Convergence rate analysis for averaged fixed point iterations in the presence of Hölder regularity

Jonathan M. Borwein* Guoyin Li† Matthew K. Tam‡

October 23, 2015

Abstract

In this paper, we establish sublinear and linear convergence of fixed point iterations generated by averaged operators in a Hilbert space. Our results are achieved under a bounded Hölder regularity assumption which generalizes the well-known notion of bounded linear regularity. As an application of our results, we provide a convergence rate analysis for Krasnoselskii–Mann iterations, the cyclic projection algorithm, and the Douglas–Rachford feasibility algorithm along with some variants. In the important case in which the underlying sets are convex sets described by convex polynomials in a finite dimensional space, we show that the Hölder regularity properties are automatically satisfied, from which sublinear convergence follows.

MSC(2010). Primary 41A25, 90C25; Secondary 41A50, 90C31

Keywords. averaged operator; fixed point iteration; convergence rate; Hölder regularity; semi-algebraic; Douglas–Rachford algorithm

1 Introduction

Consider the problem of finding a point in the intersection of a finite family of closed convex subsets of a Hilbert space. This problem, often referred to as the *convex feasibility problem*, frequently arises throughout diverse areas of mathematics, science and engineering. For details, we refer the reader to the surveys [5, 20], the monographs [6, 23], any of [1, 13, 19], and the references therein.

One approach to solving convex feasibility problems involves designing a nonexpansive operator whose fixed point set can be used to easily produce a point in the target intersection (in the simplest case, the fixed point set coincides with the target intersection). The operator's fixed point iteration can then be used as the basis of an iterative algorithm which, in the limit, yields a desired solution. An important class of such methods is the so-called *projection and reflection methods* which employ various combinations of *projection* and *reflection* operations with respect underlying constraint sets. Notable methods of this kind include the *alternating projection algorithm* [4, 15, 24], the *Douglas–Rachford (DR) algorithm* [33, 34, 39], along with many extensions and variants [8, 17, 18, 43]. Even in settings without convexity [1–3, 14, 37, 38], such methods remain a popular choice due largely to their simplicity, ease-of-implementation and relatively – often surprisingly – good performance.

*CARMA, Univ. of Newcastle, Callaghan, NSW 2308, Australia. E-mail: jonathan.borwein@newcastle.edu.au.
†Dept. of Applied Maths, Univ. of New South Wales, Sydney, NSW 2052, Australia. E-mail: g.li@unsw.edu.au.
‡CARMA, Univ. of Newcastle, Callaghan, NSW 2308, Australia. E-mail: matthew.tam@uon.edu.au.

The origins of the Douglas–Rachford (DR) algorithm can be traced to [33] where it was used to solve problems arising in nonlinear heat flow. In its full generality, the method finds zeros of the sum of two maximal monotone operators. Weak convergence of the scheme was originally proven by Lions and Mercier [39], and the result was recently strengthened by Svaiter [41]. Specialized to feasibility problems, Svaiter's result implies that the iterates generated by the DR algorithm are always weakly convergent, and that the *shadow sequence* converges weakly to a point in the intersection of the two closed convex sets. The scheme has also been examined in [34] where its relationship with another popular method, the *proximal point algorithm*, was revealed and explained.

Motivated by the computational observation that the Douglas–Rachford algorithm sometimes outperforms other projection methods, in the convex case many researchers have studied the actual convergence rate of the algorithm. By *convergence rate*, we mean how *fast* the sequences generated by the algorithm converges to their limit points. For the Douglas–Rachford algorithm, the first such result, which appeared in [36] and was later extended by [9], showed that the algorithm converges linearly whenever the two constraint sets are closed subspaces with a closed sum, and, further, that the rate is governed exactly by the cosine of the *Friedrichs angle* between the subspaces. When the sum of the two subspaces is not closed, convergence of the method – while still assured – need not be linear [9, Sec. 6]. For most projection methods, it is typical that there exists instances in which the rate of convergence is arbitrarily slow and not even sublinear or arithmetic [11, 21]. Most recently, a preprint of Davis and Yin shows that indeed the Douglas–Rachford method may converge arbitrarily slowly in infinite dimensions [22, Th. 9].

In potentially nonconvex settings, a number of recent works [35, 36, 40] have established local linear convergence rates for the DR algorithm using commonly used constraint qualifications. When specialized to the convex case, these results state that the DR algorithm exhibits locally linear convergence for convex feasibility problems in a finite dimensional space whenever the relative interiors of the two convex sets have a non-empty intersection. On the other hand, when such a regularity condition is not satisfied, the DR algorithm can fail to exhibit linear convergence, even in simple two dimensional cases as observed by [10, Ex. 5.4(iii)] (see Section 6 for further examples and discussion). This therefore calls for further research aimed at answering the question: *Can the global convergence rate of the DR algorithm and its variants be established or estimated for some reasonable class of convex sets without the above mentioned regularity condition?*

The goal of this paper is to provide some partial answers to the above question, as well as simple tools for establishing sublinear and linear convergence of the Douglas–Rachford algorithm and variants. Our analysis is performed within the more general setting of *fixed point iterations* described by *averaged nonexpansive operators*. We pay special attention to the case in which the underlying sets are *convex basic semi-algebraic sets* in a finite dimensional space. Such sets comprise a broad sub-class of convex sets that we shall show satisfy *Hölder regularity properties* without requiring any further assumptions; they capture all polyhedra and all convex sets described by convex quadratic functions. Furthermore, convex basic semi-algebraic structure can often be relatively easily identified.

1.1 Content and structure of the paper

The detailed contributions of this paper are summarized as follows:

(I) We first examine an abstract algorithm which we refer to as the *quasi-cyclic algorithm*. This algorithm covers many iterative fixed-point methods including various Krasnoselskii–Mann iterations, the cyclic projection algorithm, and Douglas–Rachford feasibility algorithms. We show that the norm of the successive change in iterates of the quasi-cyclic algorithm converges with the order at least $o(1/\sqrt{t})$ (Proposition 3.2). This is a quantitative form of asymptotic regularity [5]. In the presence of so-called *bounded Hölder regular properties*, sublinear and linear convergence of the quasi-cyclic algorithm is then established (Theorem 3.6).

(II) We next specialise our results regarding the quasi-cyclic algorithm, in particular, to the Douglas–Rachford algorithm and its variants (Section 4). We show that the results apply, for instance, to the important case of feasibility problems for which the underlying sets are convex basis semi-algebraic in a finite dimensional space.

(III) We also examine a damped variant of the Douglas–Rachford algorithm. In the case where again the underlying sets are convex basic semi-algebraic sets in a finite dimensional space, we obtain a more *explicit estimate of the sublinear convergence rate* in terms of the dimension of the underlying space and the maximum degree of the polynomials involved (Theorem 5.5).

The remainder of the paper is organized as follows: in Section 2 we recall definitions and key facts used in our analysis. In Section 3 we investigate the rate of convergence of the *quasi-cyclic algorithm* – which encompasses an averaged fixed point iteration – in the presence of Hölder regularity. In Section 4 we specialize these results to the classical Douglas–Rachford algorithm and its cyclic variants. In Section 5 we consider a damped version of the Douglas–Rachford algorithm. In Section 6 we establish explicit convergence rates for two illustrative problems. We conclude the paper in Section 7 and mention some possible future research directions.

2 Preliminaries

Throughout this paper our setting is a (real) *Hilbert space* H with inner product $\langle \cdot, \cdot \rangle$. The *induced norm* is defined by $\|x\| := \sqrt{\langle x, x\rangle}$ for all $x \in H$. Given a closed convex subset A of H, the *(nearest point) projection* operator is the operator $\mathrm{P_A} : H \to A$ given by

$$\mathrm{P_A} x = \operatorname*{arg\,min}_{a \in A} \|x - a\|.$$

Let us now recall various definitions and facts used throughout this work, beginning with the notion of *Fejér monotonicity.*

Definition 2.1 (Fejér monotonicity)**.** *Let A be a non-empty convex subset of a Hilbert space H. A sequence $(x_k)_{k\in\mathbb{N}}$ in H is* Fejér monotone *with respect to A if, for all $a \in A$, we have*

$$\|x_{k+1} - a\| \leq \|x_k - a\| \quad \forall k \in \mathbb{N}.$$

Fact 2.2 (Shadows of Fejér monotone sequences [5, Th. 5.7(iv)])**.** *Let A be a non-empty closed convex subset of a Hilbert space H and let $(x_k)_{k\in\mathbb{N}}$ be Fejér monotone with respect to A. Then $\mathrm{P_A}(x_k) \to x$, in norm, for some $x \in A$.*

Fact 2.3 (Fejér monotone convergence [4, Th. 3.3(iv)])**.** *Let A be a non-empty closed convex subset of a Hilbert space H and let $(x_k)_{k\in\mathbb{N}}$ be Fejér monotone with respect to A with $x_k \to x \in A$, in norm. Then $\|x_k - x\| \leq 2\mathrm{dist}(x_k, A)$.*

We now turn our attention to a Hölder regularity property for usually finite collections of sets.

Definition 2.4 (Bounded Hölder regular intersection)**.** *Let $\{C_j\}_{j\in\mathbb{J}}$ be a collection of closed convex subsets in a Hilbert space H with non-empty intersection. The collection $\{C_j\}_{j\in\mathbb{J}}$ has a* bounded Hölder regular intersection *if, for each bounded set K, there exists an exponent $\gamma \in (0, 1]$ and a scalar $\beta > 0$ such that*

$$\operatorname{dist}(x, \cap_{j\in\mathbb{J}} C_j) \leq \beta \left(\max_{j\in\mathbb{J}} d(x, C_j) \right)^{\gamma} \quad \forall x \in K.$$

Furthermore, if the exponent γ does not depend on the set K, we say the collection $\{C_j\}_{j\in\mathbb{J}}$ is bounded Hölder regular with exponent γ *(*i.e., *we explicitly specify the exponent).*

It is clear, from Definition 2.4, that any collection containing only a single set trivially has a bounded Hölder regular intersection with exponent $\gamma = 1$. More generally, Definition 2.4 with $\gamma = 1$ is well-studied in the literature where it appears, amongst other names, as *bounded linear regularity* [5]. For a recent study, the reader is referred to [29, Remark 7]. The local counterpart to Definition 2.4 has been characterized in [29, Th. 1] under the name of *metric* $[\gamma]$*-subregularity.*

We next turn our attention to a nonexpansivity notion for operators.

Definition 2.5. *An operator $T : H \to H$ is:*

(a) non-expansive *if for all $x, y \in H$,*

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

(b) firmly non-expansive *if for all $x, y \in H$,*

$$\|T(x) - T(y)\|^2 + \|(I - T)(x) - (I - T)(y)\|^2 \leq \|x - y\|^2;$$

(c) α-averaged *for some $\alpha \in (0, 1)$, if there exists a non-expansive mapping $R : H \to H$ such that*

$$T = (1 - \alpha)I + \alpha R.$$

The class of firmly non-expansive mappings comprises precisely the 1/2-averaged mappings, and any α-averaged operator is non-expansive [6, Ch. 4]. The following fact provides a characterization of averaged maps that is useful for our purposes.

Fact 2.6 (Characterization of averaged maps [6, Prop. 4.25(iii)])**.** *Let $T : H \to H$ be an α-averaged operator on a Hilbert space with $\alpha \in (0, 1)$. Then, for all $x, y \in H$,*

$$\|T(x) - T(y)\|^2 + \frac{1 - \alpha}{\alpha} \|(I - T)(x) - (I - T)(y)\|^2 \leq \|x - y\|^2.$$

Denote the set of *fixed points* of an operator $T : H \to H$ by

$$\operatorname{Fix} T = \{x \in H \mid T(x) = x\}.$$

The following definition is of a Hölder regularity property for operators.

Definition 2.7 (Bounded Hölder regular operators)**.** *Let D be a subset of a Hilbert space H. An operator $T\colon H \to H$ is* bounded Hölder regular *if, for each bounded set $K \subseteq H$, there exists an exponent $\gamma \in (0,1]$ and a scalar $\mu > 0$ such that*

$$d(x, \operatorname{Fix} T) \leq \mu \|x - T(x)\|^{\gamma} \quad \forall x \in K.$$

Furthermore, if the exponent γ does not depend on the set K, we say that T is bounded Hölder regular with exponent γ *(*i.e., *we explicitly specify the exponent).*

Note that, in the case when $\gamma = 1$, Definition 2.7 collapses to the well studied concept of bounded linear regularity [5] and has been used in [8] to analyze linear convergence of algorithms involving non-expansive mappings. Moreover, it is also worth noting that if an operator T is bounded Hölder regular with exponent $\gamma \in (0,1]$ then the mapping $x \mapsto x - T(x)$ is bounded Hölder metric subregular with exponent γ. Hölder metric subregularity – which is a natural extension of metric subregularity – and Hölder type error bounds, have recently been studied in [28, 30–32].

Finally, we recall the definitions of *semi-algebraic functions* and *semi-algebraic sets*

Definition 2.8 (Semi-algebraic sets and functions [12])**.** *A set $D \subseteq \mathbb{R}^n$ is* semi-algebraic *if*

$$D := \bigcap_{j=1}^{s} \bigcup_{i=1}^{l} \{x \in \mathbb{R}^n \mid f_{ij}(x) = 0, h_{ij}(x) < 0\}$$

for some integers l, s and some polynomial functions f_{ij}, h_{ij} on $\mathbb{R}^n$ ($1 \leq i \leq l$, $1 \leq j \leq s$). A mapping $F\colon \mathbb{R}^n \to \mathbb{R}^p$ is said to be semi-algebraic *if its* graph, $\operatorname{gph} F := \{(x, F(x)) \mid x \in \mathbb{R}^n\}$*, is a semi-algebraic set in $\mathbb{R}^n \times \mathbb{R}^p$.*

The next fact summarises some fundamental properties of semi-algebraic sets and functions.

Fact 2.9 (Properties of semi-algebraic sets/functions)**.** *The following statements hold.*

(P1) Any polynomial is a semi-algebraic function.

(P2) Let D be a semi-algebraic set. Then $\operatorname{dist}(\cdot, D)$ is a semi-algebraic function.

(P3) If f, g are semi-algebraic functions on $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$ then $f + g$, λf, $\max\{f, g\}$, fg are semi-algebraic.

(P4) If f_i are semi-algebraic functions, $i = 1, \ldots, m$, and $\lambda \in \mathbb{R}$, then the sets $\{x \mid f_i(x) = \lambda, i = 1, \ldots, m\}$, $\{x \mid f_i(x) \leq \lambda,\ i = 1, \ldots, m\}$ are semi-algebraic sets.

(P5) If $F : \mathbb{R}^n \to \mathbb{R}^p$ and $G : \mathbb{R}^p \to \mathbb{R}^q$ are semi-algebraic mappings, then their composition $G \circ F$ is also a semi-algebraic mapping.

(P6) (Lojasiewicz's inequality) If ϕ, ψ are two continuous semi-algebraic functions on a compact semi-algebraic set $K \subseteq \mathbb{R}^n$ such that $\emptyset \neq \phi^{-1}(0) \subseteq \psi^{-1}(0)$ then there exist constants $c > 0$ and $\tau \in (0,1]$ such that

$$|\psi(x)| \leq c|\phi(x)|^{\tau} \quad \forall x \in K.$$

Proof. (P1) and (P4) follow directly from the definitions. See [12, Prop. 2.2.8] for (P2), [12, Prop. 2.2.6] for (P3) and (P5), and [12, Cor. 2.6.7] for (P6). □

Definition 2.10 (Basic semi-algebraic convex sets in $\mathbb{R}^n$)**.** *A set $C \subseteq \mathbb{R}^n$ is a* basic semi-algebraic convex *set if there exist $\gamma \in \mathbb{N}$ and convex polynomial functions, $g_j, j = 1, \ldots, \gamma$ such that $C = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, j = 1, \cdots, \gamma\}$.*

Any basic semi-algebraic convex set is clearly convex and semi-algebraic. On the other hand, there exist sets which are both convex and semi-algebraic but fail to be basic semi-algebraic convex set, see [15].

It transpires out that any finite collection of basic semi-algebraic convex sets has an intersection which is boundedly Hölder regular (without requiring further regularity assumptions). In the following lemma, $B(n)$ denotes the *central binomial coefficient* with respect to n given by $\binom{n}{[n/2]}$ where $[\cdot]$ denotes the integer part of a real number.

Lemma 2.11 (Hölder regularity of basic semi-algebraic convex sets in $\mathbb{R}^n$ [15])**.** *Let C_i be basic convex semi-algebraic sets in $\mathbb{R}^n$ given by $C_i = \{x \in \mathbb{R}^n \mid g_{ij}(x) \leq 0, j = 1, \ldots, m_i\}, i = 1, \ldots, m$ where g_{ij} are convex polynomials on $\mathbb{R}^n$ with degree at most d. Let $\theta > 0$ and $K \subseteq \mathbb{R}^n$ be a compact set. Then there exists $c > 0$ such that*

$$\operatorname{dist}^\theta(x, C) \leq c \left(\sum_{i=1}^{m} \operatorname{dist}^\theta(x, C_i) \right)^\gamma \quad \forall x \in K,$$

where $\gamma = \left[\min \left\{ \frac{(2d-1)^n + 1}{2}, \, B(n-1)d^n \right\} \right]^{-1}$.

We also recall the following useful recurrence relationship established in [15].

Lemma 2.12 (Recurrence relationship [15])**.** *Let $p > 0$, and let $\{\delta_t\}_{t \in \mathbb{N}}$ and $\{\beta_t\}_{t \in \mathbb{N}}$ be two sequences of nonnegative numbers such that*

$$\beta_{t+1} \leq \beta_t (1 - \delta_t \beta_t^p) \quad \forall t \in \mathbb{N}.$$

Then

$$\beta_t \leq \left(\beta_0^{-p} + p \sum_{i=0}^{t-1} \delta_i \right)^{-\frac{1}{p}} \quad \forall t \in \mathbb{N}.$$

[We use the convention that $\frac{1}{0} = +\infty$.]

3 The rate of convergence of the quasi-cyclic algorithm

In this section we investigate the rate of convergence of an abstract algorithm we call *quasi-cyclic*. To define the algorithm, let J be a finite set, and let $\{T_j\}_{j \in J}$ be a finite family of operators on a Hilbert space H. Given an initial point $x^0 \in H$, the quasi-cyclic algorithm generates a sequence according to

$$x^{t+1} = \sum_{j \in J} w_{j,t} T_j(x^t) \quad \forall t \in \mathbb{N}, \tag{1}$$

where $w_{j,t} \geq 0$ for all $j \in J$, and $\sum_{j \in J} w_{j,t} = 1$.

The quasi-cyclic algorithm was proposed in [8] where linear convergence of the algorithm was established under suitable regularity conditions. We note that, as we will see later, the quasi-cyclic algorithm provides a broad framework which covers many important existing algorithms such as Douglas-Rachford algorithms, cyclic projection algorithm and the Krasnoselskii–Mann method.

We first examine the complexity of the successive change of the sequence generated by the quasi-cyclic algorithm. To establish this, we prove a more general result which shows that the *the successive change* of a sequence generated by iterating averaged operators is at worst of order $o(1/\sqrt{t})$.

Proposition 3.1 (Complexity of the successive change)**.** *Let $\{\overline{T}_t\}_{t\in\mathbb{N}}$ be a family of α-averaged operators on a Hilbert space H with $\cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t \neq \emptyset$ and $\alpha \in (0,1)$. Let $x^0 \in H$ and set $x^{t+1} = \overline{T}_t x^t$ for all $t \in \mathbb{N}$. Then there exists a sequence $\epsilon_t \in [0,1]$ with $\epsilon_t \to 0$ such that*

$$\|x^{t+1} - x^t\| \le \frac{\epsilon_t}{\sqrt{t+1}}\sqrt{\frac{\alpha}{1-\alpha}}\mathrm{dist}(x^0, \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t) \quad \forall\, t \in \mathbb{N}.$$

Proof. First of all, if $x^0 \in \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t$, then $x^t = x^0$ for all $t \in \mathbb{N}$ and so the conclusion follows trivially. Thus, we can assume that $x^0 \notin \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t$. Fix $x^* \in \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t$. For all $t \in \mathbb{N}$, since $\overline{T}_t$ is α-averaged, Fact 2.6 with $x = x^t$ and $y = x^*$ gives us that

$$\|x^{t+1} - x^*\|^2 + \frac{1-\alpha}{\alpha}\|x^t - x^{t+1}\|^2 \le \|x^t - x^*\|^2.$$

This implies that, for all $t \in \mathbb{N}$, $\|x^{t+1} - x^*\| \le \|x^t - x^*\| \le \ldots \le \|x^0 - x^*\|$ and

$$\frac{1-\alpha}{\alpha}\sum_{t=0}^{\infty}\|x^t - x^{t+1}\|^2 \le \|x^0 - x^*\|^2.$$

Since $\|x^t - x^*\|^2$ is nonincreasing, we have that

$$\|x^{n+1} - x^n\|^2 \le \frac{1}{n+1}\sum_{t=0}^{n}\|x^t - x^{t+1}\|^2 \le \frac{1}{n+1}\left(\frac{\alpha}{1-\alpha}\right)\|x^0 - x^*\|^2.$$

Thus, by replacing n with t and taking infimum over all $x^* \in \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t$, we see that

$$\|x^{t+1} - x^t\| \le \frac{1}{\sqrt{t+1}}\sqrt{\frac{\alpha}{1-\alpha}}\mathrm{dist}(x^0, \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t) \tag{2}$$

Now, as $\sum_{t=0}^{+\infty}\|x^{t+1} - x^t\|^2 < +\infty$, observe that $\sum_{N=t}^{2t-1}\|x^{N+1} - x^N\|^2 \to 0$ as $t \to \infty$. Note that, for each $t \in \mathbb{N}$

$$\|x^{t+2} - x^{t+1}\| = \|\overline{T}_t x^{t+1} - \overline{T}_t x^t\| \le \|x^{t+1} - x^t\|,$$

where the last inequality follows from the fact that $\overline{T}_t$ is an α-averaged operator (and in particular, is nonexpansive). Then, we have

$$t\,\|x^{2t} - x^{2t-1}\|^2 \le \|x^{t+1} - x^t\|^2 + \ldots + \|x^{2t} - x^{2t-1}\|^2 = \sum_{N=t}^{2t-1}\|x^{N+1} - x^N\|^2 \to 0 \text{ as } t \to \infty.$$

This implies that $\sqrt{t}\,\|x^t - x^{t-1}\| \to 0$ as $t \to \infty$. Let

$$\epsilon_t := \frac{\|x^{t+1} - x^t\|}{\frac{1}{\sqrt{t+1}}\sqrt{\frac{\alpha}{1-\alpha}}\mathrm{dist}(x^0, \cap_{t\in\mathbb{N}}\mathrm{Fix}\,\overline{T}_t)}.$$

Then, $\epsilon_t \to 0$ as $t \to \infty$. Moreover, from (2), we see that $\epsilon_t \in [0,1]$. So, the conclusion follows. □

Corollary 3.2 (Complexity of quasi-cyclic algorithm)**.** *Let J be a finite set and let $\{T_j\}_{j\in J}$ be a finite family of α-averaged operators on a Hilbert space H with $\cap_{j\in J}\mathrm{Fix}\, T_j \neq \emptyset$ and $\alpha \in (0,1)$. For each $t \in \mathbb{N}$, let $w_{j,t} \in \mathbb{R}$, $j \in J$, be such that $w_{j,t} \geq 0$ and $\sum_{j\in J} w_{j,t} = 1$. Let $x^0 \in H$ and consider the quasi-cyclic algorithm generated by (1). Then there exists a sequence $\epsilon_t \in (0,1]$ with $\epsilon_t \to 0$ such that*

$$\|x^{t+1} - x^t\| \leq \frac{\epsilon_t}{\sqrt{t+1}}\sqrt{\frac{\alpha}{1-\alpha}}\mathrm{dist}(x^0, \cap_{j\in J}\mathrm{Fix}\, T_j) \quad \forall\, t \in \mathbb{N}.$$

Proof. Define $\overline{T}_t = \sum_{j\in J} w_{j,t}T_j$. As each T_j is an α-averaged operator, $w_{j,t} \geq 0$ and $\sum_{j\in J} w_{j,t} = 1$, it can be verified that $\overline{T}_t$ is also an α-averaged operator. Note that $\cap_{j\in J}\mathrm{Fix}\, T_j \subseteq \cap_{t\in\mathbb{N}}\overline{T}_t$. Thus the conclusion follows immediately by applying Proposition 3.1. □

Remark 3.3 (Comments on the convergence order of the successive change). Corollary 3.2 shows that the *norm square of the successive change*, $\|x^{t+1} - x^t\|^2$, generated by the quasi-cyclic algorithm is at worst of the order $o(1/t)$, whenever the operators T_j are averaged operators. As we shall see, this proposition applies, in particular, to the classical Douglas–Rachford algorithm where an $O(1/t)$ complexity of $\|x^{t+1} - x^t\|^2$ has been proved recently for the DR algorithm (see [26,27]). Herein, we obtain a slightly improved complexity result for a more general algorithm framework. Moreover, we note that, very recently, an $o(1/t)$ complexity of $\|x^{t+1} - x^t\|^2$ has also been established in [21] for the forward-Douglas–Rachford splitting method under an additional Lipschitz gradient assumption. We note that, as pointed out in [21, Section 1.4], this method is a special case of the Krasnoselskii–Mann iterative method, and so, is a particular case of the quasi-cyclic algorithm.

We also note that the norm square of the successive change, $\|x^{t+1} - x^t\|^2$, provides a straightforward numerical necessary condition for convergence of x^t to a point in $\cap_{j\in J}\mathrm{Fix}\, T_j$ or not. Of course $\|x^{t+1} - x^t\| \to 0$ does not necessarily guarantee that the sequence $\{x^t\}$ converges[1], and so, the convergence order of the successive change is not enough to establish convergence of $\{x^t\}$. ◇

To establish the convergence rate of the quasi-cyclic algorithm, we require the following two lemmas.

Lemma 3.4. *Let J be a finite set and let $\{T_j\}_{j\in J}$ be a finite family of α-averaged operators on a Hilbert space H with $\cap_{j\in J}\mathrm{Fix}\, T_j \neq \emptyset$ and $\alpha \in (0,1)$. For each $t \in \mathbb{N}$, let $w_{j,t} \in \mathbb{R}$, $j \in J$, be such that $w_{j,t} \geq 0$ and $\sum_{j\in J} w_{j,t} = 1$. Let $x^0 \in H$ and consider the quasi-cyclic algorithm generated by (1). Suppose that*

$$\sigma := \inf_{t\in\mathbb{N}} \inf_{j\in J_+(t)} \{w_{j,t}\} > 0 \text{ where } J_+(t) = \{j \in J : w_{j,t} > 0\} \text{ for each } t \in \mathbb{N}.$$

Then, for each $j \in J$, we have that $\|x^t - T_j(x^t)\| \to 0$ as $t \to \infty$.

Proof. Let $y \in \cap_{j\in J}\mathrm{Fix}T_j$. Then, for all $t \in \mathbb{N}$, convexity of $\|\cdot\|^2$ yields

$$\|x^{t+1} - y\|^2 = \|\sum_{j\in J} w_{j,t}T_j(x^t) - y\|^2 \leq \sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2 \leq \|x^t - y\|^2, \tag{3}$$

where the last inequality follows by the fact that each T_j is α-averaged (and so, is nonexpansive). Thus $(\|x^t - y\|^2)_{t\in\mathbb{N}}$ is a decreasing and hence convergence sequence. Furthermore,

$$\lim_{t\to\infty} \sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2 = \lim_{t\to\infty} \|x^t - y\|^2. \tag{4}$$

[1] A simple example is $0, \frac{1}{2}, 1, \frac{1}{3}, \frac{2}{3}, 1, \ldots, \frac{1}{t}, \frac{2}{t}, \ldots, \frac{t-1}{t}, 1, \ldots$

Since T_j is α-averaged for each $j \in J$, Fact 2.6 implies, for all $t \in \mathbb{N}$,

$$\|T_j(x^t) - y\|^2 + \frac{1-\alpha}{\alpha}\|x^t - T_j(x^t)\|^2 \le \|x^t - y\|^2.$$

from which, for sufficiently large t, we deduce

$$\frac{1-\alpha}{\alpha}\sigma\|x^t - T_j(x^t)\|^2 \le \frac{1-\alpha}{\alpha}\sum_{j\in J} w_{j,t}\|x^t - T_j(x^t)\|^2 \le \|x^t - y\|^2 - \sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2.$$

Together with (4), this implies $\|x^t - T_j(x^t)\| \to 0$ for all $j \in J$. □

The following proposition gives a convergence rate for Fejér monotone sequences which satisfy an additional property, which property we will later show is satisfied in the presence of Hölder regularity.

Proposition 3.5. *Let F be a non-empty closed convex set in a Hilbert space H. Suppose the sequence $\{x^t\}$ is Fejér monotone with respect to F and satisfies*

$$\operatorname{dist}^2(x^{t+1}, F) \le \operatorname{dist}^2(x^t, F) - \delta \operatorname{dist}^{2\theta}(x^t, F),\ \forall\, t \in \mathbb{N}, \tag{5}$$

for some $\delta > 0$ and $\theta \ge 1$. Then $x^t \to \bar{x}$ for some $\bar{x} \in F$. Moreover, there exist $M > 0$ and $r \in [0, 1)$ such that

$$\|x^t - \bar{x}\| \le \begin{cases} Mt^{-\frac{1}{2(\theta-1)}} & \theta > 1, \\ Mr^t & \theta = 1. \end{cases}$$

Further, when $\theta = 1$, δ necessarily lies in $(0, 1]$.

Proof. Let $\beta_t = \operatorname{dist}^2(x^t, F)$ and $p = \theta - 1 \ge 0$. Then (5) becomes

$$\beta_{t+1} \le \beta_t\Big(1 - \delta\beta_t^p\Big) \tag{6}$$

We now distinguish two cases based on the value of θ.

Case 1: Suppose $\theta \in (1, +\infty)$. Then Lemma 2.12 implies

$$\beta_t \le \Big(\beta_0^{-p} + (\theta - 1)\delta t\Big)^{-\frac{1}{\theta-1}} \quad \text{for all } \ t \in \mathbb{N}.$$

So, we see that, there exists $M_1 > 0$ such that $\operatorname{dist}(x^t, F) = \sqrt{\beta_t} \le M_1 t^{-\frac{1}{2(\theta-1)}}$. In particular, $\|x^t - \mathrm{P}_F(x^t)\| = \operatorname{dist}(x^t, F) \to 0$. By Fact 2.2, $\mathrm{P}_F(x^t) \to \bar{x}$ for some $\bar{x} \in F$ and hence $x^t \to \bar{x} \in F$. This together with Fact 2.3 implies that

$$\|x^t - \bar{x}\| \le 2\operatorname{dist}(x^t, F) \le 2M_1 t^{-\frac{1}{2(\theta-1)}}.$$

Case 2: Suppose $\theta = 1$. Then (6) simplifies to $\beta_{t+1} \le (1 - \delta)\beta_t$ for all $t \in \mathbb{N}$. Moreover, this shows that $\delta \in (0, 1]$ and that

$$\operatorname{dist}(x^t, F) = \sqrt{\beta_t} \le \sqrt{\beta_0}\left(\sqrt{1-\delta}\right)^t.$$

Then, by the same argument as used in Case 1, for some $\bar{x} \in F$, we see that

$$\|x^t - \bar{x}\| \le 2\operatorname{dist}(x^t, F) \le 2\sqrt{\beta_0}(\sqrt{1-\delta})^t.$$

The conclusion follows on setting $M = \max\{2M_1, 2\sqrt{\beta_0}\}$ and $r = \sqrt{1-\delta} \in [0, 1)$. □

We are now in a position to state our first main convergence result, which we simultaneously prove for both variants of the Hölder regularity assumption (with and without the Hölder exponents being independent of the choice of bounded set).

Theorem 3.6 (Rate of convergence of the quasi-cyclic algorithm)**.** *Let J be a finite set and let $\{T_j\}_{j\in J}$ be a finite family of α-averaged operators on a Hilbert space H with $\cap_{j\in J}\mathrm{Fix}\, T_j \neq \emptyset$ and $\alpha \in (0,1)$. For each $t \in \mathbb{N}$, let $w_{j,t} \in \mathbb{R}$, $j \in J$, be such that $w_{j,t} \geq 0$ and $\sum_{j\in J} w_{j,t} = 1$. Let $x^0 \in H$ and consider the quasi-cyclic algorithm generated by (1). Suppose the following assumptions hold:*

(a) For each $j \in J$, the operator T_j is bounded Hölder regular .
(b) $\{\mathrm{Fix}\ T_j\}_{j\in J}$ has a boundedly Hölder regular intersection.
(c) $\sigma := \inf\limits_{t\in\mathbb{N}} \inf\limits_{j\in J_+(t)} \{w_{j,t}\} > 0$ where $J_+(t) = \{j \in J : w_{j,t} > 0\}$ for each $t \in \mathbb{N}$.

Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \cap_{j\in J}\mathrm{Fix}\, T_j \neq \emptyset$ with sublinear rate $O(t^{-\rho})$.

Furthermore, if we assume the following stronger assumptions:

(a′) For each $j \in J$, the operator T_j is bounded Hölder regular with exponent $\gamma_{1,j} \in (0,1]$;
(b′) $\{\mathrm{Fix}\ T_j\}_{j\in J}$ has a bounded Hölder regular intersection with exponent $\gamma_2 \in (0,1]$.

Then there exist $M > 0$ and $r \in [0,1)$ such that

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0,1), \\ M\, r^t, & \gamma = 1, \end{cases}$$

where $\gamma := \gamma_1\gamma_2$ and $\gamma_1 := \min\{\gamma_{1,j} \mid j \in J\}$.

Proof. Suppose first that the assumptions (a), (b) and (c) hold. Since T_j is α-averaged for each $j \in J$, Fact 2.6 implies that, for all $x, y \in H$,

$$\|T_j(x) - T_j(y)\|^2 + \frac{1-\alpha}{\alpha}\|(I - T_j)(x) - (I - T_j)(y)\|^2 \leq \|x - y\|^2.$$

Set $F = \cap_{j\in J}\mathrm{Fix}\, T_j$. Then, for all $x \in H$ and for all $y \in F$,

$$\|T_j(x) - y\|^2 + \frac{1-\alpha}{\alpha}\|x - T_j(x)\|^2 \leq \|x - y\|^2.$$

Hence for all $x \in H$ and for all $y \in F$,

$$\begin{aligned} \|\sum_{j\in J} w_{j,t}T_j(x) - y\|^2 &= \|\sum_{j\in J} w_{j,t}\big(T_j(x) - y\big)\|^2 \\ &\leq \sum_{j\in J} w_{j,t}\|T_j(x) - y\|^2 \\ &\leq \sum_{j\in J} w_{j,t}\big(\|x - y\|^2 - \frac{1-\alpha}{\alpha}\|x - T_j(x)\|^2\big) \\ &= \|x - y\|^2 - \frac{1-\alpha}{\alpha}\sum_{j\in J} w_{j,t}\|x - T_j(x)\|^2 \\ &\leq \|x - y\|^2 - \sigma\left(\frac{1-\alpha}{\alpha}\right)\|x - T_j(x)\|^2 \quad \forall j \in J, \end{aligned}$$

where the first inequality follows from convexity of $\|\cdot\|^2$, and the final inequality follows from Assumption (c). Setting $x = x^t$ yields, for all $y \in F$,

$$\|x^{t+1} - y\|^2 \leq \|x^t - y\|^2 - \sigma\left(\frac{1-\alpha}{\alpha}\right)\|x^t - T_j(x^t)\|^2 \quad \forall j \in J.$$

In particular, the sequence $(x^t)_{t\in\mathbb{N}}$ is bounded and Fejér monotone with respect to F. Further, setting $y = \mathrm{P}_F(x^t)$ gives

$$\mathrm{dist}^2(x^{t+1}, F) \leq \|x^{t+1} - \mathrm{P}_F(x^t)\|^2 \leq \mathrm{dist}^2(x^t, F) - \sigma\left(\frac{1-\alpha}{\alpha}\right)\|x^t - T_j(x^t)\|^2 \quad \forall j \in J. \tag{7}$$

Let K be a bounded set such that $\{x^t \mid t \in \mathbb{N}\} \subseteq K$. For each $j \in J$, since the operator T_j is bounded Hölder regular, there exist exponents $\gamma_{1,j} > 0$ and scalars $\mu_j > 0$ such that

$$\mathrm{dist}(x, \mathrm{Fix}\, T_j) \leq \mu_j \|x - T_j(x)\|^{\gamma_{1,j}} \quad \forall x \in K. \tag{8}$$

Setting $\gamma_1 = \min\{\gamma_{1,j} \mid j \in J\}$ gives

$$\|x - T_j(x)\|^{\gamma_{1j}} \leq \|x - T_j(x)\|^{\gamma_1} \quad \forall x \in K \cap \{x \mid \|x - T_j(x)\| \leq 1\}.$$

By Lemma 3.4, there exists $t_0 \in \mathbb{N}$ such that $\|x^t - T_j(x^t)\| \leq 1$ for all $t \geq t_0$ and $j \in J$. Hence, for all $t \geq t_0$, it follows from (7) and (8) that

$$\sigma\left(\frac{1-\alpha}{\alpha}\right)\mu_j^{-\frac{2}{\gamma_1}}\mathrm{dist}^{\frac{2}{\gamma_1}}(x^t, \mathrm{Fix}\, T_j) \leq \mathrm{dist}^2(x^t, F) - \mathrm{dist}^2(x^{t+1}, F) \quad \forall j \in J.$$

Taking the maximum over all $j \in J$ and letting $\mu = \max\{\mu_j \mid j \in J\}$, we have

$$\sigma\left(\frac{1-\alpha}{\alpha}\right)\mu^{-\frac{2}{\gamma_1}}\max_{j\in J}\mathrm{dist}^{\frac{2}{\gamma_1}}(x^t, \mathrm{Fix}\, T_j) \leq \mathrm{dist}^2(x^t, F) - \mathrm{dist}^2(x^{t+1}, F). \tag{9}$$

Since $\{\mathrm{Fix}\ T_j\}_{j\in J}$ has a bounded Hölder regular intersection, there exists $\beta > 0$ such that

$$\frac{1}{\beta}\mathrm{dist}^{2\theta}(x^t, F) \leq \left(\max_{j\in J}\mathrm{dist}(x^t, \mathrm{Fix}\ T_j)\right)^{2/\gamma_1} = \max_{j\in J}\mathrm{dist}^{2/\gamma_1}(x^t, \mathrm{Fix}\ T_j), \tag{10}$$

where $\theta := 1/(\gamma_1\gamma_2) \geq 1$. Altogether, combining (9) and (10), we see that, for all $t \geq t_0$,

$$\mathrm{dist}^2(x^{t+1}, F) \leq \mathrm{dist}^2(x^t, F) - \delta\,\mathrm{dist}^{2\theta}(x^t, F),$$

where $\delta := \sigma(\frac{1-\alpha}{\alpha})\mu^{-\frac{2}{\gamma_1}}\beta^{-1} > 0$. Then, the first assertion follows from Proposition 3.5.

To see the second assertion, we suppose that the assumptions (a′), (b′) and (c) hold. Proceed with the same proof as above, and note that the exponent γ_{1j} and γ_2 are now independent of the choice of K, we see that the second assertion also follows. □

Remark 3.7. Theorem 3.6 is a generalization of [8, Th. 6.1] which considered the case in which the Hölder exponents are independent of the bounded set K and given by $\gamma_{1j} = \gamma_2 = 1$, $j = 1, \ldots, m$. ◇

We next provide three important specializations of Theorem 3.6. The first result is concerned with a simple fixed point iteration, the second with a Kransnoselskii–Mann scheme, and the third with the method of cyclic projections.

Corollary 3.8 (Averaged fixed point iterations with Hölder regularity)**.** *Let T be an α-averaged operators on a Hilbert space H with $\operatorname{Fix} T \neq \emptyset$ and $\alpha \in (0,1)$. Suppose T is bounded Hölder regular. Let $x^0 \in H$ and set $x^{t+1} = Tx^t$. Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \operatorname{Fix} T \neq \emptyset$ with sublinear rate $O(t^{-\rho})$. Furthermore, if T is bounded Hölder regular with exponent $\gamma \in (0,1]$ then there exist $M > 0$ and $r \in [0,1)$ such that*

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0,1), \\ M\, r^t, & \gamma = 1. \end{cases}$$

Proof. The conclusion follows immediately from Theorem 3.6. □

Corollary 3.9 (Krasnoselskii–Mann iterations with Hölder regularity)**.** *Let T be an α-averaged operator on a Hilbert space H with $\operatorname{Fix} T \neq \emptyset$ and $\alpha \in (0,1)$. Suppose T is bounded Hölder regular. Let $\sigma_0 \in (0,1)$ and let $(\lambda_t)_{t\in\mathbb{N}}$ be a sequence of real numbers with $\sigma_0 := \inf_{t\in\mathbb{N}}\{\lambda_t(1-\lambda_t)\} > 0$. Given an initial point $x^0 \in H$, set*

$$x^{t+1} = x^t + \lambda_t(Tx^t - x^t).$$

Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \operatorname{Fix} T \neq \emptyset$ with sublinear rate $O(t^{-\rho})$. Furthermore, if T is bounded Hölder regular with exponent $\gamma \in (0,1]$ then there exist $M > 0$ and $r \in [0,1)$ such that

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0,1), \\ M\, r^t, & \gamma = 1. \end{cases}$$

Proof. First observe that the sequence $(x^t)_{t\in\mathbb{N}}$ is given by $x^{t+1} = T_t x^t$ where $T_t = (1-\lambda_t)I + \lambda_t T$. Here, $1 - \lambda_t \geq \sigma_0 > 0$ and $\lambda_t \geq \sigma_0 > 0$ for all $t \in \mathbb{N}$ by our assumption.

A straightforward manipulation shows that the identity map, I, is bounded Hölder regular with exponent $\gamma_{1,1} \leq 1$. Since $\operatorname{Fix} I = H$, the collection $\{\operatorname{Fix} I, \operatorname{Fix} T\}$ has a bounded Hölder regular intersection with exponent 1. The result now follows from Theorem 3.6. □

The following result includes [15, Th. 4.4] and [5, Th. 3.12] as a special cases.

Corollary 3.10 (Cyclic projection algorithm with Hölder regularity)**.** *Let $J = \{1, 2, \ldots, m\}$ and let $\{C_j\}_{j\in J}$ a collection of closed convex subsets of a Hilbert space H with non-empty intersection. Given $x^0 \in H$ set*

$$x^{t+1} = \mathrm{P}_{C_j} x^t \text{ where } j = t \bmod m.$$

Suppose that $\{C_j\}_{j\in J}$ has a bounded Hölder regular intersection. Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \cap_{j\in J} C_j \neq \emptyset$ with sublinear rate $O(t^{-\rho})$. Furthermore, if the collection $\{C_j\}_{j\in J}$ is bounded Hölder regular with exponent $\gamma \in (0,1]$ there exist $M > 0$ and $r \in [0,1)$ such that

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0,1), \\ M\, r^t, & \gamma = 1. \end{cases}$$

Proof. First note that the projection operator over a closed convex set is 1/2-averaged. Now, for each $j \in J$, $C_j = \operatorname{Fix} \mathrm{P}_{C_j}$, and hence

$$d(x, C_j) = d(x, \operatorname{Fix} \mathrm{P}_{C_j}) = \|x - \mathrm{P}_{C_j} x\| \quad \forall x \in H.$$

That is, for each $j \in J$, the projection operator P_{C_j} is bounded Hölder regular with exponent 1. The result follows from Theorem 3.6. □

4 The rate of convergence of DR algorithms

We now specialize our convergence results to the classical DR algorithm and its variants, and so obtain a convergence rate under the Hölder regularity condition. Recall that the basic *Douglas–Rachford algorithm* for two set feasibility problems can be stated as follows:

Algorithm 1: Basic Douglas–Rachford algorithm

Data: Two closed and convex sets $C, D \subseteq H$
Choose an initial point $x^0 \in H$;
for $t = 0, 1, 2, 3, \ldots$ **do**
  Set:

$$\begin{cases} y^{t+1} := \mathrm{P}_C(x^t), \\ z^{t+1} := \mathrm{P}_D(2y^{t+1} - x^t), \\ x^{t+1} := x^t + (z^{t+1} - y^{t+1}). \end{cases} \tag{11}$$

end

Direct verification shows that the relationship between consecutive terms in the sequence (x^t) of (11) can be described in terms of the firmly nonexpansive *(two-set) Douglas–Rachford operator* which is of the form

$$T_{C,D} = \frac{1}{2} \left(I + R_D R_C \right), \tag{12}$$

where I is the identity mapping and $R_C := 2P_D - I$ is the *reflection operator* with respect to the set C ('reflect-reflect-average') .

We shall also consider the following abstraction which chooses two constraint sets from some finite collection at each iteration. Note that iterations (11) and (13) have the same structure.

Algorithm 2: A multiple-sets Douglas–Rachford algorithm

Data: A family of m closed and convex sets $C_1, C_2, \ldots, C_m \subseteq H$
Choose a list of 2-tuples $\Omega_1, \ldots, \Omega_s \in \{(i,j) : i, j = 1, 2, \ldots, m \text{ and } i \neq j\}$ with $\cup_{j=1}^{s} \Omega_j = \{1, \ldots, m\}$;
Choose an initial point $x^0 \in H$;
for $t = 0, 1, 2, 3, \ldots$ **do**
  Set the indices $(i, j) := \Omega_{t'}$ where $t' = t \bmod m$;
  Set

$$\begin{cases} y^{t+1} := \mathrm{P}_{C_i}(x^t), \\ z^{t+1} := \mathrm{P}_{C_j}(2y^{t+1} - x^t), \\ x^{t+1} := x^t + (z^{t+1} - y^{t+1}). \end{cases} \tag{13}$$

end

The motivation for studying Algorithm 2 is that, beyond Algorithm 1, it include two further DR-type schemes from the literature. The first scheme is the *cyclic DR algorithm* and is generated according to:

$$x^{t+1} = (T_{C_m,C_1}\, T_{C_{m-1},C_m}\, \dots\, T_{C_2,C_3}\, T_{C_1,C_2})(x^t) \quad \forall t \in \mathbb{N},$$

which corresponds the Algorithm 2 with $s = m$ and $\Omega_j = (j, j+1)$, $j = 1, \dots, m-1$, and $\Omega_m = (m, 1)$. The second scheme the *cyclically anchored DR algorithm* and is generated according to:

$$x^{t+1} = (T_{C_1,C_m}\, \dots\, T_{C_1,C_3}\, T_{C_1,C_2})(x^t) \quad \forall t \in \mathbb{N},$$

which corresponds the Algorithm 2 with $s = m-1$ and $\Omega_j = (1, j+1)$, $j = 1, \dots, m-1$. The following lemma shows that underlying operators both these methods are also averaged.

Lemma 4.1 (Compositions of DR operators)**.** *Let s be a positive integer. The composition of s Douglas–Rachford operators is $\frac{s}{s+1}$–averaged.*

Proof. The two-set Douglas–Rachford operator of (12) is firmly nonexpansive, and hence 1/2-averaged. The result follows by [6, Prop. 4.32]. □

As a consequence of Lemma 4.1 and Proposition 3.2, we now obtain the complexity of the successive change for the many set DR algorithm. For convenience, in the following result, for a set $\Omega_j = \{(i_1^j, i_2^j)\}$ with indices $i_1^j, i_2^j \in \{1, \dots, m\}$, $j = 1, \dots, s$, we denote

$$T_{\Omega_j} := T_{C_{i_1}^j, C_{i_2}^j}.$$

Corollary 4.2 (Complexity of the multiple-sets DR algorithm)**.** *Let $C_1, C_2, \dots, C_m$ be closed convex sets in a Hilbert space H with non-empty intersection. Let $\{\Omega_j\}_{j=1}^s$ and $\{(y^t, z^t, x^t)\}$ be as used by the multiple-sets Douglas–Rachford algorithm (13). Then, for each $t \in \mathbb{N}$, there exists a sequence $\epsilon_t \to 0$ with $0 \le \epsilon_t \le 1$ such that*

$$\|x^{t+1} - x^t\| \le \frac{\epsilon_t \left(\sqrt{s}\, \mathrm{dist}(x^0, \cap_{j=1}^s \mathrm{Fix}\, T_{\Omega_j}) \right)}{\sqrt{t+1}}.$$

Proof. Let $J = \{1\}$, $\overline{T}_1 = T_{\Omega_s} T_{\Omega_{s-1}} \dots T_{\Omega_1}$ and $w_{t,1} \equiv 1$. From Lemma 4.1, $\overline{T}_1$ is $s/(s+1)$-averaged.

Note that $\cap_{j=1}^s \mathrm{Fix}\, T_{\Omega_j} \subseteq \mathrm{Fix}\, \overline{T}_1$. Thus, the conclusion follows by applying Proposition 3.1 with $\alpha = s/(s+1)$. □

Remark 4.3. The previous corollary holds with $s = m$ for the cyclic DR algorithm, and with $s = (m-1)$ for the cyclically anchored DR algorithm. ◇

Corollary 4.4 (Convergence rate for the multiple-sets DR algorithm)**.** *Let $C_1, C_2, \dots, C_m$ be closed convex sets in a Hilbert space H with non-empty intersection. Let $\{\Omega_j\}_{j=1}^s$ and $\{(y^t, z^t, x^t)\}$ be generated by the multiple-sets Douglas–Rachford algorithm (11). Suppose that:*

(a) For each $j \in \{1, \dots, s\}$, the operator T_{Ω_j} is bounded Hölder regular.

(b) The collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^s$ has a bounded Hölder regular intersection.

Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \cap_{j=1}^s \mathrm{Fix}\, T_{\Omega_j}$ with sublinear rate $O(t^{-\rho})$. Furthermore, suppose we assume the stronger assumptions:

(a′) For each $j \in \{1,\ldots,s\}$, the operator T_{Ω_j} is bounded Hölder regular with exponent $\gamma_{1,j}$.

(b′) The collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{s}$ has a bounded Hölder regular intersection with exponent $\gamma_2 \in (0,1]$.

Then there exist $M > 0$ and $r \in (0,1)$ such that

$$\|x^t - \bar{x}\| \le \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}} & \text{if} \quad \gamma \in (0,1), \\ M\, r^t & \text{if} \quad \gamma = 1. \end{cases}$$

where $\gamma := \gamma_1\gamma_2$ where $\gamma_1 := \min\{\gamma_{1,j} \mid 1 \le j \le s\}$.

Proof. Let $J = \{1,2,\ldots,s\}$. For for all $j \in J$, set $T_j = T_{\Omega_j}$ and

$$w_{t,j} \equiv \begin{cases} 1 & j = t \bmod m, \\ 0 & \text{otherwise.} \end{cases}.$$

Since T_{Ω_j} is firmly nonexpansive (that is, 1/2-averaged), the conclusion follows from Corollary 4.4. □

We next observe that bounded Hölder regularity of the Douglas-Rachford operator T_{Ω_j} and Hölder regular intersection of the collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{s}$ are automatically satisfied for the semi-algebraic convex case, and so, sublinear convergence analysis follows in this case without any further regularity conditions. This follows from:

Proposition 4.5 (Semi-algebraicity implies Hölder regularity & sublinear convergence)**.** *Let C_1, C_2, …, C_m be basic convex semi-algebraic sets in $\mathbb{R}^n$ with non-empty intersection, given by*

$$C_j = \{x \in \mathbb{R}^n \mid g_{ij}(x) \le 0, i = 1,\ldots,m_j\}$$

where g_{ij}, $i = 1,\ldots,m_j$, $j = 1,\ldots,m$, are convex polynomials on $\mathbb{R}^n$ with degree d. Let $\{\Omega_j\}_{j=1}^{s}$ be a list of 2-tuples with $\cup_{j=1}^{s}\Omega_j = \{1,\ldots,m\}$ Then,

(a) For each $j \in \{1,\ldots,s\}$, the operator T_{Ω_j} is bounded Hölder regular. Moreover, if $d = 1$, then T_{Ω_j} is bounded Hölder regular with exponent 1.

(b) The collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{s}$ has a bounded Hölder regular intersection.

In particular, let $\{(y^t, z^t, x^t)\}$ be generated by the multiple-sets Douglas–Rachford algorithm (11). Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \cap_{j=1}^{s}\mathrm{Fix}\, T_{\Omega_j}$ with sublinear rate $O(t^{-\rho})$.

Proof. Fix any $j \in \{1,\ldots,s\}$. We first verify that the operator T_{Ω_j} is bounded Hölder regular. Without loss of generality, we assume that $\Omega_j = \{1,2\}$ and so, $T_{\Omega_j} = T_{C_1,C_2}$ where T_{C_1,C_2} is the Douglas-Rachford operator for C_1 and C_2. Recall that for each $x \in \mathbb{R}^n$, $T_{C_1,C_2}(x) - x = \mathrm{P}_{C_2}(\mathrm{R}_{C_1}(x)) - \mathrm{P}_{C_1}(x)$. We now distinguish two cases depending on the value of the degree d of the polynomials which describes C_j, $j = 1,2$.

Case 1 ($d > 1$): We first observe that, for a closed convex semi-algebraic set $C \subseteq \mathbb{R}^n$, the projection mapping $x \mapsto \mathrm{P}_C(x)$ is a semi-algebraic mapping. This implies that, for $i = 1,2$, $x \mapsto \mathrm{P}_{C_i}(x)$ and $x \mapsto \mathrm{R}_{C_i}(x) = 2\mathrm{P}_{C_i}(x) - x$ are all semi-algebraic mappings. Since the composition of semi-algebraic maps remains semi-algebraic ((P5) of Fact 2.9), we deduce that $f\colon x \mapsto \|T_{C_1,C_2} - x\|^2$ is a continuous semi-algebraic function. By (P4) of Fact 2.9, $\mathrm{Fix}\, T_{C_1,C_2} = \{x \mid f(x) = 0\}$ which is therefore a semi-algebraic set. By (P2) of Fact 2.9, the function $\operatorname{dist}(\cdot, \operatorname{Fix} T_{C_1,C_2})$ is semi-algebraic, and clearly $\operatorname{dist}(\cdot, \operatorname{Fix} T_{C_1,C_2})^{-1}(0) = f^{-1}(0)$.

By the Łojasiewicz inequality for semi-algebraic functions ((P6) of Fact 2.9), we see that for every $\rho > 0$, one can find $\mu > 0$ and $\gamma \in (0, 1]$ such that

$$\operatorname{dist}(x, \operatorname{Fix} \mathrm{T}_{\mathrm{C}_1,\mathrm{C}_2}) \leq \mu \|x - T_{C_1,C_2}x\|^{\gamma} \quad \forall x \in \mathbb{B}(0, \rho).$$

So, the Douglas–Rachford operator T_{C_1,C_2} is bounded Hölder regular in this case.

Case 2 ($d = 1$): In this case, both C_1 and C_2 are polyhedral, hence their projections P_{C_1} and P_{C_2} are piecewise affine mappings. Noting that composition of piecewise affine mappings remains piecewise affine [42], we deduce that $F\colon x \mapsto T_{C_1,C_2}(x) - x$ is continuous and piecewise affine. Then, Robinson's theorem on metric subregularity of piecewise affine mappings [44] implies that for all $a \in \mathbb{R}^n$, there exist $\mu > 0$, $\epsilon > 0$ such that

$$\operatorname{dist}(x, \operatorname{Fix} T_{C_1,C_2}) = \operatorname{dist}(x, F^{-1}(0)) \leq \mu \|F(x)\| = \mu \|x - T_{C_1,C_2}(x)\| \quad \forall x \in \mathbb{B}(a, \epsilon).$$

Then, a standard compactness argument shows that the Douglas–Rachford operator T_{C_1,C_2} is bounded linear regular, that is, bounded Hölder regular with exponent 1.

Next, we assert that the collection $\{\operatorname{Fix} T_{\Omega_j}\}_{j=1}^{s}$ has a bounded Hölder regular intersection. To see this, as in the proof of part (a), we can show that for each $j = 1, \ldots, s$, $\operatorname{Fix} T_{\Omega_j}$ is a semi-algebraic set. Then, their intersection $\cap_{j=1}^{s} \operatorname{Fix} T_{\Omega_j}$ is also a semi-algebraic set. Thus, $\psi(x) = \operatorname{dist}(x, \cap_{j=1}^{s} \operatorname{Fix} T_{\Omega_j})$ and $\phi(x) = \max_{1 \leq j \leq s} \operatorname{dist}(x, \operatorname{Fix} T_{\Omega_j})$ are semi-algebraic functions. It is easy to see that $\phi^{-1}(0) = \psi^{-1}(0)$ and hence the Łojasiewicz inequality for semi-algebraic functions ((P6) of Fact 2.9) implies that the collection $\{\operatorname{Fix} T_{\Omega_j}\}_{j=1}^{s}$ has a bounded Hölder regular intersection.

The final conclusion follows by Theorem 3.6. □

Next, we establish the convergence rate for DR algorithm assuming bounded Hölder regularity of the Douglas–Rachford operator $T_{C,D}$. Indeed, by supposing that $T_{C,D}$ is bounded Hölder regular with exponent $\gamma \in (0, 1]$, it is immediate from Proposition 3.2 and the firmly nonexpansive property of $T_{C,D}$ that $\operatorname{dist}(x^t, \operatorname{Fix} \mathrm{T}_{\mathrm{C,D}})$ converges to 0 at the order of $o(t^{-\frac{\gamma}{2}})$. Below, under the same assumption we show a stronger result: x^t converges to $\bar{x}$ for some $\bar{x} \in \operatorname{Fix} T_{C,D}$ at least in the order of $O(t^{-\frac{\gamma}{2(1-\gamma)}})$ if $\gamma \in (0, 1)$, and x^t converges to $\bar{x}$ linearly if $\gamma = 1$.

Corollary 4.6 (Convergence rate for the DR algorithm)**.** *Let C, D be two closed convex sets in a Hilbert space H with $C \cap D \neq \emptyset$, and let $T_{C,D}$ be the Douglas–Rachford operator. Let $\{(y^t, z^t, x^t)\}$ be generated by the Douglas–Rachford algorithm* (11)*. Suppose that $T_{C,D}$ is bounded Hölder regular. Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \operatorname{Fix} T_{C,D}$ with sublinear rate $O(t^{-\rho})$. Furthermore, if $T_{C,D}$ is bounded Hölder regular with exponent $\gamma \in (0, 1]$ then there exist $M > 0$ and $r \in (0, 1)$ such that*

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}} & \text{if} \quad \gamma \in (0, 1), \\ M\, r^t & \text{if} \quad \gamma = 1. \end{cases}$$

Proof. Let $J = \{1\}$, $T_1 = T_{C,D}$ and $w_{t,1} \equiv 1$. Note that $T_{C,D}$ is firmly nonexpansive (that is, 1/2-averaged) and any collection containing only one set has Hölder regular intersection with exponent one. Then the conclusion follows immediately from Theorem 3.6. □

Similar to Proposition 4.5, if C and D are basic convex semi-algebraic sets, then DR algorithm exhibits a sublinear convergence rate.

Remark 4.7 (Linear convergence of the DR algorithm). We note that if $H = \mathbb{R}^n$ and $\mathrm{ri}C \cap \mathrm{riD} \neq \emptyset$, then $T_{C,D}$ is bounded linear regular, that is, bounded Hölder regular with exponent 1. Thus, the Douglas–Rachford algorithm converges linearly in this case. This had been shown in [8]. Also, if C and D are both subspaces such that $C + D$ is closed (as is automatic in finite-dimensions), then $T_{C,D}$ is also bounded linear regular, and so, the DR algorithm converges linearly in this case as well. This was been established in [9]. It should be noted that [9] deduced the stronger result that the linear convergence rate is exactly the cosine of the Friedrichs angle. ◇

5 The rate of convergence of the damped DR algorithm

We now investigate a variant of Algorithm 1 which we refer to as the *damped Douglas–Rachford algorithm*. To proceed, let $\eta > 0$, let A be a closed convex set in H, and define the operator P_A^η by

$$\mathrm{P}_A^\eta = \left(\frac{1}{2\eta+1} I + \frac{2\eta}{2\eta+1} \mathrm{P}_A\right),$$

where I denotes the identity operator on H. The operator P_A^η can be considered as a relaxation of the projection mapping. Further, a direct verification shows that

$$\lim_{\eta \to \infty} \mathrm{P}_A^\eta(x) = \mathrm{P}_A(x) \quad \forall x \in H,$$

in norm, and

$$\mathrm{P}_A^\eta(x) = \mathrm{prox}_{\eta \mathrm{dist}_A^2}(x) = \arg\min_{y \in H} \left\{ \mathrm{dist}_A^2(y) + \frac{1}{2\eta}\|y - x\|^2 \right\} \quad \forall x \in H, \tag{14}$$

where prox_f denotes the *proximity operator* of the function f. The damped variant can be stated as follows:

Algorithm 3: Damped Douglas–Rachford algorithm

Data: Two closed sets $C, D \subseteq H$
Choose $\eta > 0$ and step-size parameters $\lambda_t \in (0, 2]$;
Choose an initial point $x^0 \in H$;
for $t = 0, 1, 2, 3, \ldots$ **do**
Choose $\lambda_t \in (0, 2]$ with $\lambda_t \geq \lambda$ and set:

$$\begin{cases} y^{t+1} := \mathrm{P}_C^\eta(x^t), \\ z^{t+1} := \mathrm{P}_D^\eta(2y^{t+1} - x^t), \\ x^{t+1} := x^t + \lambda_t(z^{t+1} - y^{t+1}). \end{cases} \tag{15}$$

end

Remark 5.1. Algorithm 2 can be found, for instance, in [6, Cor. 27.4]. A similar relaxation of the Douglas–Rachford algorithm for lattice cone constraints has been proposed and analyzed in [16]. ◇

Whilst it is possible to analyze the damped Douglas–Rachford algorithm within the quasi-cyclic framework, we learn more by proving the following result directly.

Theorem 5.2 (Convergence Rate for the damped Douglas–Rachford algorithm)**.** *Let C, D be two closed and convex sets in a Hilbert space H with $C \cap D \neq \emptyset$. Let $\lambda = \inf_{t\in\mathbb{N}} \lambda_t > 0$ with $\lambda_t \in (0, 2]$ and let $\{(y^t, z^t, x^t)\}$ be generated by the damped Douglas–Rachford algorithm (15). Suppose that the pair of sets $\{C, D\}$ has a bounded Hölder regular intersection. Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in C \cap D$ with sublinear rate $O(t^{-\rho})$. Furthermore, if the pair $\{C, D\}$ has a bounded Hölder regular intersection with exponent $\gamma \in (0, 1]$ then there exist $M > 0$ and $r \in (0, 1)$ such that*

$$\|x^t - \bar{x}\| \leq \begin{cases} M t^{-\frac{\gamma}{2(1-\gamma)}} & \text{if} \quad \gamma \in (0,1) \\ M\, r^t & \text{if} \quad \gamma = 1. \end{cases}$$

Proof. Step 1 (A Fejér monotonicity type inequality for x^t): Let $x^* \in C \cap D$. We first show that

$$2\eta\lambda\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \leq \|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2. \tag{16}$$

To see this, note that, for any closed and convex set A, $\mathrm{dist}^2(\cdot, A)$ is a differentiable convex function satisfying $\nabla(\mathrm{dist}^2)(x, A) = 2(x - \mathrm{P}_A(x))$ which is 2-Lipschitz. Using the convex subgradient inequality, we have

$$\begin{aligned}
& 2\eta\lambda\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \\
\leq\ & 2\eta\lambda_t\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \\
=\ & 2\eta\lambda_t\big(\mathrm{dist}^2(y^{t+1}, C) - \mathrm{dist}^2(x^*, C) + \mathrm{dist}^2(z^{t+1}, D) - \mathrm{dist}^2(x^*, D)\big) \\
\leq\ & 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - x^*\rangle\big) \\
=\ & 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - y^{t+1}\rangle \\
& \quad + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle\big) \\
=\ & 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}) + z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - y^{t+1}\rangle\big) \\
=\ & 4\eta\big(\lambda_t\langle y^{t+1} - \mathrm{P_C}(y^{t+1}) + z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), x^{t+1} - x^t\rangle\big)
\end{aligned} \tag{17}$$

where the last equality follows from the last relation in (15). Now using (14), we see that

$$\begin{aligned}
0 &= \nabla(\mathrm{dist}^2(\cdot, C) + \frac{1}{2\eta}\|\cdot - x^t\|^2)(y^{t+1}) = 2\left(y^{t+1} - \mathrm{P_C}(y^{t+1})\right) + \frac{1}{\eta}(y^{t+1} - x^t), \text{ and} \\
0 &= \nabla(\mathrm{dist}^2(\cdot, D) + \frac{1}{2\eta}\|\cdot - (2y^{t+1} - x^t)\|^2)(z^{t+1}) = 2\left(z^{t+1} - \mathrm{P_D}(z^{t+1})\right) + \frac{1}{\eta}(z^{t+1} - 2y^{t+1} + x^t).
\end{aligned}$$

Summing these two equalities and multiplying by λ_t yields

$$\lambda_t\left(y^{t+1} - \mathrm{P_C}(y^{t+1}) + z^{t+1} - \mathrm{P}_D(z^{t+1})\right) = -\frac{\lambda_t}{2\eta}(z^{t+1} - y^{t+1}) = -\frac{1}{2\eta}(x^{t+1} - x^t).$$

Note also that

$$x^t + z^{t+1} - y^{t+1} = x^{t+1} + (1 - \lambda_t)(z^{t+1} - y^{t+1}) = x^{t+1} + \frac{1 - \lambda_t}{\lambda_t}(x^{t+1} - x^t).$$

Substituting the last two equations into (17) gives

$$\begin{aligned}
2\eta\lambda\big(\mathrm{dist}^2(y^{t+1},C)+\mathrm{dist}^2(z^{t+1},D)\big) &\le 4\eta\langle z^{t+1}-\mathrm{P}_D(z^{t+1})-\frac{1}{2\eta}(y^{t+1}-x^*),x^{t+1}-x^t\rangle \\
&= 4\eta\langle -\frac{1}{2\eta}(z^{t+1}-2y^{t+1}+x^t)-\frac{1}{2\eta}(y^{t+1}-x^*),x^{t+1}-x^t\rangle \\
&= -2\langle z^{t+1}-y^{t+1}+x^t-x^*,x^{t+1}-x^t\rangle \\
&= -2\langle x^{t+1}-x^*,x^{t+1}-x^t\rangle-2\frac{1-\lambda_t}{\lambda_t}\|x^{t+1}-x^t\| \\
&= \big(\|x^t-x^*\|^2-\|x^{t+1}-x^*\|^2-\|x^{t+1}-x^t\|^2\big)-2\frac{1-\lambda_t}{\lambda_t}\|x^{t+1}-x^t\| \\
&= \|x^t-x^*\|^2-\|x^{t+1}-x^*\|^2-\frac{2-\lambda_t}{\lambda_t}\|x^{t+1}-x^t\|.
\end{aligned}$$

Step 2 (establishing a recurrence for $\mathrm{dist}^2(x^t,C\cap D)$*):* First note that

$$y^{t+1}=\mathrm{P}_C^\eta(x^t)=\frac{1}{2\eta+1}x^t+\frac{2\eta}{2\eta+1}\mathrm{P}_C(x^t).$$

This shows that y^{t+1} lies in the line segment between x^t and its projection onto C. So, $P_C(y^{t+1})=\mathrm{P}_C(x^t)$ and hence,

$$\mathrm{dist}^2(y^{t+1},C)=\|y^{t+1}-\mathrm{P}_C(x^t)\|^2=\left(\frac{1}{2\eta+1}\right)^2\|\mathrm{P}_C(x^t)-x^t\|^2=\left(\frac{1}{2\eta+1}\right)^2\mathrm{dist}^2(x^t,C).$$

Similarly, as

$$z^{t+1}=\mathrm{P}_D^\eta(2y^{t+1}-x^t)=\frac{1}{2\eta+1}(2y^{t+1}-x^t)+\frac{2\eta}{2\eta+1}\mathrm{P}_D(2y^{t+1}-x^t),$$

the point z^{t+1} lies in the line segment between $2y^{t+1}-x^t$ and its projection onto D. Thus $P_D(z^{t+1})=\mathrm{P}_D(2y^{t+1}-x^t)$ and so,

$$\begin{aligned}
\mathrm{dist}^2(z^{t+1},D) &= \|z^{t+1}-\mathrm{P}_D(2y^{t+1}-x^t)\|^2 \\
&= \left(\frac{1}{2\eta+1}\right)^2\|\mathrm{P}_D(2y^{t+1}-x^t)-(2y^{t+1}-x^t)\|^2 \\
&= \left(\frac{1}{2\eta+1}\right)^2\mathrm{dist}^2(2y^{t+1}-x^t,D).
\end{aligned}$$

Now, using the non-expansiveness of $\mathrm{dist}(\cdot,D)$, we have

$$\begin{aligned}
\mathrm{dist}^2(x^t,D) &\le \Big(\|x^t-(2y^{t+1}-x^t)\|+\mathrm{dist}(2y^{t+1}-x^t,D)\Big)^2 \\
&= \Big(2\|x^t-y^{t+1}\|+\mathrm{dist}(2y^{t+1}-x^t,D)\Big)^2 \\
&= \Big(\frac{4\eta}{2\eta+1}\mathrm{dist}(x^t,C)+\mathrm{dist}(2y^{t+1}-x^t,D)\Big)^2 \\
&\le c\Big(\mathrm{dist}^2(x^t,C)+\mathrm{dist}^2(2y^{t+1}-x^t,D)\Big),
\end{aligned}$$

where $c = 2(\max\{\frac{4\eta}{2\eta+1}, 1\})^2$, and where the last inequality above follows from the following elementary inequalities: for all $\alpha, x, y \in \mathbb{R}_+$,

$$\alpha x + y \le \max\{\alpha, 1\}(x+y), \quad (x+y)^2 \le 2(x^2+y^2).$$

Therefore, we have

$$\mathrm{dist}^2(2y^{t+1} - x^t, D) \ge c^{-1}\mathrm{dist}^2(x^t, D) - \mathrm{dist}^2(x^t, C).$$

So, using (16), we have

$$\begin{aligned} \|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2 &\ge 2\eta\lambda\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \\ &= 2\eta\lambda\left(\frac{1}{2\eta+1}\right)^2\left(\mathrm{dist}^2(x^t, C) + \mathrm{dist}^2(2y^{t+1} - x^t, D)\right). \end{aligned}$$

Note that

$$\mathrm{dist}^2(x^t, C) + \mathrm{dist}^2(2y^{t+1} - x^t, D) \ge \mathrm{dist}^2(x^t, C) + c^{-1}\mathrm{dist}^2(x^t, D) - \mathrm{dist}^2(x^t, C) = c^{-1}\mathrm{dist}^2(x^t, D)$$

and

$$\mathrm{dist}^2(x^t, C) + \mathrm{dist}^2(2y^{t+1} - x^t, D) \ge \mathrm{dist}^2(x^t, C).$$

It follows that

$$\begin{aligned} \|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2 &\ge 2\eta\left(\frac{1}{2\eta+1}\right)^2 c^{-1}\max\{\mathrm{dist}^2(x^t, C), \mathrm{dist}^2(x^t, D)\} \\ &= 2\eta\left(\frac{1}{2\eta+1}\right)^2 c^{-1}(\max\{\mathrm{dist}(x^t, C), \mathrm{dist}(x^t, D)\})^2. \end{aligned} \tag{18}$$

In particular, we see that the sequence $\{x^t\}$ is bounded and Fejér monotone with respect to $C \cap D$. Thence, letting K be a bounded set containing $\{x^t\}$, by bounded Hölder regularity of $\{C, D\}$, there exists $\mu > 0$ and $\gamma \in (0, 1]$ such that

$$\mathrm{dist}(x, C \cap D) \le \mu \max\{\mathrm{dist}(x, C), \mathrm{dist}(x, D)\}^\gamma \quad \forall x \in K.$$

Thus there exists a $\delta > 0$ such that

$$\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2 \ge \delta\,\mathrm{dist}^{2\theta}(x^t, C \cap D)$$

where $\theta = \frac{1}{\gamma} \in [1, \infty)$. So, Fact 2.2 implies that $\mathrm{P}_{C\cap D}(x^t) \to \bar{x}$ for some $\bar{x} \in C \cap D$. Setting $x^* = \mathrm{P}_{C\cap D}(x^t)$ in (18) we therefore obtain

$$\mathrm{dist}^2(x^{t+1}, C \cap D) \le \mathrm{dist}^2(x^t, C \cap D) - \delta\mathrm{dist}^{2\theta}(x^t, C \cap D).$$

Now, the conclusion follows by applying Proposition 3.5 with $\theta = 1/\gamma$. □

Remark 5.3 (DR versus damped DR). Note that Theorem 5.2 only requires Hölder regularity of the underlying collection of constraint sets, rather than the damped DR operator explicitly. A careful examination of the proof of Theorem 5.2 shows that the inequality (17) does not hold for the basic DR algorithm (which would require setting $\eta = +\infty$). ◇

Remark 5.4 (Comments on linear convergence). In the case when $0 \in \operatorname{sri}(C - D)$, where sri is the *strong relative interior*, then the Hölder regularity result holds with exponent $\gamma = 1$ (see [5]). The preceding proposition therefore implies that the damped Douglas–Rachford method converges linearly in the case where $0 \in \operatorname{sri}(C - D)$. ◇

We next show that an explicit sublinear convergence rate estimate can be achieved in the case where $H = \mathbb{R}^n$ and C, D are basic convex semi-algebraic sets.

Theorem 5.5 (Convergence rate for the damped DR algorithm with semi-algebraic sets)**.** *Let C, D be two basic convex semi-algebraic sets in $\mathbb{R}^n$ with $C \cap D \neq \emptyset$, where C, D are given by*

$$C = \{x \in \mathbb{R}^n \mid g_i(x) \le 0, i = 1, \dots, m_1\} \text{ and } D = \{x \in \mathbb{R}^n \mid h_j(x) \le 0, j = 1, \dots, m_2\}$$

where g_i, h_j, $i = 1, \dots, m_1$, $j = 1, \dots, m_2$, are convex polynomials on $\mathbb{R}^n$ with degree at most d. Let $\lambda := \inf_{t\in\mathbb{N}} \lambda_t > 0$ with $\lambda_t \in (0, 2]$ and let $\{(y^t, z^t, x^t)$ be generated by the damped Douglas–Rachford algorithm (15). Then, $x^t \to \bar{x} \in C \cap D$. Moreover, there exist $M > 0$ and $r \in (0, 1)$ such that

$$\|x^t - \bar{x}\| \le \begin{cases} Mt^{-\frac{\gamma}{2(1-\gamma)}} & \text{if} \quad d > 1 \\ M\, r^t & \text{if} \quad d = 1. \end{cases}$$

where $\gamma = [\min\left\{\frac{(2d-1)^n+1}{2},\, B(n-1)d^n\right\}]^{-1}$ and $\beta(n-1)$, is the central binomial coefficient with respect to $n - 1$ which is given by $\binom{n-1}{[(n-1)/2]}$.

Proof. By Lemma 2.11 with $\theta = 1$, we see that for any compact set K, there exists $c > 0$ such that for all $x \in K$,

$$\begin{aligned} \operatorname{dist}(x, C \cap D) &\le c\Big(\operatorname{dist}(x, C) + \operatorname{dist}(x, D)\Big)^{\gamma} \\ &\le 2^{\gamma} c \max\{\operatorname{dist}(x, C), \operatorname{dist}(x, D)\}^{\gamma}. \end{aligned}$$

where $\gamma = [\min\left\{\frac{(2d-1)^n+1}{2},\, B(n-1)d^n\right\}]^{-1}$. Note that $\gamma = 1$ if $d = 1$; while $\gamma \in (0, 1)$ if $d > 1$. The conclusion now follows from Theorem 5.2. □

6 Examples

In this section we fully examine two concrete problems which illustrate the difficulty of establishing optimal rates. We begin with an example consisting of two sets having an intersection which is bounded Hölder regular but not bounded linearly regular. In the special case where $n = 1$, it has previously been examined in detail as part of [10, Ex. 5.4].

Example 6.1 (Half-space and epigraphical set described by $\|x\|^d$). Consider the sets

$$C = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \mid r \le 0\} \text{ and } D = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \mid r \ge \|(x_1, \dots, x_n)\|^d\},$$

where $d > 0$ is an even number. Clearly, $C \cap D = \{0_{\mathbb{R}^{n+1}}\}$ and $\operatorname{ri} C \cap \operatorname{ri} D = \emptyset$. It can be directly verified that $\{C, D\}$ does not has a bounded linearly regular intersection because, for $x_k = (\frac{1}{k}, 0, \dots, 0) \in \mathbb{R}^n$ and $r_k = \frac{1}{k^d}$,

$$\operatorname{dist}\big((x_k, r_k), C \cap D\big) = O\left(\frac{1}{k}\right) \text{ and } \max\{\operatorname{dist}\big((x_k, r_k), C\big), \operatorname{dist}\big((x_k, r_k), D\big)\} = \frac{1}{k^d}.$$

Let $T_{C,D}$ be the Douglas–Rachford operator with respect to the sets C and D. We will verify that $T_{C,D}$ is bounded Hölder regular with exponent $\frac{1}{d}$. Granting this, by Corollary 4.4, the sequence (x^t, r^t) generated by the Douglas–Rachford algorithm converges to a point in $\mathrm{Fix}\, T_{C,D} = \{0_{\mathbb{R}^n}\} \times \mathbb{R}_+$ at least at the order of $t^{\frac{-1}{2(d-1)}}$, regardless of the chosen initial point

Firstly, it can be verified (see also [7, Cor. 3.9]) that

$$\mathrm{Fix}\, T_{C,D} = C \cap D + N_{\overline{C-D}}(0) = \{0_{\mathbb{R}^n}\} \times \mathbb{R}_+,$$

and so,

$$\mathrm{dist}((x,r), \mathrm{Fix}\, T) = \begin{cases} \|x\| & \text{if} \quad r \geq 0 \\ \|(x,r)\| & \text{if} \quad r < 0. \end{cases}$$

Moreover, for all $(x,r) \in \mathbb{R}^n \times \mathbb{R}$,

$$(x,r) - T_{C,D}(x,r) = \mathrm{P}_D(\mathrm{R}_C(x,r)) - \mathrm{P}_C(x,r) = \mathrm{P}_D(x, -|r|) - (x, \min\{r, 0\}).$$

Note that, for any $(z,s) \in \mathbb{R}^n \times \mathbb{R}$, denote $(z^+, s^+) = P_D(z,s)$. Then we have

$$s^+ = \|z^+\|^d \text{ and } (z^+ - z) + d\|z^+\|^{d-2}\big(\|z^+\|^d - s\big)z^+ = 0.$$

Let $(a, \gamma) = \mathrm{P}_D(x, -|r|)$. Then, $a = 0_{\mathbb{R}^n}$ if and only if $x = 0_{\mathbb{R}^n}$,

$$a - x = -d\|a\|^{d-2}(\|a\|^d + |r|)a \text{ and } \gamma = \|a\|^d.$$

It follows that

$$a = \frac{1}{1 + d\|a\|^{2d-2} + d\|a\|^{d-2}|r|} x. \tag{19}$$

So,

$$\begin{aligned} (x,r) - T_{C,D}(x,r) &= (-d\|a\|^{d-2}(\|a\|^d + |r|)a, \|a\|^d - \min\{r, 0\}) \\ &= \begin{cases} (-d\|a\|^{d-2}(\|a\|^d + r)a, \|a\|^d) & \text{if } r \geq 0 \\ (-d\|a\|^{d-2}(\|a\|^d - r)a, \|a\|^d - r) & \text{if } r < 0. \end{cases} \end{aligned}$$

Let K be any bounded set of $\mathbb{R}^{n+1}$ and consider any $(x,r) \in K$. By the nonexpansivity of the projection mapping, $(a,\gamma) = \mathrm{P}_D(x, -|r|)$ is also bounded for any $(x_1, x_2) \in K$. Let $M > 0$ be such that $\|(a,\gamma)\| \leq M$ and $\|(x,r)\| \leq M$ for all $(x,r) \in K$. To verify the bounded Hölder regularity, we divide the discussion into two cases depending on the sign of r.

Case 1 ($r \geq 0$)*:* As d is even, it follows that for all $(x,r) \in K$ with $x \neq 0_{\mathbb{R}^n}$

$$\begin{aligned} \frac{\|(x,r) - T_{C,D}(x,r)\|^2}{\|x\|^{2d}} = \frac{d^2\|a\|^{2(d-1)}(\|a\|^d + r)^2 + \|a\|^{2d}}{\|x\|^{2d}} &\geq \frac{\|a\|^{2d}}{\|x\|^{2d}} \\ &= \left(\frac{1}{1 + d\|a\|^{2d-2} + d\|a\|^{d-2}|r|}\right)^{2d} \\ &\geq \left(\frac{1}{1 + dM^{2d-2} + dM^{d-1}}\right)^{2d}, \end{aligned}$$

where the equality follows from (19). This shows that, for all $(x,r) \in K$,

$$\mathrm{dist}((x,r), \mathrm{Fix}\, T_{C,D}) \leq (1 + dM^{2d-2} + dM^{d-1})\|(x,r) - T_{C,D}(x,r)\|^{\frac{1}{d}}.$$

Case 2 ($r<0$)*:* As d is even, it follows that for all $(x,r)\in K\backslash\{0_{\mathbb{R}^{n+1}}\}$,

$$\begin{aligned}\frac{\|(x,r)-T_{C,D}(x,r)\|^2}{\|x\|^{2d}+r^{2d}} &= \frac{(1+d^2\|a\|^{2(d-1)})(\|a\|^d-r)^2}{\|x\|^{2d}+r^{2d}}\\ &\geq \frac{\|a\|^{2d}+r^2}{\|x\|^{2d}+r^{2d}}\\ &\geq \frac{\|a\|^{2d}+r^{2d}M^{2-2d}}{\|x\|^{2d}+r^{2d}}\\ &= \frac{\left(\frac{1}{1+d\|a\|^{2d-2}+d\|a\|^{d-2}|r|}\right)^{2d}\|x\|^{2d}+r^{2d}M^{2-2d}}{\|x\|^{2d}+r^{2d}}\\ &\geq \min\{\left(\frac{1}{1+dM^{2d-2}+dM^{d-1}}\right)^{2d},M^{2-2d}\},\end{aligned}$$

where the equality follows from (19). Therefore, there exists $\mu>0$ such that, for all $(x,r)\in K$,

$$\mathrm{dist}((x,r),\mathrm{Fix}\,T_{C,D})\leq\mu\|(x,r)-T_{C,D}(x,r)\|^{\frac{1}{d}}.$$

Combining these two cases, we see that $T_{C,D}$ is bounded Hölder regular with exponent $\frac{1}{d}$, and so, the sequence (x^t,r^t) generated by the Douglas–Rachford algorithm converges to a point in $\mathrm{Fix}\,T_{C,D}=\{0_{\mathbb{R}^n}\}\times\mathbb{R}_+$ at least at the order of $t^{\frac{-1}{2(d-1)}}$.

We note that, for $n=1$, it was shown in [10] (by examining the generated DR sequence directly) that the sequence x^t converges to zero at the order $t^{-\frac{1}{d-2}}$ where $d>2$. Note that $r^t=\|x^t\|^d$. It follows that the actual convergence rate for (x^t,r^t) for this example is $t^{-\frac{1}{d-2}}$ in the case $n=1$. Thus, our convergence rate estimate is not tight for this example in the case $n=1$. On the other hand, as noted in [10], their analysis is largely limited to the 2-dimensional case and it is not clear how it can be extended to the higher dimensional setting. ◇

We now examine an even more concrete example involving a subspace and a lower level set of a convex quadratic function in the plane.

Example 6.2 (Hölder regularity of the DR operator involving a ball and a tangent line)*.* Consider the following basic convex semi-algebraic sets in $\mathbb{R}^2$:

$$C=\{x\in\mathbb{R}^2\mid x_1=0\}\text{ and }D=\{x\in\mathbb{R}^2\mid\|x+(1,0)\|^2\leq1\},$$

which have intersection $C\cap D=\{0\}$. We now show that the DR operator $T_{C,D}$ is bounded Hölder regular. Since $C-D=[0,1]\times\mathbb{R}$, by [7, Cor. 3.9], the fixed point set is given by

$$\mathrm{Fix}\,T_{C,D}=C\cap D+N_{C-D}(0)=(-\infty,0]\times\{0\}.$$

We therefore have that

$$\mathrm{dist}(x,\mathrm{Fix}\,T_{C,D})=\begin{cases}\|x\| & x_1>0,\\ |x_2| & x_1\leq0.\end{cases}$$

Setting $\alpha=1/\max\{1,\|x-(1,0)\|\}$, a direct computation shows that

$$T_{C,D}x:=\left(\frac{I+R_DR_C}{2}\right)x=(\alpha-1-\alpha x_1,\alpha x_2)\,, \tag{20}$$

and thus

$$\|x - T_{C,D}x\|^2 = ((1-\alpha) + x_1(1+\alpha))^2 + (x_2(1-\alpha))^2. \tag{21}$$

Now, fix an arbitrary compact set K and let $M > 0$ such that $\|x\| \le M$ for all $x \in K$. For all $x \in K$, there exists $m \in (0,1]$ such that $\alpha = 1/\max\{1, \|x-(1,0)\|\} \in [m,1]$ for all $x \in K$. By shrinking m if necessary, we may assume that

$$\frac{\sqrt{m^2+2m}}{2} \ge M\frac{m^2}{1+m}. \tag{22}$$

We now distinguish two cases depending on α.

Case 1 ($\alpha = 1$)*:* In this case, we have

$$\|x-(1,0)\| \le 1 \implies \|x\|^2 \le 2x_1.$$

In particular, this shows that $x_1 \ge 0$. Now (21) gives

$$\|x - T_{C,D}x\| = 2x_1 \ge \|x\|^2 = \operatorname{dist}^2(x, \operatorname{Fix} T_{C,D}).$$

Case 2 ($\alpha < 1$)*:* Fix $x \in K$. In this case, we show that

$$\|x - T_{C,D}x\| \ge \frac{m^2}{2(1+m)}\|x\|^3 = \frac{m^2}{2(1+m)}\operatorname{dist}^3(x, T_{C,D}x). \tag{23}$$

To do this, we further divide the discussion into two subcases depending on the sign of x_1.

Subcase I ($x_1 > 0$)*:* In this case, $\operatorname{dist}(x, \operatorname{Fix} T_{C,D}) = \|x\|$. Note that

$$\begin{aligned}
\|x - T_{C,D}x\|^2 &= ((1-\alpha) + x_1(1+\alpha))^2 + (x_2(1-\alpha))^2 \\
&\ge (x_1(1+\alpha))^2 + (x_2(1-\alpha))^2 \\
&\ge (m^2+2m)x_1^2 + (1-\alpha)^2\|x\|^2,
\end{aligned}$$

where the last inequality follows by the fact that $\alpha \ge m$. So, the elementary inequality $\sqrt{a^2+b^2} \ge (a+b)/2$ for all $a, b \ge 0$ implies that

$$\|x - T_{C,D}x\| \ge \frac{\sqrt{m^2+2m}}{2}x_1 + \frac{1-\alpha}{2}\|x\|. \tag{24}$$

From the definition of α, we see that

$$1-\alpha = \frac{\|x-(1,0)\|-1}{\|x-(1,0)\|} = \frac{x_1^2 - 2x_1 + x_2^2}{\|x-(1,0)\|(\|x-(1,0)\|+1)}.$$

As $m \le \alpha < 1$, $\|x-(1,0)\| \le \frac{1}{m}$. So,

$$1-\alpha \ge \frac{m^2}{1+m}(x_1^2 - 2x_1 + x_2^2) = \frac{m^2}{1+m}\|x\|^2 - 2\frac{m^2}{1+m}x_1.$$

Then, by combining with (24), we deduce

$$\begin{aligned}
\|x - T_{C,D}x\| &\ge \frac{\sqrt{m^2+2m}}{2}x_1 + \frac{1}{2}\|x\|\left(\frac{m^2}{1+m}\|x\|^2 - 2\frac{m^2}{1+m}x_1\right) \\
&= \frac{m^2}{2(1+m)}\|x\|^3 + x_1\left(\frac{\sqrt{m^2+2m}}{2} - \frac{m^2}{1+m}\|x\|\right) \\
&= \frac{m^2}{2(1+m)}\|x\|^3 + x_1\left(\frac{\sqrt{m^2+2m}}{2} - \frac{m^2}{1+m}M\right).
\end{aligned}$$

The claimed equation (23) now follows from (22).

Subcase II ($x_1 \leq 0$)*:* In this case, $\mathrm{dist}(x, \mathrm{Fix}\, T_{C,D}) = |x_2|$ and

$$\begin{aligned} \|x - T_{C,D}x\| &= \sqrt{((1-\alpha) + x_1(1+\alpha))^2 + (x_2(1-\alpha))^2} \\ &\geq (1-\alpha)x_2. \end{aligned}$$

Similar to Subcase I, we can show that

$$1 - \alpha \geq \frac{m^2}{1+m}(x_1^2 - 2x_1 + x_2^2) \geq \frac{m^2}{1+m}x_2^2.$$

where the last inequality follows from $x_1 \leq 0$. Thus, (23) also follows in this subcase.

Combining the two cases we have

$$\mathrm{dist}(x, \mathrm{Fix}\, T_{C,D}) \leq \|x - T_{C,D}x\|^{1/3} \quad \forall x \in K.$$

That is, $T_{C,D}$ is bounded Hölder regular with exponent $\gamma = 1/3$. Therefore, for this example, Corollary 4.6 implies that the DR algorithm generated a sequence $\{x^t\}$ which converges to $\bar{x} \in \mathrm{Fix} T_{C,D} = (-\infty, 0] \times \{0\}$ at least in a sublinear convergence rate $O(\frac{1}{\sqrt[4]{t}})$. Let $x^t = (x_1^t, x_2^t)$ and $\bar{x} = (\bar{x}_1, 0)$ with $\bar{x}_1 \leq 0$. As $x^{t+1} = T_{C,D}(x^t)$, by passing to the limit in (20), we have $\bar{x}_1 = \bar{\alpha} - 1 - \bar{\alpha}\bar{x}_1$ where $\bar{\alpha} = 1/\max\{1, |\bar{x}_1 - 1|\}$. If $\bar{x}_1 < 0$, then $|\bar{x}_1 - 1| > 1$, and so, $\bar{\alpha} = 1/(1 - \bar{x}_1)$. This implies that $\bar{x}_1 = (\bar{\alpha} - 1)/(1 + \bar{\alpha}) = \bar{x}_1/(2 - \bar{x}_1)$ and hence, $\bar{x}_1 = 1$ or $\bar{x}_1 = 0$ which is impossible. This shows that $\bar{x}_1 = 0$, and so, $\{x^t\}$ converges to $\bar{x} = (0, 0)$ at worst in a sublinear convergence rate $O(\frac{1}{\sqrt[4]{t}})$ regardless the choice of the initial points.

We now illustrate the sublinear convergence rate by numerical simulation. To do this, we first randomly generate an initial point in $[-100, 100]^2$. We then ran the DR algorithm for this example (starting with the corresponding random starting point) whilst tracking the value of $\sqrt[4]{t}\,\|x^t - \bar{x}\|$ and $\frac{-\log(\|x^t - \bar{x}\|)}{\log(t)}$. The experiment was repeated 200 times, and the results plotted in Figure 1.

From the first graph, we see that the value of $\sqrt[4]{t}\,\|x^t - \bar{x}\|$ quickly decrease with increasing t. This supports the conclusion that x^t converges at least in the order of $O(1/\sqrt[4]{t})$. From the second graph, the value of $\frac{-\log(\|x^t - \bar{x}\|)}{\log(t)}$ appears to approach $1/2$. This suggests that the actual sublinear convergence rate for this example is $O(1/\sqrt{t})$, regardless of the choice of the initial point. ◇

Furthermore, the following example shows that, whenever the initial point is chosen in the region specified below, the sequence in Example 6.2 converges with an exact order $O(1/\sqrt{t})$ and thus supports the conjectured rate of convergence.

Example 6.3 (The sequence in Example 6.2 with specific initial points). Consider the setting of Example 6.2, and suppose that the initial point $x^0 = (u_0, v_0) \in \mathbb{R}_{--} \times (0, 1)$. If $x^t = (u_t, v_t) \in \mathbb{R}_{--} \times (0, 1)$, then using (20) we deduce that

$$x^{t+1} = T_{C,D}(x^t) = \frac{(1 - u_t, v_t)}{\sqrt{(1 - u_t)^2 + v_t^2}} - (1, 0) \in \mathbb{R}_{--} \times (0, 1).$$

Inductively, the Douglas–Rachford sequence $\{x^t\}$ is contained in $\mathbb{R}_{--} \times \mathbb{R}_{++}$. By Example 6.2, the sequence $x^t = (u_t, v_t) \to (0, 0)$. Below we verify that the sequence with an exact sublinear convergence order $O(1/\sqrt{t})$.

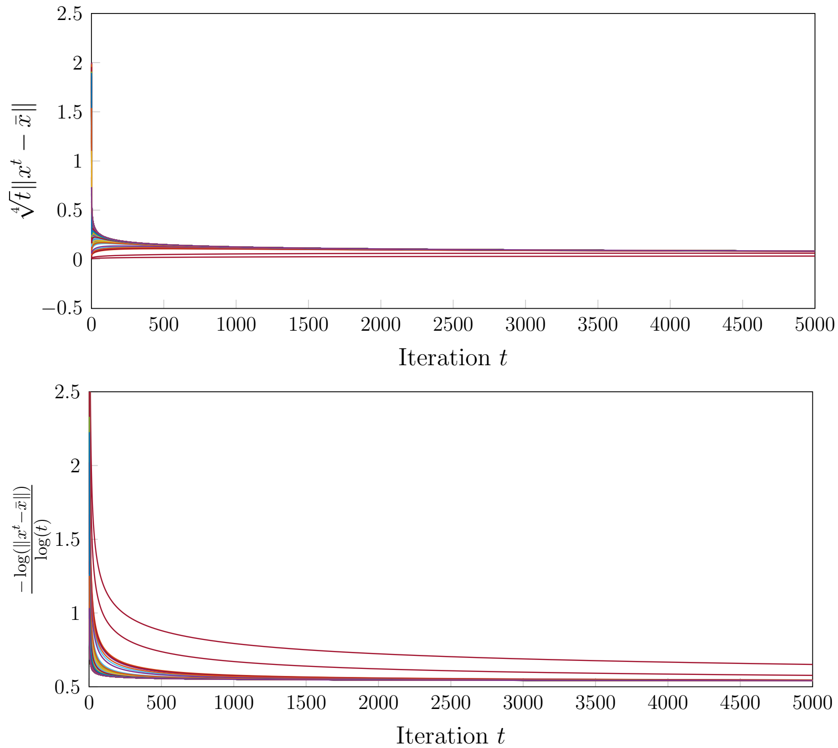

Figure 1: Numerical simulation results: (top) the successive change $\sqrt[4]{t}\,\|x^t - \bar{x}\|$ and (bottom) the ratio $\frac{-\log(\|x^t-\bar{x}\|)}{\log(t)}$ as a function of the number of iterations, t.

To see this, we note from $u_t < 0$ that

$$v_{t+1} = \frac{v_t}{\sqrt{(1-u_t)^2 + v_t^2}} < \frac{v_t}{\sqrt{1+v_t^2}}.$$

Setting $w_t = v_t^2$, we deduce

$$w_{t+1} < \frac{w_t}{1+w_t} = w_t - w_t^2 + O(w_t^3).$$

Since $w_t \to 0$, for sufficiently large t, we have

$$w_{t+1} < w_t - \frac{1}{2}w_t^2 \implies \frac{1}{w_{t+1}} - \frac{1}{w_t} > \frac{1}{2-w_t} \implies \liminf_{t\to\infty}\left(\frac{1}{w_{t+1}} - \frac{1}{w_t}\right) \geq \frac{1}{2}.$$

It now follows that

$$\left(\liminf_{t\to\infty}\frac{1/\sqrt{t}}{v_t}\right)^2 = \liminf_{t\to\infty}\frac{1}{t}\frac{1}{w_t} = \liminf_{t\to\infty}\frac{1}{t}\left(\frac{1}{w_t} - \frac{1}{w_0}\right) = \liminf_{t\to\infty}\frac{1}{t}\sum_{n=0}^{t-1}\left(\frac{1}{w_{n+1}} - \frac{1}{w_n}\right) \geq \frac{1}{2}.$$

Square rooting and inverting both sides we obtain

$$\limsup_{t\to\infty}\frac{v_t}{1/\sqrt{t}} \leq \sqrt{2}. \tag{25}$$

Now, recall that

$$u_{t+1} = \frac{1-u_t}{\sqrt{(1-u_t)^2+v_t^2}} - 1 \quad = \quad \frac{(1-u_t)-\sqrt{(1-u_t)^2+v_t^2}}{\sqrt{(1-u_t)^2+v_t^2}}$$
$$= \quad \frac{-v_t^2}{\sqrt{(1-u_t)^2+v_t^2}\big((1-u_t)+\sqrt{(1-u_t)^2+v_t^2}\big)}.$$

Since $\sqrt{(1-u_t)^2+v_t^2}\big((1-u_t)+\sqrt{(1-u_t)^2+v_t^2}\big) \to 2$ as $t \to \infty$, whenever t is sufficiently large we have

$$0 > u_{t+1} \geq -v_t^2. \tag{26}$$

Combining (25) and (26), we see that there exists $C > 0$ such that $\|(u^t, v^t)\| \leq C\frac{1}{\sqrt{t}}$ for all $t \in \mathbb{N}$. In particular, this also shows that $u_t \to 0$.

Noting that $\frac{v_{t+1}}{v_t} = \frac{1}{\sqrt{(1-u_t)^2+v_t^2}} \to 1$ as $t \to \infty$ and $v_t > 0$, we therefore deduce that $v_{t-1} < 2v_t$ for all sufficiently large t. Combining with (26), this yields

$$v_{t+1} = \frac{v_t}{\sqrt{(1-u_t)^2+v_t^2}} \geq \frac{v_t}{\sqrt{(1+v_{t-1}^2)^2+v_t^2}} = \frac{v_t}{\sqrt{1+9v_t^2+16^2v_t^4}} > \frac{v_t}{1+\frac{9}{2}v_t^2}.$$

As before, we set $w_t = v_t^2$. Since $w_t \to 0$ and $(1+\frac{9}{2}w_t)^2(1-10w_t) = 1 - w_t - \frac{279}{4}w_t^2 - \frac{405}{2}w_t^3 < 1$, we deduce

$$w_{t+1} > \frac{w_t}{(1+\frac{9}{2}w_t)^2} > w_t(1-10w_t) \implies w_{t+1} > w_t - 10w_t^2.$$

Proceeding as before, we obtain

$$\liminf_{t\to\infty} \frac{v_t}{1/\sqrt{t}} \geq \frac{1}{10}.$$

This shows that $\|(u_t, v_t)\| \geq \frac{1}{10}\frac{1}{\sqrt{t}}$. Altogether, we have proven that $(u^t, v^t) \to (0,0)$ with an exact sublinear convergence order $O(1/\sqrt{t})$. ◇

7 Conclusions

In this paper, using a Hölder regularity assumption, sublinear and linear convergence of fixed point iterations described by averaged nonexpansive operators has been established. The framework was then specialized to various fixed point algorithms including Krasnoselskii–Mann iterations, the cyclic projection algorithm, and the Douglas–Rachford feasibility algorithm along with some variants. In the case where the underlying sets are basic convex semi-algebraic, in a finite dimensional space, the results apply without any further regularity assumptions.

In particular, for our damped Douglas–Rachford algorithm, an explicit estimate for the sublinear convergence rate has been provided in terms of the dimension and the maximum degree of the polynomials which define the convex sets. We emphasise that, unlike the for damped Douglas–Rachford algorithm, we were not able to provide an explicit estimate of the sublinear convergence rate for the classical Douglas–Rachford algorithm when the two convex sets are described by convex polynomials. Our approach relies on the Łojasiewicz's inequality which gives no quantitative information regarding the Hölder exponent. Providing explicit estimates is left as an open question for future research.

Another area for future research involves characterization of the convergence rate in the absence of Hölder regularity properties. For instance, it is known that the alternating projection method can exhibit arbitrarily slow convergence when applied to two subspaces in infinite dimensional spaces without closed sum [11]. As shown in [18, Cor. 3.1], if only two sets are involved and the initial point is chosen in a specific way, the cyclic Douglas–Rachford method can coincide with the alternating projection method, and so, it may exhibit arbitrarily slow convergence. On the other hand, it was shown in Proposition 4.5 that the basic/cyclic Douglas–Rachford method enjoys a sublinear convergence rate if the underlying sets are basic convex semi-algebraic sets in finite dimensional spaces. It would be interesting to see whether an arbitrarily slow convergence can happen for these two methods for general closed and convex sets in finite dimensional spaces.

Finally, the current definition of basic semi-algebraic convex sets only applies to finite dimensional spaces. It would interesting to see if a suitable extension of the notion can be profitably used in infinite dimensional spaces using, for instance, polynomials as defined in [25].

Acknowledgements. JMB is supported, in part, by the Australian Research Council. GL is supported, in part, by the Australian Research Council. MKT is supported by an Australian Postgraduate Award.

References

[1] F.J. Aragón Artacho, J.M. Borwein and M.K. Tam, *Douglas–Rachford feasibility methods for matrix completion problems,* ANZIAM J., 55 (2014), pp. 299–326.

[2] F.J. Aragón Artacho, J.M. Borwein and M.K. Tam, *Recent results on Douglas–Rachford methods for combinatorial optimization problems,* J. Optim. Theory Appl., 163 (2014), pp. 1–30.

[3] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality,* Math. Oper. Res., 35 (2010), pp. 438-457.

[4] H.H. Bauschke and J.M. Borwein, *On the convergence of von Neumann's alternating projection algorithm for two sets,* Set-Valued Anal., 1 (1993), pp. 185–212.

[5] H.H. Bauschke and J.M. Borwein, *On projection algorithms for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.

[6] H.H. Bauschke and P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces,* Springer-Verlag, New York, 2011.

[7] H.H. Bauschke, P.L. Combettes and D.R. Luke, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces,* J. Approx. Theory, 127 (2004), pp. 179–192.

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

[9] H. Bauschke, J. B. Cruz, T.T. Nghia, H.M. Phan and X. Wang, *The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle,* J. Approx. Theory, 185 (2014), pp. 63–79.

[10] H.H. Bauschke, M.N. Dao, D. Noll and H.M. Phan, *Proximal point algorithm, Douglas–Rachford algorithm and alternating projections: a case study,* J. Convex Anal., in press.

[11] H.H. Bauschke, F. Deutsch and H. Hundal, *Characterizing arbitrarily slow convergence in the method of alternating projections,* Int. Trans. Oper. Res., 16 (2009), pp. 413–425.

[12] J. Bochnak, M. Coste and M.F. Roy, *Real algebraic geometry,* Springer-Verlag, Berlin, 1998.

[13] J.M. Borwein, *Maximum entropy and feasibility methods for convex and nonconvex inverse problems,* Optim. 61 (2012), pp. 1–33.

[14] J.M. Borwein and B. Sims, *The Douglas–Rachford algorithm in the absence of convexity,* in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, H.H. Bauschke, R. Burachik, P.L. Combettes, V. Elser, D.R. Luke and H. Wolkowicz, eds., Springer-Verlag, New York, 2011, pp. 93–109.

[15] J.M. Borwein, G.Li and L.J. Yao, *Analysis of the convergence rate for the cyclic projection algorithm applied to basic semi-algebraic convex sets,* SIAM J. Optim., 24 (2014), pp. 498–527.

[16] J.M. Borwein, B. Sims and M.K. Tam, *Norm convergence of realistic projection and reflection methods,* Optim. 64 (2015), pp. 161–178.

[17] J.M. Borwein and M.K. Tam, *The cyclic Douglas–Rachford method for inconsistent feasibility problems,* J. Nonlinear Convex Anal., 16 (2015), pp. 537–584.

[18] J.M. Borwein and M.K. Tam, *A cyclic Douglas–Rachford iteration scheme,* J. Optim. Theory Appl., 160 (2014), pp. 1–29.

[19] P.L. Combettes, *The convex feasibility problem in image recovery,* Adv. Imaging Electron Phys., 95 (1996), pp. 155–270.

[20] Y. Censor and A. Cegielski, *Projection methods: an annotated bibliography of books and reviews,* Optim., 64 (2015), pp. 2343–2358. Then

[21] D. Davis, *Convergence rate analysis of the forward-Douglas–Rachford splitting scheme,* SIAM J. Optim., 25 (2015), pp. 1760–1786.

[22] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, arXiv:1406.4834v3, (2015).

[23] R. Escalante and M. Raydan, *Alternating projection methods,* SIAM, Philadelphia, PA, 2011.

[24] L.G. Gubin, B.T. Polyak, and E.V. Raik, *The method of projections for finding the common point of convex sets,* USSR Comp. Math. Math.+, 7 (1967), pp. 1–24.

[25] J.M. Gutiérrez, J.A. Jaramillo, and J.G. Llavona, *Polynomials and geometry of Banach spaces,* Extracta Math., 10 (1995), p. 77–114.

[26] B.S. He and X.M. Yuan, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method,* SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[27] B.S. He and X.M. Yuan, *On the convergence rate of Douglas–Rachford operator splitting method,* Math. Program., 2 (2015), pp. 715-722.

[28] A.Y. Kruger. *Error bounds and metric subregularity,* Optim., 64 (2015), pp. 49–79.

[29] A.Y. Kruger and Nguygen H.T, *About $[q]$-regularity properties of collections of sets,* J. Math. Anal. Appl., 416 (2014), pp. 471–496.

[30] G. Li, *On the asymptotic well behaved functions and global error bound for convex polynomials,* SIAM J. Optim., 20 (2010), pp. 1923–1943.

[31] G. Li, *Global error bounds for piecewise convex polynomials,* Math. Program., 137 (2013), pp. 37–64.

[32] G. Li and B.S. Mordukhovich, *Hölder metric subregularity with applications to proximal point method,* SIAM J. Optim., 22 (2012), pp. 1655–1684.

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

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

[35] R. Hesse and D.R. Luke, *Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems,* SIAM J. Optim., 23 (2013), pp. 2397–2419.

[36] R. Hesse, D.R. Luke and P. Neumann, *Alternating projections and Douglas–Rachford for sparse affine feasibility,* IEEE Trans. Signal Process. 62 (2014), pp. 4868–4881.

[37] A.S. Lewis, D.R. Luke and J. Malick, *Local convergence for alternating and averaged nonconvex projections,* Found. Comput. Math., 9 (2009), pp. 485–513.

[38] A.S. Lewis and J. Malick, *Alternating projections on manifolds,* Math. Oper. Res., 33 (2008), pp. 216–234.

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

[40] H. Phan, *Linear convergence of the Douglas–Rachford method for two closed sets,* Optim., online first (2015). doi: 10.1080/02331934.2015.1051532.

[41] B.F. Svaiter, *On weak convergence of the Douglas–Rachford method,* SIAM J. Control Optim., 49 (2011), pp. 280–287.

[42] E.D. Sontag, *Remarks on piecewise-linear algebra,* Pacific J. Math., 98 (1982), pp. 183–201.

[43] S. Reich and R. Zalas, *A modular string averaging procedure for solving the common fixed point problem for quasi-nonexpansive mappings in Hilbert space,* Numer. Algorithms, online first (2015). doi: 10.1007/s11075-015-0045-z.

[44] S.M. Robinson, *Some continuity properties of polyhedral multifunctions,* in Mathematical Programming at Oberwolfach vol. 14, Springer Berlin Heidelberg, 1981, pp. 206–214.