Conical averagedness and convergence analysis of fixed point algorithms

Sedi Bartz,∗ Minh N. Dao † and Hung M. Phan ‡

October 30, 2019

Abstract

In this paper, we consider a conical extension of averaged nonexpansive operators and its usefulness in the convergence analysis of fixed point algorithms. Various properties of conically averaged operators are systematically studied, in particular, the stability under relaxations, convex combinations, and compositions. We then derive the conical averagedness of the resolvents of operators that possess certain types of generalized monotonicity. These properties are used to analyze the convergence of fixed point algorithms including the proximal point algorithm, the forward-backward algorithm, and the adaptive Douglas–Rachford algorithm. Our results not only unify and improve recent studies but also provide a different perspective on the topic.

AMS Subject Classifications: Primary: 47H10, 49M27; Secondary: 65K05, 65K10.

Keywords: Adaptive Douglas–Rachford algorithm, cocoercivity, conically averaged operator, forward-backward algorithm, proximal point algorithm, strong monotonicity, weak monotonicity.

1. Introduction

Averaged nonexpansive operators were originally introduced by [1] and they are well known to be useful in convergence theories of various fixed point algorithms, see [2, 4, 6, 8, 20] and the references therein. Particularly, iterative sequences generated by many fixed point algorithms can be expressed in terms of Krasnosel’skiī–Mann iterations whose convergence relies on averagedness property. Although frequently understood in the single-valued setting, some averaged nonexpansive structure can also be explored in the set-valued framework. For instance, the notion of union averaged nonexpansive operators has been recently studied in [11] with applications to the local convergence of proximal algorithms.

Apparently, each averaged operator is an under-relaxation of some nonexpansive operator. As we will see later on, over-relaxations of nonexpansive operators also arise in several situations. In this paper, we consider the conically averaged operators that unify both types of relaxations for nonexpansive operators. We then show that this class of operators plays a significant role in

∗Department of Mathematical Sciences, Kennedy College of Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA. E-mail: sedi_bartz@uml.edu
†Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: daonminh@gmail.com
‡Department of Mathematical Sciences, Kennedy College of Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA. E-mail: hung_phan@uml.edu.
the convergence of several fixed point algorithms, in particular, the proximal point algorithm, the forward–backward algorithm, and the adaptive Douglas–Rachford (DR) algorithm.

The proximal point algorithm was first introduced by Martinet [18] and further studied by Rockafellar [22] for finding a zero of a maximally monotone operator. Since then, it has become an indispensable component of optimization theory and applications. In fact, several other iterative optimization algorithms can be reformulated as special cases of the proximal point algorithm, see [13] and the references therein. The forward-backward algorithm, on the other hand, was first proposed by Lions and Mercier [17] and Passty [21] for finding a zero of the sum of two maximally monotone operators. Indeed, this splitting idea can be traced back to the projected gradient method [15]. Another famous splitting algorithm is the DR algorithm, which was initially studied by Douglas and Rachford [12] in the setting of linear operators, and was later generalized for maximally monotone operators by Lions and Mercier also in [17]. It is worth to mention that both forward-backward and DR algorithms reduce to the proximal point algorithm when one operator is zero. Recently, the so-called adaptive DR algorithm has been proposed in [9] to deal with the lack of the classical monotonicity assumption.

The main contributions of the paper are two folds. On the one hand, we systematically study the conically averagedness property, which is stable under relaxations, convex combinations, and compositions. It is also showed that the relaxed resolvents of generalized monotone operators either belong to or directly link to the class of conically averaged operators. On the other hand, we analyze the convergence of all the aforementioned algorithms by means of the conical averagedness. This approach not only provides simple convergence proofs but also brings more flexibility for the algorithm parameters. In particular, we prove the global convergence and the rate of asymptotic regularity of the relaxed proximal point algorithm, the relaxed forward-backward algorithm, and the adaptive DR algorithm when one operator is no longer monotone. An application to the strongly and weakly convex optimization problem is also discussed. Our analysis also improves several contemporary results on this topic.

The remainder of the paper is organized as follows. Section 2 presents the conically averaged operators with numerous interesting properties that are beneficial to our analysis. In Section 3, we study the relaxed resolvents of generalized monotone operators in connection with conical averagedness properties. Based on these developments, in Sections 4 and 5, we provide our main results on the averagedness of operators associated with those algorithms, which leads to the convergence and the rate of asymptotic regularity. Finally, we warp up the paper in Section 6.

The notations in the paper are fairly standard and follow largely [2]. Throughout, $X$ is a real Hilbert space with inner product $\langle \cdot , \cdot \rangle$ and induced norm $\| \cdot \|$. The set of nonnegative integers is denoted by $\mathbb{N}$, the set of real numbers by $\mathbb{R}$, the set of nonnegative real numbers by $\mathbb{R}_+ := \{ x \in \mathbb{R} \mid x \geq 0 \}$, and the set of positive real numbers by $\mathbb{R}_{++} := \{ x \in \mathbb{R} \mid x > 0 \}$. We use the notation $A : X \rightrightarrows X$ to indicate that $A$ is a set-valued operator on $X$ and the notation $A : X \rightarrow X$ to indicate that $A$ is a single-valued operator on $X$. Given an operator $A$ on $X$, its domain is denoted by $\text{dom} A := \{ x \in X \mid Ax \neq \emptyset \}$, its set of zeros by $\text{zer} A := \{ x \in X \mid 0 \in Ax \}$, and its set of fixed points by $\text{Fix} A := \{ x \in X \mid x \in Ax \}$. As usual, $\text{Id}$ denotes the identity operator.

We will frequently use the following identity, for every $\sigma, \tau \in \mathbb{R}$ and $s, t \in X$,

$$
\| \sigma s + \tau t \|^2 = \sigma (\sigma + \tau) \| s \|^2 + \tau (\sigma + \tau) \| t \|^2 - \sigma \tau \| s - t \|^2
$$

(1)

and if $\sigma + \tau \neq 0$, then

$$
\sigma \| s \|^2 + \tau \| t \|^2 = \frac{1}{\sigma + \tau} \| \sigma s + \tau t \|^2 + \frac{\sigma \tau}{\sigma + \tau} \| s - t \|^2.
$$

(2)
2. Conically averaged operators

We recall that $T: X \to X$ is nonexpansive if it is Lipschitz continuous with constant 1 on its domain, i.e.,

$$\forall x, y \in \text{dom } T, \quad \|Tx - Ty\| \leq \|x - y\|. \tag{3}$$

The operator $T$ is said to be $\theta$-averaged if $\theta \in [0, 1]$ and $T = (1 - \theta) \text{Id} + \theta N$ for some nonexpansive operator $N: X \to X$, see, e.g., [2, Definition 4.33]. We now extend this concept to allow for $\theta \in \mathbb{R}^+$.  

**Definition 2.1 (conically averaged operator).** We say that an operator $T: X \to X$ is conically averaged with constant $\theta \in \mathbb{R}^+$, or conically $\theta$-averaged, if there exists a nonexpansive operator $N: X \to X$ such that

$$T = (1 - \theta) \text{Id} + \theta N. \tag{4}$$

Let $T$ be conically $\theta$-averaged. Then $T$ is nonexpansive when $\theta = 1$, and $T$ is $\theta$-averaged when $\theta \in ]0, 1[$. As one would expect, conically averaged operators also possess properties similar to averaged operators. Indeed, we now present numerous properties which generalize and extend the corresponding ones in [2, Chapter 4], see also [3] for a related development where conically averaged operators were referred to as conically nonexpansive operators.

**Proposition 2.2.** Let $T: X \to X$, $\theta \in \mathbb{R}^+$, and $\lambda \in \mathbb{R}^+$. Then the following are equivalent:

(i) $T$ is conically $\theta$-averaged.

(ii) $(1 - \lambda) \text{Id} + \lambda T$ is conically $\lambda \theta$-averaged.

(iii) For all $x, y \in \text{dom } T$,

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{1 - \theta}{\theta} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2. \tag{5}$$

(iv) For all $x, y \in \text{dom } T$,

$$\|Tx - Ty\|^2 + (1 - 2\theta)\|x - y\|^2 \leq 2(1 - \theta) \langle x - y, Tx - Ty \rangle. \tag{6}$$

**Proof.** Set $N := (1 - 1/\theta) \text{Id} + (1/\theta)T$. Then $T = (1 - \theta) \text{Id} + \theta N$ and $(1 - \lambda) \text{Id} + \lambda T = (1 - \lambda \theta) \text{Id} + \lambda \theta N$. By definition,

$$T \text{ is conically } \theta \text{-averaged} \quad \iff \quad N \text{ is nonexpansive} \quad \tag{7a}$$

$$\iff \quad (1 - \lambda) \text{Id} + \lambda T \text{ is conically } \lambda \theta \text{-averaged}, \tag{7c}$$

which implies the equivalence between (i) and (ii).

Next, we note that $\text{Id} - N = (\text{Id} - T)/\theta$ and $\text{dom } T = \text{dom } N =: D$. Now using (1), we have for all $x, y \in D$ that

$$\|Tx - Ty\|^2 = \|(1 - \theta)(x - y) + \theta(Nx - Ny)\|^2$$

$$= (1 - \theta)\|x - y\|^2 + \theta\|Nx - Ny\|^2$$

$$- \theta(1 - \theta)\|(\text{Id} - N)x - (\text{Id} - N)y\|^2$$

$$= \|x - y\|^2 - \frac{1 - \theta}{\theta} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2$$

$$+ \theta(\|Nx - Ny\|^2 - \|x - y\|^2). \tag{8e}$$
Therefore,

\[ T \text{ is conically } \theta\text{-averaged} \quad (9a) \]
\[ \iff N \text{ is nonexpansive} \quad (9b) \]
\[ \iff \forall x, y \in D, \ |N x - N y| \leq |x - y| \quad (9c) \]
\[ \iff \forall x, y \in D, \ |T x - T y|^2 \leq |x - y|^2 - \frac{1 - \theta}{\theta} |(\text{Id} - T)x - (\text{Id} - T)y|^2 \quad (9d) \]
\[ \iff \forall x, y \in D, \ |T x - T y|^2 + (1 - 2\theta)|x - y|^2 \leq 2(1 - \theta) (x - y, T x - T y), \quad (9e) \]

and we get the equivalence between (i), (iii), and (iv). The proof is complete. \hfill \blacksquare

In view of Proposition 2.2, \( T \) is 1/2-averaged if and only if, for all \( x, y \in \text{dom} \, T \),

\[ |T x - T y|^2 \leq |x - y|^2 - |(\text{Id} - T)x - (\text{Id} - T)y|^2, \quad (10) \]

in which case, \( T \) is also said to be firmly nonexpansive.

**Proposition 2.3.** Let \( T : X \to X \) and \( \lambda \in \mathbb{R}_{++} \). Then \( T \) is firmly nonexpansive if and only if \( \text{Id} - \lambda T \) is conically \( \lambda/2 \)-averaged.

**Proof.** Set \( N := 2T - \text{Id} \). Then, by definition, \( T = (1/2) \text{Id} + (1/2)N \) is firmly nonexpansive \( \iff \)
\( N \) is nonexpansive \( \iff \text{Id} - \lambda T = (1 - \lambda/2) \text{Id} + (\lambda/2)(-N) \) is conically \( \lambda/2 \)-averaged. \hfill \blacksquare

The following results reiterate and extend the corresponding results for nonexpansive operators and averaged operators.

**Proposition 2.4 (convex combination).** Let \( I \) be a finite index set, let \( T_i : X \to X \) be conically \( \theta_i \)-averaged for each \( i \in I \), and let \( \{\omega_i\}_{i \in I} \subseteq \mathbb{R}_{++} \) be such that \( \sum_{i \in I} \omega_i = 1 \). Then \( \sum_{i \in I} \omega_i T_i \) is conically \( \theta \)-averaged with \( \theta := \sum_{i \in I} \omega_i \theta_i \).

**Proof.** For each \( i \in I \), there exists a nonexpansive operator \( N_i \) such that \( T_i = (1 - \theta_i) \text{Id} + \theta_i N_i \). It follows that

\[ \sum_{i \in I} \omega_i T_i = \sum_{i \in I} \omega_i(1 - \theta_i) \text{Id} + \sum_{i \in I} \omega_i \theta_i N_i = (1 - \theta) \text{Id} + \theta \sum_{i \in I} \omega_i \theta_i N_i. \quad (11) \]

As \( \sum_{i \in I} \frac{\omega_i \theta_i}{\theta} N_i \) is a nonexpansive operator due to [2, Proposition 4.9(i)], the proof is complete. \hfill \blacksquare

**Proposition 2.5 (composition of two conically averaged operators).** Let \( T_1 : X \to X \) and \( T_2 : X \to X \) be respectively conically \( \theta_1 \)- and \( \theta_2 \)-averaged. Suppose that either \( \theta_1 = \theta_2 = 1 \) or \( \theta_1 \theta_2 < 1 \). Let \( \omega \in \mathbb{R} \setminus \{0\} \) and set

\[ T := \left( \frac{1}{\omega} T_2 \right) (\omega T_1) \quad \text{and} \quad \theta := \begin{cases} 1 & \text{if } \theta_1 = \theta_2 = 1, \\ \frac{\theta_1 + \theta_2 - 2\theta_1 \theta_2}{1 - \theta_1 \theta_2} & \text{if } \theta_1 \theta_2 < 1. \end{cases} \quad (12) \]

Then the following hold:

(i) \( T \) is conically \( \theta \)-averaged, where \( \theta = 1 \) if and only if \( \theta_1 = 1 \) or \( \theta_2 = 1 \).

(ii) \( T \) is nonexpansive if and only if \( T_1 \) and \( T_2 \) are nonexpansive.

(iii) \( T \) is averaged if and only if \( T_1 \) and \( T_2 \) are averaged.
Proof. (i): Let \( x, y \in \text{dom} T_2(\omega T_1) \). Applying Proposition 2.2 to \( T_2 \) and then to \( T_1 \), we have

\[
\left\| \left( \frac{1}{\omega} T_2 \right)(\omega T_1)x - \left( \frac{1}{\omega} T_2 \right)(\omega T_1)y \right\| \leq \frac{1}{\omega^2} \left( \left\| T_2(\omega T_1)x - T_2(\omega T_1)y \right\| \right)^2
\]

(ii): Setting \( \theta \) combining with (i) and (ii) yields \( \theta \)

\[
\text{Case 2: } \theta = 1. \text{ Then, (13) immediately implies that } \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \text{ is nonexpansive, i.e., conically 1-averaged.}
\]

Case 2: \( \theta_1 \theta_2 < 1 \). By the inequality of arithmetic and geometric means (AM-GM inequality),

\[
\frac{1}{\theta_1} + \frac{1}{\theta_2} \geq \frac{2}{\sqrt{\theta_1 \theta_2}} > 2,
\]

which yields

\[
\frac{1 - \theta_1}{\theta_1} + \frac{1 - \theta_2}{\theta_2} > 0.
\]

Setting \( s := (\text{Id} - T_1)x - (\text{Id} - T_1)y \) and \( t := \left( T_1 - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)x - \left( T_1 - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)y \), we have from (2) and (15) that

\[
\frac{1 - \theta_1}{\theta_1} \left\| s \right\|^2 + \frac{1 - \theta_2}{\theta_2} \left\| t \right\|^2 = \frac{1 - \theta_1}{\theta_1} \left\| s + \frac{\theta_1 - \theta_2}{\theta_1 + \theta_2} t \right\|^2 + \frac{1 - \theta_2}{\theta_2} \left\| \frac{\theta_1 - \theta_2}{\theta_1 + \theta_2} t + \frac{1}{\theta_1} (s - \frac{\theta_1 - \theta_2}{\theta_1 + \theta_2} t) \right\|^2 \]

\[
\geq \frac{1 - \theta_1}{\theta_1} \left\| \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)x - \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)y \right\|^2 \]

\[
= \frac{(1 - \theta_1)(1 - \theta_2)}{\theta_1 + \theta_2 - 2\theta_1 \theta_2} \left\| \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)x - \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)y \right\|^2 \]

\[
= \frac{1 - \theta}{\theta} \left\| \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)x - \left( \text{Id} - \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \right)y \right\|^2,
\]

where

\[
\theta := \frac{\theta_1 + \theta_2 - 2\theta_1 \theta_2}{1 - \theta_1 \theta_2} > 0.
\]

Combining with (13) and Proposition 2.2(i) and (iii), we deduce that \( \left( \frac{1}{\omega} T_2 \right)(\omega T_1) \) is conically \( \theta \) averaged. In this case, we observe that

\[
\theta = 1 \iff \theta_1 + \theta_2 - 2\theta_1 \theta_2 = 1 - \theta_1 \theta_2 \iff (1 - \theta_1)(1 - \theta_2) = 0.
\]

Thus, in both cases, \( \theta = 1 \) if and only if \( \theta_1 = 1 \) or \( \theta_2 = 1 \).

(ii): It follows from the definition of \( \theta \) that

\[
\theta > 1 \iff \theta_1 \theta_2 < 1 \text{ and } \theta_1 + \theta_2 - 2\theta_1 \theta_2 > 1 - \theta_1 \theta_2 \quad (19a)
\]

\[
\iff \theta_1 \theta_2 < 1 \text{ and } (1 - \theta_1)(1 - \theta_2) > 0 \quad (19b)
\]

\[
\iff \theta_1 \theta_2 < 1 \text{ and } (\theta_1 > 1 \text{ or } \theta_2 > 1). \quad (19c)
\]

5
As a result, \( \theta \leq 1 \) if and only if \( \theta_1 \leq 1 \) and \( \theta_2 \leq 1 \). The conclusion then follows.

(iii): This follows from (i) and (ii). \( \blacksquare \)

For convenience in later applications, we derive an equivalent presentation of Proposition 2.5(i). This result generalizes [2, Proposition 4.44] and interestingly, also recaptures [14, Proposition 3.12], in which the composition of an averaged operator and a so-called negatively averaged operator was considered.

**Proposition 2.6.** Let \( T_1, T_2 : X \to X, \omega_1, \omega_2 \in \mathbb{R} \setminus \{0\} \), and \( \theta_1, \theta_2 \in \mathbb{R}_{++} \). Suppose that \( \omega_1 T_1 \) and \( \omega_2 T_2 \) are respectively conically \( \theta_1 \)- and \( \theta_2 \)-averaged such that either \( \theta_1 = 1 \) or \( \theta_1 \theta_2 < 1 \). Then \( \omega_1 \omega_2 T_2 T_1 \) and \( \omega_1 \omega_2 T_1 T_2 \) are both conically \( \theta \)-averaged, where

\[
\theta := \begin{cases} 
1 & \text{if } \theta_1 = 1, \\
\frac{\theta_1 + \theta_2 - 2 \theta_1 \theta_2}{1 - \theta_1 \theta_2} & \text{if } \theta_1 \theta_2 < 1.
\end{cases}
\] (20)

**Proof.** Note that \( \omega_1 \omega_2 T_2 T_1 = \omega_1 (\omega_2 T_2) \left( \frac{1}{\omega_1} (\omega_1 T_1) \right) \) and \( \omega_1 \omega_2 T_1 T_2 = \omega_2 (\omega_1 T_1) \left( \frac{1}{\omega_2} (\omega_2 T_2) \right) \), and then apply Proposition 2.5. \( \blacksquare \)

**Proposition 2.7 (finite composition of conically averaged operators).** Let \( m \geq 2 \) be an integer, let \( T_i : X \to X \) be conically \( \theta_i \)-averaged for each \( i \in I := \{1, \ldots, m\} \), and let \( \{\omega_i\}_{i \in I} \subseteq \mathbb{R} \) be such that \( \omega_1 \omega_2 \cdots \omega_m = 1 \). Set

\[
T := (\omega_m T_m) (\omega_{m-1} T_{m-1}) \cdots (\omega_1 T_1).
\] (21)

Then the following hold:

(i) If \( \max_{i \in I} \theta_i \leq 1 \), then \( T \) is nonexpansive.

(ii) If \( \theta_i \neq 1 \) for each \( i \in I \) and

\[
\forall k \in \{2, \ldots, m\}, \quad \theta_k < 1 + \frac{1}{\sum_{i=1}^{k-1} \frac{\theta_i}{1-\theta_i}}.
\] (22)

then \( T \) is conically \( \theta \)-averaged with

\[
\theta := \frac{1}{1 + \sum_{i \in I} \frac{1}{1-\theta_i}}.
\] (23)

(iii) If \( \max_{i \in I} \theta_i < 1 \), then \( T \) is \( \theta \)-averaged with \( \theta < 1 \) given by (23).

**Proof.** (i): The proof is straightforward by using Proposition 2.5 repeatedly.

(ii): We will prove by induction on \( m \). For \( m = 2 \), the conclusion is straightforward by Proposition 2.5. Suppose that the statement is true for \( m - 1 \). Let \( T_i \) be conically \( \theta_i \)-averaged for \( i \in \{1, \ldots, m\} \) such that

\[
\forall k \in \{2, \ldots, m\}, \quad \theta_k < 1 + \frac{1}{\sum_{i=1}^{k-1} \frac{\theta_i}{1-\theta_i}}.
\] (24)

Let also \( \omega_i \in \mathbb{R} \) satisfy \( \omega_1 \omega_2 \cdots \omega_m = 1 \). We will show the operator

\[
T = (\omega_m T_m) (\omega_{m-1} T_{m-1}) \cdots (\omega_1 T_1)
\] (25)
is conically $\theta$-averaged with $\theta$ given by (23).

Since the statement is true for $m - 1$ conically averaged operators, we have

$$T^* := (\omega_m \omega_{m-1} T_{m-1})(\omega_{m-2} T_{m-2}) \cdots (\omega_1 T_1)$$

is $\theta^*$-averaged with

$$\theta^* := \frac{1}{1 + \sum_{i=1}^{m-1} \frac{\theta_i}{1-\theta_i}}.$$  

Using (24) with $k = m$, we have

$$\theta_m < \frac{1}{\theta^*}.$$  

Now apply Proposition 2.5 to two operators $T^*$ and $T_m$, we have $T = (\omega_m T_m)(\frac{1}{\omega_m} T^*)$ is conically $\theta_0$-averaged with

$$\theta_0 := \frac{\theta^* + \theta_m - 2\theta^* \theta_m}{1 - \theta^* \theta_m}.$$  

Then

$$\frac{\theta_0}{1 - \theta_0} = \frac{\theta^*}{1 - \theta^*} + \frac{\theta_m}{1 - \theta_m} = \sum_{i=1}^{m-1} \frac{\theta_i}{1 - \theta_i} + \frac{\theta_m}{1 - \theta_m} = \sum_{i=1}^{m} \frac{\theta_i}{1 - \theta_i}.$$  

It follows that

$$\theta_0 = \frac{1}{1 + \sum_{i=1}^{m} \frac{\theta_i}{1 - \theta_i}} = \theta.$$  

So, the statement is true for $m$. By the mathematical induction principle, the statement is true for all $m \geq 2$.

(iii): For each $i \in I$, since $\theta_i < 1$, we have $\frac{\theta_i}{1-\theta_i} > 0$. Therefore,

$$\forall k \in \{2, \ldots, m\}, \quad \theta_k < 1 < 1 + \frac{1}{\sum_{i=1}^{k-1} \frac{\theta_i}{1-\theta_i}},$$  

which fulfills (22). The conclusion then follows by applying (ii).

Remark 2.8. On the one hand, Proposition 2.7 indeed recovers [2, Propositions 4.9(i) and 4.46]. On the other hand, it expands these results beyond the usual composition of conically averaged operators by applying scalar multiplications on each of these operators, see (21).

We conclude this section by establishing the convergence of sequences generated by an averaged/conically averaged operator using Fejér monotonicity. Recall that a sequence $(x_n)_{n \in \mathbb{N}}$ is said to be Fejér monotone with respect to a nonempty subset of $C$ of $X$ if

$$\forall c \in C, \forall n \in \mathbb{N}, \quad \|x_{n+1} - c\| \leq \|x_n - c\|.$$  

The following result whose proof is included for completeness is a slight extension of [2, Theorem 5.15].

Proposition 2.9 (Krasnosel'skiǐ–Mann iterations). Let $T$ be a conically $\theta$-averaged operator with full domain and $\text{Fix} T \neq \emptyset$. Given $x_0 \in X$, define

$$\forall n \in \mathbb{N}, \quad x_{n+1} = (1 - \lambda_n)x_n + \lambda_n T x_n,$$  

where $(\lambda_n)_{n \in \mathbb{N}}$ is a sequence in $[0, 1/\theta]$ such that $\sum_{n=0}^{+\infty} \lambda_n(1 - \theta \lambda_n) = +\infty$. Then the following hold:

7
(i) \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(\text{Fix} \ T\).

(ii) \((x_n - Tx_n)_{n \in \mathbb{N}}\) converges strongly to 0.

(iii) \((x_n)_{n \in \mathbb{N}}\) converges weakly to a point in \(\text{Fix} \ T\).

(iv) If \(\liminf_{n \to +\infty} \lambda_n (1 - \theta \lambda_n) > 0\), then \(\|x_n - Tx_n\| = o(1/\sqrt{n})\) as \(n \to +\infty\).

**Proof.** By definition, \(T = (1 - \theta)\text{Id} + \theta N\) for some nonexpansive operator \(N: X \to X\). Then \(\text{Id} - T = \theta(\text{Id} - N)\), \(\text{Fix} \ T = \text{Fix} \ N\), and

\[
\forall n \in \mathbb{N}, \quad x_{n+1} = (1 - \theta \lambda_n)x_n + \theta \lambda_n N x_n. \tag{35}
\]

(i): For all \(y \in \text{Fix} \ N\) and all \(n \in \mathbb{N}\), using (1) yields

\[
\|x_{n+1} - y\|^2 = \|(1 - \theta \lambda_n)(x_n - y) + \lambda_n (N x_n - y)\|^2 \\
= (1 - \lambda_n)\|x_n - y\|^2 + \lambda_n \|N x_n - y\|^2 - \lambda_n (1 - \theta \lambda_n)\|x_n - N x_n\|^2 \\
\leq \|x_n - y\|^2 - \lambda_n (1 - \theta \lambda_n)\|x_n - N x_n\|^2 	ag{36c}
\]

where we have used the nonexpansiveness of \(N\). Since \(\lambda_n (1 - \theta \lambda_n) \geq 0\) for all \(n\), \((x_n)_{n \in \mathbb{N}}\) is Fejér monotone with respect to \(\text{Fix} \ N = \text{Fix} \ T\).

(ii): By telescoping (36) over \(n \in \mathbb{N}\),

\[
\theta \sum_{n=0}^{+\infty} \lambda_n (1 - \theta \lambda_n)\|x_n - N x_n\|^2 \leq \|x_0 - y\|^2 < +\infty. \tag{37}
\]

Since \(\sum_{n=0}^{+\infty} \lambda_n (1 - \theta \lambda_n) = +\infty\), it follows that

\[
\liminf_{n \to +\infty} \|x_n - N x_n\| = 0. \tag{38}
\]

Moreover, for all \(n \in \mathbb{N}\), since \(x_{n+1} = (1 - \theta \lambda_n)x_n + \theta \lambda_n N x_n\) and since \(N\) is nonexpansive,

\[
\|x_{n+1} - N x_{n+1}\| = \|(1 - \theta \lambda_n)(x_n - N x_n) + (N x_n - N x_{n+1})\| \\
\leq (1 - \theta \lambda_n)\|x_n - N x_n\| + \|N x_n - N x_{n+1}\| \tag{39c}
\]

We deduce that \(\|x_n - N x_n\|\) is decreasing and bounded below by 0, hence it converges. Combining with (38) gives

\[
x_n - Tx_n = \theta(x_n - N x_n) \to 0 \quad \text{as} \quad n \to +\infty. \tag{40}
\]

(iii): Let \(x^*\) be a weak cluster point of \((x_n)_{n \in \mathbb{N}}\). Then there exists a subsequence \((x_{k_n})_{n \in \mathbb{N}}\) such that \(x_{k_n} \to x^*\). By (40) and [2, Corollary 4.28], \(x^* \in \text{Fix} \ N\). In turn, [2, Theorem 5.5] implies that \((x_n)_{n \in \mathbb{N}}\) converges weakly to a point in \(\text{Fix} N = \text{Fix} T\).

(iv): It follows from \(\liminf_{n \to +\infty} \lambda_n (1 - \theta \lambda_n) > 0\) and (37) that \(\sum_{n=0}^{+\infty} \|x_n - N x_n\|^2 < +\infty\), which combined with (39) yields

\[
\frac{n}{2}\|x_n - N x_n\|^2 \leq \sum_{k=[n/2]}^{n} \|x_k - N x_k\|^2 \to 0 \quad \text{as} \quad n \to +\infty, \tag{41}
\]

where \([n/2]\) is the largest integer not exceeding \(n/2\). Therefore, \(\|x_n - Tx_n\| = \theta\|x_n - N x_n\| = o(1/\sqrt{n})\) as \(n \to +\infty\). 

\[\square\]
Corollary 2.10 (convergence of averaged operators). Let $T$ be a $\theta$-averaged operator with full domain and $\text{Fix } T \neq \emptyset$. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence generated by $T$. Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } T$ and the rate of asymptotic regularity of $T$ is $o(1/\sqrt{n})$, i.e., $\|x_n - Tx_n\| = o(1/\sqrt{n})$ as $n \to +\infty$.

Proof. Apply Proposition 2.9 with $\theta < 1$ and all $\lambda_n = 1$. 

3. Generalized monotonicity

Let $A : X \rightrightarrows X$. The graph of $A$ is $\text{gra } A := \{(x, u) \in X \times X \mid u \in Ax\}$ and the inverse of $A$, denoted by $A^{-1}$, is the operator with graph $\text{gra } A^{-1} := \{(u, x) \in X \times X \mid u \in Ax\}$. The resolvent of $A$ is defined by

$$J_A := (\text{Id} + A)^{-1}$$

and the relaxed resolvent of $A$ with parameter $\lambda \in \mathbb{R}_+$ is defined by

$$J^\lambda_A := (1 - \lambda) \text{Id} + \lambda J_A.$$

Note that $R_A := 2J_A - \text{Id}$ is the reflected resolvent of $A$.

For $\alpha \in \mathbb{R}$, recall that $A$ is $\alpha$-monotone [9] if

$$\forall (x, u), (y, v) \in \text{gra } A, \quad \langle x - y, u - v \rangle \geq \alpha \|x - y\|^2$$

and $\alpha$-comonotone [3] if

$$\forall (x, u), (y, v) \in \text{gra } A, \quad \langle x - y, u - v \rangle \geq \alpha \|u - v\|^2.$$

We say that $A$ is maximally $\alpha$-monotone (respectively, maximally $\alpha$-comonotone) if it is $\alpha$-monotone (respectively, $\alpha$-monotone) and there is no $\alpha$-monotone (respectively, $\alpha$-monotone) operator $B : X \rightrightarrows X$ such that $\text{gra } B$ strictly contains $\text{gra } A$.

Notice that both 0-monotonicity and 0-comonotonicity simply mean monotonicity. If $\alpha > 0$, then $\alpha$-monotonicity is actually $\alpha$-strong monotonicity in [2, Definition 22.1(iv)], while $\alpha$-comonotonicity coincides with $\alpha$-cocoercivity in [2, Definition 4.10]. If $\alpha < 0$, then $\alpha$-monotonicity and $\alpha$-comonotonicity are respectively $\alpha$-hypomonotonicity and $\alpha$-cohypomonotonicity as [7, Definition 2.2]. Additionally, $\alpha$-monotonicity with $\alpha < 0$ is also referred to as weak monotonicity in [9]. We refer the reader to [2, 5] for more discussions on maximal monotonicity and its variants.

Remark 3.1. Several properties are immediate from the definition.

(i) $A$ is $\alpha$-comonotone if and only if $A^{-1}$ is $\alpha$-monotone.

(ii) $A$ is maximally $\alpha$-comonotone if and only if $A^{-1}$ is maximally $\alpha$-monotone.

(iii) If $A$ is $\alpha$-comonotone with $\alpha \geq 0$, then $A$ is also monotone.

(iv) If $A$ is $\alpha$-comonotone with $\alpha > 0$, then $A$ is single-valued and Lipschitz continuous with constant $1/\alpha$.

When $\alpha$ is nonnegative, further characterizations can be derived for maximal $\alpha$-monotonicity and maximal $\alpha$-comonotonicity.
Proposition 3.2 (maximal \(\alpha\)-monotonicity and \(\alpha\)-comonotonicity). Let \(A : X \rightrightarrows X\) and \(\alpha \in \mathbb{R}_+\). Then the following hold:

(i) \(A\) is maximally \(\alpha\)-monotone if and only if \(A\) is \(\alpha\)-monotone and maximally monotone.

(ii) \(A\) is maximally \(\alpha\)-comonotone if and only if \(A\) is \(\alpha\)-comonotone and maximally monotone.

Proof. (i): See [9, Proposition 3.5(i)].

(ii): \(A\) is maximally \(\alpha\)-comonotone \iff \(A^{-1}\) is maximally \(\alpha\)-monotone \iff \(A^{-1}\) is \(\alpha\)-monotone and maximally monotone (by (i)) \iff \(A\) is \(\alpha\)-comonotone and maximally monotone. ■

We now collect important and useful properties of relaxed resolvents of \(\alpha\)-monotone and \(\alpha\)-comonotone operators. Although part of the results appeared in [2, 3, 9, 10], we include here for the readers’ convenience. In particular, we will show that if an operator is either \(\alpha\)-monotone or \(\alpha\)-comonotone, then its relaxed resolvents are, to some extent, related to conically averaged operators. These results are crucial in the convergence analysis of several iterative algorithms that make use of the resolvents. We begin with some auxiliary properties.

Proposition 3.3. Let \(A : X \rightarrow X\) and \(\alpha, \lambda \in \mathbb{R}^{++}\). Then the following are equivalent:

(i) \(A\) is \(\alpha\)-comonotone (i.e., \(\alpha\)-coercive).

(ii) \(\lambda A\) is firmly nonexpansive.

(iii) \(\text{Id} - \lambda A\) is conically \(\frac{\lambda}{2\alpha}\)-averaged.

Proof. The equivalence between (i) and (ii) follows, e.g., from [2, Remark 4.34(iv)]. The equivalence between (ii) and (iii) follows from Proposition 2.3 by noting that \(\text{Id} - \lambda A = \text{Id} - \frac{\lambda}{\alpha}(\alpha A)\). ■

Proposition 3.4 (single-valuedness and full domain). Let \(A : X \rightrightarrows X\) be \(\alpha\)-monotone and let \(\gamma \in \mathbb{R}^{++}\) such that \(1 + \gamma \alpha > 0\). Then the following hold:

(i) \(J_{\gamma A}\) is single-valued.

(ii) \(\text{dom} J_{\gamma A} = X\) if and only if \(A\) is maximally \(\alpha\)-monotone.

Proof. See [9, Proposition 3.4]. ■

Proposition 3.5 (relaxed resolvents of \(\alpha\)-monotone operators). Let \(A : X \rightrightarrows X\) be \(\alpha\)-monotone and let \(\gamma \in \mathbb{R}^{++}\) be such that \(1 + \gamma \alpha > 0\). Set \(R := (1 - \lambda)\text{Id} + \lambda J_{\gamma A}\) with \(\lambda \in ]1, +\infty[\). Then the following hold:

(i) \(J_{\gamma A}\) is \((1 + \gamma \alpha)\)-comonotone. Consequently, \((1 + \gamma \alpha)J_{\gamma A}\) is firmly nonexpansive.

(ii) \(\frac{1}{1 - \lambda} R\) is conically \(\frac{\lambda}{2(\lambda - 1)(1 + \gamma \alpha)}\)-averaged.

Proof. (i): By Proposition 3.4(i), \(J_{\gamma A}\) is single-valued, and by [9, Lemma 3.3], it is \((1 + \gamma \alpha)\)-cocoercive, i.e., \((1 + \gamma \alpha)\)-comonotone. Since \(1 + \gamma \alpha > 0\), we have from Proposition 3.3 that \((1 + \gamma \alpha)J_{\gamma A}\) is firmly nonexpansive.

(ii): As \((1 + \gamma \alpha)J_{\gamma A}\) is firmly nonexpansive, Proposition 2.3 implies that

\[
\frac{1}{1 - \lambda} R = \text{Id} + \frac{\lambda}{1 - \lambda} J_{\gamma A} = \text{Id} - \frac{\lambda}{(\lambda - 1)(1 + \gamma \alpha)}(1 + \gamma \alpha)J_{\gamma A}
\] (46)

is conically \(\frac{\lambda}{2(\lambda - 1)(1 + \gamma \alpha)}\)-averaged. ■
Lemma 3.6 (metric inequalities for resolvents of α-comonotone operators). Let $A : X \rightrightarrows X$ and let $\gamma \in \mathbb{R}_{++}$. Then $A$ is α-comonotone if and only if, for all $(x, a), (y, b) \in \text{gra} J_{\gamma A}$,

$$\langle x - y, a - b \rangle \geq \alpha \|x - y\|^2 + (\gamma + \alpha)\|a - b\|^2.$$  \hspace{1cm} (47)

Consequently, if $A$ is α-comonotone and $J_{\gamma A}$ is single-valued, then, for all $x, y \in \text{dom} J_{\gamma A}$,

$$(\gamma + 2\alpha) \langle x - y, J_{\gamma A}x - J_{\gamma A}y \rangle \geq \alpha \|x - y\|^2 + (\gamma + \alpha)\|J_{\gamma A}x - J_{\gamma A}y\|^2.$$  \hspace{1cm} (48)

Proof. Let $(a, u), (b, v) \in \text{gra} A$ and set $x := a + \gamma u$ and $y := b + \gamma v$. We have the equivalence

$$\langle a - b, u - v \rangle \geq \alpha \|u - v\|^2 \hspace{1cm} \iff \hspace{1cm} \gamma \langle a - b, \gamma u - \gamma v \rangle \geq \alpha \|\gamma u - \gamma v\|^2 \hspace{1cm} \iff \hspace{1cm} \gamma \langle a - b, (x - y) - (a - b) \rangle \geq \alpha \|(x - y) - (a - b)\|^2 \hspace{1cm} \iff \hspace{1cm} \gamma \langle x - y, a - b \rangle - \gamma \|a - b\|^2 \geq \alpha \left(\|x - y\|^2 + \|a - b\|^2 - 2 \langle x - y, a - b \rangle\right) \hspace{1cm} \iff \hspace{1cm} (\gamma + 2\alpha) \langle x - y, a - b \rangle \geq \alpha \|x - y\|^2 + (\gamma + \alpha)\|a - b\|^2.$$  \hspace{1cm} (49a-c)

The conclusion then follows. \hfill \blacksquare

Proposition 3.7 (single-valuedness and full domain). Let $A : X \rightrightarrows X$ be α-comonotone and let $\gamma \in \mathbb{R}_{++}$ be such that $\gamma + \alpha > 0$. Then the following hold:

(i) $J_{\gamma A}$ is single-valued.

(ii) $\text{dom} J_{\gamma A} = X$ if and only if $A$ is maximally α-comonotone.

Proof. (i): This follows from (47) in Lemma 3.6 and the fact that $\gamma + \alpha > 0$.

(ii): Since $A$ is α-comonotone, $A' := A^{-1} - \alpha \text{Id}$ is monotone. Using the fact that $(\beta B)^{-1} = B^{-1}(\frac{1}{\beta} \text{Id})$ for any operator $B$ and any $\beta \in \mathbb{R} \setminus \{0\}$, we have that

$$(\text{Id} - J_{\gamma A}) = \text{Id} - (\text{Id} + \gamma A)^{-1} = (\text{Id} + \gamma A)^{-1} - \text{Id},$$  \hspace{1cm} (50a)

$$= ((\text{Id} + \gamma A) - \text{Id}) (\text{Id} + \gamma A)^{-1} \hspace{1cm} \text{(50b)}$$

$$= \left(A^{-1} \left(\frac{1}{\gamma} \text{Id}\right)\right)^{-1} (\text{Id} + \gamma A)^{-1} \hspace{1cm} \text{(50c)}$$

$$= \left(A^{-1} + \gamma \text{Id} \left(\frac{1}{\gamma} \text{Id}\right)\right)^{-1} \hspace{1cm} \text{(50d)}$$

$$= \gamma (A^{-1} + \gamma \text{Id})^{-1} \hspace{1cm} \text{(50e)}$$

and that

$$(A^{-1} + \gamma \text{Id})^{-1} = (A' + (\gamma + \alpha) \text{Id})^{-1} \hspace{1cm} \text{(51a)}$$

$$= \left(\text{Id} + \frac{1}{\gamma + \alpha} A'\right)^{-1} \left(\frac{1}{\gamma + \alpha} \text{Id}\right) \hspace{1cm} \text{(51b)}$$

$$= J_{\frac{1}{\gamma + \alpha} A'} \left(\frac{1}{\gamma + \alpha} \text{Id}\right) \hspace{1cm} \text{(51c)}$$

Therefore,

$$\text{Id} - J_{\gamma A} = \gamma J_{\frac{1}{\gamma + \alpha} A'} \left(\frac{1}{\gamma + \alpha} \text{Id}\right). \hspace{1cm} \text{(52)}$$
We deduce that \( \text{dom} J_{\gamma A} = X \) if and only if \( \text{dom} J_{\frac{1}{\gamma + \alpha} A} \), which, due to [2, Theorem 21.1 and Proposition 20.22], happens if and only if \( A' \) is maximally monotone. In turn, it is clear that the maximal monotonicity of \( A' \) is equivalent to the maximal \( \alpha \)-comonotonicity of \( A \).

**Proposition 3.8 (relaxed resolvents of \( \alpha \)-comonotone operators).** Let \( A : X \rightrightarrows X \) be \( \alpha \)-comonotone and let \( \gamma \in \mathbb{R}^+ \) be such that \( \gamma + \alpha > 0 \). Set \( R := (1 - \lambda) \text{Id} + \lambda J_{\gamma A} \) for \( \lambda \in \mathbb{R}^+ \).

Then the following hold:

1. \( J_{\gamma A} \) is conically \( \frac{\gamma}{2(\gamma + \alpha)} \)-averaged.
2. \( R \) is conically \( \frac{\lambda \gamma}{2(\gamma + \alpha)} \)-averaged.

**Proof.** First, Proposition 3.7(i) implies that \( J_{\gamma A} \) is single-valued and then Lemma 3.6 implies that, for all \( x, y \in \text{dom} J_{\gamma A} \),

\[
(\gamma + 2\alpha) \langle x - y, J_{\gamma A} x - J_{\gamma A} y \rangle \geq \alpha \|x - y\|^2 + (\gamma + \alpha) \|J_{\gamma A} x - J_{\gamma A} y\|^2,
\]

which is equivalent to

\[
2 \left( 1 - \frac{\gamma}{2(\gamma + \alpha)} \right) \langle x - y, J_{\gamma A} x - J_{\gamma A} y \rangle \geq \left( 1 - \frac{\gamma}{\gamma + \alpha} \right) \|x - y\|^2 + \|J_{\gamma A} x - J_{\gamma A} y\|^2.
\]

Therefore, \( J_{\gamma A} \) is conically \( \frac{\gamma}{2(\gamma + \alpha)} \)-averaged due to Proposition 2.2(iv). In turn, \( R \) is conically \( \frac{\lambda \gamma}{2(\gamma + \alpha)} \)-averaged due to Proposition 2.2(i)&(ii).

In the sequel, we will present the connection between the conical averagedness and several fixed point algorithms including the forward-backward algorithm and the adaptive Douglas–Rachford algorithm.

### 4. Relaxed forward-backward algorithm

In this section, we let \( A : X \rightrightarrows X, B : X \rightarrow X \), and consider the problem

\[
\text{find } x \in X \text{ such that } 0 \in Ax + Bx.
\]

Also let \( \gamma \in \mathbb{R}^+ \) and \( \kappa \in \mathbb{R}^+ \). Given \( x_0 \in X \), the relaxed forward-backward algorithm (rFB) for (55) generates sequences \((x_n)_{n \in \mathbb{N}}\) via

\[
\forall n \in \mathbb{N}, \quad x_{n+1} \in T_{\text{FB}}x_n \quad \text{with} \quad T_{\text{FB}} := (1 - \kappa) \text{Id} + \kappa J_{\gamma A}(\text{Id} - \gamma B).
\]

When \( \kappa = 1 \), this is actually the well-known forward-backward algorithm, see, e.g., [2, Section 26.5].

When an iterative fixed point algorithm is used to solve a problem, an important concern is the connection between the fixed points of that algorithm and the solutions of the given problem. For completeness, we include the proof of the following known result, see, e.g., [2, Proposition 26.1(iv)(a)], which states that the fixed points of the forward-backward algorithm actually solve (55).

**Lemma 4.1.** It holds that

\[
\text{Fix } T_{FB} = \text{Fix } (J_{\gamma A}(\text{Id} - \gamma B)) = \text{zer}(A + B).
\]
Proof. Noting that \( \text{Id} - T_{FB} = \kappa (\text{Id} - J_{\gamma A} (\text{Id} - \gamma B)) \), we have \( \text{Fix} T_{FB} = \text{Fix} (J_{\gamma A} (\text{Id} - \gamma B)) \). Now observe that \( x \in \text{Fix} (J_{\gamma A} (\text{Id} - \gamma B)) \iff x \in J_{\gamma A} (x - \gamma Bx) \iff x - \gamma Bx \in x + \gamma Ax \iff 0 \in Ax + Bx \). ■

Remark 4.2. In the light of Lemma 4.1 and Corollary 2.10, if \( T_{FB} \) is averaged, then, as soon as \( \text{zer}(A + B) \neq \emptyset \), we deduce that every sequence \( (x_n)_{n \in \mathbb{N}} \) generated by \( T_{FB} \) converges weakly to a point in \( \text{Fix} T_{FB} = \text{zer}(A + B) \) and that the rate of asymptotic regularity of \( T_{FB} \) is \( o(1/\sqrt{n}) \). Therefore, we will only investigate the averagedness of the operator \( T_{FB} \).

When \( B = 0 \), problem (55) reduces to finding a zero of operator \( A : X \rightrightarrows X \), i.e.,

\[
\text{find } x \in X \text{ such that } 0 \in Ax
\]

and the corresponding relaxed forward-backward algorithm reduces to the relaxed proximal point algorithm of the form

\[
\forall n \in \mathbb{N}, \quad x_{n+1} \in T_{PP} x_n \quad \text{with } T_{PP} := (1 - \kappa) \text{Id} + \kappa J_{\gamma A}.
\]

In this case, we have from Lemma 4.1 that

\[
\text{Fix} T_{PP} = \text{Fix} J_{\gamma A} = \text{zer} A.
\]

We arrive at the main results of this section, in which the averagedness of \( T_{PP} \) and \( T_{FB} \) is obtained when the operator \( A \) is not necessarily monotone. For classical results, we refer the reader to, e.g., [2, Example 23.40 and Proposition 26.1(iv)(d)].

Theorem 4.3 (relaxed proximal point algorithm). Suppose that \( A \) is maximally \( \alpha \)-comonotone with \( \alpha \in \mathbb{R} \) and that

\[
\kappa < \kappa^* := \frac{2(\gamma + \alpha)}{\gamma}.
\]

Then \( T_{PP} \) is \( \frac{\kappa}{\kappa^*} \)-averaged and has full domain.

Proof. To see the existence of \( \gamma, \kappa \in \mathbb{R}_{++} \) satisfying (61), we first take \( \gamma > \max\{0, -\alpha\} \) which implies \( 2(\gamma + \alpha)/\gamma > 0 \), and then choose \( \kappa \) between 0 and \( 2(\gamma + \alpha)/\gamma \).

Now, since (61) implies \( \gamma + \alpha > 0 \), we derive \( J_{\gamma A} \) and hence \( T_{PP} \) are single-valued and have full domain due to Proposition 3.7. Using Proposition 3.8(ii), we obtain that \( T_{PP} \) is conically \( \theta \)-averaged, where

\[
\theta := \frac{\kappa \gamma}{2(\gamma + \alpha)} = \frac{\kappa}{\kappa^*} < 1.
\]

This completes the proof. ■

Theorem 4.4 (rFB for \( \alpha \)- and \( \beta \)-comonotone operators with \( \alpha + \beta = 0 \)). Suppose that \( A \) is maximally \( \alpha \)-comonotone with \( \alpha \in \mathbb{R} \), that \( B \) is \( \beta \)-comonotone with \( \beta \in \mathbb{R}_{++} \), and that

\[
\alpha + \beta = 0, \quad \gamma = 2\beta, \quad \text{and } \kappa < 1.
\]

Then \( T_{FB} \) is \( \kappa \)-averaged and has full domain.

Proof. By assumption, \( \gamma + \alpha = 2\beta - \beta = \beta > 0 \). Since \( A \) is maximally \( \alpha \)-comonotone, we derive from Proposition 3.7 that \( J_{\gamma A} \) is single-valued and has full domain, so is \( T_{FB} \). Next, by Proposition 3.8(i), \( J_{\gamma A} \) is conically \( \theta_1 \)-averaged, where

\[
\theta_1 := \frac{\gamma}{2(\gamma + \alpha)} = 1.
\]
On the other hand, since $B$ is $\beta$-comonotone, we have from Proposition 3.3 that $\text{Id} - \gamma B$ is conically $\theta_2$-averaged, where
\[
\theta_2 := \frac{\gamma}{2\beta} = 1.
\] (65)

By Proposition 2.5, $J_{\gamma A}(\text{Id} - \gamma B)$ is nonexpansive, and so $T_{FB}$ is $\kappa$-averaged.

**Theorem 4.5 (rFB for $\alpha$- and $\beta$-comonotone operators with $\alpha + \beta > 0$).** Suppose that $A$ is maximally $\alpha$-comonotone with $\alpha \in \mathbb{R}$, that $B$ is $\beta$-comonotone with $\beta \in \mathbb{R}_{++}$, and that $\alpha + \beta > 0$ and
\[
\kappa < \kappa^* := \frac{4(\gamma + \alpha)\beta - \gamma^2}{2\gamma(\alpha + \beta)}.
\] (66)

Then $T_{FB}$ is $\frac{4\beta - \gamma}{\kappa^*}$-averaged and has full domain.

**Proof.** First, we observe that
\[
4(\gamma + \alpha)\beta - \gamma^2 > 0 \iff \alpha + \beta > 0 \quad \text{and} \quad 2\beta - 2\sqrt{\beta(\alpha + \beta)} < \gamma < 2\beta + 2\sqrt{\beta(\alpha + \beta)},
\] (67)

which ensures the existence of $\gamma, \kappa \in \mathbb{R}_{++}$ satisfying (66).

Next, it also follows from (66) that $4(\gamma + \alpha)\beta > \gamma^2 > 0$, which yields $\gamma + \alpha > 0$ since $\beta > 0$. By arguing as in the proof of Theorem 4.4, we obtain that $J_{\gamma A}$ and hence $T_{FB}$ are single-valued and has full domain, that $J_{\gamma A}$ is conically $\theta_1$-averaged, and that $\text{Id} - \gamma B$ is conically $\theta_2$-averaged, where
\[
\theta_1 := \frac{\gamma}{2(\gamma + \alpha)} > 0 \quad \text{and} \quad \theta_2 := \frac{\gamma}{2\beta} > 0.
\] (68)

Thanks to (66), it is clear that $\theta_1 \theta_2 < 1$. Applying Proposition 2.5 implies that $J_{\gamma A}(\text{Id} - \gamma B)$ is conically $\theta$-averaged, where
\[
\theta := \frac{\theta_1 + \theta_2 - 2\theta_1 \theta_2}{1 - \theta_1 \theta_2} = \frac{2\gamma(\alpha + \beta)}{4(\gamma + \alpha)\beta - \gamma^2} = \frac{1}{\kappa^*}. \tag{69}
\]

The conclusion then follows.

**Corollary 4.6.** Suppose that $A$ is maximally monotone, that $B$ is $\beta$-comonotone with $\beta \in \mathbb{R}_{++}$, and that
\[
\kappa < \kappa^* := \frac{4\beta - \gamma}{2\beta}.
\] (70)

Then $T_{FB}$ is $\frac{4\beta - \gamma}{\kappa^*}$-averaged and has full domain.

**Proof.** Note that $A$ is maximally 0-comonotone and then apply Theorem 4.5 with $\alpha = 0$.

**Remark 4.7 (range of parameter $\gamma$).** We recall that the classical convergence analysis for the forward-backward algorithm requires $\gamma \in ]0, 2\beta[$, see, for example, [2, Proposition 26.1(iv)(d)]. Interestingly, in Corollary 4.6, the existence of all parameters is ensured if and only if $\gamma \in ]0, 4\beta[$, which is indeed an improvement.

### 5. Adaptive Douglas–Rachford algorithm

In this section, we consider problem (55) with $A : X \rightrightarrows X$ and $B : X \rightrightarrows X$. Let $(\gamma, \delta) \in \mathbb{R}_{++}^2$ and $(\lambda, \mu, \kappa) \in \mathbb{R}_{++}^3$. The *adaptive DR operator* first introduced in [9] is given by
\[
T := T_{A,B} := (1 - \kappa) \text{Id} + \kappa R_2 R_1,
\] (71)
where
\[ R_1 := (1 - \lambda) \text{Id} + \lambda J_{\gamma A} \quad \text{and} \quad R_2 := (1 - \mu) \text{Id} + \mu J_{\delta B}. \] (72)

Given \( x_0 \in X \), the \textit{adaptive DR algorithm (aDR)} for (55) generates a sequence \((x_n)_{n \in \mathbb{N}}\), also called a \textit{DR sequence}, according to
\[ \forall n \in \mathbb{N}, \quad x_{n+1} \in T x_n. \] (73)

As usual, we will refer to the case when \( \delta = \gamma > 0 \) and \( \lambda = \mu = 2 \) as the \textit{classical DR algorithm} (or simply DR).

Unlike the forward-backward counterpart, the fixed points of the adaptive DR algorithm, in general, do not directly solve (55). Nevertheless, by choosing appropriate parameters, we are able to show that the images of the fixed points under the resolvent will do the job. For this purpose, as in [9], we make the \textit{standing assumption}
\[ \lambda = 1 + \frac{\delta}{\gamma} \quad \text{and} \quad \mu = 1 + \frac{\gamma}{\delta}, \] (74)
which is equivalent to
\[ (\lambda - 1)(\mu - 1) = 1 \quad \text{and} \quad \delta = \gamma(\lambda - 1), \] (75)
and which clearly holds for classical DR algorithm. The following fact justifies our setting.

\textbf{Lemma 5.1 (fixed point set of aDR operator).} Suppose that (74) holds, then \( \text{Fix } T \neq 0 \) if and only if \( \text{zer}(A + B) \neq \emptyset \). Moreover, if \( J_{\gamma A} \) is single-valued, then
\[ J_{\gamma A}(\text{Fix } T) = \text{zer}(A + B). \] (76)

\textit{Proof.} See [9, Lemma 4.1(iii)]. \hfill \Box

Therefore, we only study the convergence of the adaptive DR algorithm to the fixed point set of the operator \( T \) under condition (74). In turn, such a convergence can be guaranteed by the averagedness property as shown in the next result.

\textbf{Proposition 5.2 (convergence of aDR algorithm via averagedness).} Suppose that \( T \) is \( \theta \)-averaged and has full domain, that \( \text{zer}(A + B) \neq \emptyset \), and that (74) holds. Then the rate of asymptotic regularity of \( T \) is \( o(1/\sqrt{n}) \) and every sequence \((x_n)_{n \in \mathbb{N}}\) generated by \( T \) converges weakly to a point \( \overline{x} \in \text{Fix } T \). Moreover, if \( J_{\gamma A} \) is single-valued, then \( J_{\gamma A}\overline{x} \in \text{zer}(A + B) \).

\textit{Proof.} Since \( \text{zer}(A + B) \neq \emptyset \), Lemma 5.1 implies that \( \text{Fix } T \neq \emptyset \). Now, by Corollary 2.10, the rate of asymptotic regularity of \( T \) is \( o(1/\sqrt{n}) \) and every sequence \((x_n)_{n \in \mathbb{N}}\) generated by \( T \) converges weakly to a point \( \overline{x} \in \text{Fix } T \). The remaining conclusion follows from Lemma 5.1. \hfill \Box

Based on these observations, we will focus on the averagedness of the adaptive DR operator. To achieve such goal, we need to find appropriate parameters \( \gamma, \delta, \lambda, \mu, \kappa \). In fact, it suffices to determine only \( \gamma, \delta, \kappa > 0 \), since the parameters \( \lambda, \mu \) will be automatically defined by (74).

\textbf{5.1. The case of } \alpha \textit{- and } \beta \textit{-monotone operators revisited}

The convergence of the adaptive DR algorithm for \( \alpha \)- and \( \beta \)-monotone operators was originally provided in [9]. In this section, we revisit part of the results using the conical averagedness. In comparison to [9], the new results (see Theorems 5.3 and 5.5) \textit{enlarge} the acceptable range for the parameters \( \gamma, \delta, \lambda, \mu, \) and \( \kappa \), which guarantees the averagedness of \( T \), and hence, the convergence of the adaptive DR algorithm. In addition, the assumptions on the parameters are \textit{unified} into a \textit{single} condition, see, for example, (82). For clarity, we will split the main results of this section into two cases: \( \alpha + \beta = 0 \) and \( \alpha + \beta > 0 \).
Theorem 5.3 (aDR for $\alpha$- and $\beta$-monotone operators with $\alpha + \beta = 0$). Suppose that $A$ and $B$ are respectively maximally $\alpha$- and $\beta$-monotone with $\alpha + \beta = 0$, that $\gamma, \delta, \kappa \in \mathbb{R}_{++}$ satisfy

\[ 1 + 2\gamma\alpha > 0, \quad \delta = \frac{\gamma}{1 + 2\gamma\alpha}, \quad \kappa < 1, \]

and that $\lambda, \mu$ are given by (74). Then the adaptive DR operators $T_{A,B}$ and $T_{B,A}$ are both $\kappa$-averaged and have full domain.

Proof. By assumption, $1 + \gamma\alpha = 1/2 + (1 + 2\gamma\alpha)/2 > 1/2 > 0$ and $1 + \delta\beta = 1 - \delta\alpha = 1 - \gamma\alpha/(1 + 2\gamma\alpha) = (1 + \gamma\alpha)/(1 + 2\gamma\alpha) > 0$. Thus, Proposition 3.4 implies that $J_{A,B}$, $J_{B,A}$, and hence $T_{A,B}$ and $T_{B,A}$ are single-valued and have full domain. Next, by Proposition 3.5, $\frac{1}{1 - \alpha} R_1$ and $\frac{1}{1 - \mu} R_2$ are respectively conically $\theta_1$- and $\theta_2$-averaged with

\[ \theta_1 := \frac{\lambda}{2(\lambda - 1)(1 + \gamma\alpha)} = 1 \quad \text{and} \quad \theta_2 := \frac{\mu}{2(\mu - 1)(1 + \delta\beta)} = 1. \]

Applying Proposition 2.6 to $\frac{1}{1 - \alpha} R_1$ and $\frac{1}{1 - \mu} R_2$, we derive that $R_2 R_1 = \frac{1}{(1 - \lambda)(1 - \mu)} R_2 R_1$ and $R_1 R_2 = \frac{1}{(1 - \lambda)/(1 - \mu)} R_1 R_2$ are conically $1$-averaged, which implies that $T_{A,B} = (1 - \kappa) \text{Id} + \kappa R_2 R_1$ and $T_{B,A} = (1 - \kappa) \text{Id} + \kappa R_1 R_2$ are both $\kappa$-averaged by Proposition 2.2(i)&(ii).

Clearly, Theorem 5.3 recovers the convergence of the classical DR algorithm for two maximally monotone operators, see, e.g., [17]. While Theorem 5.3 is the case $\alpha + \beta = 0$ in [9, Theorem 4.5], we have presented an alternate proof using conical averagedness that applies concurrently to both adaptive DR operators $T_{A,B}$ and $T_{B,A}$.

For the case $\alpha + \beta > 0$, we will need the following technical lemma.

Lemma 5.4 (existence of parameters). Let $\alpha, \beta \in \mathbb{R}$ be such that $\alpha + \beta > 0$. Then, for every $\gamma, \delta \in \mathbb{R}_{++}$, the following statements are equivalent:

(i) $(\gamma + \delta)^2 < 4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta).

(ii) $1 + \gamma\alpha > 0$ and

\[ 2 + 2\gamma\alpha - 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)} < 1 + \frac{\gamma}{\delta} < 2 + 2\gamma\alpha + 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)}. \]

(iii) $1 + \gamma\alpha > 0$ and

\[ \frac{1}{\gamma} \left(1 + 2\gamma\alpha - 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)}\right) < \frac{1}{\delta} < \frac{1}{\gamma} \left(1 + 2\gamma\alpha + 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)}\right). \]

Consequently, given $\alpha + \beta > 0$, there always exist $\gamma, \delta \in \mathbb{R}_{++}$ that satisfy all the three statements.

Proof. We have that

\[ (\gamma + \delta)^2 < 4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta) \]

\[ \iff (1 - 4\gamma\beta - 4\gamma^2\alpha\beta)\delta^2 - 2\gamma(1 + 2\gamma\alpha)\delta + \gamma^2 < 0 \]

\[ \iff (1 - 4\gamma\beta - 4\gamma^2\alpha\beta) - 2(1 + 2\gamma\alpha)\frac{\gamma}{\delta} + \frac{\gamma^2}{\delta^2} < 0 \]

\[ \iff \begin{bmatrix} 1 + \gamma\alpha > 0 \& 1 + 2\gamma\alpha - 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)} < \frac{\gamma}{\delta} < 1 + 2\gamma\alpha + 2\sqrt{\gamma(1 + \gamma\alpha)(\alpha + \beta)}, \end{bmatrix} \]
which implies that the three statements are equivalent.

To see the existence of \( \gamma \) and \( \delta \), we choose \( \gamma > 0 \) such that \( 1/\gamma > -\alpha \) and then choose \( \delta > 0 \) that satisfies the second condition in (iii).

**Theorem 5.5 (aDR for \( \alpha \)- and \( \beta \)-monotone operators with \( \alpha + \beta > 0 \)).** Suppose that \( A \) and \( B \) are respectively maximally \( \alpha \)- and \( \beta \)-monotone with \( \alpha + \beta > 0 \), that \( \gamma, \delta, \kappa \in \mathbb{R}_{++} \) satisfy

\[
\kappa < \kappa^* := \frac{4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta) - (\gamma + \delta)^2}{2\gamma\delta(\gamma + \delta)(\alpha + \beta)},
\]

and that \( \lambda, \mu \) are given by (74). Then the adaptive DR operators \( T_{A,B} \) and \( T_{B,A} \) are both \( \frac{1}{\kappa} \)-averaged and have full domain.

**Proof.** We first verify the existence of \( \gamma, \delta, \kappa \in \mathbb{R}_{++} \) satisfying (82). According to Lemma 5.4, we can choose \( \gamma, \delta \in \mathbb{R}_{++} \) such that

\[
(\gamma + \delta)^2 < 4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta),
\]

which ensures the choice of \( \kappa \in \mathbb{R}_{++} \) in (82).

Next, (82) implies (83), and again by Lemma 5.4, we have \( 1 + \gamma\alpha > 0 \) and then also \( 1 + \delta\beta > 0 \). On the one hand, by Proposition 3.4, \( J_{\gamma A}, J_{\delta B} \) and hence \( T_{A,B} \) and \( T_{B,A} \) are single-valued and have full domain. On the other hand, Proposition 3.5 implies that \( \frac{1}{1+\alpha}R_1 \) and \( \frac{1}{1+\beta}R_2 \) are respectively conically \( \theta_1 \)- and \( \theta_2 \)-averaged with

\[
\theta_1 := \frac{\gamma + \delta}{2(\gamma + \delta)(1 + \gamma\alpha)} > 0 \quad \text{and} \quad \theta_2 := \frac{\mu + \delta}{2(\mu + \delta)(1 + \delta\beta)} > 0.
\]

Now, it follows from (83) that \( \theta_1 \theta_2 < 1 \). Thus, Proposition 2.6 implies that \( R_1 R_2 \) and \( R_2 R_1 \) are both conically \( \theta \)-averaged with

\[
\theta := \frac{\theta_1 + \theta_2 - 2\theta_1\theta_2}{1 - \theta_1\theta_2} = \frac{1 + 2\gamma\beta}{1 + 2\gamma\alpha} - 1 = \frac{2\gamma\delta(\gamma + \delta)(\alpha + \beta)}{4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta) - (\gamma + \delta)^2} = \frac{1}{\kappa^*}.
\]

Finally, we derive from Proposition 2.2(i)\&(ii) that \( T_{A,B} = (1 - \kappa)\text{Id} + \kappa R_2 R_1 \) and \( T_{B,A} = (1 - \kappa)\text{Id} + \kappa R_1 R_2 \) are both \( \frac{1}{\kappa} \)-averaged.

It is worth noting that Theorems 5.3 and 5.5 indeed present a different perspective on the convergence analysis of the adaptive DR algorithm for two generalized monotone operators. Moreover, they also improve recent results as presented in the next remark.

**Remark 5.6 (aDR for the case \( \alpha + \beta > 0 \)).** We will show that Theorem 5.5 readily implies and extends [9, Theorem 4.5(i)] in terms of parameter ranges. For this purpose, we first claim that

\[
\kappa^* \geq 1 \iff \frac{1 - 2\gamma\beta}{\gamma} \leq \frac{1 + 2\gamma\alpha}{\gamma} \iff \mu \in [2 - 2\gamma\beta, 2 + 2\gamma\alpha].
\]

Indeed, as \( \alpha + \beta > 0 \), we have the following equivalences

\[
\begin{align*}
\kappa^* \geq 1 & \iff 2\gamma\delta(\gamma + \delta)(\alpha + \beta) \leq 4\gamma\delta(1 + \gamma\alpha)(1 + \delta\beta) - (\gamma + \delta)^2 \\
& \iff (1 + 2\gamma\alpha - 2\gamma\beta - 4\gamma^2\alpha\beta)\delta^2 - 2\gamma(1 + \gamma\alpha - \gamma\beta)\delta + \gamma^2 \leq 0 \\
& \iff (1 + 2\gamma\alpha - 2\gamma\beta - 4\gamma^2\alpha\beta) - 2(1 + \gamma\alpha - \gamma\beta)\frac{\gamma}{\delta} + \frac{\gamma^2}{\delta^2} \leq 0 \\
& \iff 1 - 2\gamma\beta \leq \frac{\gamma}{\delta} \leq 1 + 2\gamma\alpha.
\end{align*}
\]
Since \( \mu = 1 + \frac{\gamma}{2} \), our claim (86) is true.

For the case \( \alpha + \beta > 0 \), recall that [9, Theorem 4.5(i)] requires \( \kappa \in [0, 1[ \) and \( (\gamma, \delta, \lambda, \mu) \in \mathbb{R}_+^2 \times [1, +\infty[^2 \) satisfying

\[
1 + 2\gamma\alpha > 0, \\
\mu \in [2 - 2\gamma\beta, 2 + 2\gamma\alpha], \\
(\lambda - 1)(\mu - 1) = 1, \delta = (\lambda - 1)\gamma.
\]

(88a) \hspace{1cm} (88b) \hspace{1cm} (88c)

We note that (88c) is equivalent to (74), and that (88b) is equivalent to \( \kappa^* \geq 1 \) due to (86). Therefore, the assumption of [9, Theorem 4.5(i)] for the case \( \alpha + \beta > 0 \) can be rewritten as

\[
\kappa < 1 \leq \kappa^*,
\]

(89)

which is more restrictive than the assumption \( \kappa < \kappa^* \) of Theorem 5.5. In summary, Theorem 5.5 not only recovers but also extends the parameter ranges for [9, Theorem 4.5(i)].

**Corollary 5.7 (DR for \( \alpha \)- and \( \beta \)-monotone operators with \( \alpha + \beta > 0 \)).** Suppose that \( A \) and \( B \) are respectively maximally \( \alpha \)- and \( \beta \)-monotone with \( \alpha + \beta > 0 \), that \( \gamma = \delta \in \mathbb{R}_+^2 \) and \( \lambda = \mu = 2 \), and that

\[
0 < \kappa < \kappa^* := 1 + \gamma \frac{\alpha\beta}{\alpha + \beta}.
\]

(90)

Then the DR operators \( T_{A,B} \) and \( T_{B,A} \) are both \( \frac{\alpha}{\kappa} \)-averaged and have full domain.

**Proof.** Apply Theorem 5.5 with \( \gamma = \delta \) and \( \lambda = \mu = 2 \). \( \blacksquare \)

**Remark 5.8 (DR for the case \( \alpha + \beta > 0 \)).** Corollary 5.7 is essentially [9, Theorem 4.5(ii)]. In the standard setup, the DR operator is defined as a strict convex combination of \( \text{Id} \) and \( R_2R_1 \) (or \( R_1R_2 \)), i.e., \( \kappa \in [0, 1[ \). In Corollary 5.7, on the one hand, we note that when both operators \( A \) and \( B \) are strongly monotone (\( \alpha > 0 \) and \( \beta > 0 \)), the upper bound for \( \kappa \) is \( \kappa^* = 1 + \gamma \frac{\alpha\beta}{\alpha + \beta} > 1 \), that means, \( \kappa \) can be chosen to be larger than 1 while still guarantees the convergence. On the other hand, if either \( A \) or \( B \) is weakly monotone (i.e., \( \alpha < 0 \) or \( \beta < 0 \)), then \( \kappa^* < 1 \), that means, one needs to further restrict \( \kappa \) from the standard range \( [0, 1[ \) in order to maintain the convergence.

In the rest of this section, we consider the adaptive DR algorithm for the problem of minimizing the sum of two functions. Let \( f: X \to ]-\infty, +\infty[ \). Then \( f \) is proper if \( \text{dom} f := \{ x \in X \mid f(x) < +\infty \} \neq \emptyset \), and lower semicontinuous if \( \forall x \in \text{dom} f, f(x) \leq \liminf_{z \to x} f(z) \). Given \( \alpha \in \mathbb{R} \), the function \( f \) is \( \alpha \)-convex (see [23, Definition 4.1]) if \( \forall x, y \in \text{dom} f, \forall \kappa \in [0, 1[, \)

\[
f((1 - \kappa)x + \kappa y) + \frac{\alpha}{2} \kappa(1 - \kappa)\|x - y\|^2 \leq (1 - \kappa)f(x) + \kappa f(y).
\]

(91)

We say that \( f \) is convex if \( \alpha = 0 \), strongly convex if \( \alpha > 0 \), and weakly convex if \( \alpha < 0 \). The proximity operator of a proper function \( f \) with parameter \( \gamma \in \mathbb{R}_+^2 \) is the mapping \( \text{Prox}_{\gamma f}: X \rightrightarrows X \) given by

\[
\forall x \in X, \quad \text{Prox}_{\gamma f}(x) := \arg\min_{z \in X} \left( f(z) + \frac{1}{2\gamma}\|z - x\|^2 \right).
\]

(92)

Now, we consider the \((\alpha, \beta)\)-convex minimization problem of the form

\[
\min_{x \in X} (f(x) + g(x)), \quad (93)
\]
where $f$ and $g$ are respectively $\alpha$- and $\beta$-convex functions. This problem arises in several important applications, see [16]. The adaptive DR algorithm for (93) is based on the operators

$$
T_{f,g} := (1 - \kappa) \text{Id} + \kappa R_g R_f \quad \text{and} \quad T_{g,f} := (1 - \kappa) \text{Id} + \kappa R_f R_g, \tag{94}
$$

where

$$
R_f := (1 - \lambda) \text{Id} + \lambda \text{Prox}_{\gamma f} \quad \text{and} \quad R_g := (1 - \mu) \text{Id} + \mu \text{Prox}_{\delta g}. \tag{95}
$$

The following analysis shows the connection between the proximal operators and resolvents of the subdifferentials and how the problem of finding zeros of sum of the operators can help in solving the minimization problems. Recall that the Fréchet subdifferential of $f$ at $x$ is given by

$$
\hat{\partial} f(x) := \left\{ u \in X \mid \liminf_{z \to x} \frac{f(z) - f(x) - \langle u, z - x \rangle}{\|z - x\|} \geq 0 \right\}. \tag{96}
$$

For more discussion on various subdifferentials and related properties, we refer interested readers to the monograph [19].

**Lemma 5.9 (proximity operators of $\alpha$-convex functions).** Let $f : X \to ]-\infty, +\infty]$ be a proper, lower semicontinuous, and $\alpha$-convex function and let $\gamma \in \mathbb{R}_+$ be such that $1 + \gamma \alpha > 0$. Then the following hold:

(i) $\hat{\partial} f$ is maximally $\alpha$-monotone.

(ii) $\text{Prox}_{\gamma f} = J_{\gamma \hat{\partial} f}$ is single-valued and has full domain.

**Proof.** See [9, Lemma 5.2]. ■

**Theorem 5.10 (aDR for $(\alpha, \beta)$-convex minimization).** Let $f : X \to ]-\infty, +\infty]$ and $g : X \to ]-\infty, +\infty]$ be proper and lower semicontinuous. Suppose that $f$ and $g$ are respectively $\alpha$- and $\beta$-convex, that $\alpha, \beta \in \mathbb{R}$ and $\gamma, \delta, \kappa \in \mathbb{R}_+$ satisfy either

(i) $\alpha + \beta = 0$, $1 + 2\gamma \alpha > 0$, $\delta = \frac{\gamma}{1 + 2\gamma \alpha}$, $\kappa < \kappa^* := 1$, or

(ii) $\alpha + \beta > 0$, $\kappa < \kappa^* := \frac{4\delta(1+\gamma\alpha)(1+\delta\beta)-(\gamma+\delta)^2}{2\delta(\gamma+\delta)(\alpha+\beta)}$,

and that $\lambda, \mu$ are given by (74). Then the adaptive DR operators $T_{f,g}$ and $T_{g,f}$ are both $\frac{\kappa}{\kappa^*}$-averaged and have full domain. Additionally, if $\text{zer}(\hat{\partial} f + \hat{\partial} g) \neq \emptyset$, then, for any $T \in \{T_{f,g}, T_{g,f}\}$, every sequence $(x_n)_{n \in \mathbb{N}}$ generated by $T$ converges weakly to a point $\overline{x} \in \text{Fix} T$ with $\text{Prox}_{\gamma f}(\overline{x}) \in \text{zer}(\hat{\partial} f + \hat{\partial} g) \subseteq \text{argmin}(f + g)$ and the rate of asymptotic regularity of $T$ is $o(1/\sqrt{n})$.

**Proof.** As in the proof of Theorems 5.3 and 5.5, we have from the assumption that $1 + \gamma \alpha > 0$ and $1 + \delta \beta > 0$. Now, Lemma 5.9 implies that $\hat{\partial} f$ and $\hat{\partial} g$ are respectively maximally $\alpha$- and $\beta$-monotone, with $\text{Prox}_{\gamma f} = J_{\gamma \hat{\partial} f}$ and $\text{Prox}_{\gamma g} = J_{\gamma \hat{\partial} g}$. So applying Theorems 5.3 and 5.5 to $A = \hat{\partial} f$ and $B = \hat{\partial} g$ yields the first conclusion. The remaining conclusion follows from Proposition 5.2 with noting that $\text{zer}(\hat{\partial} f + \hat{\partial} g) \subseteq \text{argmin}(f + g)$ (see [9, Lemma 5.3]). ■

As a consequence, we retrieve [9, Theorem 5.4(ii)], which unifies and extends [16, Theorems 4.4 and 4.6] to the Hilbert space setting.

**Corollary 5.11 (DR for $(\alpha, \beta)$-convex minimization).** Let $f : X \to ]-\infty, +\infty]$ and $g : X \to ]-\infty, +\infty]$ be proper and lower semicontinuous. Suppose that $f$ and $g$ are respectively $\alpha$- and $\beta$-convex, that $\gamma = \delta \in \mathbb{R}_+$ and $\lambda = \mu = 2$, and that $\alpha, \beta \in \mathbb{R}$ and $\kappa \in \mathbb{R}_+$ satisfy either
(i) \(\alpha = \beta = 0, \kappa < \kappa^* := 1, \) or
(ii) \(\alpha + \beta > 0, \kappa < \kappa^* := 1 + \gamma \frac{\alpha \beta}{\alpha + \beta}.\)

Then the DR operators \(T_{f,g}\) and \(T_{g,f}\) are both \(\frac{2}{\kappa^*}\)-averaged and have full domain. Additionally, if \(\text{zer}(\tilde{\partial}f + \tilde{\partial}g) \neq \emptyset\), then, for any \(T \in \{T_{f,g}, T_{g,f}\}\), every sequence \(\{x_n\}_{n \in \mathbb{N}}\) generated by \(T\) converges weakly to a point \(\bar{x} \in \text{Fix} T\) with \(\text{Prox}_{\gamma f}(\bar{x}) \in \text{zer}(\tilde{\partial}f + \tilde{\partial}g) \subseteq \text{argmin}(f + g)\) and the rate of asymptotic regularity of \(T\) is \(o(1/\sqrt{n})\).

**Proof.** Apply Theorem 5.10 with \(\gamma = \delta\) and \(\lambda = \mu = 2.\)

5.2. **The case of \(\alpha\) - and \(\beta\)-comonotone operators**

In this section, we consider the adaptive DR operators for two comonotone operators. Particularly, we derive the convergence results by using the conical averagedness.

**Theorem 5.12 (aDR for \(\alpha\)- and \(\beta\)-comonotone operators with \(\alpha + \beta = 0\).** Suppose that \(A\) and \(B\) are respectively maximally \(\alpha\)- and \(\beta\)-comonotone with \(\alpha + \beta = 0\), that

\[\gamma > \max\{0, -2\alpha\}, \quad \delta = \gamma + 2\alpha, \quad 0 < \kappa < 1,\]

and that \(\lambda, \mu\) are given by (74). Then the adaptive DR operators \(T_{A,B}\) and \(T_{B,A}\) are both \(\kappa\)-averaged and have full domain.

**Proof.** First, we see that \(\gamma + \alpha = \gamma/2 + (\gamma + 2\alpha)/2 > \gamma/2 > 0\) and

\[\delta + \beta = (\gamma + 2\alpha) - \alpha = \gamma + \sigma > 0.\]  

By Proposition 3.7, \(J_{\gamma A}, J_{\delta B}\), and hence \(T_{A,B}\) and \(T_{B,A}\) are single-valued and have full domain. Next, Proposition 3.8 implies that \(R_1\) and \(R_2\) are respectively conically \(\theta_1\)- and \(\theta_2\)-averaged with

\[\theta_1 := \frac{\lambda \gamma}{2(\gamma + \sigma)} = 1 \quad \text{and} \quad \theta_2 := \frac{\mu \delta}{2(\delta + \beta)} = 1.\]

That is, \(R_1\) and \(R_2\) are nonexpansive, so are \(R_2R_1\) and \(R_1R_2\). The conclusion then follows. 

Analogous to Lemma 5.4, we also present the existence result for the parameters.

**Lemma 5.13 (existence of parameters).** Let \(\alpha, \beta \in \mathbb{R}\) be such that \(\alpha + \beta > 0\). Then for every \(\gamma, \delta \in \mathbb{R}_{++}\), the following statements are equivalent:

(i) \((\gamma + \delta)^2 < 4(\gamma + \alpha)(\delta + \beta).\)
(ii) \(\gamma + \alpha > 0\) and \(\gamma + 2\alpha - 2\sqrt{(\gamma + \alpha)(\alpha + \beta)} < \delta < \gamma + 2\alpha + 2\sqrt{(\gamma + \alpha)(\alpha + \beta)}.\)

Consequently, there always exist \(\gamma, \delta \in \mathbb{R}_{++}\) that satisfy the above two statements.

**Proof.** We see that

\[(\gamma + \delta)^2 < 4(\gamma + \alpha)(\delta + \beta)\]  

\[\iff \delta^2 - 2(\gamma + 2\alpha)\delta + \gamma^2 - 4(\gamma + \alpha)\beta < 0\]  

\[\iff \begin{cases} \gamma + \alpha > 0 \quad \text{and} \\ \gamma + 2\alpha - 2\sqrt{(\gamma + \alpha)(\alpha + \beta)} < \delta < \gamma + 2\alpha + 2\sqrt{(\gamma + \alpha)(\alpha + \beta)}, \end{cases}\]

\[\text{\textbf{Remark 5.14.}}\]
which yields the equivalence of the two statements.

To show the existence of $\gamma$ and $\delta$, we first take $\gamma > \max\{0, -\alpha\}$ and then pick $\delta > 0$ so that the second condition in (ii) is satisfied.

**Theorem 5.14 (aDR for $\alpha$- and $\beta$-comonotone operators with $\alpha + \beta > 0$).** Suppose that $A$ and $B$ are respectively maximally $\alpha$- and $\beta$-comonotone with $\alpha + \beta > 0$, that $\gamma, \delta, \kappa \in \mathbb{R}^{++}$ satisfy

$$\kappa < \kappa^* := \frac{4(\gamma + \alpha)(\delta + \beta) - (\gamma + \delta)^2}{2(\gamma + \delta)(\alpha + \beta)},$$

and that $\lambda, \mu$ are given by (74). Then the adaptive DR operators $T_{A,B}$ and $T_{B,A}$ are both $\frac{\kappa}{\kappa^*}$-averaged and have full domain.

**Proof.** To show the existence of $\gamma, \delta, \kappa \in \mathbb{R}^{++}$ satisfying (101), we derive from Lemma 5.13 that there exist $\gamma, \delta \in \mathbb{R}^{++}$ such that

$$\kappa < \kappa^* := \frac{4(\gamma + \alpha)(\delta + \beta) - (\gamma + \delta)^2}{2(\gamma + \delta)(\alpha + \beta)},$$

hence the choice of $\kappa \in \mathbb{R}^{++}$ satisfying (101) is possible.

From (101), we have (102) and, again by Lemma 5.13, $\gamma + \alpha > 0$, which also implies $\delta + \beta > 0$. Next, by Proposition 3.7, $J_{\gamma A}, J_{\delta B}$, and hence $T_{A,B}$ and $T_{B,A}$ are single-valued and have full domain. By Proposition 3.8, $R_1$ and $R_2$ are respectively conically $\theta_1$- and $\theta_2$-averaged with

$$\theta_1 := \frac{\lambda \gamma}{2(\gamma + \alpha)} = \frac{\gamma + \delta}{2(\gamma + \alpha)} > 0 \quad \text{and} \quad \theta_2 := \frac{\mu \delta}{2(\delta + \beta)} = \frac{\gamma + \delta}{2(\delta + \beta)} > 0.$$

Note that $\theta_1 \theta_2 < 1$ due to (102). Using Proposition 2.6 implies that $R_1 R_2$ and $R_2 R_1$ are both conically $\theta$-averaged with

$$\theta := \frac{\theta_1 + \theta_2 - 2\theta_1 \theta_2}{1 - \theta_1 \theta_2} = \frac{4(\gamma + \delta)(\alpha + \beta)}{2(\gamma + \delta)(\alpha + \beta) - (\gamma + \delta)^2} = \frac{1}{\kappa^*}.$$

Now apply Proposition 2.2(i) & (ii).

**Corollary 5.15 (DR for $\alpha$- and $\beta$-comonotone operators with $\alpha + \beta > 0$).** Suppose that $A$ and $B$ are respectively maximally $\alpha$- and $\beta$-comonotone with $\alpha + \beta > 0$, that $\gamma = \delta \in \mathbb{R}^{++}$ and $\lambda = \mu = 2$, and that

$$0 < \kappa < \kappa^* := 1 + \frac{\alpha \beta}{\gamma(\alpha + \beta)}.$$

Then the adaptive DR operators $T_{A,B}$ and $T_{B,A}$ are both $\frac{\kappa}{\kappa^*}$-averaged and have full domain.

**Proof.** Apply Theorem 5.14 with $\gamma = \delta$ and $\lambda = \mu = 2$.

### 6. Conclusion

We have studied the conical averagedness and its characterizations, especially, the stability under relaxations, convex combinations, and compositions. We have also explored its connection to relaxed resolvents of generalized monotone operators. This property facilitates a new perspective on the convergence analysis of several fixed point algorithms including the proximal point, the forward-backward, and the adaptive DR algorithms.
Acknowledgement

SB was partially supported by a UMass Lowell faculty startup grant. MND was partially supported by Discovery Projects 160101537 and 190100555 from the Australian Research Council. HMP was partially supported by Autodesk, Inc. via a gift made to the Department of Mathematical Sciences, UMass Lowell.

References

[1] J.-B. Baillon, R.E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups, *Houston J. Math.* 4(1), 1–9 (1978).

[2] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, Cham (2017).

[3] H.H. Bauschke, W.M. Moursi, and X. Wang, Generalized monotone operators and their averaged resolvents, preprint 2019, arXiv:1902.09827.

[4] H.H. Bauschke, D. Noll, and H.M. Phan, Linear and strong convergence of algorithms involving averaged nonexpansive operators, *J. Math. Anal. Appl.* 421(1), 1–20 (2015).

[5] R.S. Burachik and A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer, New York (2008).

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

[7] P.L. Combettes and T. Pennanen, Proximal methods for cohypomonotone operators, *SIAM J. Control Optim.* 43(2), 731–742 (2004).

[8] P.L. Combettes and I. Yamada, Compositions and convex combinations of averaged nonexpansive operators, *J. Math. Anal. Appl.* 425(1), 55–70 (2015).

[9] M.N. Dao and H.M. Phan, Adaptive Douglas–Rachford splitting algorithm for the sum of two operators, *SIAM J. Optim.* 29(4), 2697–2724 (2019).

[10] M.N. Dao and H.M. Phan, Computing the resolvent of the sum of operators with application to best approximation problems, *Optim. Letters* (2019), DOI:10.1007/s11590-019-01432-x.

[11] M.N. Dao and M.K. Tam, Union averaged operators with applications to proximal algorithms for min-convex functions, *J. Optim. Theory Appl.* 181(1), 61–94 (2019).

[12] J. Douglas and H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, *Trans. Amer. Math. Soc.* 82, 421–439 (1956).

[13] J. Eckstein and D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program., Ser. A* 55(3), 293–318 (1992).

[14] P. Giselsson, Tight global linear convergence rate bounds for Douglas–Rachford splitting, *J. Fixed Point Theory Appl.* 19(4), 2241–2270 (2017).
[15] A.A. Goldstein, Convex programming in Hilbert space. *Bull. Amer. Math. Soc.* **70**, 709–710 (1964).

[16] K. Guo, D. Han, and X. Yuan, Convergence analysis of Douglas–Rachford splitting method for “strongly + weakly” convex programming, *SIAM J. Numer. Anal.* **55**(4), 1549–1577 (2017).

[17] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* **16**(6), 964–979 (1979).

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

[19] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I. Basic Theory*, Springer, Berlin (2006).

[20] N. Ogura and I. Yamada, Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping, *Numer. Funct. Anal. Optim.* **23**(1–2) 113–137 (2002).

[21] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, *J. Math. Anal. Appl.* **72**(2), 383–390 (1979).

[22] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14**(5), 877–898, (1976).

[23] J.-P. Vial, Strong and weak convexity of sets and functions. *Math. Oper. Res.* **8**(2), 231–259 (1983).