Fixed Point Theory, 21(2020), No. 2, 685-692
DOI: 10.24193/fpt-ro.2020.2.47
http://www.math.ubbcluj.ro/~nodeacj/sfptcj.html

A THREE-OPERATOR SPLITTING ALGORITHM FOR NULL-POINT PROBLEMS

ABDELLATIF MOUDAFI

Aix-Marseille Université, LIS UMR CNRS 7020
E-mail: abdellatif.moudafi@univ-amu.fr

Abstract. The aim of this paper is to present and investigate the asymptotic behavior of a novel splitting algorithm for solving a class of null-point problems governed by three maximal monotone operators. Two of which are assumed to be proximable and one verified a cocoercive property. The proposed algorithm is based on a duality principle and the convergence proofs rely on classical arguments of nonlinear analysis and properties of the resolvent mappings of maximal monotone operators. The convex optimization case is also addressed with its related algorithm and convergence result.
Key Words and Phrases: Maximal monotone operators, splitting algorithms, Yosida approximate operator, cocoercivity, strong monotonicity, convex minimization, conjugate function.
2010 Mathematics Subject Classification: 49J53, 65K10, 49M37, 47H10, 90C25.

1. Introduction and preliminaries

Throughout, H is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. Notations not explicitly defined here are standard.

Recall that the graph, $gphA$, of a set-valued operator $A : H \to 2^H$ is given by $gphA = \{(x, y) \in H \times H; y \in A(x)\}$, that the mapping A is monotone if

$$\forall (x, y) \in gphA \ \forall (x', y') \in gphA \ \langle x - x', y - y' \rangle \geq 0,$$

and maximal monotone if it cannot be properly extended without destroying monotonicity. A is μ-strongly monotone if there exists $\mu > 0$ such that

$$\forall (x, y) \in gphA \ \forall (x', y') \in gphA \ \langle x - x', y - y' \rangle \geq \mu \|x - x'\|^2.$$

The inverse of A is defined via its graph by $gphA^{-1} = \{(y, x) \in H \times H; y \in A(x)\}$. The resolvent of A with parameter $\gamma > 0$ is $J_\gamma^A = (I + \gamma A)^{-1}$, it is firmly monotone (and thus Lipschitz continuous), namely

$$\forall x, y \in H \quad \langle J_\gamma^A(x) - J_\gamma^A(y), x - y \rangle \geq \|J_\gamma^A(x) - J_\gamma^A(y)\|^2.$$

Moreover, the resolvent has full domain H precisely when A is maximal monotone. Remember also that the Yosida approximate of the operator A is defined by

$$A_\gamma = (\gamma I + A^{-1})^{-1}.$$

Finally, remember that an operator A is τ-cocoercive if there exists $\tau > 0$ such that

$$\forall x, y \in H \quad \langle A(x) - A(y), x - y\rangle \geq \tau \|A(x) - A(y)\|^2.$$

We will assume that the inverse of one of the three operators is cocoercive, which is the case, for example, when $A = \partial f$, the subdifferential of a proper closed strongly convex function f.
It is then well-known, see [2, Theorem 18.5], that we have $A^{-1} = \nabla f^*$, the function $f^*(y) = \sup_x(\langle x, y\rangle - f(x))$ being the conjugate of f. It is worth mentioning that there are several works suggesting that the notion of strong convexity is a fundamental tool in designing and analyzing (the regret or generalization ability of) a wide class of learning algorithms, see [4], [6] or [7] and references therein.
Splitting algorithms have been successfully used in computational sciences to reduce complex problems into a series of simpler ones. Recently, these algorithms have been widely used to solve problems in machine learning, signal processing, and imaging, see [4], [3], [5] and have created a resurgence of interest due of its often simple descriptions and (nearly) state-of-the-art performance for large-scale optimization problems. In the spirit of these works, we will focus our attention on the problem

$$\text{find } x^* \in H \text{ such that } 0 \in (A + B + C)(x^*), \tag{1.1}$$

where A, B and C are three maximal monotone operators with one of which, say A, is such that A^{-1} is τ-cooercive.

Note that the inclusion problem (1.1) subsumes a wide spectrum of problems in applied nonlinear analysis and it has recently enjoyed a surge in popularity thanks to its ability to nicely model many optimization problems that arise in machine learning, signal processing and statistics. The most straightforward example of (1.1) arises from the optimization problem, namely

$$\min_x \big(f(x) + g(x) + h(x)\big), \tag{1.2}$$

where f, g and h are proper closed and convex functions. The first-order optimality condition of this problem, under certain qualification assumptions see for example [2, Corollary 16.39], reduces to (1.1) with $A = \partial f$, $B = \partial g$ and $C = \partial h$, the subdifferentials of f, g and h.

In this paper, based on a key observation in [1, Remark 3.4], we develop a novel algorithmic approach relying on the following nice dual formulation of (1.1). Setting $\xi^* \in Bx^*$ and $\eta^* \in Cx^*$, (1.1) is nothing but $Ax^* + \xi^* + \eta^* \ni 0$ which is equivalent to the following dual system

$$\begin{cases} -A^{-1}(-\xi^* - \eta^*) + B^{-1}(\xi^*) \ni 0; \\ -A^{-1}(-\xi^* - \eta^*) + C^{-1}(\eta^*) \ni 0. \end{cases} \tag{1.3}$$

To obtain a fixed-point formulation from (1.3) which will amounts us to design a splitting algorithm and go on to establish its convergence to a primal an a dual solutions, we state the following key facts that will be needed in the sequel.
Fact 1. The Yosida approximate A_γ is γ-cocoercive, which is equivalent to

$$\forall x, y \in H \quad \gamma^2 \|A_\gamma x - A_\gamma y\|^2 \leq \|x - y\|^2 - \|J_\gamma^A x - J_\gamma^A y\|^2.$$

Fact 2. The Yosida approximate is related to the resolvent by the formula

$$\forall x \in H \quad J_\gamma^A(x) + \gamma A_\gamma(x) = x.$$

Fact 3. Using the Yosida approximate, we can write down system (1.3) as a fixed point system, namely

$$\begin{cases} \xi^* = B_\alpha(\alpha\xi^* + A^{-1}(-\xi^* - \eta^*)); \\ \eta^* = C_\alpha(\alpha\eta^* + A^{-1}(-\xi^* - \eta^*)). \end{cases} \tag{1.4}$$

Indeed,

$$\begin{aligned} -A^{-1}(-\xi^* - \eta^*) + B^{-1}(\xi^*) \ni 0 \quad &\Leftrightarrow \quad \alpha\xi^* + A^{-1}(-\xi^* - \eta^*) \in \alpha\xi^* + B^{-1}(\xi^*) \\ &\Leftrightarrow \quad \xi^* = B_\alpha(\alpha\xi^* + A^{-1}(-\xi^* - \eta^*)). \end{aligned}$$

Likewise, we obtain

$$\eta^* = C_\alpha(\alpha\eta^* + A^{-1}(-\xi^* - \eta^*)), \quad \text{for all } \alpha > 0.$$

The algorithm is then to solve this fixed point system via the iteration

$$\begin{cases} \xi_{k+1} = B_{\alpha_k}(\alpha_k\xi_k + A^{-1}(-\xi_k - \eta_k)); \\ \eta_{k+1} = C_{\alpha_k}(\alpha_k\eta_k + A^{-1}(-\xi_k - \eta_k)), \end{cases} \tag{1.5}$$

where for all $k \in I\!N$, $\alpha_k > 0$ is a stepsize.
Using Fact 2, we can rewrite algorithm (1.5) by means of the resolvent operators of B and C, namely

$$\begin{cases} \xi_{k+1} = \frac{(I - J^B_{\alpha_k})}{\alpha_k}(\alpha_k\xi_k + A^{-1}(-\xi_k - \eta_k)); \\ \eta_{k+1} = \frac{(I - J^C_{\alpha_k})}{\alpha_k}(\alpha_k\eta_k + A^{-1}(-\xi_k - \eta_k)), \end{cases} \tag{1.6}$$

which is efficient when the resolvent operators (proximal mappings) of B and C (g and h) are available in closed form or at least fast to compute as well as the inverse of A (the conjugate function of f or at least its gradient operator).

2. Main convergence results

Now, we are in a position to state and prove our main convergence results. To begin with, let us establish the following key property.

Proposition 2.1. *Let $\alpha_k \in]0, \frac{1}{\tau}[$, A, B and C be three maximal monotone operators with A^{-1} a τ-cocoercive operator and assume that (1.1) possesses at least one solution, say x^*. Let $(\xi_k, \eta_k)_{k\in\mathbb{N}}$ be the sequence defined by (1.5). Then, we have the following estimate*

$$\Gamma_{k+1} \leq \Gamma_k - \frac{2}{\alpha_k}(\tau - \frac{1}{\alpha_k})\|\upsilon_k - x^*\|^2 - \frac{\Theta_k}{\alpha_k^2}, \tag{2.1}$$

where

$$\Gamma_k := \|(\xi_k - \xi^*)\|^2 + \|\eta_k - \eta^*\|^2 = \|(\xi_k, \eta_k) - (\xi^*, \eta^*)\|^2 \quad (\textit{norm of } H \times H)$$

and

$$\Theta_k := \|\alpha_k(\xi_k - \xi_{k+1}) + \upsilon_k - x^*\|^2 + \|\alpha_k(\eta_k - \eta_{k+1}) + \upsilon_k - x^*\|^2$$

with $\upsilon_k := A^{-1}(-\xi_k - \eta_k)$, and remember that $x^* = A^{-1}(-\xi^* - \eta^*)$.

Proof. Let x^* be a solution of (1.1), $\xi^* \in Bx^*$ and $\eta^* \in Cx^*$. In view of relations (1.4) and (1.5) and according to the α_k-cocoercive property of the Yosida approximate B_{α_k}, we can successively write

$$\begin{aligned}
\|\xi_{k+1}-\xi^*\|^2 &\le \frac{1}{\alpha_k^2}\|\alpha_k(\xi_k-\xi^*)+A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2 \\
&- \frac{1}{\alpha_k^2}\|\alpha_k(\xi_k-\xi_{k+1})+A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2 \\
&\le \|(\xi_k-\xi^*)\|^2+\frac{2}{\alpha_k}\langle A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*),\xi_k-\xi^*\rangle \\
&+ \frac{1}{\alpha_k^2}\|A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2 \\
&- \frac{1}{\alpha_k^2}\|\alpha_k(\xi_k-\xi_{k+1})+A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2.
\end{aligned}$$

Likewise, using the α_k-cocoercivenes property of the Yosida approximate C_{α_k}, we obtain

$$\begin{aligned}
\|\eta_{k+1}-\eta^*\|^2 &\le \|\eta_k-\eta^*\|^2+\frac{2}{\alpha_k}\langle A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*),\eta_k-\eta^*\rangle \\
&+ \frac{1}{\alpha_k^2}\|A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2 \\
&- \frac{1}{\alpha_k^2}\|\alpha_k(\eta_k-\eta_{k+1})+A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|^2.
\end{aligned}$$

Summing the two last inequality, using the τ-cocoerciveness of the operator $-A^{-1}(-\cdot)$ which shares this property with the operator A^{-1} and the fact that

$$\upsilon_k := A^{-1}(-\xi_k-\eta_k) \text{ and } x^* = A^{-1}(-\xi^*-\eta^*),$$

we get

$$\begin{aligned}
\|(\xi_{k+1},\eta_{k+1})-(\xi^*,\eta^*)\|^2 &\le \|(\xi_k,\eta_k)-(\xi^*,\eta^*)\|^2 \\
&+ \frac{2}{\alpha_k}\langle \upsilon_k-x^*,(\xi_k+\eta_k)-(\xi^*+\eta^*)\rangle+\frac{2}{\alpha_k^2}\|\upsilon_k-x^*\|^2 \\
&- \frac{1}{\alpha_k^2}\|\alpha_k(\xi_k-\xi_{k+1})+\upsilon_k-x^*\|^2 \\
&- \frac{1}{\alpha_k^2}\|\alpha_k(\eta_k-\eta_{k+1})+\upsilon_k-x^*\|^2 \\
&\le \|(\xi_k,\eta_k)-(\xi^*,\eta^*)\|^2-\frac{2}{\alpha_k}(\tau-\frac{1}{\alpha_k})\|\upsilon_k-x^*\|^2-\frac{\Theta_k}{\alpha_k^2}.
\end{aligned}$$

This completes the proof.

Theorem 2.2. *If in addition to the assumptions of Proposition 2.1, we assume that* $\alpha_k \in]\varepsilon, \frac{1}{\tau}-\varepsilon[$ *for some given* $\varepsilon > 0$ *small enough, then the sequence* $(\xi_k,\eta_k)_{k\in\mathbb{N}}$ *is Fejer monotone with respect to the solution set of (1.3) and the sequence* $(\upsilon_k)_{k\in\mathbb{N}}$ *norm converges to* x^* *solution of (1.1). Moreover,* $(\xi_k)_{k\in\mathbb{N}}$ *and* $(\eta_k)_{k\in\mathbb{N}}$ *are asymptotically regular and the sequence* $(\xi_k,\eta_k)_{k\in\mathbb{N}}$ *weakly converges to a solution of (1.3).*

Proof. Relation (2.1) implies that the sequence $(\Gamma_k)_{k\in\mathbb{N}}$ is no increasing and thus it converges, which implies in turn the boundedness of the sequence $(\xi_k,\eta_k)_{k\in\mathbb{N}}$. Taking into account conditions on the sequence of parameters (α_k), we also directly obtain that

$$\lim_k \|A^{-1}(-\xi_k-\eta_k)-A^{-1}(-\xi^*-\eta^*)\|=0 \text{ and } \lim_k \Theta_k=0. \tag{2.2}$$

These in turn ensure that

$$\lim_{k\to+\infty}\|\xi_k-\xi_{k+1}\|=\lim_{k\to+\infty}\|\eta_k-\eta_{k+1}\|=0, \tag{2.3}$$

in other words, $(\xi_k)_{k\in\mathbb{N}}$ and $(\eta_k)_{k\in\mathbb{N}}$ are asymptotically regular. On the other hand, (1.5) can be rewritten as

$$\begin{cases} \alpha_k(\xi_k-\xi_{k+1})+A^{-1}(-\xi_k-\eta_k)\in B^{-1}(\xi_{k+1}); \\ \alpha_k(\eta_k-\eta_{k+1})+A^{-1}(-\xi_k-\eta_k)\in C^{-1}(\eta_{k+1}). \end{cases} \tag{2.4}$$

By passing to the limit on a subsequence of $(\xi_k,\eta_k)_{k\in\mathbb{N}}$ converging to a weak-cluster point $(\tilde\xi,\tilde\eta)$ in (2.4), in the light of both (2.2) and (2.3) and by taking also into account the fact that the graphs of B and C are strongly-weakly closed, we obtain

$$\begin{cases} A^{-1}(-\tilde\xi-\tilde\eta)-B^{-1}(\tilde\xi)\ni 0; \\ A^{-1}(-\tilde\xi-\tilde\eta)-C^{-1}(\tilde\eta)\ni 0. \end{cases} \tag{2.5}$$

Thus $(\tilde\xi,\tilde\eta)$ solves (1.3) and the weak convergence of the whole sequence $(\xi_k,\eta_k)_{k\in\mathbb{N}}$ follows then by the celebrate Opial's Lemma.

3. Convex minimization

To begin with, let us recall that the partial differential of a convex function h is defined as $\partial h(x):=\{z\in H;\ h(y)\geq h(x)+\langle z,y-x\rangle\ \forall y\in H\}$ and focus on the minimization problem (1.3), namely

$$\min_x (f+g+h)(x), \tag{3.1}$$

which is equivalent, under a qualification condition see for instance [2, Corollary 16.39], to

$$0\in\partial f(x^*)+\partial g(x^*)+\partial h(x^*). \tag{3.2}$$

This is equivalent in turn to the following dual system

$$\begin{cases} -\nabla f^*(-\xi^*-\eta^*)+\partial g^*(\xi^*)\ni 0; \\ -\nabla f^*(-\xi^*-\eta^*)+\partial h^*(\eta^*)\ni 0, \end{cases} \tag{3.3}$$

which amounts to

$$\begin{cases} \xi^*=argmin_\xi\big(f^*(-\xi-\eta^*)+g^*(\xi)\big); \\ \eta^*=argmin_\eta\big(f^*(-\xi^*-\eta)+h^*(\eta)\big). \end{cases}$$

Observe that both

$$\min_\xi\big(f^*(-\xi-\eta^*)+g^*(\xi)\big)=(f+g)^*(-\eta)$$

and

$$\min_{\eta}\big(f^*(-\xi-\eta^*)+h^*(\eta)\big)=(f+h)^*(-\xi)$$

are infimal convolutions (i.e., epigraphical sums).
In this context, the dual minimization algorithm reduces to

$$\begin{cases} \xi_{k+1}=prox_{\alpha_k^{-1}g^*}(\xi_k+\alpha_k^{-1}\nabla f^*(-\xi_k-\eta_k)); \\ \eta_{k+1}=prox_{\alpha_k^{-1}h^*}(\eta_k+\alpha_k^{-1}\nabla f^*(-\xi_k-\eta_k)), \end{cases} \tag{3.4}$$

which can be rewritten as

$$\begin{cases} \xi_{k+1}=\frac{(I-prox_{\alpha_k g})}{\alpha_k}(\alpha_k\xi_k+\nabla f^*(-\xi_k-\eta_k)); \\ \eta_{k+1}=\frac{(I-prox_{\alpha_k h})}{\alpha_k}(\alpha_k\eta_k+\nabla f^*(-\xi_k-\eta_k)). \end{cases} \tag{3.5}$$

The corresponding convergence result can then be summarized as follows.
Theorem 3.1. *Let $\alpha_k\in]\varepsilon,\tau-\varepsilon[$ for some given $\varepsilon>0$ small enough, f,g and h be three proper closed convex functions with ∇f^* τ-Lipschitz continuous and assume that (3.1) possesses at least one solution, say x^*. Then, the sequence $(\xi_k,\eta_k)_{k\in\mathbb{N}}$ generated by (3.5) weakly converges to some solution of (3.3) and the sequence*

$$\big(\upsilon_k:=\nabla f^*(-\xi_k-\eta_k)\big)_{k\in\mathbb{N}}$$

norm converges to x^* solution of (3.1).
Proof. Follows by applying Theorem 2.2 with $A=\partial f, B=\partial g$ and $C=\partial h$ and using the fact that, for any closed proper convex function h, we have

$$(\partial h)^{-1}=\partial h^* \text{ and } prox_{\gamma h}(x)=argmin_y\left(h(y)+\frac{1}{2\gamma}\|y-x\|^2\right)=J_\gamma^{\partial h}(x),$$

and that τ-Lipschitz continuity of ∇f^* is equivalent to its $\frac{1}{\tau}$-cocovercivity, see for example [2, Theorem 18.15].

A straightforward example is the Elastic net regularization which combines both the l_1 and l_2 penalties. The optimization problem is

$$\min_{x\in\mathbb{R}^n}\big(\mu_2\|x\|_2^2+\mu_1\|x\|_1+l(Ax,b)\big),$$

where $A\in\mathbb{R}^{m\times n}, b\in\mathbb{R}^m$ and l is the loss function, which may be nondifferentiable. The conjugate function of the l_2 regularization term, equal to $\frac{1}{4\mu_2}\|x\|_2^2$, is differentiable and has a Lipschitz continuous gradient. The algorithm has the remarkable property that the operators involved are evaluated separately in each iteration, either by the forward step using here the gradient of f^*, or by the backward steps for the nonsmooth terms, by using the corresponding proximal mappings, here $prox_{\mu_1\|\cdot\|}$ a simple operation called soft thresholding as well as $prox_l$, the matrices A and A^* with a loss function l that is proximable.

4. Concluding remarks

To conclude, it worth mentioning that, in the multi-valued case of all inverse operators, the dual key system (1.3) can be solved sequentially, for example, by the Douglas-Rachford splitting algorithm. Moreover, using a partial regularization by replacing in (1.3) the inverse of the operator A by its Yosida approximate or else the inverses of B and C by their Yosida approximates, it is then easy to see that we obtain simultaneous Primal, dual or primal-dual Passty's type algorithms. More precisely, we obtain in the first case

$$\begin{cases} \xi_{k+1} = J_{\alpha_k}^{B^{-1}}(-\eta_k - J_{\alpha_k}^{A^{-1}}(-\xi_k - \eta_k)); \\ \eta_{k+1} = J_{\alpha_k}^{C^{-1}}(-\xi_k - J_{\alpha_k}^{A^{-1}}(-\xi_k - \eta_k)), \end{cases} \tag{4.1}$$

which is equivalent to

$$\begin{cases} \xi_{k+1} = J_{\alpha_k}^{B^{-1}}(\xi_k + \alpha_k (A^{-1})_{\alpha_k}(-\xi_k - \eta_k)); \\ \eta_{k+1} = J_{\alpha_k}^{C^{-1}}(\eta_k + \alpha_k (A^{-1})_{\alpha_k}(-\xi_k - \eta_k)), \end{cases} \tag{4.2}$$

and can be rewritten with A, B and C using the fact that $M_\lambda(x) = J_{\frac{1}{\lambda}}^{M^{-1}}(\frac{x}{\lambda})$ for all $x \in H$ and $\lambda > 0$. This will be detailed in a forthcoming paper. The convergence results in the case of partial regularization may be of ergodic type.

Acknowledgment. This work was supported by I&M team of Systems and Computer Sciences Laboratory (LIS) at Aix Marseille University.

References

[1] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, J. Convex Analysis, **3**(1996) no. 1, 1-24.

[2] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011.

[3] R.I. Bot, E.R. Csetnek, C. Hendrich, *Recent developments on primal-dual splitting methods with applications to convex minimization*, Mathematics Without Boundaries, (2014), 57-99.

[4] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality,* Optimization, **63**(2014) no. 9, 1289-1318.

[5] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optimization Theory and Applications, **158**(2013) no. 2, 460-79.

[6] D. Davis, W. Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued and Variational Analysis, **25**(2017), no. 4, 829-858.

[7] Sh.M. Kakade, Sh. Shalev-Shwartz, *On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization*, Unpublished Manuscript, 2009.

[8] M. Yan, *A new primal-dual algorithm for minimizing the sum of three functions with a linear operator*, J. Scientific Computing, (2018), 1-20.

Received: June 28, 2018; Accepted: February 2, 2019.