Discussiones Mathematicae
Differential Inclusions, Control and Optimization 27 (2007) 51–69

NUMERICAL CONSIDERATIONS OF A HYBRID PROXIMAL PROJECTION ALGORITHM FOR SOLVING VARIATIONAL INEQUALITIES

Christina Jager

Department of Mathematics, University of Trier
54286 Trier, Germany

e-mail: christina.jager@uni-trier.de

Abstract

In this paper, some ideas for the numerical realization of the hybrid proximal projection algorithm from Solodov and Svaiter [22] are presented. An example is given which shows that this hybrid algorithm does not generate a Fejér-monotone sequence. Further, a strategy is suggested for the computation of inexact solutions of the auxiliary problems with a certain tolerance. For that purpose, ε-subdifferentials of the auxiliary functions and the bundle trust region method from Schramm and Zowe [20] are used. Finally, some numerical results for non-smooth convex optimization problems are given which compare the hybrid algorithm to the inexact proximal point method from Rockafellar [17].

Keywords: variational inequality, proximal point algorithm, bundle method.

2000 Mathematics Subject Classification: 47J20, 90C25, 65K10, 65J20.

1. Introduction

Let $(\mathcal{H}, \|\cdot\|)$ be a real Hilbert space with the topological dual $\mathcal{H}$' and the duality pairing $\langle\cdot,\cdot\rangle$ between $\mathcal{H}$ and $\mathcal{H}'$. The following variational inequality problem (VI) is considered:

$$\text{(VI)} \qquad \begin{array}{c} \text{Find } x^* \in X \text{ and } v^* \in T(x^*) \text{ with} \\ \langle v^*, x - x^*\rangle \geq 0 \quad \forall x \in X \end{array}$$

assuming that $X \subseteq \mathcal{H}$ is a closed, convex set and $T\colon \mathcal{H} \to 2^{\mathcal{H}'}$ is a set-valued maximal monotone operator. We suppose that $\mathrm{dom}T \cap \mathrm{int}X \neq \emptyset$, where $\mathrm{dom}T := \{x \in \mathcal{H}\colon T(x) \neq \emptyset\}$ and $\mathrm{int}X$ denotes the interior of the set X. Let $\mathcal{N}_X$ denote the normality operator to X. Then the operator

$$\hat{T} := T + \mathcal{N}_X$$

is maximal monotone (see [18], Theorem 1), and problem (VI) is equivalent to the following inclusion problem:

$$\text{(IP)} \qquad \text{Find } x^* \in \mathcal{H} \text{ with } 0 \in \hat{T}(x^*).$$

The classical proximal point algorithm (PPA) for solving maximal monotone inclusions goes back to Martinet [15] and was further developed by Rockafellar [17], who considered the inexact version. Starting with an arbitrary $x^0 \in \mathcal{H}$ the inexact PPA solves the following auxiliary problem in each iteration k:

$$\text{(1)} \qquad \begin{gathered} \text{Find } x^{k+1} \in \mathcal{H} \text{ and } v^{k+1} \in \hat{T}(x^{k+1}) \text{ with} \\ e^k = v^{k+1} + \mu_k(x^{k+1} - x^k), \end{gathered}$$

where x^k is the current iterate, e^k is an error vector, and $\mu_k > 0$ is a regularization parameter. Rockafellar proved a weak global convergence of the sequence $\{x^k\}$ towards a solution of (IP) under the conditions that the solution set is not empty, the sequence $\{\mu_k\}$ of the positive regularization parameters is bounded from above, and each error vector e^k satisfies the condition

$$\text{(2)} \qquad \| e^k \| \leq \sigma_k \mu_k, \ \forall k, \ \text{ and } \ \sum_{k=0}^{\infty} \sigma_k < \infty.$$

A survey of some recent developments concerning the proximal point method can be found, e.g., in Kaplan and Tichatschke [9]. Decomposition methods based on proximal methods are investigated by Chen and Teboulle [5]. A number of recent papers are devoted to generalized proximal methods with non-quadratic proximal regularization (see, e.g., Burachik and Iussem [2], Auslender, Teboulle and Ben-Tiba [1], Kaplan and Tichatschke [10, 11]). Solodov and Svaiter [22] introduced a hybrid proximal point algorithm (HPPA) in order to weaken the error tolerance criterion of the PPA, which converges strongly under mild assumptions.

The paper is organized as follows: Chapter 2 briefly introduces the hybrid proximal projection algorithm from Solodov and Svaiter [22] and summarizes the theoretical results. In Chapter 3, we present an example which shows that the HPPA cannot preserve the monotonicity properties which are valid for the PPA. The implementation of the HPPA and the inexact PPA is described in Chapter 4. Some numerical examples are summarized in Chapter 5 and concluding remarks are given in Chapter 6.

2. The hybrid proximal projection algorithm

In this section, the ideas of Solodov and Svaiter will be summarized briefly. We consider the general problem:

$$\text{(P)} \qquad \text{Find } x^* \in \mathcal{H} \text{ with } 0 \in T(x^*),$$

where T is a set-valued maximal monotone operator on a Hilbert space $\mathcal{H}$. In [22] a new definition of inexact solutions of the regularized auxiliary problems of (P) is used to get a weakened error tolerance criterion for the corresponding algorithm. A strong convergence of the iterates is achieved by adding a projection step onto the intersection of two halfspaces which contain the solution set. The overall method is described in Algorithm 1.

Algorithm 1 (HPPA)**.** *Choose an arbitrary $x^0 \in \mathcal{H}$ and $\sigma \in [0,1)$. Given x^k in iteration k, choose $\mu_k > 0$ and solve the auxiliary problem*

$$\text{(3)} \qquad \begin{aligned} &\textit{find } y^k \in \mathcal{H} \quad \textit{and} \quad v^k \in T(y^k) \quad \textit{such that} \\ &e^k = v^k + \mu_k(y^k - x^k), \end{aligned}$$

where the error vector e^k satisfies

$$\text{(4)} \qquad \| e^k \| \leq \sigma \max\{\| v^k \|,\ \mu_k \| y^k - x^k \|\}.$$

Stop if $v^k = 0$ or $y^k = x^k$. Otherwise define two halfspaces

$$\begin{aligned} H_k &= \{z \in \mathcal{H} \colon \langle z - y^k, v^k \rangle \leq 0\} \textit{ and} \\ W_k &= \{z \in \mathcal{H} \colon \langle z - x^k, x^0 - x^k \rangle \leq 0\} \end{aligned}$$

and perform the projection

$$x^{k+1} = P_{H_k \cap W_k}(x^0).$$

The pair (y^k, v^k) is called an inexact solution to the problem $0 \in T(y) + \mu_k(y - x^k)$ with tolerance σ.

As refers to the convergence analysis of the HPPA, the main result is given by the following theorem:

Theorem 2 ([22], Corollary 2, Theorem 1)**.** *Suppose that the solution set* S *of problem* (P) *is not empty. Then the* HPPA *generates infinite sequences* $\{x^k\}, \{y^k\}$, *and* $\{v^k\}$ *such that* $S \subseteq H_k \cap W_k$ *for all* k. *Suppose further that the sequence* $\{\mu_k\}$ *of positive regularization parameters is bounded from above. Then* $\{x^k\}$ *converges strongly to* $x^* = P_S(x^0)$.

3. Monotonicity properties

We consider problem (P) with $T = \partial f$, where $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, lower semi-continuous and convex function and ∂f denotes the subdifferential operator of f. Then we can prove the following monotonicity properties.

Lemma 3. *Let* $f\colon \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *be a proper, lower semi-continuous and convex function and* $\{x^k\}$ *be generated by the exact PPA (i.e.,* (1) *with* $e^k = 0$ *for all* k*) for problem* (P) *with* $T = \partial f$. *Let* S *denote the solution set of problem* (P)*. Then*

1. $f(x^k) \leq f(x^{k-1}) \quad \forall k \in \mathbb{N}$,
2. $\| x^k - z \| < \| x^{k-1} - z \| \quad \forall z \in S, x^{k-1} \notin S$, *i.e., the sequence* $\{x^k\}$ *is strictly Fejér-monotone.*

Proof.
Inequality 1. Let $\{\mu_k\}$ be the sequence of regularization parameters used in the exact PPA. For all k we have that $\mu_{k-1}(x^{k-1} - x^k) \in \partial f(x^k)$. Since f is convex we get

$$f(x^{k-1}) - f(x^k) \geq \langle \mu_{k-1}(x^{k-1} - x^k), x^{k-1} - x^k \rangle \geq 0.$$

Inequality 2 follows as a special case of Kaplan and Tichatschke [7], Proposition 8.3. ■

The function values of the iterates generated by the inexact PPA are also monotonically decreasing (see, for instance, [7], Theorem 13.9).

We will now present an example which shows that the above monotonicity properties are in general not valid for a sequence $\{x^k\}$ generated by the exact version of the HPPA. Hence, in the case of an inexact version the same happens.

Example 1. We consider problem (P) with $T = \partial f$ and function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x) = \max\{f_1(x), f_2(x)\}, \quad x = (x_1, x_2) \in \mathbb{R}^2$$

with $f_1(x) = x_1^2 + x_2^2 - \frac{16}{25}$ and $f_2(x) = \frac{9}{25}x_1^2$. The solution set is

$$S = \left\{ x \in \mathbb{R}^2\colon x_1 = 0, x_2 \in \left[-\frac{4}{5}, \frac{4}{5} \right] \right\}$$

and the optimal value is $f^* = 0$.

We choose $x^0 = (128, 1)$ and present the first eight iterates of the exact HPPA in Table 1.

k	x^k	$f(x^k)$	k	x^k	$f(x^k)$
1	$(64, 2^{-1})$	4095.61	5	$(4, 2^{-5})$	15.36
2	$(32, 2^{-2})$	1023.42	6	$(2, 2^{-6})$	3.36
3	$(16, 2^{-3})$	255.37	7	$(1, 2^{-7})$	0.36
4	$(8, 2^{-4})$	63.36	8	$(\frac{25}{34}, 1)$	0.90

Table 1. First iterates of the exact HPPA show violation of monotonicity properties.

Looking at iterates x^7 and x^8, it can be seen that the function values are not monotonically decreasing. Furthermore, if we take $z = (0, 2^{-7}) \in S$, we get

$$\| x^7 - z \| = 1 < 1.235 \approx \| x^8 - z \|$$

showing that the Fejér-monotonicity is violated.

4. Implementation of the HPPA

We now discuss the numerical realization of the HPPA. The calculation of the projection $P_{H_k \cap W_k}(x^0)$ is easy: If the point

$$\bar{z} := P_{H_k}(x^0) = x^0 - \frac{\langle v^k, x^0 - v^k \rangle}{\| v^k \|^2} v^k$$

is an element of W_k, then $P_{H_k \cap W_k}(x^0) = \bar{z}$. Otherwise, we have

$$P_{H_k \cap W_k}(x^0) = x^0 + \lambda_1^* v^k + \lambda_2^*(x^0 - x^k)$$

where λ_1^*, λ_2^* are the solutions of the following linear system:

$$\begin{aligned} &\lambda_1 \| v^k \|^2 + \lambda_2 \langle v^k, x^0 - x^k \rangle = -\langle x^0 - y^k, v^k \rangle, \\ &\lambda_1 \langle v^k, x^0 - x^k \rangle + \lambda_2 \| x^0 - x^k \|^2 = - \| x^0 - x^k \|^2 . \end{aligned}$$

For more details see [22] or [6].

The crucial step is to find a practicable rule for the determination of an inexact solution (y^k, v^k) with tolerance σ according to (3)–(4). The requirements in (3)–(4) offer some degrees of freedom. Since the operator T is set-valued, there is in general more than one possible choice for each $v^k \in T(y^k)$. Furthermore, from a numerical point of view, there is no obvious rule for calculating y^k and a suitable v^k, such that the error tolerance criterion (4) is satisfied. Of course, it is easy to check requirement (4) if a pair (y^k, v^k) is given. But it is not clear how we should proceed if the candidate (y^k, v^k) cannot be chosen as an inexact solution with tolerance σ. It can easily be seen that the error tolerance criterion would always be fulfilled if we chose $\sigma = 2$, but for the convergence of the method it is important to have $\sigma \in [0, 1)$.

In our implementation we follow the idea to use the error tolerance condition

$$\| e^k \| \leq \sigma \max\{\| v^k \|, \mu_k \| y^k - x^k \|\}$$

as a stopping criterion for the determination of the vector y^k in the auxiliary problem (3). That is, we determine y^k as an approximate solution to the auxiliary problem $0 \in T(y) + \mu_k(y - x^k)$ with a given accuracy ε.

Then we choose an element v^k in $T(y^k)$ and check, whether the error tolerance criterion holds. If so, we accept (y^k, v^k) as an inexact solution with tolerance σ, otherwise, we increase the accuracy ε and determine a new solution y^k. Note that for the determination of the element v^k there should also be an adequate strategy.

4.1. HPPA for unrestricted non-smooth convex Optimization Problems

To get things more concrete, let us concentrate on a special problem class: The unrestricted minimization of a max-function, i.e., we consider non-smooth convex optimization problems with objective functions $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ of the form

$$f(x) = \max\{f_i(x)\colon 1 \le i \le m\}$$

where the functions $f_i\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ $(1 \le i \le m)$ are convex and continuously differentiable. In this situation, the subdifferential of f is easy to calculate in each $x \in \mathbb{R}^n$:

$$\partial f(x) = \operatorname{conv}\{\nabla f_i(x)\colon i \in I(x)\},$$

where $I(x) = \{i \in \{1, \dots, m\}\colon f_i(x) = f(x)\}$. The corresponding regularized auxiliary problems of the HPPA are according to (3):

$$(P_k) \qquad \begin{aligned}&\text{Find } y^k \in \mathbb{R}^n, v^k \in \partial f(y^k) \text{ and } e^k \in \mathbb{R}^n \text{ such that}\\ &e^k = v^k + \mu_k(y^k - x^k),\end{aligned}$$

where the error tolerance criterion is the same as in (4). Obviously, problem (P_k) is equivalent to:

$$\begin{aligned}&\text{Find } y^k \in \mathbb{R}^n \text{ and } e^k \in \partial f_k(y^k), \text{ where}\\ &f_k(y) := f(y) + \frac{\mu_k}{2}\,\|y - x^k\|^2 .\end{aligned}$$

To find an approximate solution y^k of the non-smooth auxiliary problem (P_k), we implemented the bundle trust region method (BT-method) from Schramm and Zowe [20] with some modifications concerning the adaptation of the trust region parameter described by Kiwiel [12].

The BT-method performs a piecewise linear approximation of the objective function f_k with the help of a bundle $(f_k(z^i), g^i)_{i\in J_s}$, where the z^i are test-points from earlier iterations, $g^i \in \partial f_k(z^i)$, and J_s is the set of the current bundle indices. With s we denote the iteration index within the BT-method. The information in the bundle is weighted with the help of the linearization errors

$$\alpha_i^s := f_k(x^s) - [f_k(z^i) + \langle g^i, x^s - z^i\rangle],$$

where x^s is the current iterate in the BT-method. In each iteration s of the BT-method a search direction d^s is defined as a convex combination of the subgradients in the bundle:

$$d^s = -t_s \sum_{i\in J_s} \lambda_i^s g^i$$

where t_s is the trust region parameter and $\lambda^s = (\lambda_i^s)_{i\in J_s}$ is the solution to the following quadratic problem

$$\begin{aligned} \min_{\lambda_i} \quad & \frac{1}{2}\Big\| \sum_{i\in J_s} \lambda_i g^i \Big\|^2 + \frac{1}{t_s} \sum_{i\in J_s} \lambda_i \alpha_i^s \\ s.t. \quad & \sum_{i\in J_s} \lambda_i = 1, \quad \lambda_i \geq 0, \quad i \in J_s. \end{aligned}$$

In our implementation this quadratic problem is solved by the NAG-routine `nag_opt_qp` (e04nfc) of the NAG-library [16]. To determine the next iterate x^{s+1} and to improve the subgradient information, the following two cases are considered in the BT-method:

1. If $f_k(x^s + d^s)$ is sufficiently smaller than $f_k(x^s)$, then either
   (a) increase t_s and determine a new direction d^s, or
   (b) make a serious step: $x^{s+1} := z^{s+1} := x^s + d^s$, compute $g^{s+1} \in \partial f_k(z^{s+1})$, set $J_{s+1} := J_s \cup \{s+1\}$.
2. If $f_k(x^s + d^s)$ is not sufficiently smaller than $f_k(x^s)$, then either
   (a) decrease t_s and determine a new direction d^s, or
   (b) make a null step: $x^{s+1} = x^s$, compute $g^{s+1} \in \partial f_k(z^{s+1})$, where $z^{s+1} = x^s + d^s$, set $J_{s+1} := J_s \cup \{s+1\}$.

The set J_s is cleaned up from time to time to keep it at a reasonable size. The iterative process in the BT-method is stopped in iteration $\hat{s}$ if

$$\Big\| \sum_{i \in J_{\hat{s}}} \lambda_i^{\hat{s}} g^i \Big\| \le \varepsilon \quad \text{and} \quad \sum_{i \in J_{\hat{s}}} \lambda_i^{\hat{s}} \alpha_i^{\hat{s}} \le \varepsilon \,, \tag{5}$$

where $\varepsilon \ge 0$ is a given accuracy parameter. Then $y^k := x^{\hat{s}}$ is an approximate minimum of the auxiliary function f_k (see [20], Lemma 2.2).

Now we are ready to present our ideas to determine an inexact solution (y^k, v^k) with tolerance σ. Let y^k be calculated by the BT-method with accuracy ε. We want to discuss two strategies to determine $v^k \in \partial f(y^k)$.

4.1.1. The use of $v^k = \arg\min\{\| v \| \colon v \in \partial f(y^k)\}$

The first idea is to take v^k as the minimal-norm element of the subdifferential:

$$v^k = \arg \min_{v \in \partial f(y^k)} \| v \| \,.$$

For our special problem class this is realized by calculating the solution $\lambda^* = (\lambda_i^*)_{i \in I(y^k)}$ to the following quadratic problem

$$\begin{aligned} \min_{\lambda_i} \quad & \frac{1}{2} \Big\| \sum_{i \in I(y^k)} \lambda_i \nabla f_i(y^k) \Big\|^2 \\ s.t. \quad & \sum_{i \in I(y^k)} \lambda_i = 1, \quad \lambda_i \ge 0, \quad i \in I(y^k), \end{aligned}$$

with $I(y^k) = \{i \in \{1, \dots, m\} \colon f_i(y^k) = f(y^k)\}$. This λ^* defines

$$v^k = \sum_{i \in I(y^k)} \lambda_i^* \nabla f_i(y^k).$$

Our numerical experiments show that a pair (y^k, v^k) calculated in the above described way often does not fulfill requirement (4). As a consequence, a new solution y^k is determined with the BT-method under an increased accuracy ε. In the test examples this process is often repeated several times and in most of the cases the determination of an inexact solution with tolerance σ has to be stopped, since the accuracy ε cannot be realized any more. So this strategy does not work very well in practice.

4.1.2. The use of $e^k \in \partial_\varepsilon f(y^k)$

The second idea is more promising. Here we use the BT-method to determine both y^k and v^k. If the BT-method stops at iteration $\hat{s}$ we define

$$e^k := \sum_{i \in J_{\hat{s}}} \lambda_i^{\hat{s}} g^i, \tag{6}$$

$$v^k := e^k - \mu_k (y^k - x^k). \tag{7}$$

Then we get the following relation:

Lemma 4. *Let f_k be the objective function of the auxiliary problem (P_k), $\varepsilon \geq 0$ the chosen accuracy for the BT-method, $y^k := x^{\hat{s}}$ the approximate solution with accuracy ε determined by the BT-method using stopping criterion* (5)*, $\hat{s}$ the iteration index of the last iteration of the BT-method and $J_{\hat{s}}$, $\lambda_i^{\hat{s}}$, g^i, $i \in J_{\hat{s}}$, the associated information in the last BT-iteration. Then*

$$f_k(y) \geq f_k(y^k) + \left\langle \sum_{i \in J_{\hat{s}}} \lambda_i^{\hat{s}} g^i, y - y^k \right\rangle - \varepsilon \quad \forall y \in \mathbb{R}^n, \tag{8}$$

i.e., $\sum_{i \in J_{\hat{s}}} \lambda_i^{\hat{s}} g^i \in \partial_\varepsilon f_k(y^k)$.

Proof. For $g^i \in \partial f_k(z^i)$, $i \in J_{\hat{s}}$, we have the subgradient inequality

$$\langle g^i, y - z^i \rangle \leq f_k(y) - f_k(z^i) \quad \forall y \in \mathbb{R}^n.$$

Subtracting $\langle g^i, y^k - z^i \rangle = f_k(y^k) - f_k(z^i) - \alpha_i^{\hat{s}}$ one gets

$$\langle g^i, y - y^k \rangle \leq f_k(y) - f_k(y^k) + \alpha_i^{\hat{s}} \quad \forall y \in \mathbb{R}^n, i \in J_{\hat{s}},$$

multiplying this inequality with λ_i and summing up over i, we have

$$\left\langle \sum_{i \in J_{\hat{s}}} \lambda_i g^i, y - y^k \right\rangle \leq f_k(y) - f_k(y^k) + \sum_{i \in J_{\hat{s}}} \lambda_i \alpha_i^{\hat{s}}, \quad \forall y \in \mathbb{R}^n,$$

remembering that $\sum_{i \in J_{\hat{s}}} \lambda_i = 1$. Rearranging the terms and using (5) we get the desired inequality. ∎

In other words, determining y^k and v^k with the BT-method in the suggested way makes the error vector e^k be an element of the ε-subdifferential of the function f_k. Although, for $\varepsilon \neq 0$, this result differs from the original requirement $e^k \in \partial f_k(y^k)$ in the auxiliary problem (P_k), we get good numerical results. In most of our examples we can accept the first calculated pair (y^k, v^k) in iteration k of the HPPA as an inexact solution with tolerance σ. As a consequence, increasing the accuracy ε is not necessary. Therefore, we believe that the idea (6)–(7) is a promising strategy for the numerical realization of the HPPA. In order to achieve conformity in the limit, one can decrease ε by a constant factor $\theta_1 \in (0,1)$ in each iteration: $\varepsilon_k = \theta_1 \cdot \varepsilon_{k-1}$.

With reference to the stopping criterion for the HPPA, we suggest to choose $\delta > 0$ as a given tolerance and terminate the algorithm if

$$\| y^k - x^k \| < \delta \text{ or } \| v^k \| < \delta.$$

4.2. Implementation of the inexact PPA

In order to get a glimpse of the *numerical* performance of the HPPA, we compare it with the inexact PPA (1). This seems to be natural because for problems in a finite dimensional space weak and strong convergence coincide. Therefore, we briefly describe our implementation of the classical inexact PPA (1) for the same class of unrestricted non-smooth convex optimization problems. As in the HPPA, in each iteration k, a solution x^{k+1} of the auxiliary problem

$$\min \left\{ f(x) + \frac{\mu_k}{2} \| x - x^k \|^2 : x \in \mathbb{R}^n \right\}$$

is calculated by the BT-method with accuracy ε_k. The required error tolerance criterion (2) is implied by the condition

$$\text{dist}(0, \partial f_k(x^{k+1})) \leq \sigma_k \mu_k, \quad \sum_{k=0}^{\infty} \sigma_k < \infty \tag{9}$$

as explained in Rockafellar [17]. Using the BT-method to calculate the iterates, x^{k+1} realizes the above condition with ∂f_k replaced by $\partial_{\varepsilon_k} f_k$, as will be explained now. We update ε_k and μ_k according to

$$\varepsilon_k = \theta_1 \cdot \varepsilon_{k-1}, \quad \mu_k = \theta_2 \cdot \mu_{k-1},$$

where $\varepsilon_0 > 0$, $\mu_0 > 0$, $\theta_1 \in (0,1)$ and $\theta_2 \in (0,1]$. With (5) and (8) one gets

$$\operatorname{dist}(0, \partial_{\varepsilon_k} f_k(x^{k+1})) \leq \Big\| \sum_{i \in J_s} \lambda_i^s g^i \Big\| \leq \varepsilon_k, \text{ with}$$

$$\sum_{k=0}^{\infty} \sigma_k := \sum_{k=0}^{\infty} \frac{\varepsilon_k}{\mu_k} = \frac{\varepsilon_0}{\mu_0} \sum_{k=0}^{\infty} \left(\frac{\theta_1}{\theta_2}\right)^k < \infty \quad \Leftrightarrow \quad 0 < \theta_1 < \theta_2.$$

5. NUMERICAL EXAMPLES

In [6], we tested many self-constructed examples being of the type described in section 4.1 as well as examples from Shor [21] and Lemarechal [13].

The calculations were done with a C++-implementation of the HPPA and the inexact PPA. All experiments basically show that in the HPPA the monotonicity properties are violated and that the iteration numbers are higher than in the inexact PPA. Here, we are going to present three examples.

In the tables we give the following information: x^0 denotes the start iterate and $\| x^0 - x^* \|$ gives the distance of the start iterate to the solution. $\sharp$iter. means the number of outer iterations. One outer iteration in the HPPA mainly consists of two parts: the solution of the regularized auxiliary problem with the BT-method and the projection step. So in each outer iteration there is a number of inner iterations caused by the BT-method. In the inexact PPA we have the same situation, except that we leave out the projection step. Therefore, we also give information about the inner iterations: $\sharp$in.iter. stands for the total number of BT-iterations needed to solve all auxiliary problems and $\sharp f/g$ denotes the number of objective/subgradient evaluations in a total run of the algorithm. Column x^* contains the calculated optimal point of accuracy δ, $f(x^*)$ the corresponding optimal value.

The parameter settings for all examples are summarized in Table 2. We start with the two-dimensional Example 1 from Section 3 in order to demonstrate the geometrical behavior of the HPPA and the inexact PPA.

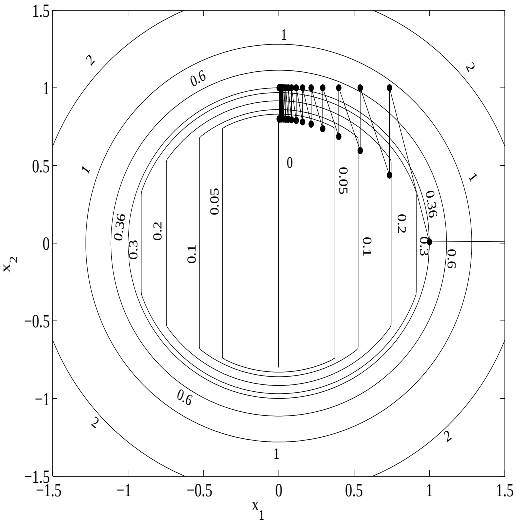
description	parameter	Ex1	Shor	Maxquad
accuracy of the HPPA/inexact PPA	δ	10^{-4}	10^{-4}	10^{-3}
accuracy of the BT-method	ε_0	10^{-4}	10^{-4}	10^{-4}
changing factor for accuracy ε_k	θ_1	0.89	0.89	0.89
regularization parameter	μ_0	2.0	30.0	1.0
changing factor for parameter μ_k	θ_2	0.9	0.9	0.9
tolerance parameter	σ	0.8	0.9	0.9

Table 2. Parameter settings for Examples 1–3.

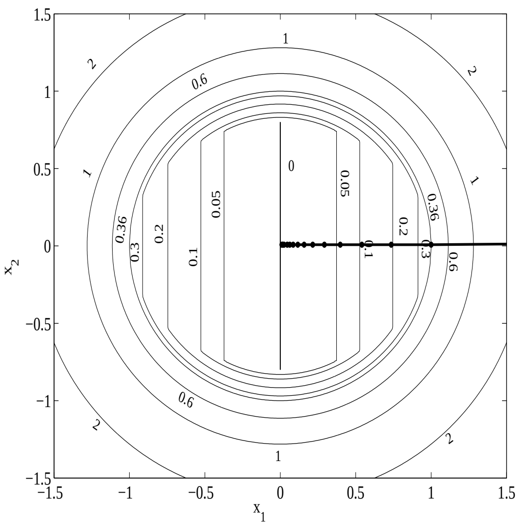
The results are given in Table 3 for different start iterates x^0. In each case, the algorithms converge to different optimal points. Figure 1 shows the geometrical behavior of the two algorithms. The level sets of function f are plotted and the corresponding levels are indicated on the lines. The points, connected by lines, show the iterates x^k ($k \geq 7$) of the HPPA and the inexact PPA, respectively, using the start iterate (128,1). The iterates generated by the HPPA "jump" up and down from a lower level set to a higher one and back again. The violation of the monotonicity properties becomes obvious. In contrast, the iterates of the inexact PPA get closer to the point (0,0) in each iteration and the monotonicity properties are fulfilled.

		HPPA				
x^0	$\|x^0 - x^*\|$	♯iter.	♯in.iter.	$\sharp f/g$	x^*	$f(x^*)$
$\binom{128}{1}$	128.0001	24	299	410	$\binom{0.0000}{0.7999}$	3.22e-10
$\binom{-150}{100}$	179.8349	82	30748	31498	$\binom{-0.0001}{0.7998}$	5.84e-09
$\binom{10}{-10000}$	9999.2	51	815	1141	$\binom{0.0000}{-0.8000}$	1.90e-09
		inexact PPA				
x^0	$\|x^0 - x^*\|$	♯iter.	♯in.iter.	$\sharp f/g$	x^*	$f(x^*)$
$\binom{128}{1}$	128.0001	18	192	225	$\binom{0.0000}{0.0086}$	2.63e-11
$\binom{-150}{100}$	179.8349	18	210	242	$\binom{-0.0000}{0.6489}$	2.54e-11
$\binom{10}{-10000}$	9999.2	11	162	185	$\binom{0.0001}{-0.7999}$	3.64e-09

Table 3. Numerical results for Example 1, solution set $S = \{x \in \mathbb{R}^2 \colon x_1 = 0, x_2 \in [-4/5, 4/5]\},\ f^* = 0.$



(a) HPPA



(b) inexact PPA

Figure 1. Geometrical behavior in Example 1: level sets and iterates.

Example 2. This example is extracted from Shor [21], p. 138. The function to be minimized is

$$\begin{aligned} f(x) &= \max\{f_i(x)\colon i = 1, \ldots, 10\}, \quad x \in \mathbb{R}^5, \text{ with} \\ f_i(x) &= b_i \sum_{j=1}^{5} (x_j - a_{ij})^2, \quad i = 1, \ldots, 10, \\ (b_i) &= (1, 5, 10, 2, 4, 3, 1.7, 2.5, 6, 3.5), \\ (a_{ij})^T &= \begin{pmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 2 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}. \end{aligned}$$

The optimal set consists of the unique element

$$x^* = (1.12434, 0.97945, 1.47770, 0.92023, 1.12429),$$

the optimal value is $f^* = 22.60016$.

The results of the HPPA and the inexact PPA for Example 2 are shown in Table 4. As starting points we use

$$\begin{aligned} x_I^0 &= (0, 0, 0, 0, 1), \\ x_{II}^0 &= (1.5, 1.5, 1.5, 1.5, 1.0), \\ x_{III}^0 &= (0.5, 2.5, 1.0, 1.7, 1.0). \end{aligned}$$

We additionally list the number α of monotonicity violations concerning the function values. Independent of the start iterate and the algorithm, the calculated solution is (1.1243, 0.9794, 1.4777 0.9202, 1.1242), hence an approximation of the optimal point with accuracy 10^{-4}.

		HPPA				inexact PPA			
x^0	$\lVert x^0 - x^* \rVert$	♯iter.	♯in.iter.	$\sharp f/g$	α	♯iter.	♯in.iter.	$\sharp f/g$	α
x_I^0	2.2955	83	2873	3842	40	18	536	755	0
x_{II}^0	0.8741	76	2543	3353	37	16	417	628	0
x_{III}^0	1.8850	81	2054	3076	38	15	451	451	0

Table 4. Numerical results for Example 2 (Shor).

Example 3. Finally, we present the numerical results for the Maxquad example taken from Lemarechal and Mifflin [13], p.151. The problem data are as follows:

$$
\begin{aligned}
&f(x) = \max\left\{\langle A^{(k)}x, x\rangle - \langle b^k, x\rangle : k = 1,\dots,5\right\}, \quad x \in \mathbb{R}^{10},\\
&A^{(k)}_{ij} = \exp(i/j)\cos(i\cdot j)\sin(k), \quad i < j,\\
&A^{(k)}_{ji} = A^{(k)}_{ij},\\
&A^{(k)}_{ii} = 0.1\cdot i\cdot|\sin(k)| + \sum_{j\neq i}|A^{(k)}_{ij}|,\\
&b^k_i = \exp(i/k)\sin(i\cdot k),\\
&i,j = 1,\dots,10, \quad k = 1,\dots,5.
\end{aligned}
$$

The unique optimal point is $x^* = (-0.1263,\ -0.0346,\ -0.0067,\ 0.2668,\ 0.0673,\ -0.2786,\ 0.0744,\ 0.1387,\ 0.0839,\ 0.0385)$ and the optimal value is $f^* = -0.8414$.

As the start iterates we choose

$$
\begin{aligned}
x^0_I &= (1,1,1,1,1,1,1,1,1,1),\\
x^0_{II} &= (0,0,0,0,0,0,0,0,0,0),\\
x^0_{III} &= (-1,-1,0,-1,0,0,-1,-1,-1,0).
\end{aligned}
$$

In each case, both algorithms converge to the solution (-0.126, -0.034, -0.006, 0.026, 0.067, -0.278, 0.074, 0.138, 0.084, 0.038), which coincides approximately with the optimal solution. The number of iterations and monotonicity violations are given in Table 5.

		HPPA				inexact PPA			
x^0	$\|x^0 - x^*\|$	♯iter.	♯in.iter.	$\sharp f/g$	α	♯iter.	♯in.iter.	$\sharp f/g$	α
x^0_I	3.1885	23	1264	1753	11	3	146	221	0
x^0_{II}	0.3648	11	529	667	5	3	153	194	0
x^0_{III}	2.5412	27	1578	2060	13	3	212	276	0

Table 5. Numerical results for Example 3 (Maxquad).

Other algorithms for the solution of non-smooth convex optimization problems can be found, e.g., in [3, 4, 13, 14, 19]. However, a comparison with the numerical performance of the algorithms described in these references is not reasonable since we focus here on the comparison of the performance of two proximal-based algorithms and especially the influence of the projection step.

6. Conclusion

It has been shown that the hybrid proximal projection algorithm from Solodov and Svaiter [22] does not keep up the monotonicity properties, which are known to be valid for the inexact PPA from Rockafellar [17]. Further, a possible numerical realization of the HPPA and the inexact PPA has been suggested, especially, strategies for the computation of inexact solutions with tolerance σ were discussed. Using the BT-method from Schramm and Zowe [20] and Kiwiel [12] for the determination of an inexact solution proved to be adequate, although it works with the ε-subdifferential instead of the subdifferential of the auxiliary function f_k. Numerical examples showed that the inexact PPA is superior to the HPPA concerning the iteration number.

Acknowledgement

I would like to thank the referee for the valuable hints on this work.

References

[1] A. Auslender, M. Teboulle and S. Ben-Tiba, *A logarithmic-quadratic proximal method for variational inequalities*, Computational Optimization and Applications **12** (1–3) (1999), 31–40.

[2] R.S. Burachik and A.N. Iusem, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM Journal on Optimization **8** (1) (1998), 197–216.

[3] A. Cegielski and R. Dylewski, *Selection strategies in projection methods for convex minimization problems*, Discrete Math. **22** (1) (2002), 97–123.

[4] A. Cegielski and R. Dylewski, *Residual selection in a projection method for convex minimization problems*, Optimization **52** (2) (2003), 211–220.

[5] G. Chen and M. Teboulle, *A proximal-based decomposition method for convex minimization problems*, Mathematical Programming **64** (1994), 81–101.

[6] C. Jager, Numerische Analyse eines proximalen Projektions-Algorithmus, Diploma Thesis, University of Trier 2004.

[7] A. Kaplan and R. Tichatschke, Stable Methods for Ill-Posed Variational Problems-Prox-Regularization of Elliptic Variational Inequalities and Semi-Infinite Problems, Akademie Verlag 1994.

[8] A. Kaplan and R. Tichatschke, *Multi-step-prox-regularization method for solving convex variational problems*, Optimization **33** (4) (1995), 287–319.

[9] A. Kaplan and R. Tichatschke, *A general view on proximal point methods to variational inequalities in Hilbert spaces—iterative regularization and approximation*, Journal of Nonlinear and Convex Analysis **2** (3)(2001), 305–332.

[10] A. Kaplan and R. Tichatschke, *Convergence analysis of non-quadratic proximal methods for variational inequalities in Hilbert spaces*, Journal of Global Optimization **22** (1–4) (2002), 119–136.

[11] A. Kaplan and R. Tichatschke, *Interior proximal method for variational inequalities: case of nonparamonotone operators*, Set-Valued Analysis **12** (4) (2004), 357–382.

[12] K. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable minimization*, Mathematical Programming **46** (1990), 105–122.

[13] C. Lemaréchal and R. Mifflin, eds, Nonsmooth Optimization, volume 3 of IIASA Proceedings Series, Oxford, 1978. Pergamon Press.

[14] C. Lemaréchal, A. Nemirovski and Y. Nesterov, *New variants of bundle methods*, Mathematical Programming **69** (1) (B) (1995), 111–147.

[15] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle **4** (R–3) (1970), 154–158.

[16] Numerical Algorithms Group, NAG-Library, `http://www.nag.co.uk/`.

[17] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization **14** (1976), 877–898.

[18] R.T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Transactions of the American Mathematical Society **149** (1970), 75–88.

[19] H. Schramm, *Eine Kombination von Bundle-und Trust-Region-Verfahren zur Lösung nichtdifferenzierbarer Optimierungsprobleme*, Bayreuth. Math. Schr. **30** (1989), viii+205.

[20] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results*, SIAM Journal on Optimization **2** (1992), 121–152.

[21] N.Z. Shor, Minimization Methods for Nondifferentiable Functions, Springer-Verlag 1985.

[22] M.V. Solodov and B.F. Svaiter, *Forcing strong convergence of proximal point iterations in a Hilbert space*, Mathematical Programming **A87** (2000), 189–202.

Received 16 February 2006