A Family of Operator Splitting Methods Revisited

Yunda Dong[a], Andreas Fischer[*,b]

[a]*Department of Mathematics, Zhengzhou University, 450052 Zhengzhou, P. R. of China*
[b]*Institute of Numerical Mathematics, Technische Universität Dresden, 01062 Dresden, Germany*

Abstract

A family of operator splitting methods for maximal monotone operators is investigated. It generalizes the Douglas–Peaceman–Rachford–Varga class of methods in the way that it allows the scaling parameters to vary from iteration to iteration non-monotonically. Conditions for convergence of methods within this family and for obtaining a linear rate of convergence are given. These conditions cover more general cases than existing ones.

Key words: Operator splitting, monotone operator, variable scaling parameter, convergence, convergence rate
2000 MSC: 65K05, 49M30, 46N10, 90C31

---

1. Introduction

For a real Hilbert space H and maximal monotone operators $A : H \rightrightarrows H$ and $B : H \rightrightarrows H$ let us consider the inclusion

$$0 \in A(x) + B(x). \tag{1}$$

This problem serves as a framework for the design and analysis of methods for definite linear systems, convex programs, monotone variational inequalities, and more general monotone inclusions, for examples see [14, 15].

The Douglas-Peaceman-Rachford-Varga (DPRV) class of iterative methods is one of the fundamental approaches for solving (1). In each iteration a method of this class performs the following two steps to obtain $x^{k+1} \in H$ from $x^k \in H$, where $\gamma_k > 0$ and $\mu > 0$ are parameters discussed later.

*Corresponding author. Tel.: +49 351 463 34148; fax: +49 351 463 34268
Email addresses: `ydong@zzu.edu.cn` (Yunda Dong),
`Andreas.Fischer@tu-dresden.de` (Andreas Fischer)

Preprint submitted to Nonlinear Analysis *January 11, 2017*

(a) Choose $a^k \in A(x^k)$ and determine $(y^k, b^k) \in \operatorname{graph} B$ so that

$$y^k + \mu b^k = x^k - \mu a^k. \tag{2}$$

(b) Choose $\gamma_k > 0$ and determine $(x^{k+1}, a^{k+1}) \in \operatorname{graph} A$ so that

$$x^{k+1} + \mu a^{k+1} = x^k + \mu a^k - \gamma_k(x^k - y^k). \tag{3}$$

For the case of systems of positive definite linear systems the above scheme is also called alternating direction implicit method. In particular, it is known as Douglas-Rachford method if $\gamma_k = \gamma := 1$ and as Peaceman-Rachford method for $\gamma_k = \gamma := 2$, see [1, 12]. Motivated by Wachspress and Habetler [16], Varga [15, Chapter 7] suggested to combine both methods by means of the parameter $\gamma_k = \gamma \in (0, 2]$. Later, Kellogg [7] and Lieutaud [9] extended the Peaceman-Rachford and the Douglas-Rachford method, respectively, to single-valued nonlinear maps. In [10], Lions and Mercier analyzed convergence properties of these and further methods in the DPRV class in the case when A and B are multi-valued maximal monotone maps in a Hilbert space. The above exposition of the DPRV class (that is different but equivalent to the exposition in [10]) can be found in [3, 5].

The use of variable scaling parameters μ_k instead of a fixed parameter μ can accelerate the convergence speed. Therefore, it would be desirable to have a convergence theory that allows to modify the scaling parameter μ from iteration to iteration. Recently, He et al. [6] provided a technique for varying the scaling parameter including convergence properties for a special operator splitting for continuous monotone variational inequalities. More in detail, given a nonempty closed convex set $\Omega \in \mathbb{R}^n$ and a continuous monotone map $F : \mathbb{R}^n \to \mathbb{R}^n$, the operator splitting in [6] is done by rewriting the variational inequality as the inclusion

$$0 \in F(x) + N_\Omega(x),$$

where $N_\Omega : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ denotes the normal cone operator.

For problem settings more general than the one used by He [6], known convergence results related to the DPRV class keep μ fixed for all iterations, see [2, 4, 8, 10]. In this paper we aim at filling this gap. The DPRV family with variable scaling we are going to analyze is based on the variable scaling technique introduced in [6]. However, in contrast to the restricted setting in [6] we will deal with possibly multi-valued maximal monotone operators A and B in finite or infinite dimensional Hilbert spaces. By proof techniques different from that in [6] we obtain convergence results which are valid for a much broader class of inclusions and a larger variety of splitting possibilities, see Section 4. Moreover, in Section 5, the first time for DPRV methods with variable scaling parameters, results on the convergence rate are given.

2. Preliminaries

We first review some basic definitions and then provide some auxiliary results for later use.

Let $\langle\ ,\ \rangle : H \times H \to \mathbb{R}$ denote the inner product of the Hilbert space H and $\|\cdot\| : H \to [0, \infty)$ the norm induced by the inner product. For a given (possibly multi-valued) operator $T : H \rightrightarrows H$, $\operatorname{graph} T := \{(x, y) \in H \times H \,|\, y \in T(x)\}$ denotes the graph of T and T^{-1} its inverse, i.e., $T^{-1}(y) := \{x \in H \,|\, y \in T(x)\}$ for any $y \in H$. By $I : H \to H$ the identity operator is denoted.

Definition 1. An operator $T : H \rightrightarrows H$ is called *monotone* if

$$\langle x - y,\ \xi - \eta\rangle \geq 0$$

holds for all $(x, \xi), (y, \eta) \in \operatorname{graph} T$. A monotone operator $T : H \rightrightarrows H$ is said to be *maximal monotone* if, for any $(x^0, \xi^0) \in H \times H$,

$$\langle x^0 - y,\ \xi^0 - \eta\rangle \geq 0 \quad \text{for all } (y, \eta) \in \operatorname{graph} T \qquad \text{implies} \qquad (x^0, \xi^0) \in \operatorname{graph} T,$$

i.e., the graph of T cannot be enlarged without destroying monotonicity.

Definition 2. An operator $T : H \rightrightarrows H$ is called *α-monotone* if $\alpha \geq 0$ exists so that

$$\langle x - y,\ \xi - \eta\rangle \geq \alpha\|x - y\|^2$$

holds for all $(x, \xi), (y, \eta) \in \operatorname{graph} T$.

Obviously, an operator $T : H \rightrightarrows H$ is monotone if and only if it is 0-monotone. If T is α-monotone with $\alpha > 0$ then it is usually called *strongly monotone with modulus α* in the literature. To say that $A : H \rightrightarrows H$ is α-monotone and $B : H \rightrightarrows H$ is β-monotone we say, for short, that (A, B) is (α, β)-monotone.

Definition 3. A single-valued map $T : H \to H$ is called *Lipschitz continuous with modulus $L > 0$* if

$$\|T(x) - T(y)\| \leq L\|x - y\|$$

holds for all $x, y \in H$.

Lemma 1. *Let a summable sequence* $\{\tau_k\} \subset [0, 1)$ *and a sequence* $\{\mu_k\} \subset (0, \infty)$ *be given so that*

$$(1 - \tau_k)\mu_k \leq \mu_{k+1} \leq (1 + \tau_k)\mu_k \tag{4}$$

holds for all $k \in \mathbb{N}$. *Then there are numbers* $l > 0$ *and* $u > 0$ *so that*

$$l\,\mu_0 \leq \mu_k \leq u\,\mu_0 \quad \textit{for all } k \in \mathbb{N}. \tag{5}$$

Proof. Since $\{\tau_k\}$ is summable there is $k_0 \in \mathbb{N}$ so that

$$(1 - \tau_k) \geq (1 + 2\tau_k)^{-1} \quad \text{for all } k \geq k_0.$$

Therefore, by (4),

$$\mu_{k+1} \geq \mu_0 \prod_{i=0}^{k}(1 - \tau_i) \geq \mu_0 \prod_{i=0}^{k_0-1}(1 - \tau_i) \prod_{i=k_0}^{k}(1 + 2\tau_i)^{-1} =: \tilde{\mu}_k$$

follows for all $k \in \mathbb{N}$ with $k \geq k_0 \geq 1$. Moreover, we have

$$\ln \prod_{i=k_0}^{k}(1 + 2\tau_i)^{-1} = -\sum_{i=k_0}^{k} \ln(1 + 2\tau_i) \geq -\sum_{i=k_0}^{k} 2\tau_i$$

for $k \geq k_0$. Thus, because $\{\tau_k\}$ is summable, $\{\tilde{\mu}_k\}$ is bounded below by some positive number. Since $\{\tilde{\mu}_k\}$ decreases monotonically it converges to some $\tilde{\mu} > 0$. Therefore, the left inequality in (5) follows. The right one can be proved similarly taking into account that $\ln(1 + \tau_i) \leq \tau_i$. □

Lemma 2. *Let sequences* $\{\alpha_k\}$, $\{\beta_k\}$, $\{c_k\} \subset [0, \infty)$, *and a summable sequence* $\{\lambda_k\} \subset [0, \infty)$ *be given. If*

$$\alpha_{k+1} \leq (1 + \lambda_k)\alpha_k - c_k\beta_k \tag{6}$$

holds for all $k \in \mathbb{N}$ *then*

(i) *the sequence* $\{\alpha_k\}$ *converges and*

(ii) $\liminf\limits_{k\to\infty} c_k > 0$ *implies* $\lim\limits_{k\to\infty} \beta_k = 0$.

Proof. (i) Because the nonnegative sequence $\{\lambda_k\}$ is summable, it follows that

$$p := \prod_{i=0}^{\infty}(1 + \lambda_i) < \infty$$

and furthermore, by (6), $\{c_k\} \subset [0, \infty)$, and $\{\beta_k\} \subset [0, \infty)$, that

$$\alpha_{k+1} \leq \alpha_0 \prod_{i=0}^{k} (1 + \lambda_i) \leq \alpha_0 p \quad \text{for all } k \in \mathbb{N}. \tag{7}$$

Thus, the sequence $\{\alpha_k\}$ is bounded. By (6), the nonnegativity of c_k and β_k, and (7), we obtain

$$\alpha_{k+1} - \alpha_k \leq \lambda_k \alpha_k \leq \alpha_0 p \lambda_k.$$

Therefore,

$$\alpha_{k+\ell} - \alpha_k \leq \alpha_0 p \sum_{i=k}^{k+\ell-1} \lambda_i \tag{8}$$

holds for all $k, \ell \in \mathbb{N}$ with $\ell \geq 1$. Since we already know that the sequence $\{\alpha_k\}$ is bounded it has at least one cluster point, say $\hat{\alpha}_1$. Let us assume that there is another cluster point $\hat{\alpha}_2$ with $\hat{\alpha}_2 > \hat{\alpha}_1$. Then, taking into account the summability of $\{\lambda_k\}$, we can choose $k, \ell \in \mathbb{N}$ so that

$$\frac{1}{2}(\hat{\alpha}_2 - \hat{\alpha}_1) \leq \alpha_{k+\ell} - \alpha_k, \quad \text{and} \quad \alpha_0 p \sum_{i=k}^{k+\ell-1} \lambda_i \leq \frac{1}{4}(\hat{\alpha}_2 - \hat{\alpha}_1).$$

This contradicts (8). Furthermore, if we assumed $\hat{\alpha}_1 > \hat{\alpha}_2$ similar arguments would lead to a contradiction as well. Hence, $\hat{\alpha}_1 = \hat{\alpha}_2$ follows, i.e., the sequence $\{\alpha_k\}$ converges.

(ii) According to (6), we have

$$c_k \beta_k \leq (1 + \lambda_k)\alpha_k - \alpha_{k+1} \quad \text{for all } k \in \mathbb{N}.$$

Taking into account $\{\alpha_k\}$, $\{\lambda_k\} \subset [0, \infty)$, (7), $p \geq 1$, and the summability of $\{\lambda_k\}$ we obtain

$$\sum_{i=0}^{k} c_i \beta_i \leq \alpha_0 + \sum_{i=0}^{k} \lambda_i \alpha_i \leq \alpha_0 + \alpha_0 p \sum_{i=0}^{\infty} \lambda_i < +\infty.$$

Let us assume that there are $\hat{\beta} > 0$ and an infinite set $N \subseteq \mathbb{N}$ so that $\beta_i \geq \hat{\beta}$ for all $i \in N$. Then, by $\{c_k\}, \{\beta_k\} \subset [0, \infty)$,

$$\hat{\beta} \sum_{i \in N} c_i \leq \alpha_0 + \alpha_0 p \sum_{i=0}^{\infty} \lambda_i < +\infty$$

follows. Obviously, this contradicts the assumption that $\liminf_{k\to\infty} c_k > 0$. Hence, $\lim_{k\to\infty} \beta_k = 0$ follows. □

The next lemma can be derived from [4, Proposition 4] for the case when H has infinite dimension. If H is finite-dimensional the weak convergence of a sequence is equivalent to its strong convergence so that the result in the lemma follows immediately from the closedness of $\operatorname{graph} A$ and $\operatorname{graph} B$.

Lemma 3. *Let $A, B : H \rightrightarrows H$ be maximal monotone operators and suppose that $A + B$ is maximal monotone or that H is finite-dimensional. Moreover, suppose that there are sequences $\{(x^k, a^k)\} \subset \operatorname{graph} A$ and $\{(y^k, b^k)\} \subset \operatorname{graph} B$ and elements $x^\infty, a^\infty \in H$ so that*

$$\begin{array}{lcl} (x^k, a^k) & \rightharpoonup & (x^\infty, a^\infty), \\ (y^k, b^k) & \rightharpoonup & (x^\infty, -a^\infty), \\ (x^k - y^k, a^k + b^k) & \to & (0, 0) \end{array}$$

for $k \to \infty$. Then, $0 \in A(x^\infty) + B(x^\infty)$.

3. The DPRV family with variable scaling

We now formally describe an algorithm for solving problem (1) which, due to the possible choices of parameters, describes the DPRV family with variable scaling.

Variable Scaling DPRV Algorithm

Step 0. Choose $x^0 \in H$ and $\mu_0 \in (0, \infty)$. Set $k := 0$.

Step 1. Choose $a^k \in A(x^k)$ and determine $(y^k, b^k) \in \operatorname{graph} B$ so that

$$y^k + \mu_k b^k = x^k - \mu_k a^k. \tag{9}$$

If $x^k = y^k$ then stop.

Step 2. Choose $\gamma_k > 0$ and determine $(x^{k+1}, a^{k+1}) \in \operatorname{graph} A$ so that

$$x^{k+1} + \mu_k a^{k+1} = x^k + \mu_k a^k - \gamma_k (x^k - y^k). \tag{10}$$

Step 3. Choose $\tau_k \in [0, 1)$ and $\mu_{k+1} \in [(1 - \tau_k)\mu_k,\ (1 + \tau_k)\mu_k]$.
Set $k := k + 1$, and go to Step 1.

Remark 1. By a result of Minty [11] it is well known that, for any maximal operator $T : H \rightrightarrows H$, the inclusion $r \in z + \mu T(z)$ has a unique solution for arbitrarily chosen $r \in H$ and $\mu > 0$. Thus, since A and B are assumed to be maximal monotone, y^k in Step 1 and x^{k+1} in Step 2 are uniquely defined. Moreover, the above algorithm is always well defined.

Remark 2. Due to Lemma 1, the scaling parameters μ_k generated by the above algorithm stay in the interval $[\mu_0 l,\ \mu_0 u]$, where l and u depend on the sequence $\{\tau_k\}$. In [6] the scaling parameters were allowed to vary according to

$$\mu_{k+1} \in [\mu_k/(1+\tau_k),\ (1+\tau_k)\mu_k]$$

Since $(1-\tau_k)\mu_k \leq \mu_k/(1+\tau_k)$ for any $\tau_k \in [0,1)$, the scaling condition in Step 3 of the algorithm is a bit more flexible than the one in [6].

Finally, we note that it is even possible to replace the rule for choosing μ_{k+1} in Step 3 by

$$\mu_{k+1} \in [\mu_0 l, (1+\tau_k)\mu_k],$$

where $l > 0$ is arbitrarily small but fixed. Then, all results in Sections 4 and 5 remain valid.

4. Convergence

In this section we will provide conditions under which the Variable Scaling DPRV Algorithm converges to a solution of the inclusion (1). Our approach extends He's convergence result [6] to a much more general case. In particular, none of the operators A and B has to be single-valued or even continuous.

The next lemma characterizes fundamental relations between some of the sequences generated by the Variable Scaling DPRV Algorithm.

Lemma 4. *Let $A, B : H \rightrightarrows H$ be maximal monotone operators. Moreover, suppose that the pair (A,B) is (α,β)-monotone with $\alpha,\beta \geq 0$ and that the inclusion (1) has at least one solution. Then, the following assertions hold:*

If the Variable Scaling DPRV Algorithm stops in Step 1 then x^k is a solution of (1). Otherwise, for any solution x^* of (1) there is $a^* \in A(x^*)$ so that the iterates generated by this algorithm satisfy

$$\begin{aligned} \|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \ \leq\ & (1+\tau_k)^2\|x^k - x^* + \mu_k(a^k - a^*)\|^2 \\ & -\gamma_k(2-\gamma_k)\|x^k - y^k\|^2 - 2\gamma_k\mu_k\Delta_k \end{aligned} \tag{11}$$

for all $k \in \mathbb{N}$ with Δ_k given by

$$\Delta_k := \alpha\|x^k - x^*\|^2 + \beta\|y^k - x^*\|^2. \tag{12}$$

Proof. If the algorithm stops in Step 1 we have $x^k = y^k$ for some $k \in \mathbb{N}$. By (9) this implies $b^k = -a^k$ so that $0 = a^k + b^k \in A(x^k) + B(x^k)$ follows and, hence, x^k is a solution of (1).

Let us now consider the case when the algorithm never stops and let $k \in \mathbb{N}$ be arbitrarily chosen. Since x^* is a solution of (1), there are a^*, b^* so that

$$a^* + b^* = 0.$$

By (9), we also have

$$b^k = \mu_k^{-1}(x^k - y^k) - a^k.$$

Thus, the β-monotonicity of B yields

$$\beta\|y^k - x^*\|^2 \leq \langle y^k - x^*, b^k - b^* \rangle = \langle y^k - x^*, \mu_k^{-1}(x^k - y^k) - (a^k - a^*)\rangle.$$

Using the latter and the α-monotonicity of A, we obtain

$$\begin{aligned}
&\langle x^k - y^k, x^k - x^* + \mu_k(a^k - a^*)\rangle \\
&\quad = \langle x^k - y^k, x^k - y^k\rangle + \langle x^k - y^k, y^k - x^* + \mu_k(a^k - a^*)\rangle \\
&\quad = \|x^k - y^k\|^2 + \mu_k\langle x^k - x^*, a^k - a^*\rangle + \mu_k\langle x^* - y^k, a^k - a^*\rangle + \langle x^k - y^k, y^k - x^*\rangle \\
&\quad \geq \|x^k - y^k\|^2 + \mu_k\alpha\|x^k - x^*\|^2 + \mu_k\langle y^k - x^*, \mu_k^{-1}(x^k - y^k) - (a^k - a^*)\rangle \\
&\quad \geq \|x^k - y^k\|^2 + \mu_k\Delta_k.
\end{aligned}$$

By (10) and the inequality just derived, we get

$$\begin{aligned}
&\|x^{k+1} - x^* + \mu_k(a^{k+1} - a^*)\|^2 \\
&\quad = \|x^k - x^* + \mu_k(a^k - a^*) - \gamma_k(x^k - y^k)\|^2 \\
&\quad \leq \|x^k - x^* + \mu_k(a^k - a^*)\|^2 + \gamma_k^2\|x^k - y^k\|^2 \\
&\qquad -2\gamma_k\langle x^k - x^* + \mu_k(a^k - a^*),\ x^k - y^k\rangle \\
&\quad \leq \|x^k - x^* + \mu_k(a^k - a^*)\|^2 - \gamma_k(2 - \gamma_k)\|x^k - y^k\|^2 - 2\gamma_k\mu_k\Delta_k.
\end{aligned} \tag{13}$$

The condition for choosing μ_{k+1} in Step 3 of the algorithm provides $0 < \mu_{k+1} \leq (1 + \tau_k)\mu_k$. This and the monotonicity of A imply

$$\begin{aligned}
&\|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \\
&\quad = \|x^{k+1} - x^*\|^2 + \mu_{k+1}^2\|a^{k+1} - a^*\|^2 + 2\mu_{k+1}\langle x^{k+1} - x^*, a^{k+1} - a^*\rangle \\
&\quad \leq (1 + \tau_k)^2\|x^{k+1} - x^*\|^2 + (1 + \tau_k)^2\mu_k^2\|a^{k+1} - a^*\|^2 \\
&\qquad +2(1 + \tau_k)\mu_k\langle x^{k+1} - x^*, a^{k+1} - a^*\rangle \\
&\quad = (1 + \tau_k)^2\|x^{k+1} - x^* + \mu_k(a^{k+1} - a^*)\|^2.
\end{aligned}$$

Combining this and (13) yields the desired result. □

Theorem 1. *Let the assumptions of Lemma 4 be satisfied. Moreover, suppose that the sequence $\{\tau_k\}$ is summable and that the Variable Scaling DPRV Algorithm never stops. Then, the following assertions hold:*

(a) *If $A + B$ is maximal monotone and if*

$$\liminf_{k\to\infty} \gamma_k(2 - \gamma_k + \mu_k \min\{\alpha,\beta\}) > 0, \tag{14}$$

then the sequence $\{x^k\}$ converges weakly to a solution of (1).

(b) *If $\bar{\gamma} \leq \gamma_k \leq 2$ for all $k \in \mathbb{N}$ with some $\bar{\gamma} > 0$ and if $\alpha + \beta > 0$, then the sequence $\{x^k\}$ converges strongly to the unique solution of (1).*

Proof. (a) Let x^* denote any solution of (1). Then, with Δ_k defined in Lemma 4, it can easily be verified that

$$\frac{1}{2}\min\{\alpha,\beta\}\|x^k - y^k\|^2 \leq \alpha\|x^k - x^*\|^2 + \beta\|y^k - x^*\|^2 = \Delta_k$$

holds for all $k \in \mathbb{N}$. Thus, Lemma 4 implies for some $a^* \in A(x^*)$ that

$$\begin{aligned}&\|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \\ &\quad\leq (1+\tau_k)^2\|x^k - x^* + \mu_k(a^k - a^*)\|^2 - \gamma_k(2 - \gamma_k + \mu_k \min\{\alpha,\beta\})\|x^k - y^k\|^2\end{aligned}$$

for all $k \in \mathbb{N}$. To apply Lemma 2 let us define

$$\begin{aligned} \alpha_k &:= \|x^k - x^* + \mu_k(a^k - a^*)\|^2, & \beta_k &:= \|x^k - y^k\|^2, \\ c_k &:= \gamma_k(2 - \gamma_k + \mu_k \min\{\alpha,\beta\}), & \lambda_k &:= 2\tau_k + \tau_k^2 \end{aligned}$$

for all $k \in \mathbb{N}$. Since the sequence $\{\tau_k\}$ is assumed to be summable, the sequence $\{\lambda_k\}$ must be summable, too. Hence, taking into account (14), Lemma 2 yields that

(i') the sequence $\{\|x^k - x^* + \mu_k(a^k - a^*)\|\}$ converges and

(ii') $\lim_{k\to\infty}(x^k - y^k) = 0$.

Since

$$\alpha_k = \|x^k - x^*\|^2 + \mu_k^2\|a^k - a^*\|^2 + 2\mu_k\langle x^k - x^*, a^k - a^*\rangle$$

the monotonicity of A, (i'), and $\mu_k \geq l\mu_0 > 0$ for all $k \in \mathbb{N}$ (by Lemma 1) imply that $\{x^k\}$ and $\{a^k\}$ are bounded. Hence, there is a weak cluster point of $\{(x^k, a^k)\}$, say (x^∞, a^∞), and an infinite set $N \subseteq \mathbb{N}$ so that

$$(x^k, a^k) \rightharpoonup (x^\infty, a^\infty) \qquad \text{for } k \in N,\ k \to \infty. \tag{15}$$

Using (9), (ii'), and again $\mu_k \geq l\mu_0$ for all $k \in \mathbb{N}$ we obtain

$$\lim_{k\to\infty}(x^k - y^k, a^k + b^k) = (0, 0). \tag{16}$$

Thus, because of (15), we have

$$(y^k, b^k) \rightharpoonup (x^\infty, -a^\infty) \qquad \text{for } k \in N,\ k \to \infty. \tag{17}$$

Taking into account (15), (17), (16), and that $A + B$ is assumed to be maximal monotone, we get from Lemma 3 that $0 \in A(x^\infty)+B(x^\infty)$, i.e., x^∞ solves inclusion (1).

To show that the sequence $\{x^k\}$ weakly converges to x^∞ let x_1^∞ and x_2^∞ denote two weak cluster points of $\{x^k\}$. Then, repeating the arguments above yields that x_1^∞ and x_2^∞ solve inclusion (1). Moreover, (i') still holds true if (x^*, a^*) is replaced by (x_i^∞, a_i^∞) with an appropriate $a_i^\infty \in A(x_i^\infty)$, $i = 1, 2$. Thus, the limits

$$l_i := \lim_{k\to\infty} \|x^k - x_i^\infty + \mu_k(a^k - a_i^\infty)\|^2 \quad \text{for } i = 1, 2$$

exist. By the monotonicity of A, it follows that

$$\begin{aligned}
&\|x^k - x_2^\infty + \mu_k(a^k - a_2^\infty)\|^2 \\
&\quad = \|x^k - x_1^\infty + \mu_k(a^k - a_1^\infty)\|^2 + \|x_1^\infty - x_2^\infty + \mu_k(a_1^\infty - a_2^\infty)\|^2 \\
&\qquad +2\langle x^k - x_1^\infty + \mu_k(a^k - a_1^\infty), x_1^\infty - x_2^\infty + \mu_k(a_1^\infty - a_2^\infty)\rangle \\
&\quad \geq \|x^k - x_1^\infty + \mu_k(a^k - a_1^\infty)\|^2 + \|x_1^\infty - x_2^\infty\|^2 \\
&\qquad +2\langle x^k - x_1^\infty + \mu_k(a^k - a_1^\infty), x_1^\infty - x_2^\infty + \mu_k(a_1^\infty - a_2^\infty)\rangle.
\end{aligned}$$

Taking the limit for those $k \in \mathbb{N}$ belonging to the subsequence of $\{(x^k, a^k)\}$ that weakly converges to (x_1^∞, a_1^∞), we get

$$l_2 \geq l_1 + \|x_1^\infty - x_2^\infty\|^2.$$

A similar reasoning yields

$$l_1 \geq l_2 + \|x_1^\infty - x_2^\infty\|^2.$$

This implies $x_1^\infty = x_2^\infty$. Therefore, $\{x^k\}$ converges weakly to some solution x^∞ of inclusion (1).

(b) For $\gamma_k \in [\bar{\gamma}, 2]$ Lemma 4 provides

$$\|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \leq (1 + \tau_k)^2\|x^k - x^* + \mu_k(a^k - a^*)\|^2 - 2\bar{\gamma}\mu_k\Delta_k$$

for any solution x^* of (1) with some $a^* \in A(x^*)$.

If $\alpha > 0$ then the above definitions of α_k and λ_k, the definition of Δ_k in Lemma 4, and the fact that μ_k is bounded below by $l\mu_0$ (Lemma 1) lead us to

$$\alpha_{k+1} \leq (1 + \lambda_k)\alpha_k - 2\bar{\gamma}l\mu_0\alpha\|x^k - x^*\|^2.$$

Setting $\beta_k := \|x^k - x^*\|^2$ and $c_k := 2\bar{\gamma}l\mu_0$ for all $k \in \mathbb{N}$ we obtain from (ii) of Lemma 2 that $\{\beta_k\}$ converges to 0, i.e., $\{x^k\}$ converges strongly to x^*.

If $\beta > 0$, it follows with a similar reasoning (just $\|x^k - x^*\|$ has to be replaced by $\|y^k - x^*\|$) that $\{y^k\}$ converges strongly to x^*. Because of (ii') within part (a) of this proof, we also obtain strong convergence of $\{x^k\}$ to x^*. □

Remark 3. If H is finite dimensional then condition (14) can be replaced by the weaker requirement that

$$\sum_{k=0}^{\infty} \gamma_k(2 - \gamma_k + \mu_k \min\{\alpha, \beta\}) = +\infty.$$

To see this we first note that under the conditions of Lemma 2 we also get that $\sum_{i=0}^{\infty} c_k = +\infty$ implies $\liminf_{k\to\infty} \beta_k = 0$. With this and ideas in part (a) of the proof of Theorem 1, one can show that a subsequence of $\{(x^k, a^k)\}$ converges to (x^∞, a^∞), where x^∞ solves (1) and $a^\infty \in A(x^\infty)$. Now, using (x^∞, a^∞) instead of (x^*, a^*) in the beginning of part (a) of the proof of Theorem 1, we obtain that $\alpha_k = \|x^k - x^\infty + \mu_k(a^k - a^\infty)\|^2$ goes to 0. The monotonicity of A then implies that $\|x^k - x^\infty\|$ converges to 0. Thus, x^k converges to x^*.

Remark 4. It is interesting to note that if $\min\{\alpha, \beta\} > 0$ part (a) of Theorem 1 provides convergence properties of $\{x^k\}$ for cases (depending on $\min\{\alpha, \beta\}$) when γ_k is chosen larger than 2.

5. Convergence rate

The following theorem will provide the basic means for obtaining global R–linear convergence rate results for the Variable Scaling DPRV Algorithm by specifying conditions on the parameters within the algorithm as well as on (α, β).

Theorem 2. *Let the assumptions of Lemma 4 be satisfied. If, in addition, $\alpha + \beta > 0$, A is single-valued and Lipschitz continuous with modulus $L > 0$, and $\{\gamma_k\} \in (0, 2]$, then there is $C > 0$ so that the iterates generated by the Variable Scaling DPRV Algorithm satisfy*

$$\|x^{k+1} - x^*\|^2 \leq C \prod_{i=0}^{k} ((1 + \tau_i)^2 - q_i) \tag{18}$$

for all $k \in \mathbb{N}$, where x^* denotes the unique solution of (1) and

$$q_i := \frac{\gamma_i(2\mu_i\alpha + \min\{2-\gamma_i,\ 2\mu_i\beta\}/2)}{(1+\mu_i L)^2} \quad \text{for } i \in \mathbb{N}.$$

Proof. By $\alpha + \beta > 0$ the operator $A + B$ is strongly monotone and, thus, the inclusion (1) has a unique solution. Throughout the proof let $k \in \mathbb{N}$ be arbitrary but fixed.

Noting that $e^2 + f^2 + 2ef \leq 2(e^2 + f^2)$ holds for all $e, f \in \mathbb{R}$ and having $x^k - x^* = x^k - y^k + y^k - x^*$ in mind, we obtain

$$\begin{aligned} \|x^k - x^*\|^2 &\leq \|x^k - y^k\|^2 + \|y^k - x^*\|^2 + 2\|x^k - y^k\|\|y^k - x^*\| \\ &\leq 2(\|x^k - y^k\|^2 + \|y^k - x^*\|^2) \end{aligned}$$

and, as a consequence,

$$-(2-\gamma_k)\|x^k - y^k\|^2 - 2\mu_k\beta\|y^k - x^*\|^2 \leq -\frac{1}{2}\min\{2-\gamma_k, 2\mu_k\beta\}\|x^k - x^*\|^2$$

since $\gamma_k \in (0, 2]$. Therefore, with the definition of Δ_k in mind, we get

$$-(2-\gamma_k)\|x^k - y^k\|^2 - 2\mu_k\Delta_k \leq -(2\mu_k\alpha + \frac{1}{2}\min\{2-\gamma_k, 2\mu_k\beta\})\|x^k - x^*\|^2.$$

Since A is assumed to be a Lipschitz continuous map with modulus L,

$$\|x^k - x^* + \mu_k(a^k - a^*)\| \leq \|x^k - x^*\| + \mu_k\|a^k - a^*\| \leq (1+\mu_k L)\|x^k - x^*\|$$

holds and, with this, we further have

$$-\gamma_k(2-\gamma_k)\|x^k - y^k\|^2 - 2\gamma_k\mu_k\Delta_k \leq -q_k\|x^k - x^* - \mu_k(a^k - a^*)\|.$$

By Lemma 4,

$$\|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \leq ((1+\tau_k)^2 - q_k)\|x^k - x^* + \mu_k(a^k - a^*)\|^2$$

follows. Hence,

$$\|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \leq \|x^0 - x^* + \mu_0(a^0 - a^*)\|^2 \prod_{i=0}^{k}\left((1+\tau_i)^2 - q_i\right)$$

must be valid. Exploiting the monotonicity of A, we finally get

$$\|x^{k+1} - x^*\|^2 \leq \|x^{k+1} - x^* + \mu_{k+1}(a^{k+1} - a^*)\|^2 \leq C\prod_{i=0}^{k}((1+\tau_i)^2 - q_i)$$

with $C := \|x^0 - x^* + \mu_0(a^0 - a^*)\|^2$. □

Remark 5. In this remark we consider cases where $\tau_k = 0$ for all $k \in \mathbb{N}$, i.e., $\mu_k := \mu$ is fixed throughout the algorithm. If we set $\gamma_k = 1$ for all $k \in \mathbb{N}$ (Douglas-Rachford method), then

$$\|x^k - x^*\|^2 \leq C\left(1 - \frac{2\mu\alpha}{(1+\mu L)^2}\right)^k$$

follows from Theorem 2. The rate of convergence just coincides with the one in [10]. For $\gamma_k = 2$ for all $k \in \mathbb{N}$ (Peaceman-Rachford) Theorem 2 yields the rate $1 - 4\mu\alpha(1+\mu L)^{-2}$. If a Lipschitz modulus L of A is known, then setting $\mu = 1/L$ yields

$$\|x^k - x^*\|^2 \leq C\left(1 - \gamma(\frac{\alpha}{4L} + \frac{1}{8}\min\{2-\gamma, 2\frac{\beta}{L}\})\right)^k.$$

Furthermore, according to Theorem 2, if A is Lipschitz continuous and B is strongly monotone then $\gamma_k = \gamma \in (0,2)$ for all $k \in \mathbb{N}$ provides an R–linear convergence rate less than 1. This observation seems absent in the literature.

6. Acknowledgements

The first author is greatly indebted to Prof. Bingsheng He for his insightful remarks concerning operator-splitting methods.

References

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

[2] J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators, Math. Program. 55 (1992) 293–318.

[3] J. Eckstein, M.C. Ferris, Operator splitting methods for monotone affine variational inequalities with a parallel application to optimal control, INFORMS J. Comput. 10 (1998) 218–235.

[4] J. Eckstein, B.F. Svaiter, A family of projective splitting methods for the sum of two maximal monotone operators, Math. Program. 111 (2008) 173–199.

[5] R. Glowinski, P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM, Philadelphia, PA, 1989

[6] B.S. He, L.Z. Liao, S.L. Wang, Self-adaptive operator splitting methods for monotone variational inequalities, Numer. Math. 94 (2003) 715–737.

[7] R.B. Kellogg, A nonlinear alternating direction method, Math. Comput. 23 (1969) 23–27.

[8] J. Lawrence, J.E. Spingarn, On fixed points of non-expansive piecewise isometric mappings, Proceedings of the London Mathematical Society, 3rd Series, vol. 55, Oxford University Press, Oxford, 1987, pp. 605–624.

[9] J. Lieutaud, Approximation d'opérateurs par des méthodes de décomposition, Doctoral Thesis, Université de Paris, 1969.

[10] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979) 964–979.
[11] G.J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29 (1962) 341–346.
[12] D. Peaceman, H.H. Rachford, The numerical solution of parabolic and elliptic differential equations, J. Soc. Indust. Appl. Math. 3 (1955) 28–41.
[13] J.E. Spingarn, Partial inverse of a monotone operator, Appl. Math. Optim. 10 (1983) 247–265.
[14] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976) 877–898.
[15] R.S. Varga, Matrix Iterative Analysis. Prentice-Hall, Englewood Cliffs, NJ, 1962.
[16] E.L. Wachspress, G.J. Habetler, An alternating direction implicit iteration technique, J. Soc. Indust. Appl. Math. 8 (1960) 403–424.