A note on the convergence of the monotone inclusion version of the primal-dual hybrid gradient algorithm

Levon Nurbekyan
Department of Mathematics, Emory University
lnurbek@emory.edu

November 8, 2023

Abstract

The note contains a direct extension of the convergence proof of the primal-dual hybrid gradient (PDHG) algorithm in [3] to the case of monotone inclusions.

1 Introduction

Assume that $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces, and $A : \mathcal{H}_1 \to 2^{\mathcal{H}_1}$, $B : \mathcal{H}_2 \to 2^{\mathcal{H}_2}$ are maximally monotone maps. Furthermore, assume that $C : \mathcal{H}_1 \to \mathcal{H}_2$ is a non-zero bounded linear operator, and consider the following pair of primal-dual monotone inclusions

$$\begin{aligned} &\text{find } x \in \mathcal{H}_1 \text{ s.t. } 0 \in Ax + C^*(B(Cx)) \quad \text{(P)}\\ &\text{find } y \in \mathcal{H}_2 \text{ s.t. } y \in B(Cx),\ -C^*y \in Ax,\ \text{for some } x \in \mathcal{H}_1 \quad \text{(D)} \end{aligned} \tag{1}$$

When A, B are subdifferential maps of proper convex lower semicontinuous functions, this previous problem reduces to a pair of primal-dual convex programs or a convex-concave saddle point problem. More specifically, if $A = \partial f_1$, $B = \partial f_2$ for $f_1 : \mathcal{H}_1 \to \overline{\mathbb{R}}$, $f_2 : \mathcal{H}_2 \to \overline{\mathbb{R}}$ then (1) is equivalent to

$$\begin{aligned} \inf_{x\in\mathcal{H}_1} \{f_1(x) + f_2(Cx)\} &= \inf_{x\in\mathcal{H}_1} \sup_{y\in\mathcal{H}_2} \{f_1(x) + \langle Cx, y\rangle - f_2(y)\}\\ &= \sup_{y\in\mathcal{H}_2} \{-f_1^*(-C^*y) - f_2(y)\}. \end{aligned} \tag{2}$$

In [2, 3], the authors introduced a first-order primal-dual splitting scheme for solving (2), which in its simplest form reads as

$$\begin{cases} x^{n+1} = \operatorname*{argmin}_{x\in\mathcal{H}_1} f_1(x) + \langle Cx, y^n\rangle + \frac{\|x-x^n\|^2}{2\tau},\\ \tilde{x}^{n+1} = 2x^{n+1} - x^n,\\ y^{n+1} = \operatorname*{argmax}_{y\in\mathcal{H}_2} \langle C\tilde{x}^{n+1}, y\rangle - f_2(y) - \frac{\|y-y^n\|^2}{2\sigma}, \end{cases} \tag{3}$$

where $\tau, \sigma > 0$. The main results in [2, 3] provide convergence of ergodic sequences

$$X^N = \frac{1}{N}\sum_{n=1}^N x_i, \quad Y^N = \frac{1}{N}\sum_{n=1}^N y_i, \tag{4}$$

under the assumption

$$\tau\sigma < \frac{1}{\|C\|^2}. \tag{5}$$

In [6], the author considers a more general version of (1) and introduces a splitting scheme, which in its simplest form reads as

$$\begin{cases} x^{n+1} = (I + \tau A)^{-1}(x^n - \tau C^* y^n),\\ \tilde{x}^{n+1} = 2x^{n+1} - x^n,\\ y^{n+1} = (I + \sigma B^{-1})^{-1}\left(y^n + \sigma C\tilde{x}^{n+1}\right). \end{cases} \tag{6}$$

Using techniques different from the ones in [2, 3], the author in [6] proves the convergence of the iterates in (6) to the solution of (1) under the same assumption (5). The key idea is to rewrite (6) in the form of a forward-backward splitting algorithm analyzed in [4].

In this note, we provide a direct extension of the convergence proof of (3) in [3] for the monotone inclusion version (6).

2 Notation and hypotheses

Throughout the note, we assume that $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces, A, B are maximally monotone, and C is a non-zero bounded linear operator. Furthermore, assume that $\psi_1 : \mathcal{H}_1 \to \mathbb{R}$ and $\psi_2 : \mathcal{H}_2 \to \mathbb{R}$ are continuously Fréchet differentiable convex functions, and denote by

$$\begin{aligned} D_1(x, \bar{x}) =& \psi_1(x) - \psi_1(\bar{x}) - \langle \nabla\psi_1(\bar{x}), x - \bar{x}\rangle, \quad x, \bar{x} \in \mathcal{H}_1, \\ D_2(y, \bar{y}) =& \psi_2(y) - \psi_2(\bar{y}) - \langle \nabla\psi_2(\bar{y}), y - \bar{y}\rangle, \quad y, \bar{y} \in \mathcal{H}_2, \end{aligned} \tag{7}$$

their Bregman divergences. We assume that there exists $\alpha > 0$ such that

$$D_1(x, \bar{x}) + D_2(y, \bar{y}) - \langle C(x - \bar{x}), y - \bar{y}\rangle \geq \alpha \left(\|x - \bar{x}\|^2 + \|y - \bar{y}\|^2\right), \quad \forall x, \bar{x} \in \mathcal{H}_1, \quad \forall y, \bar{y} \in \mathcal{H}_2. \tag{8}$$

Taking $y = \bar{y}$ we obtain

$$\psi_1(x) - \psi_1(\bar{x}) - \langle \nabla\psi_1(\bar{x}), x - \bar{x}\rangle = D_1(x, \bar{x}) \geq \alpha\|x - \bar{x}\|^2, \forall x, \bar{x} \in \mathcal{H}_1, \tag{9}$$

which means that ψ_1 is 2α-strongly convex. Similarly, we have that

$$\psi_2(y) - \psi_2(\bar{y}) - \langle \nabla\psi_2(\bar{y}), y - \bar{y}\rangle = D_2(y, \bar{y}) \geq \alpha\|y - \bar{y}\|^2, \forall y, \bar{y} \in \mathcal{H}_2, \tag{10}$$

and so ψ_2 is also 2α-strongly convex.

Lemma 1. *Assume that $\mathcal{H}$ is a Hilbert space, $\psi : \mathcal{H} \to \mathbb{R}$ is a continuously Fréchet differentiable strongly convex function, and $M : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximally monotone operator. Furthermore, denote by*

$$D(x, \bar{x}) = \psi(x) - \psi(\bar{x}) - \langle \nabla\psi(\bar{x}), x - \bar{x}\rangle, \quad x, \bar{x} \in \mathcal{H}.$$

Then the map

$$Tx = \nabla_x D(x, \bar{x}) + Mx, \quad x \in \mathcal{H},$$

is surjective for all $\bar{x} \in \mathcal{H}$.

Proof. Fix an arbitrary $\bar{x} \in \mathcal{H}$. Since $x \mapsto D(x, \bar{x})$ is convex and smooth [1, Theorem 20.25] yields that $x \mapsto \nabla_x D(x, \bar{x})$ is maximally monotone with a domain $\mathcal{H}$. Hence, by [5, Theorem 1] we have that T is maximally monotone.

Next, let $(x_0, y_0) \in \operatorname{gra} M$. Then for every $x \in \mathcal{H}$ we have that

$$\begin{aligned} \inf \|Tx\| =& \inf \|\nabla\psi(x) - \nabla\psi(x_0) + Mx - y_0 + (\nabla\psi(x_0) + y_0 - \nabla\psi(\bar{x}))\| \\ \geq& \inf \|\nabla\psi(x) - \nabla\psi(x_0) + Mx - y_0\| - \|\nabla\psi(x_0) + y_0 - \nabla\psi(\bar{x})\|. \end{aligned}$$

Furthermore, the strong convexity of ψ yields that

$$\langle \nabla\psi(x) - \nabla\psi(x_0) + Mx - y_0, x - x_0\rangle \geq 2\alpha\|x - x_0\|^2,$$

for some $\alpha > 0$, and from Cauchy-Schwarz inequality we obtain that

$$\inf \|\nabla\psi(x) - \nabla\psi(x_0) + Mx - y_0\| \geq 2\alpha\|x - x_0\|, \quad \forall x \in \mathcal{H}.$$

Hence

$$\inf \|T(x)\| \geq 2\alpha\|x - x_0\| - \|\nabla\psi(x_0) + y_0 - \nabla\psi(\bar{x})\|, \quad \forall x \in \mathcal{H},$$

which implies

$$\lim_{\|x\|\to\infty} \|Tx\| = \infty,$$

and [1, Corollary 21.24] concludes the proof. ◻

3 The algorithm and its convergence

Considering the following primal-dual splitting algorithm

$$\begin{cases} x^{n+1} = (\nabla_x D_1(\cdot, x^n) + A)^{-1} \left(-C^* y^n\right), \\ \tilde{x}^{n+1} = 2x^{n+1} - x^n, \\ y^{n+1} = (\nabla_y D_2(\cdot, y^n) + \sigma B^{-1})^{-1} \left(C \tilde{x}^{n+1}\right). \end{cases} \tag{11}$$

This previous algorithm is an extension of [3, Algorithm 1], where the subdifferential maps are replaced by general maximally monotone maps. When

$$\psi_1(x) = \frac{\|x\|^2}{2\tau}, \quad \psi_2(y) = \frac{\|y\|^2}{2\sigma}, \quad x \in \mathcal{H}_1,\ y \in \mathcal{H}_2,$$

we obtain

$$D_1(x, \bar{x}) = \frac{\|x - \bar{x}\|^2}{2\tau}, \quad D_2(y, \bar{y}) = \frac{\|y - \bar{y}\|^2}{2\sigma},$$

and (11) reduces to (6). Moreover the existence of an $\alpha > 0$ such that (8) holds is equivalent to (5).

Furthermore, Lemma 1 guarantees that all steps in (11) are well defined, and the algorithm will not halt.

Theorem 1. *Assume that* (1) *admits a solution* $(x^*, y^*) \in \mathcal{H}_1 \times \mathcal{H}_2$*, and* $(x^n, \tilde{x}^n, y^n)$ *are generated by* (11) *with arbitrary initial points* $(x^0, \tilde{x}^0, y^0) \in \mathcal{H}_1 \times \mathcal{H}_1 \times \mathcal{H}_2$*. Then the ergodic sequence* $\{(X_N, Y_N)\}$ *defined in* (4) *is bounded, and all its weak limits are solutions of* (1).

Proof. We introduce the following function

$$\begin{aligned} \mathcal{L}(x, \zeta; y, \eta) &= \sup_{(u,v) \in Ax \times B^{-1}y} \langle x - \zeta, -u - C^*\eta \rangle + \langle C\zeta - v, y - \eta \rangle \\ &= \sup_{(u,v) \in Ax \times B^{-1}y} \langle \zeta - x, u \rangle + \langle \eta - y, v \rangle - \langle Cx, \eta \rangle + \langle C\zeta, y \rangle, \end{aligned} \tag{12}$$

where we set the supremum of an empty set to be $-\infty$. As pointed out in [3], the basic building block of (11) is the iteration

$$\begin{cases} \hat{x} = (\nabla_x D_1(\cdot, \bar{x}) + A)^{-1} \left(-C^* \tilde{y}\right), \\ \hat{y} = (\nabla_y D_2(\cdot, \bar{y}) + \sigma B^{-1})^{-1} \left(C \tilde{x}\right), \end{cases} \tag{13}$$

for suitable choices of $\bar{x}, \hat{x}, \tilde{x}$ and $\bar{y}, \hat{y}, \tilde{y}$. In an expanded form, (13) can be written as

$$\begin{cases} \nabla_x D_1(\hat{x}, \bar{x}) + \hat{u} = -C^* \tilde{y}, \\ \nabla_y D_2(\hat{y}, \bar{y}) + \hat{v} = C \tilde{x}, \end{cases} \tag{14}$$

where $(\hat{u}, \hat{v}) \in A\hat{x} \times B^{-1}\hat{y}$. Thus, we first obtain estimates for the general iteration (14) and then apply them to (11).

Let (14) hold, and $(x, y) \in \mathcal{H}_1 \times \mathcal{H}_2$, $(u, v) \in Ax \times B^{-1}y$ be arbitrary. Then by the monotonicity of A and (14) we have that

$$\begin{aligned} \langle u, x - \hat{x} \rangle \geq& \langle \hat{u}, x - \hat{x} \rangle = \langle -C^*\tilde{y} - \nabla_x D_1(\hat{x}, \bar{x}), x - \hat{x} \rangle \\ =& \langle -C^*\tilde{y}, x - \hat{x} \rangle + D_1(\hat{x}, \bar{x}) + D_1(x, \hat{x}) - D_1(x, \bar{x}), \end{aligned} \tag{15}$$

where we also used the identity

$$\langle -\nabla_x D_1(\hat{x}, \bar{x}), x - \hat{x} \rangle = D_1(\hat{x}, \bar{x}) + D_1(x, \hat{x}) - D_1(x, \bar{x}).$$

Similarly, using the monotonicity of B^{-1} we obtain

$$\begin{aligned} \langle v, y - \hat{y} \rangle \geq& \langle \hat{v}, y - \hat{y} \rangle = \langle C\tilde{x} - \nabla_x D_2(\hat{x}, \bar{x}), x - \hat{x} \rangle \\ =& \langle C\tilde{x}, y - \hat{y} \rangle + D_2(\hat{y}, \bar{y}) + D_2(y, \hat{y}) - D_2(y, \bar{y}). \end{aligned} \tag{16}$$

Combining (15), (16), we obtain

$$\begin{aligned}&D_1(x,\bar{x})-D_1(\hat{x},\bar{x})-D_1(x,\hat{x})+D_2(y,\bar{y})-D_2(\hat{y},\bar{y})-D_2(y,\hat{y})\\ \geq&\langle x-\hat{x},-u-C^*\tilde{y}\rangle+\langle C\tilde{x}-v,y-\hat{y}\rangle\\ =&\langle x-\hat{x},-u-C^*\hat{y}\rangle+\langle C\hat{x}-v,y-\hat{y}\rangle+\langle C(x-\hat{x}),\hat{y}-\tilde{y}\rangle+\langle C(\tilde{x}-\hat{x}),y-\hat{y}\rangle.\end{aligned}$$

Since $(u,v)\in Ax\times B^{-1}y$ are arbitrary, we obtain that

$$\begin{aligned}\mathcal{L}(x,\hat{x};y,\hat{y})\leq&D_1(x,\bar{x})-D_1(\hat{x},\bar{x})-D_1(x,\hat{x})+D_2(y,\bar{y})-D_2(\hat{y},\bar{y})-D_2(y,\hat{y})\\ &+\langle C(x-\hat{x}),\tilde{y}-\hat{y}\rangle+\langle C(\tilde{x}-\hat{x}),\hat{y}-y\rangle,\quad\forall x\in\mathcal{H}_1,\ y\in\mathcal{H}_2.\end{aligned}\tag{17}$$

As in [3], this previous inequality is the key inequality in the proof. Indeed, (11) corresponds to choosing

$$\hat{x}=x^{n+1},\ \bar{x}=x^n,\ \tilde{x}^{n+1}=2x^{n+1}-x^n,\ \hat{y}=y^{n+1},\ \bar{y}=y^n,\ \tilde{y}=y^n,$$

in (13), and so (17) yields

$$\begin{aligned}\mathcal{L}(x,x^{n+1};y,y^{n+1})\leq&\{D_1(x,x^n)+D_2(y,y^n)-\langle C(x-x^n),y-y^n\rangle\}\\ &-\left\{D_1(x,x^{n+1})+D_2(y,y^{n+1})-\langle C(x-x^{n+1}),y-y^{n+1}\rangle\right\}\\ &-\left\{D_1(x^{n+1},x^n)+D_2(y^{n+1},y^n)-\langle C(x^{n+1}-x^n),y^{n+1}-y^n\rangle\right\}.\end{aligned}$$

Hence, by the convexity of $(\zeta,\eta)\mapsto\mathcal{L}(x,\zeta;y,\eta)$, we obtain

$$\begin{aligned}N\mathcal{L}(x,X^N;y,Y^N)\leq&\sum_{n=1}^{N}\mathcal{L}(x,x^n;y,y^n)\\ \leq&\left\{D_1(x,x^0)+D_2(y,y^0)-\langle C(x-x^0),y-y^0\rangle\right\}\\ &-\left\{D_1(x,x^N)+D_2(y,y^N)-\langle C(x-x^N),y-y^N\rangle\right\}\\ &-\sum_{n=1}^{N}\left\{D_1(x^n,x^{n-1})+D_2(y^n,y^{n-1})-\langle C(x^n-x^{n-1}),y^n-y^{n-1}\rangle\right\},\end{aligned}\tag{18}$$

for all $x\in\mathcal{H}_1$, $y\in\mathcal{H}_2$, and $N\in\mathbb{N}$. Note that (8) guarantees that the expressions in the curly brackets are nonnegative.

Recall that (x^*,y^*) is a solution of (1), and so

$$-C^*y^*\in Ax^*,\quad Cx^*\in B^{-1}y^*.\tag{19}$$

But then by the definition of $\mathcal{L}$ we have that

$$\mathcal{L}(x^*,\zeta;y^*,\eta)\geq\langle x^*-\zeta,C^*y^*-C^*\eta\rangle+\langle C\zeta-Cx^*,y^*-\eta\rangle=0,\quad\forall\zeta\in\mathcal{H}_1,\ \forall\eta\in\mathcal{H}_2.$$

In particular, we have that

$$\mathcal{L}(x^*,X^N;y^*,Y^N)\geq0,\tag{20}$$

and (18) yields that

$$D_1(x^*,x^N)+D_2(y^*,y^N)-\langle C(x^*-x^N),y^*-y^N\rangle\leq D_1(x^*,x^0)+D_2(y^*,y^0)-\langle C(x^*-x^0),y-y^0\rangle,$$

and (8) implies that

$$\|x^N-x^*\|^2+\|y^N-y^*\|^2\leq\frac{D_1(x^*,x^0)+D_2(y^*,y^0)-\langle C(x^*-x^0),y-y^0\rangle}{\alpha},\quad\forall N\in\mathbb{N}.$$

Therefore, $\{(x^n,y^n)\}$ is a bounded sequence, and the convexity of the norm yields the boundedness of the ergodic sequence with the same bounds; that is,

$$\|X^N-x^*\|^2+\|Y^N-y^*\|^2\leq\frac{D_1(x^*,x^0)+D_2(y^*,y^0)-\langle C(x^*-x^0),y-y^0\rangle}{\alpha},\quad\forall N\in\mathbb{N}.$$

Assume that (X, Y) is a weak (subsequential) limit of $\{(X_N, Y_N)\}$. Invoking (18) again, we obtain

$$\mathcal{L}(x, X^N; y, Y^N) \leq \frac{D_1(x, x^0) + D_2(y, y^0) - \langle C(x - x^0), y - y^0\rangle}{N}, \tag{21}$$

for all $x \in \mathcal{H}_1$, $y \in \mathcal{H}_2$, and $N \in \mathbb{N}$. Let $(u, v) \in Ax \times B^{-1}y$ be arbitrary. Then we have that

$$\langle X^N - x, u\rangle + \langle Y^N - y, v\rangle - \langle Cx, Y^N\rangle + \langle CX^N, y\rangle \leq \mathcal{L}(x, X^N; y, Y^N),$$

and so the weak convergence and (21) yield

$$\begin{aligned}
&\langle X - x, u\rangle + \langle Y - y, v\rangle - \langle Cx, Y\rangle + \langle CX, y\rangle \\
&= \lim_{N\to\infty} \langle X^N - x, u\rangle + \langle Y^N - y, v\rangle - \langle Cx, Y^N\rangle + \langle CX^N, y\rangle \\
&\leq \liminf_{N\to\infty} \mathcal{L}(x, X^N; y, Y^N) \leq 0.
\end{aligned}$$

Therefore we have that

$$\mathcal{L}(x, X; y, Y) \leq 0, \quad \forall x \in \mathcal{H}_1,\ y \in \mathcal{H}_2. \tag{22}$$

Taking $y = Y$ in (22) we obtain

$$\langle x - X, u + C^*Y\rangle \geq 0, \quad \forall (x, u) \in \operatorname{gra} A,$$

and so maximal monotonicity of A yields that

$$(X, -C^*Y) \in \operatorname{gra} A \iff -C^*Y \in AX. \tag{23}$$

Similarly, plugging in $x = X$ in (22) we find that

$$\langle y - Y, v - CX\rangle \geq 0, \quad \forall (y, v) \in \operatorname{gra} B^{-1},$$

and the maximal monotonicity of B^{-1} yields that

$$(Y, CX) \in \operatorname{gra} B^{-1} \iff Y \in B(CX). \tag{24}$$

Combining (23) and (24) we obtain that (X, Y) is a solution of (1). □

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hédy Attouch.

[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[3] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Math. Program.*, 159(1-2):253–287, 2016.

[4] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004.

[5] R. T. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc.*, 149:75–88, 1970.

[6] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, 2013.