1}-x^k\|^2 \leqslant \left(-\frac{m}{2} + \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1}-x^k\|^2, \tag{A.9}$$

and that $-m/2 + 1/4\underline{\chi}\mu < 0$. Now, we sum up the estimates for s_1, s_2, s_3 in (A.4), (A.5), (A.7) with (A.3), and insert (A.8), (A.9) in this sum. This yields

$$\Gamma^{k+1}(x^*, x^{k+1}) - \Gamma^k(x^*, x^k) \leqslant \left(-\frac{m}{2} + \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1}-x^k\|^2$$
$$+ \frac{1}{\chi_k}\frac{\mu+\eta-2\gamma}{2}\|\mathscr{F}(x^k) - \mathscr{L}^k(x^k) - \mathscr{F}(x^*) + \mathscr{L}^k(x^*)\|^2_{X'}$$
$$+ \frac{1}{\chi_k}\Bigg[\left(\frac{1}{2\eta} + \frac{l_{\mathscr{L}}}{\theta_k} + \frac{l_h^2}{2\tau}\right)\|x^* - w^k\|^2 + l_{\mathscr{L}}\theta_k\|x^{k+1}-x^*\|^2$$
$$+ \left(\frac{1}{2l_{\mathscr{L}}\theta_k} + \frac{\theta_k}{2l_{\mathscr{L}}}\right)\|\mathscr{Q}^k(w^k) - q^*(x^*)\|^2_{X'}$$
$$+ \frac{1}{\epsilon}\langle \mathscr{B}(w^k - x^*), w^k - x^*\rangle$$
$$+ \|\mathscr{F}(x^*) + q^*(x^*)\|_{X'}\|w^k - x^*\|\Bigg]. \tag{A.10}$$

But, from the first inequality in (3.6) one gets

$$\|x^{k+1}-x^*\|^2 \leqslant \frac{1}{m}\Gamma^{k+1}(x^*, x^{k+1})$$

and (A.10) leads to

$$\left(1 - \frac{l_{\mathscr{L}}\theta_k}{\underline{\chi}m}\right)\Gamma^{k+1}(x^*, x^{k+1})$$
$$\leqslant \Gamma^k(x^*, x^k) + \frac{1}{\underline{\chi}}\Bigg[\left(\frac{1}{2\eta} + \frac{l_{\mathscr{L}}}{\theta_k} + \frac{l_h^2}{2\tau}\right)\|w^k - x^*\|^2$$
$$+ \left(\frac{1}{2l_{\mathscr{L}}\theta_k} + \frac{\theta_k}{2l_{\mathscr{L}}}\right)\|\mathscr{Q}^k(w^k) - q^*(x^*)\|^2_{X'}$$

$$+\frac{1}{\epsilon}\langle \mathscr{B}(w^k - x^*), w^k - x^*\rangle + \|\mathscr{F}(x^*) + q^*(x^*)\|_{X'}\|w^k - x^*\| \Big]$$
$$+\left(-\frac{m}{2} + \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1} - x^k\|^2 . \tag{A.11}$$

Setting $\theta_k = \theta k^{-\alpha}$ for $k > 0$, where $\theta \in (0, \underline{\chi} m l_{\mathscr{L}}^{-1})$ and $\alpha > 1$ is the same as in (2-ix), we obtain with $c = l_{\mathscr{L}}\theta/\underline{\chi}m$:

$$\Gamma^{k+1}(x^*, x^{k+1}) \leqslant \frac{1}{1 - ck^{-\alpha}}\Gamma^k(x^*, x^k)$$
$$+\frac{1}{\underline{\chi}(1 - c_0)}\Big[\left(\frac{1}{2\eta} + \frac{l_{\mathscr{L}}}{\theta}k^\alpha + \frac{l_h^2}{2\tau}\right)\|w^k - x^*\|^2$$
$$+\left(\frac{k^\alpha}{2l_{\mathscr{L}}\theta} + \frac{\theta}{2l_{\mathscr{L}}}\right)\|\mathscr{Q}^k(w^k) - q^*(x^*)\|^2_{X'}$$
$$+\frac{1}{\epsilon}\langle \mathscr{B}(w^k - x^*), w^k - x^*\rangle + \|\mathscr{F}(x^*) + q^*(x^*)\|_{X'}\|w^k - x^*\| \Big]$$
$$+\left(-\frac{m}{2} + \frac{1}{4\underline{\chi}\mu}\right)\|x^{k+1} - x^k\|^2, \quad k = 1, 2, \ldots \tag{A.12}$$

Now, in view of

$$\frac{1}{1 - c_0 k^{-\alpha}} < 1 + \frac{c_0}{k^\alpha - c_0}, \quad \sum_{k=1}^{\infty}\frac{1}{k^\alpha - c_0} < \infty, \quad -\frac{m}{2} + \frac{1}{4\underline{\chi}\mu} < 0$$

and Assumption (2-ix), the convergence of the sequence $\{\Gamma^k(x^*, x^k)\}$ follows from Lemma 2.2.2 in [20], and the first inequality in (3.6) ensures the boundedness of the sequence $\{x^k\}$. Passing to the limit in (A.12), one can immediately conclude that $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$. □

References

1. Cohen, G. (1980), Auxiliary problem principle and decomposition of optimization problems. *JOTA* 32: 277–305.
2. Cohen, G. (1988), Auxiliary problem principle extended to variational inequalities. *JOTA* 59: 325–333.
3. Duvaut, G. and Lions, J.-L. (1972), *Les Inéquations en Mécanique et en Physique*. Dunod, Paris.
4. Eckstein, J. (1993), Nonlinear proximal point algorithms using Bregman functions, with application to convex programming. *Math. Oper. Res.* 18: 202–226.
5. El-Faroug, N. and Cohen, G. (1998), Progressive regularization of variational inequalities and decomposition algorithms. *JOTA* 97: 407–433.
6. Gabay, D. (1983) Applications of the method of multipliers to variational inequalities. In M. Fortin and G. Glowinski (eds.): *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. North-Holland, Amsterdam.

7. Glowinski, R. (1984), *Numerical Methods for Nonlinear Variational Problems*. Springer, New York–Berlin–Heidelberg–Tokyo.
8. Kaplan, A. and Tichatschke, R, Proximal point approach and approximation of variational inequalities. *SIAM Journal on Control and Optimization*. (to appear 2000).
9. Kaplan, A. and Tichatschke, R. (1994), *Stable Methods for Ill-Posed Variational Problems – Prox-Regularization of Elliptical Variational Inequalities and Semi-Infinite Optimization Problems*. Akademie Verlag Berlin.
10. Kaplan, A. and Tichatschke, R. (1997) Prox-regularization and solution of ill-posed elliptic variational inequalities. *Applications of Mathematics* 42: 111–145.
11. Kaplan, A. and Tichatschke, R. (1998a) Proximal penalty method for ill-posed parabolic optimal control problems. In: W. Desch, F. Kappel and K. Kunisch (eds.): *Control and Estimation of Distributed Parameter Systems*, Vol. 126, pp. 169–182.
12. Kaplan, A. and Tichatschke, R. (1998b), Proximal point methods in nonconvex optimization. *J. of Global Optim*. 4: 389–406.
13. Lions, P.-L. and Mercier, B. (1979), Splitting algorithm for the sum of two nonlinear operators. *SIAM J. Numer. Anal*. 16: 964–979.
14. Makler-Scheimberg, S., Nguyen, V. and Strodiot, J. (1996), Family of perturbation methods for variational inequalities. *JOTA* 89: 423–452.
15. Mosco, U. (1969), Convergence of convex sets and of solutions of variational inequalities. *Advances in Mathematics* 3: 510–585.
16. Panagiotopoulos, P.D. (1975), A nonlinear programming approach to the unilateral contact- and friction-boundary value problem in the theory of elasticity. *Ing. Archiv*. 44: 421–432.
17. Panagiotopoulos, P.D. (1985), *Inequality Problems in Mechanics and Applications. Convex and Nonconvex Energy Functionals*. Birkhäuser-Verlag, Boston–Basel–Stuttgart.
18. Pang, J.S. and Chan, D. (1982), Iterative methods for variational and complementarity problems. *Math. Programming* 24: 284–313.
19. Passty, G.B. (1979), Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl*. 72: 383–390.
20. Polyak, B.T. (1987), *Introduction to Optimization*. Optimization Software, Inc. Publ. Division, New York.
21. Renaud, A. and Cohen, G. (1997), An extension of the auxiliary problem principle to nonsymmetric auxiliary operators. *ESAIM: Control. Optimization and Calculus of Variations* 2: 281–306.
22. Rockafellar, R.T. (1970), On the maximality of sums of nonlinear monotone operators. *Trans. Am. Math. Soc*. 149: 75–88.
23. Rockafellar, R.T. (1976), Monotone operators and the proximal point algorithm. *SIAM J. Control and Optim*. 14: 877–898.
24. Salmon, G., Nguyen, V.H. and Strodiot, J.J., Coupling the auxiliary problem principle and the epiconvergence theory for solving general variational inequalities. Vol. Lecture Notes in Economics and Mathem. Systems.
25. Spingarn, J.E. (1981–1982), Submonotone mappings and the proximal point algorithm. *Numer. Funct. Anal. and Optimiz*. 4: 123–150.
26. Tseng, P. (1990), Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming. *Math. Program*. 48: 249–263.
27. Tseng, P. (1991), Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control Optim*. 29: 119–138.
28. Zhou, D. and Marcotte, P. (1996), Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *SIAM J. Optimization* 6: 714–726.