Indian J. Pure Appl. Math., **47**(1): 1-8, March 2016

DOI: 10.1007/s13226-015-0162-3

INERTIAL PROXIMAL ALGORITHM FOR DIFFERENCE OF TWO MAXIMAL MONOTONE OPERATORS

M. Alimohammady and M. Ramazannejad

Department of Mathematics, Faculty of Mathematical Sciences,
University of Mazandaran, Babolsar, Iran, 47416-1468
e-mails: {*m.alimohammady, m.ramzannezhad*}*@gmail.com*

(*Received* 30 *November* 2013; *after final revision* 24 *February* 2015;
accepted 6 *July* 2015)

In this note, a new algorithm is presented for finding a zero of difference of two maximal monotone operators T and S, i.e., $T - S$ in finite dimensional real Hilbert space H in which operator S has local boundedness property. This condition is weaker than Moudafi's condition on operator S in [13]. Moreover, applying some conditions on inertia term in new algorithm, one can improve speed of convergence of sequence.

Key words : Maximal monotone operator; proximal point algorithm.

1. PRELIMINARIES

Let H be a Hilbert space. The notation $\langle .,. \rangle$ will be used for inner product in $H \times H$ and $\|.\|$ for the corresponding norm. A set valued operator $T : H \to 2^H$ is said to be monotone if

$$\langle x^* - y^*, x - y \rangle \geq 0, \quad \forall (x, x^*),\ (y, y^*) \in G(T),$$

wherein $G(T) := \{(x, y) \in H \times H; y \in Tx\}$ is graph of T. The domain of T is $D(T) := \{x \in H; T(x) \neq \emptyset\}$.

A monotone operator T is called *maximal monotone* if its graph is maximal in the sense of inclusion.

Associated with a given monotone operator T, the resolvent operator for T and parameter $\lambda > 0$ is $J_\lambda^T := (I + \lambda T)^{-1}$. The resolvent J_λ^T of a monotone operator T is a single valued nonexpansive map from $Im(I + \lambda T)$ to H [5, Proposition 3.5.3]. Moreover, the resolvent has full domain precisely when T is maximal monotone. For any $x \in H$, $\lim_{\lambda \to 0} J_\lambda^T(x) = Proj_{\overline{D(T)}}x$, wherein $Proj_{\overline{D(T)}}$ is the orthogonal projection on the closure of the domain of T. One of the best known approaches in the theory of optimization that is related to resolvent operators is *Yosida approximate* $T_\lambda := \frac{(I-J_\lambda^T)}{\lambda}$ of a maximal monotone operator T which satisfies in:

(i) For all $x \in H$, $T_\lambda(x) \in T(J_\lambda^T(x))$,

(ii) T_λ is Lipschitz with constant $\frac{1}{\lambda}$ and maximal monotone,

(iii) $T_\lambda(x)$ converges strongly to $T(x)$ as $\lambda \to 0$, for $x \in D(T)$,

(iv) $\|T_\lambda(x)\| \leq \|T^0(x)\|$ for every $x \in D(T)$, $\lambda > 0$, where T^0 is *minimal selection*

$$T^0(x) := \{y \in T(x);\ \|y\| = \min_{z \in T(x)} \|z\|\}, \qquad x \in D(T).$$

The aim of this note is offering the inertial proximal algorithm for the problem

$$\text{find } \ x \in H \ \text{ such that } \ 0 \in T(x) - S(x), \tag{1.1}$$

where $T, S : H \to 2^H$ are two maximal monotone operators on finite dimensional real Hilbert space H and it is equivalent to the problem

$$\text{find } \ x \in H \ \text{ such that } \ T(x) \cap S(x) \neq \emptyset. \tag{1.2}$$

This study is important, because finding the critical points of the difference of two convex functions is the special case of finding the zeros of difference of two maximal monotone operators. Actually, an algorithm for difference of two maximal monotone operators plays a central role in the study of DC programming [8, 9]. Moreover, it is valuable to mention that the variational inclusions corresponding to the difference of two monotone operators have grown from prox-regularity, multicommodity network, image restoring processing, tomography, molecular biology and optimization, see [1, 4, 6, 10] and the references therein.

The problem (1.1) did not study extensively. The latter studies are limited to Moudafi [12, 13]. By [13], a regularization of the problem (1.1) is

$$\text{find } \ x \in H \ \text{ such that } \ 0 \in T(x) - S_\lambda(x). \tag{1.3}$$

For finding a solution of (1.1) Moudafi [13] suggested a sequence $\{x_n\}$ by

$$x_{n+1} = J_{\mu_n}^T(x_n + \mu_n S_{\lambda_n} x_n) \quad \forall n \in \mathbb{N}, \tag{1.4}$$

where $\mu_n > 0$ and x_0 is an initial point.

Here, the problem (1.1) is studied via generalization of Moudafi's algorithm in [13] as the following:

$$x_{k+1} = J^T_{\beta_k}(x_k + \alpha_k(x_k - x_{k-1}) + \beta_k S_{\mu_k} x_k) \qquad \forall k \in \mathbb{N}, \tag{1.5}$$

with starting points $x_0, x_1 \in H$ and sequences $\{\mu_k\}$, $\{\alpha_k\}$ and $\{\beta_k\} \subset [0, +\infty)$ such that

(a) $\lim_{k\to+\infty} \mu_k = 0$;

(b) $\sum_{k=1}^{+\infty} \frac{\beta_k}{\mu_k} < +\infty$;

(c) $\lim_{k\to+\infty} \frac{\alpha_k}{\beta_k} = 0$;

also we suppose that

(d) $\sum_{k=1}^{+\infty} \alpha_k \|x_k - x_{k-1}\| < +\infty$;

(e) $\lim_{k\to+\infty} \frac{\|x_{k+1} - x_k\|}{\beta_k} = 0$.

We note that (1.5) is emanated from the evolution equation

$$x''(t) + \gamma x'(t) + \nabla f(x(t)) - \nabla g(x(t)) = 0, \tag{1.6}$$

where $\gamma > 0$ and algorithm (1.4) can be inspired from

$$x'(t) + \nabla f(x(t)) - \nabla g(x(t)) = 0, \tag{1.7}$$

in which both $f, g : H \to \mathbb{R}$ are differentiable convex functions and $\nabla f(x(t))$ and $\nabla g(x(t))$ are operators T and S in (1.1), respectively.

If $\nabla g(x(t)) = 0$, then (1.6) is *heavy ball with friction* system or (HBF) and (1.5) is equivalent to the standard gradient descent iteration (1.4) with an additional *inertia term* or *momentum term* $\alpha_k(x_k - x_{k-1})$. By the inertia term, convergence of the solution trajectories of the (HBF) system to a stationary point of f can be faster than those of the first order system (1.7) when $\nabla g(x(t)) = 0$ [14].

Another important advantage of algorithm (1.5) over algorithm (1.4) is using condition of local boundedness of S instead of boundedness in (1.4).

In this note, we present different conditions under which (1.5) converges to a solution of (1.1).

Now, we recall some required results and definitions.

Definition 1.1 — A set valued operator $T : H \to 2^H$ is locally bounded at $\bar{x}$ if there exists a neighborhood U of $\bar{x}$ such that the set $T(U)$ is bounded.

Lemma 1.2 [16] — Suppose that E is a reflexive Banach space. A maximal monotone operator $T : E \to 2^{E^*}$ is locally bounded at a point $\bar{x} \in D(T)$ if and only if $\bar{x}$ belongs to interior of $D(T)$.

Defintion 1.3 — A set valued operator $T : H \to 2^H$ is upper semicontinuous at $\bar{x}$ if for any positive $\epsilon > 0$ there exists $\delta > 0$ such that

$$\|x - \bar{x}\| \leq \delta \Rightarrow T(x) \subseteq T(\bar{x}) + B(0, \epsilon). \tag{1.8}$$

Lemma 1.4 [2] — Suppose that E is a Banach space. The maximal monotone operator $T : E \to 2^{E^*}$ is demiclosed, i.e., the following conditions hold.

(1) If $\{x_k\} \subset E$ converges strongly to x_0 and $\{u_k \in T(x_k)\}$ converges weak* to u_0 in E^*, then $u_0 \in T(x_0)$.

(2) If $\{x_k\} \subset E$ converges weakly to x_0 and $\{u_k \in T(x_k)\}$ converges strongly to u_0 in E^*, then $u_0 \in T(x_0)$.

Lemma 1.5 [11] — Suppose that $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ are three sequences of nonnegative numbers such that

$$a_{n+1} \leq (1 + b_n)a_n + c_n \quad \text{for all} \quad n \geq 1.$$

If $\sum_{n=1}^{\infty} b_n < +\infty$ and $\sum_{n=1}^{\infty} c_n < +\infty$, then $\lim_{n\to\infty} a_n$ exists.

2. Main Results

In the following, we improve the conditions of Theorem 2.1 in [13].

***Theorem* 2.1** — *Assume that S is locally bounded on $\overline{D(S)}$ and the solution set Ω of problem (1.1) is nonempty. If the conditions $(a), ..., (e)$ satisfy and $D(T) \subset D(S)$, then the sequence $\{x_k\}$ generated by (1.5) converges to a solution of (1.1).*

Proof : Take $x^* \in \Omega$. According to (1.2), there exists $y^* \in T(x^*) \cap S(x^*)$ and from (1.5), $x^* = J_{\beta_k}^T(x^* + \beta_k y^*)$. From the triangular inequality, (iv), nonexpansivity of $J_{\beta_k}^T$ and the fact that S_{μ_k} is also nonexpansive with constant $\frac{1}{\mu_k}$, one quickly deduces that

$$\begin{aligned}
\|x_{k+1} - x^*\| &= \|J_{\beta_k}^T(x_k + \alpha_k(x_k - x_{k-1}) + \beta_k S_{\mu_k} x_k) - J_{\beta_k}^T(x^* + \beta_k y^*)\| \\
&\leq \|x_k + \alpha_k(x_k - x_{k-1}) + \beta_k S_{\mu_k}(x_k) - x^* - \beta_k y^*\| \\
&\leq \|x_k - x^*\| + \alpha_k\|x_k - x_{k-1}\| + \beta_k\|S_{\mu_k}(x_k) - y^*\| \\
&\leq \|x_k - x^*\| + \alpha_k\|x_k - x_{k-1}\| + \beta_k(\|S_{\mu_k}(x_k) - S_{\mu_k}(x^*)\| + \|S_{\mu_k}(x^*) - y^*\|) \\
&\leq (1 + \frac{\beta_k}{\mu_k})\|x_k - x^*\| + \alpha_k\|x_k - x_{k-1}\| + \beta_k(\|S^0 x^*\| + \|y^*\|).
\end{aligned}$$

Applying (a) and (b), $\sum_{k=0}^{\infty}\beta_k < \infty$. Also by (d) and Lemma 1.5, we have $\lim_{k\to+\infty}\|x_k - x^*\|$ exists. Hence, $\{x_k\}$ is bounded. Notice that there exist $\widetilde{x}$ and a subsequence $\{x_{k_\nu}\}$ such that $\lim_{\nu\to\infty} x_{k_\nu} = \widetilde{x}$, since H is a finite dimensional space. We see $J^S_{\mu_{k_\nu}} x_{k_\nu}$ tends to $\widetilde{x}$, because

$$\begin{aligned}\|J^S_{\mu_{k_\nu}} x_{k_\nu} - \widetilde{x}\| &\le \|J^S_{\mu_{k_\nu}} x_{k_\nu} - J^S_{\mu_{k_\nu}}\widetilde{x}\| + \|J^S_{\mu_{k_\nu}}\widetilde{x} - \widetilde{x}\| \\ &\le \|x_{k_\nu} - \widetilde{x}\| + \|J^S_{\mu_{k_\nu}}\widetilde{x} - \widetilde{x}\|,\end{aligned}$$

and $\lim_{\nu\to+\infty} J^S_{\mu_{k_\nu}}\widetilde{x} = Proj_{\overline{D(S)}}\widetilde{x} = \widetilde{x}$. This fact and local boundedness of S imply that

$$\left\{S_{\mu_{k_\nu}} x_{k_\nu}\right\} \subseteq S\left(\{J^S_{\mu_{k_\nu}} x_{k_\nu}\}\right) \subseteq B, \tag{2.1}$$

where B is a bounded set. Therefore, $\{S_{\mu_{k_\nu}} x_{k_\nu}\}$ is bounded and there exist $\tilde{y}$ and a subsequence $\{S_{\mu_{k_{\nu'}}} x_{k_{\nu'}}\}$ such that $\lim_{\nu'\to\infty} S_{\mu_{k_{\nu'}}} x_{k_{\nu'}} = \tilde{y}$. Then $\tilde{y} \in S(\tilde{x})$ follows from

$$S_{\mu_{k_{\nu'}}} x_{k_{\nu'}} \in S\left(J^S_{\mu_{k_{\nu'}}} x_{k_{\nu'}}\right), \tag{2.2}$$

and Lemma 1.4. In sequel by (1.5), we have

$$S_{\mu_{k_{\nu'}}} x_{k_{\nu'}} - \left(\frac{x_{k_{\nu'}+1} - x_{k_{\nu'}}}{\beta_{k_{\nu'}}}\right) + \frac{\alpha_{k_{\nu'}}}{\beta_{k_{\nu'}}}(x_{k_{\nu'}} - x_{k_{\nu'}-1}) \in T x_{k_{\nu'}+1}, \tag{2.3}$$

tending ν' to $+\infty$ in (2.3) and using conditions (c), (e), boundedness of $\{x_k\}$ and Lemma 1.4, it is obtained that $\widetilde{y} \in T\widetilde{x}$. By similar procedure to proof of Theorem 2.1 in [13], $\widetilde{x}$ is unique. Then proof is complete. □

Example 2.2 : The best example of Theorem 2.1 can be seen in digital halftoning which is a procedure for producing a sample of pixels when a limited number of colors are available with a binary system so that it is a continuous-tone image. In this context Teuber *et al*. [17] minimized difference of two functions that one is corresponding to attraction of the dots by the image gray values and the other corresponds to the repulsion between the dots. They signified black pixel with 0 and white pixel with 1 and investigated images $u : G \to [0,1]$ on an integer grid $G := \{1,...,n_x\} \times \{1,...,n_y\}$. If m be the number of black pixels generated by the dithering procedure and $p := (p_k)_{k=1}^m = ((p_{k,x}, p_{k,y})^T)_{k=1}^m \in \mathbb{R}^{2m}$ be their position vector then $|p_k| := \sqrt{p_{k,x}^2 + p_{k,y}^2}$ is the Eucilidian norm of the position of the k-th black pixel.

In [17], it is detected minimizer $\hat{p}$ of the functional

$$E(p) = \underbrace{\sum_{k=1}^{m}\sum_{(i,j)\in G} w(i,j)|p_k - \begin{pmatrix} i \\ j \end{pmatrix}|}_{F(p)} - \underbrace{\lambda\sum_{k=1}^{m}\sum_{l=k+1}^{m}|p_k - p_l|}_{G(p)}, \tag{2.4}$$

where $w := 1 - u$ is the corresponding weight distribution and $\lambda := \frac{1}{m}\sum_{(i,j)\in G} w(i,j)$.

Given two functions $F(p)$ and $G(p)$ are continuous and convex. Since ∂F and ∂G are maximal monotone operators [15] and ∂G is locally bounded on $\mathbb{R}^{2m}$ [7], the problem of finding a minimizer of (2.4) is a special case of (1.1). If conditions (a),...,(e) satisfy and $D(\partial F) \subset D(\partial G)$, then by Theorem 2.1 the generated sequence $\{x_k\}$ of (1.5) converges to a minimizer of (2.4).

In next result, the condition of local boundedness of S in Theorem 2.1 is eliminated and domain of it will be entire H.

Corollary 2.3 — Assume that the solution set Ω of problem (1.1) is nonempty, conditions $(a), \ldots, (e)$ satisfy and $D(S) = H$, then the sequence $\{x_k\}$ generated by (1.5) converges to a solution of (1.3).

PROOF : Since $D(S)$ is open, using Lemma 1.2 the operator S is locally bounded at any point of $D(S)$. The rest of proof is similar to Theorem 2.1. □

Remark 2.4 : If $D(S) = H$ and $T - S$ is a monotone operator then by [3, Theorem 2.1], $T - S$ is maximal monotone. Hence, (1.1) reduces to find a zero point of maximal monotone operator $T - S$ and iteration algorithm (1.5) changes to $x_{k+1} = J_{\beta_k}^{T-S}(x_k + \alpha_k(x_k - x_{k-1}))$.

Corollary 2.5 — Assume that S is bounded value (i.e. for all $x \in H$, Sx is a bounded set) and upper semicontinuous at any point of $\overline{D(S)}$ and the solution set Ω of problem (1.1) is nonempty. If the conditions $(a), \ldots, (e)$ satisfy and $D(T) \subset D(S)$ then the sequence $\{x_k\}$ generated by (1.5) converges to a solution of (1.1).

PROOF : Since S is bounded value and upper semicontinuous at any point of $\overline{D(S)}$, so it is locally bounded. The rest of proof is similar to Theorem 2.1. □

Two types of interesting particular instances of (1.1) are:

$$\text{find} \quad x^* \in H \quad \text{such that} \quad y^* \in T(x^*), \tag{2.5}$$

and

$$\text{find} \quad x^* \in H \quad \text{such that} \quad x^* \in T(x^*). \tag{2.6}$$

It is assumed that $G(S) := H \times \{y\}$ for an arbitrary point $y \in H$ in (2.5) and $G(S) := \{(x,x); x \in H\}$ for any point $x \in H$ in (2.6).

In the following, we present the results of these types of problems.

Corollary 2.6 — Assume that the operator $S : H \to H$ is continuous and the solution set Ω of problem (1.1) is nonempty. If the conditions (a),...,(e) satisfy and $D(T) \subset D(S)$ then the sequence $\{x_k\}$ generated by (1.5) converges to a solution of (1.1).

PROOF : It is easy to check that sequence $\{x_k\}$ is bounded and there exist $\widetilde{x}$ and a subsequence $\{x_{k_\nu}\}$ such that $\lim_{\nu\to\infty} x_{k_\nu} = \widetilde{x}$. In proof of Theorem 2.1 it has been shown that $\lim_{\nu\to\infty} J_{\mu_{k_\nu}} x_{k_\nu} = \widetilde{x}$. Consequently, from

$$S_{\mu_{k_\nu}} x_{k_\nu} - \left(\frac{x_{k_\nu+1} - x_{k_\nu}}{\beta_{k_\nu}}\right) + \frac{\alpha_{k_\nu}}{\beta_{k_\nu}}(x_{k_\nu} - x_{k_\nu-1}) \in Tx_{k_\nu+1}, \tag{2.7}$$

$S_{\mu_{k_\nu}}(x_{k_\nu}) = S(J^S_{\mu_{k_\nu}}(x_{k_\nu}))$, continuity of S and by passing to a subsequence, we can arrange that left side of (2.7) converges to $S(\widetilde{x})$. By Lemma 1.4, we see that $S(\widetilde{x}) \in T(\widetilde{x})$, i.e. $0 \in T(\widetilde{x}) - S(\widetilde{x})$. $\square$

Corollary 2.7 — Assume that $S : H \to H$ is Lipschitz continuous, the solution set Ω of problem (1.1) is nonempty and $D(T) \subset D(S)$. If conditions (c),...,(e) satisfy and if one replaces condition $\sum_{k=1}^{\infty} \beta_k < \infty$ with (a) and (b) then the generated sequence $\{x_k\}$ of method

$$x_{k+1} = J^T_{\beta_k}(x_k + \alpha_k(x_k - x_{k-1}) + \beta_k S(x_k))$$

converges to a solution of problem (1.1).

Remark 2.8 : All results of this paper has derived from Lemma 1.4. In an infinite dimensional real Hilbert space, boundedness of sequence $\{x_k\}$ in Theorem 2.1 implies that there exist subsequence $\{x_{k_\nu}\}$ and $\tilde{x} \in H$ such that $\{x_{k_\nu}\}$ converges weakly to $\tilde{x}$. The fundamental difficulties in proving $\tilde{y} \in S(\tilde{x})$ and $\tilde{y} \in T(\tilde{x})$ are showing strongly convergence of either $\{J^S_{\mu_{k_\nu}} x_{k_\nu}\}$ to $\tilde{x}$ or $\{S_{\mu_{k_\nu}} x_{k_\nu}\}$ to $\tilde{y}$ and the left side of (2.3) to $\tilde{y}$.

ACKNOWLEDGEMENT

The authors thank the referees for their pertinent and constructive comments.

REFERENCES

1. S. Adly and W. Oettli, Solvability of generalized nonlinear symmetric variational inequalities, *J. Austral. Math. Soc. Ser.*, **40** (1999), 289-300.

2. Y. Alber and I. Ryazantseva, *Nonlinear Ill-posed Problems of Monotone Type*, Springer, New York, (2006).

3. M. Alimohammady, M. Ramazannejad and M. Roohi, Notes on the difference of two monotone operators, *Optimization Letters*, **8**(1) (2014), 81-84.

4. L. T. H. An and D. T. Pham, The DC programming and DCA revised with DC models of real world nonconvex optimization problems, *Ann. Oper. Res.*, **133** (2005), 25-46.

5. J. P. Aubin and H. Frankowska, *Set-valued analysis,* Reprint of the 1990 Edition.

6. S. Chandra, Strong pseudo-convex programming, *Indian J. Pure Appl. Math.*, **3** (1972), 278-282.

7. F. H. Clarke, R. J. Stern and G. Sabidussi, *Nonlinear analysis, differential equations and control*, ser. NATO Sciences Series, Series C: Mathematical and Physical Sciences, Kluwer Academic Publishers, **528** (1999).

8. A. Hamdi, A modified Bregman proximal schemes to minimize the difference of two convex functions, *Appl. Math. E-Notes.*, **6** (2006), 132-140.

9. A. Hamdi, A Moreau-Yosida regularization of a DC functions: application to variational inequality problems, *Appl. Math. E-Notes.*, **5** (2005), 164-170.

10. S. Huda and R. Mukerjee, Minimax second-order designs over cuboidal regions for the difference between two estimated responses, *Indian J. Pure Appl. Math.*, **41**(1) (2010), 303-312.

11. D. Lei and L. Shenghong, Ishikawa iteration process with errors for nonexpansive mappings in uniformly convex Banach spaces, *Internat. J. Math. and Math. Sci.*, **24**(1) (2000), 49-53.

12. A. Moudafi, On the difference of two maximal monotone operators: Regularization and algorithmic approaches, *Appl. Math. Comput.*, **202** (2008), 446-452.

13. A. Moudafi, On critical points of the difference of two maximal monotone operators, *Afr. Mat.*, (2013). DOI 10.1007/s13370-013-0218-7.

14. N. Qian, On the momentum term in gradient descent learning algorithms, *Neural networks*, **12**(1) (1999), 145-151.

15. R. T. Rockafellar, On the maximal monotonicity of subdifferential mappings, *Pacific J. Math.*, **33** (1970), 209-216.

16. R. T. Rockafellar, Local boundedness of nonlinear monotone operators, *Michigan Math. J.*, **16** (1969), 397-407.

17. T. Teuber, G. Steidl, P. Gwosdek, Ch. Schmaltz and J. Weickert, Dithering by differences of convex functions, *SIAM Journal on Imaging Sciences*, **4**(1) (2011), 79-108.