SIAM J. OPTIM.
Vol. 27, No. 1, pp. 1–33

CONVERGENCE RATE ANALYSIS FOR AVERAGED FIXED POINT ITERATIONS IN COMMON FIXED POINT PROBLEMS*

JONATHAN M. BORWEIN†, GUOYIN LI‡, AND MATTHEW K. TAM§

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 many important iterative methods in solving broad mathematical problems such as convex feasibility problems and variational inequality problems. These include Krasnoselskii–Mann iterations, the cyclic projection algorithm, forward-backward splitting 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.

Key words. averaged operator, fixed point iteration, convergence rate, Hölder regularity, semi-algebraic, Douglas–Rachford algorithm

AMS subject classifications. Primary, 41A25, 90C25; Secondary, 41A50, 90C31

DOI. 10.1137/15M1045223

1. Introduction.

Consider the problem of finding a point in the intersection of a finite family of closed convex subsets of a Hilbert space, a problem often referred to as the *convex feasibility problem* which arises frequently throughout areas of mathematics, science, and engineering. For details, we refer the reader to the surveys [6, 21], the monographs [7, 27], any of [22, 1, 15], 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 the desired solution. An important class of such methods comprises the so-called projection *and* reflection methods which employ various combinations of projection and reflection operations with respect to underlying constraint sets.

Notable methods of this kind include the *alternating projection algorithm* [5, 29, 17], the *Douglas–Rachford (DR) algorithm* [37, 43, 38], and many extensions and variants [19, 20, 10, 48]. Even in settings without convexity [1, 2, 3, 16, 41, 42], such methods remain a popular choice due largely to their simplicity, ease of implementation, and relatively—often surprisingly—good performance.

*Received by the editors October 23, 2015; accepted for publication (in revised form) September 30, 2016; published electronically January 5, 2017.
http://www.siam.org/journals/siopt/27-1/M104522.html
Funding: The first and second authors are supported, in part, by the Australian Research Council. The third author is supported by Deutsche Forschungsgemeinschaft Research Training Grant 2088 and by an Australian Postgraduate Award.
†CARMA, University of Newcastle, Callaghan NSW 2308, Australia (jonathan.borwein@newcastle.edu.au).
‡Department of Applied Mathematics, University of New South Wales, Sydney NSW 2052, Australia (g.li@unsw.edu.au).
§Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 37083 Göttingen, Germany (m.tam@math.uni-goettingen.de).

The origins of the DR algorithm can be traced to [37], 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 [43], and the result was recently strengthened by Svaiter [46]. 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 [38], where its relationship with another popular method, the *proximal point algorithm*, was discussed.

Motivated by the computational observation that the DR 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 converge to their limit points. For the DR algorithm, the first such result, which appeared in [40] and was later extended by [11], showed the algorithm to converge 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. In finite dimensions, if the sum of the two subspaces is not closed, convergence of the method—while still ensured—need not be linear [11, Sect. 6]. See also [28] for other recent work regarding linear convergence. For most projection methods, it is typical that there exist instances in which the rate of convergence is arbitrarily slow and not even sublinear or arithmetic [24, 13]. Most recently, a preprint of Davis and Yin shows that indeed the DR method also may converge arbitrarily slowly in infinite dimensions [25, Thm. 9].

In potentially nonconvex settings, a number of recent works [39, 40, 45, 36] 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 nonempty 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 [12, Ex. 5.4(iii)] (see section 6 for further examples and discussion). This situation therefore calls for further research aimed at answering the question, *Can a global convergence rate for 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 to provide simple tools for establishing sublinear or linear convergence of the DR algorithm and variants. Our analysis is performed within the much more general setting of *fixed point iterations* described by *averaged nonexpansive operators*. This broad framework covers many iterative fixed point methods including various Krasnoselskii–Mann iterations, the cyclic projection algorithm, DR algorithms, and forward-backward splitting methods, and can be used to solve not only convex feasibility problems but also convex optimization problems and variational inequality problems. We pay special attention to the case in which the underlying sets are *convex semialgebraic sets* in a finite dimensional space. Such sets comprise a broad subclass of convex sets that we shall show satisfy *Hölder regularity properties* without requiring any further assumptions. Indeed, they capture all polyhedra and convex sets described by convex quadratic functions. Furthermore, convex semialgebraic 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 study 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, DR algorithms, and forward-backward splitting methods. In the presence of so-called *bounded Hölder regularity properties*, sublinear convergence of the algorithm is then established (Theorem 3.1).

(II) The quasi-cyclic algorithm framework is then specialized to the DR algorithm and its variants (section 4). We show the results apply, for instance, to the important case of feasibility problems for which the underlying sets are convex semialgebraic in a finite dimensional space.

(III) A damped variant of the DR algorithm is examined. Again, in the case in which the underlying sets are convex basic semialgebraic 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.* In section 4 we specialize these results to the classical DR algorithm and its cyclic variants. In section 5 we consider a damped version of the DR algorithm. In section 6 we establish explicit convergence rates for two illustrative problems. We conclude the paper in section 7 by discussing possible directions for future research.

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 nonempty 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*

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

FACT 2.1 (shadows of Fejér monotone sequences [6, Thm. 5.7(iv)]). *Let A be a nonempty 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 2.2 (Fejér monotone convergence [5, Thm. 3.3(iv)]). *Let A be a nonempty 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 $\|x_k - x\| \le 2\mathrm{dist}(x_k, A)$.*

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

DEFINITION 2.2 (bounded Hölder regular intersection). *Let $\{C_j\}_{j\in\mathbb{J}}$ be a collection of closed convex subsets in a Hilbert space H with nonempty 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*

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

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 uniform exponent γ.

It is clear, from Definition 2.2, that any collection containing only a single set trivially has a bounded Hölder regular intersection with uniform exponent $\gamma = 1$. More generally, Definition 2.2 with $\gamma = 1$ is well-studied in the literature, where it appears, among other names, as *bounded linear regularity* [6]. For a recent study, the reader is referred to [32, Rem. 7]. The local counterpart to Definition 2.2 has been characterized in [32, Thm. 1] under the name of *metric* $[\gamma]$*-subregularity*.

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

Definition 2.3. *An operator* $T\colon H \to H$ *is*

(a) nonexpansive *if, for all* $x, y \in H$*,*

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

(b) firmly nonexpansive *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 nonexpansive mapping* $R\colon H \to H$ *such that*

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

The class of firmly nonexpansive mappings comprises precisely the 1/2-averaged mappings, and any α-averaged operator is nonexpansive [7, Chap. 4]. The term "averaged mapping" was coined in [4]. The following fact provides a characterization of averaged maps that is useful for our purposes.

Fact 2.3 (characterization of averaged maps [7, Prop. 4.25(iii)]). *Let* $T\colon H \to H$ *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 $T\colon H \to H$ by

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

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

Definition 2.4 (bounded Hölder regular operators). *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*

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

Furthermore, if the exponent γ *does not depend on the set* K*, we say that* T *is* bounded Hölder regular with uniform exponent γ.

Note that, in the case when $\gamma = 1$, Definition 2.4 collapses to the well-studied concept of bounded linear regularity [6] and has been used in [10] to analyze linear convergence of algorithms involving nonexpansive 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—along with Hölder type error bounds has recently been studied in [35, 33, 34, 31].

Finally, we recall the definitions of *semialgebraic functions* and *semialgebraic sets*.

Definition 2.5 (semialgebraic sets and functions [14]). *A set $D \subseteq \mathbb{R}^n$ is* semialgebraic *if*

$$D := \bigcap_{j=1}^{s} \bigcup_{i=1}^{l} \{x \in \mathbb{R}^n \mid f_{ij}(x) = 0, h_{ij}(x) < 0\} \tag{2.1}$$

for integers l, s and 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 semialgebraic *if its* graph, $\mathrm{gph} F := \{(x, F(x)) \mid x \in \mathbb{R}^n\}$*, is a semialgebraic set in $\mathbb{R}^n \times \mathbb{R}^p$.*

The next fact summarizes some fundamental properties of semialgebraic sets and functions.

Fact 2.4 (properties of semialgebraic sets/functions). *The following statements hold:*

- (P1) *Any polynomial is a semialgebraic function.*
- (P2) *Let D be a semialgebraic set. Then $\mathrm{dist}(\cdot, D)$ is a semialgebraic function.*
- (P3) *If f, g are semialgebraic functions on $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$, then $f + g$, λf, $\max\{f, g\}$, fg are semialgebraic.*
- (P4) *If f_i are semialgebraic 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 semialgebraic sets.*
- (P5) *If $F : \mathbb{R}^n \to \mathbb{R}^p$ and $G : \mathbb{R}^p \to \mathbb{R}^q$ are semialgebraic mappings, then their composition $G \circ F$ is also a semialgebraic mapping.*
- (P6) *(Lojasiewicz's inequality) If ϕ, ψ are two continuous semialgebraic functions on a compact semialgebraic 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 [14, Prop. 2.2.8] for (P2), [14, Prop. 2.2.6] for (P3) and (P5), and [14, Cor. 2.6.7] for (P6). ∎

Definition 2.6 (basic semialgebraic convex sets in $\mathbb{R}^n$). *A set $C \subseteq \mathbb{R}^n$ is a* basic semialgebraic convex *set if there exist $\gamma \in \mathbb{N}$ and convex polynomial functions, $g_j, j = 1, \ldots, \gamma$, such that $C = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, j = 1, \ldots, \gamma\}$.*

Any basic semialgebraic convex set is clearly convex and semialgebraic. On the other hand, there exist sets which are both convex and semialgebraic but fail to be basic semialgebraic convex sets; see [17].

It transpires that any finite collection of basic semialgebraic convex sets has an intersection which is boundedly Hölder regular with uniform exponent (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 2.1 (Hölder regularity of basic semialgebraic convex sets in $\mathbb{R}^n$ [17]). *Let C_i be basic convex semialgebraic 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 $\gamma = [\min\{\frac{(2d-1)^n+1}{2}, B(n-1)d^n\}]^{-1}$.

We also recall the following useful recurrence relationship established in [17].

Lemma 2.2 (recurrence relationship [17]). *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

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

where the convention that $\frac{1}{0} = +\infty$ *is adopted.*

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

$$x^{t+1} = \sum_{j\in J} w_{j,t} T_j(x^t) \quad \forall t \in \mathbb{N} \tag{3.1}$$

for appropriately chosen weights $w_{j,t} \in \mathbb{R}$.

The quasi-cyclic algorithm appears in [10], where linear convergence of the algorithm was established under suitable regularity conditions. As we shall soon see, the quasi-cyclic algorithm provides a broad framework which covers many important existing algorithms including DR algorithms, the cyclic projection algorithm, the Krasnoselskii–Mann method, and forward-backward splitting. To establish its convergence rate, we use three preparatory results.

Lemma 3.1. *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 $\cap_{j\in J}\operatorname{Fix} T_j \neq \emptyset$ and let $\alpha \in (0, 1)$. 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*

$$x^{t+1} = \sum_{j\in J} w_{j,t} T_j\big(x^t\big) \quad \forall t \in \mathbb{N}. \tag{3.2}$$

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 $\{x^t\}_{t\in\mathbb{N}}$ is Fejér monotone with respect to $\cap_{j\in J}\operatorname{Fix} T_j$, $\{\operatorname{dist}(x^t, \cap_{j\in J}\operatorname{Fix} T_j)\}_{t\in\mathbb{N}}$ is nonincreasing (and hence convergent), and $\max_{j\in J_+(t)} \|x^t - T_j(x^t)\| \to 0$ 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

$$\|x^{t+1} - y\|^2 = \left\|\sum_{j\in J} w_{j,t}T_j(x^t) - y\right\|^2 \le \sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2 \le \|x^t - y\|^2, \tag{3.3}$$

where the last inequality follows by the fact that each T_j is α-averaged (and so is nonexpansive). Thus, $\{x^t\}_{t\in NN}$ is Fejér monotone with respect to $\cap_{j\in J}\mathrm{Fix}T_j$ and $\{\|x^t - y\|^2\}_{t\in\mathbb{N}}$ is a nonnegative, decreasing sequence and hence convergent. Furthermore, from (3.3) we obtain

$$\lim_{t\to\infty}\sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2 = \lim_{t\to\infty}\|x^t - y\|^2. \tag{3.4}$$

Since T_j is α-averaged for each $j \in J$, Fact 2.3 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

$$\begin{aligned}\|x^t - y\|^2 - \sum_{j\in J} w_{j,t}\|T_j(x^t) - y\|^2 &\ge \frac{1-\alpha}{\alpha}\sum_{j\in J} w_{j,t}\|x^t - T_j(x^t)\|^2 \\ &\ge \frac{1-\alpha}{\alpha}\sigma\max_{j\in J_+(t)}\|x^t - T_j(x^t)\|^2.\end{aligned}$$

Together with (3.4) this gives $\max_{j\in J_+(t)}\|x^t - T_j(x^t)\| \to 0$ as claimed. ∎

The following proposition provides a convergence rate for Fejér monotone sequences satisfying an additional property, which we later show to be satisfied in the presence of Hölder regularity.

PROPOSITION 3.1. *Let F be a nonempty closed convex set in a Hilbert space H and let s be a positive integer. Suppose the sequence $\{x^t\}$ is Fejér monotone with respect to F and satisfies*

$$\mathrm{dist}^2(x^{(t+1)s}, F) \le \mathrm{dist}^2(x^{ts}, F) - \delta\,\mathrm{dist}^{2\theta}(x^{ts}, F) \quad \forall\, t \in \mathbb{N} \tag{3.5}$$

for some $\delta > 0$ and $\theta \ge 1$. Then $x^t \to \bar{x}$ for some $\bar{x} \in F$ and there exist constants $M_1, M_2 \ge 0$ and $r \in [0,1)$ such that

$$\|x^t - \bar{x}\| \le \begin{cases} M_1\, t^{-\frac{1}{2(\theta-1)}}, & \theta > 1,\\ M_2\, r^t, & \theta = 1.\end{cases}$$

Furthermore, the constants may be chosen to be

$$\begin{cases} M_1 &:= \ 2\max\{(2s)^{\frac{1}{2(\theta-1)}}\ [(\theta-1)\delta]^{-\frac{1}{2(\theta-1)}}, (2s)^{\frac{1}{2(\theta-1)}}\,\mathrm{dist}(x^0, F)\},\\ M_2 &:= \ 2\max\{\left(\sqrt[4s]{1-\delta}\right)^{-2s}\mathrm{dist}(x^0, F), \sqrt{\mathrm{dist}(x^0, F)}\},\\ \quad r &:= \ \sqrt[4s]{1-\delta},\end{cases} \tag{3.6}$$

and δ necessarily lies in $(0,1]$ whenever $\theta = 1$.

Proof. Without loss of generality, we assume that $x^0 \notin F$. Let $\beta_t := \mathrm{dist}^2(x^{ts}, F)$ and $p := \theta - 1 \geq 0$. Then (3.5) becomes

$$\beta_{t+1} \leq \beta_t \left(1 - \delta \beta_t^p\right). \tag{3.7}$$

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

Case 1. Suppose $\theta \in (1, +\infty)$. Then, noting that $1/(\theta - 1) > 0$, Lemma 2.2 implies

$$\beta_t \leq \left(\beta_0^{-p} + (\theta - 1)\delta t\right)^{-\frac{1}{\theta-1}} \leq \left(0 + (\theta - 1)\delta t\right)^{-\frac{1}{\theta-1}} \quad \forall\, t \in \mathbb{N}.$$

It follows that $\mathrm{dist}(x^{ts}, F) = \sqrt{\beta_t} \leq [(\theta - 1)\delta]^{-\frac{1}{2(\theta-1)}}\, t^{-\frac{1}{2(\theta-1)}}$. In particular, we have $\|x^{ts} - \mathrm{P}_F(x^{ts})\| = \mathrm{dist}(x^{ts}, F) \to 0$. By Fact 2.1, $\mathrm{P}_F(x^{ts}) \to \bar{x}$ for some $\bar{x} \in F$ and hence $x^{ts} \to \bar{x} \in F$ as $t \to \infty$. Denote

$$\bar{M}_1 := \max\{(2s)^{\frac{1}{2(\theta-1)}}\, [(\theta - 1)\delta]^{-\frac{1}{2(\theta-1)}}, (2s)^{\frac{1}{2(\theta-1)}}\, \mathrm{dist}(x^0, F)\}.$$

On one hand, if $t \leq 2s$, then

$$\mathrm{dist}(x^t, F) \leq \mathrm{dist}(x^0, F) = \left[(2s)^{\frac{1}{2(\theta-1)}}\, \mathrm{dist}(x^0, F)\right](2s)^{-\frac{1}{2(\theta-1)}} \leq \bar{M}_1 t^{-\frac{1}{2(\theta-1)}},$$

and, on the other hand, if $t > 2s$ (and so, $\frac{t}{s} - 1 \geq \frac{t}{2s}$), then

$$\begin{aligned}
\mathrm{dist}(x^t, F) \leq \mathrm{dist}(x^{s\lfloor \frac{t}{s} \rfloor}, F) &\leq [(\theta - 1)\delta]^{-\frac{1}{2(\theta-1)}} \left(\lfloor \frac{t}{s} \rfloor\right)^{-\frac{1}{2(\theta-1)}} \\
&\leq [(\theta - 1)\delta]^{-\frac{1}{2(\theta-1)}} \left(\frac{t}{s} - 1\right)^{-\frac{1}{2(\theta-1)}} \\
&\leq [(\theta - 1)\delta]^{-\frac{1}{2(\theta-1)}} \left(\frac{t}{2s}\right)^{-\frac{1}{2(\theta-1)}} \\
&\leq \bar{M}_1\, t^{-\frac{1}{2(\theta-1)}}.
\end{aligned}$$

Here $\lfloor \frac{t}{s} \rfloor$ denotes the largest integer which is smaller than or equal to $\frac{t}{s}$, the first inequality follows from the Fejér monotonicity of $\{x^t\}$, and the last inequality follows from the definition of $\bar{M}_1$. This, together with Fact 2.2, implies that

$$\|x^t - \bar{x}\| \leq 2\mathrm{dist}(x^t, F) \leq 2\bar{M}_1 t^{-\frac{1}{2(\theta-1)}} = M_1 t^{-\frac{1}{2(\theta-1)}},$$

where the last equality follows from the definition of M_1.

Case 2. Suppose $\theta = 1$. Then (3.7) simplifies to $\beta_{t+1} \leq (1 - \delta)\beta_t$ for all $t \in \mathbb{N}$. Moreover, this shows that $\delta \in (0, 1]$ and that

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

Let $\bar{M}_2 = \max\{\left(\sqrt[4s]{1 - \delta}\right)^{-2s} \mathrm{dist}(x^0, F), \sqrt{\mathrm{dist}(x^0, F)}\}$. On one hand, if $t \leq 2s$, then

$$\mathrm{dist}(x^t, F) \leq \mathrm{dist}(x^0, F) = \left[\left(\sqrt[4s]{1 - \delta}\right)^{-2s} \mathrm{dist}(x^0, F)\right]\left(\sqrt[4s]{1 - \delta}\right)^{2s} \leq \bar{M}_2\left(\sqrt[4s]{1 - \delta}\right)^t,$$

and, on the other hand, if $t > 2s$ (and so, $\frac{t}{s} - 1 \geq \frac{t}{2s}$), then

$$\begin{aligned}\operatorname{dist}(x^t, F) \leq \operatorname{dist}(x^{s\lfloor \frac{t}{s} \rfloor}, F) &\leq \sqrt{\beta_0}\left(\sqrt{1-\delta}\right)^{\lfloor \frac{t}{s} \rfloor} \\ &\leq \sqrt{\beta_0}\left(\sqrt{1-\delta}\right)^{\frac{t}{s}-1} \\ &\leq \sqrt{\beta_0}\left(\sqrt{1-\delta}\right)^{\frac{t}{2s}} \\ &\leq \sqrt{\beta_0}\left(\sqrt[4s]{1-\delta}\right)^{t} \leq \bar{M}_2\left(\sqrt[4s]{1-\delta}\right)^{t}.\end{aligned}$$

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

$$\|x^t - \bar{x}\| \leq 2\operatorname{dist}(x^t, F) \leq 2\bar{M}_2\left(\sqrt[4s]{1-\delta}\right)^{t} = M_2\left(\sqrt[4s]{1-\delta}\right)^{t}.$$

The conclusion follows by setting $r = \sqrt[4s]{1-\delta} \in [0, 1)$. ∎

We are now in a position to state our main convergence result, which we simultaneously prove for both variants of our Hölder regularity assumption (nonuniform and uniform versions).

THEOREM 3.1 (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 $\cap_{j\in J}\operatorname{Fix} T_j \neq \emptyset$ 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* (3.1). *Suppose the following assumptions hold:*

(a) *For each $j \in J$, the operator T_j is bounded Hölder regular.*
(b) *$\{\operatorname{Fix} T_j\}_{j\in J}$ has a boundedly Hölder regular intersection.*
(c) *$\sigma := \inf_{t\in\mathbb{N}} \inf_{j\in J_+(t)}\{w_{j,t}\} > 0$, where $J_+(t) := \{j \in J : w_{j,t} > 0\}$ for each $t \in \mathbb{N}$, and there exists an $s \in \mathbb{N}$ such that*

$$J_+(t) \cup J_+(t+1) \cup \cdots \cup J_+(t+s-1) = J \quad \forall t \in \mathbb{N}.$$

Then $x^t \to \bar{x} \in \cap_{j\in J}\operatorname{Fix} T_j \neq \emptyset$ at least with a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$.

In particular, if we assume (a′), (b′), *and* (c) *hold where* (a′), (b′) *are given by*

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

then there exist constants $M_1, M_2 \geq 0$ and $r \in [0, 1)$ such that

$$\|x^t - \bar{x}\| \leq \begin{cases} M_1 t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0, 1), \\ M_2\, r^t, & \gamma = 1, \end{cases}$$

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

Proof. Denote $F := \cap_{j\in J}\operatorname{Fix} T_j$. We first consider the case in which Assumptions (a), (b), and (c) hold. We first observe that, as a consequence of Lemma 3.1, we may assume without loss of generality the following two inequalities hold:

$$\max_{j\in J_+(t)} \|x^t - T_j(x^t)\| \leq 1 \quad \forall t \in \mathbb{N}, \tag{3.8}$$

$$\operatorname{dist}^2(x^{ts}, F) - \operatorname{dist}^2(x^{(t+1)s}, F) \leq 1 \quad \forall t \in \mathbb{N}. \tag{3.9}$$

Now, 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

$$\operatorname{dist}(x, \operatorname{Fix} T_j) \leq \mu_j \|x - T_j(x)\|^{\gamma_{1,j}} \quad \forall x \in K. \tag{3.10}$$

Set $\gamma_1 = \min\{\gamma_{1,j} \mid j \in J\}$ and $\mu := \max\{\mu_j \mid j \in J\}$. By (3.10) and (3.8), for all $j \in J_+(t)$, it follows that

$$\operatorname{dist}(x^t, \operatorname{Fix} T_j) \leq \mu_j \|x^t - T_j(x^t)\|^{\gamma_{1,j}} \leq \mu \|x^t - T_j(x^t)\|^{\gamma_1} \quad \forall x \in K. \tag{3.11}$$

Also, since $\{\operatorname{Fix} T_j\}_{j \in J}$ has a boundedly Hölder regular intersection, there exist an exponent $\gamma_2 > 0$ and a scalar $\beta > 0$ such that

$$\operatorname{dist}(x, F) \leq \beta \left(\max_{j \in J} \operatorname{dist}(x, \operatorname{Fix} T_j) \right)^{\gamma_2} \quad \forall x \in K. \tag{3.12}$$

Fix an arbitrary index $j' \in J$. Assumption (c) ensures that, for any $t \in \mathbb{N}$, there exists index $t_k \in \{ts, \ldots, (t+1)s - 1\}$ such that $j' \in J_+(t_k)$. Then

$$\begin{aligned}
\operatorname{dist}^2(x^{ts}, \operatorname{Fix} T_{j'}) &\leq \left(\operatorname{dist}(x^{t_k}, \operatorname{Fix} T_{j'}) + \|x^{ts} - x^{t_k}\| \right)^2 \\
&\leq \left(\operatorname{dist}(x^{t_k}, \operatorname{Fix} T_{j'}) + \sum_{n=ts}^{t_k - 1} \|x^n - x^{n+1}\| \right)^2 \\
&\leq (t_k - ts + 1) \left(\operatorname{dist}^2(x^{t_k}, \operatorname{Fix} T_{j'}) + \sum_{n=ts}^{t_k - 1} \|x^n - x^{n+1}\|^2 \right) \\
&\leq s \left(\mu \left(\|x^{t_k} - T_{j'}(x^{t_k})\|^2 \right)^{\gamma_1} + \sum_{n=ts}^{(t+1)s-1} \|x^n - x^{n+1}\|^2 \right),
\end{aligned} \tag{3.13}$$

where the second from the last inequality follows from convexity of the function $(\cdot)^2$, and the last uses (3.11) noting that $j' \in J_+(t_k)$.

Since each T_j is α-averaged, for all $t \in \mathbb{N}$, the convex combination $\sum_{j \in J} w_{t,j} T_j$ is α-averaged (and, in particular, nonexpansive), and hence, for all $x \in H$ and $y \in F$, we have

$$\begin{aligned}
\left\| \sum_{j \in J} w_{j,t} T_j(x) - y \right\|^2 &= \left\| \sum_{j \in J} w_{j,t} \big(T_j(x) - y \big) \right\|^2 \\
&\leq \sum_{j \in J} w_{j,t} \|T_j(x) - y\|^2 \\
&\leq \sum_{j \in J} w_{j,t} \left(\|x - y\|^2 - \frac{1-\alpha}{\alpha} \|x - T_j(x)\|^2 \right) \\
&= \|x - y\|^2 - \frac{1-\alpha}{\alpha} \sum_{j \in J} w_{j,t} \|x - T_j(x)\|^2 \\
&\leq \|x - y\|^2 - \sigma \left(\frac{1-\alpha}{\alpha} \right) \|x - T_j(x)\|^2 \quad \forall j \in J_+(t).
\end{aligned} \tag{3.14}$$

We therefore have that

$$
\begin{aligned}
\sigma \frac{1-\alpha}{\alpha}\|x^{t_k}-T_{j'}(x^{t_k})\|^2 &\le \|x^{t_k}-P_F(x^{ts})\|^2-\|x^{t_k+1}-P_F(x^{ts})\|^2 \\
&\le \|x^{ts}-P_F(x^{ts})\|^2-\|x^{(t+1)s}-P_F(x^{ts})\|^2 \\
&\le \operatorname{dist}^2(x^{ts},F)-\operatorname{dist}^2(x^{(t+1)s},F).
\end{aligned} \tag{3.15}
$$

Furthermore, for each $n \in \{ts,\dots,(t+1)s-1\}$, applying $x = x^n$ and $y = P_F(x^{ts})$ in (3.14) we have

$$
\begin{aligned}
\frac{1-\alpha}{\alpha}\|x^n-x^{n+1}\|^2 &= \frac{1-\alpha}{\alpha}\|x^n-\sum_{j\in J} w_{j,n}T_j(x^n)\|^2 \\
&\le \frac{1-\alpha}{\alpha}\sum_{j\in J} w_{j,n}\|x^n-T_j(x^n)\|^2 \\
&\le \|x^n-P_F(x^{ts})\|^2-\|x^{n+1}-P_F(x^{ts})\|^2,
\end{aligned}
$$

and thus it follows that

$$
\begin{aligned}
\frac{1-\alpha}{\alpha}\sum_{n=ts}^{(t+1)s-1}\|x^n-x^{n+1}\|^2 &\le \sum_{n=ts}^{(t+1)s-1}\left(\|x^n-P_F(x^{ts})\|^2-\|x^{n+1}-P_F(x^{ts})\|^2\right) \\
&= \|x^{ts}-P_F(x^{ts})\|^2-\|x^{(t+1)s}-P_F(x^{ts})\|^2 \\
&\le \operatorname{dist}^2(x^{ts},F)-\operatorname{dist}^2(x^{(t+1)s},F) \\
&\le \left(\operatorname{dist}^2(x^{ts},F)-\operatorname{dist}^2(x^{(t+1)s},F)\right)^{\gamma_1},
\end{aligned} \tag{3.16}
$$

where the last inequality follows from (3.9). Altogether, combining (3.13), (3.15), and (3.16) gives

$$
\operatorname{dist}^2(x^{ts},\operatorname{Fix}T_{j'}) \le s\left(\mu\left(\frac{\alpha}{\sigma(1-\alpha)}\right)^{\gamma_1}+\frac{\alpha}{1-\alpha}\right)\left(\operatorname{dist}^2(x^{ts},F)-\operatorname{dist}^2(x^{(t+1)s},F)\right)^{\gamma_1}.
$$

Since $j' \in J$ was chosen arbitrarily, using (3.12) we obtain

$$
\begin{aligned}
\operatorname{dist}^2(x^{ts},F) &\le \beta^2 \max_{j\in J}\operatorname{dist}^{2\gamma_2}(x^{ts},\operatorname{Fix}T_j) \\
&\le \delta^{-1}\left(\operatorname{dist}^2(x^{ts},F)-\operatorname{dist}^2(x^{(t+1)s},F)\right)^{1/\theta},
\end{aligned} \tag{3.17}
$$

where the constant $\delta > 0$ and $\gamma > 0$ are given by

$$
\delta := \left(s^{\gamma_2}\beta^2\left(\mu\left(\frac{\alpha}{\sigma(1-\alpha)}\right)^{\gamma_1}+\frac{\alpha}{1-\alpha}\right)^{\gamma_2}\right)^{-1}, \qquad \theta := \frac{1}{\gamma_1\gamma_2}.
$$

Rearranging (3.17) gives

$$
\operatorname{dist}^2(x^{(t+1)s},F) \le \operatorname{dist}^2(x^{ts},F)-\delta\operatorname{dist}^{2\theta}(x^{ts},F),
$$

Then, the first assertion follows from Proposition 3.1.

To establish the second assertion, we suppose that the assumptions (a′), (b′), and (c) hold. We proceed with the same proof as above, and noting that the exponents γ_{1j} and γ_2 are now independent of the choice of K, we see that the second assertion also follows. ∎

Remark 3.2. A closer look at the proof of Theorem 3.1 reveals that a quantification of the constants M_1, M_2, and r is possible using the various regularity constants/exponents and (3.6). More precisely, (3.6) holds with

$$\delta := \left(s^{\gamma_2}\beta^2\left(\mu\left(\frac{\alpha}{\sigma(1-\alpha)}\right)^{\gamma_1}+\frac{\alpha}{1-\alpha}\right)^{\gamma_2}\right)^{-1}, \qquad \theta := \frac{1}{\gamma_1\gamma_2}, \quad F := \bigcap_{j\in J}\operatorname{Fix} T_j.$$

Here μ is the max of the constants of bounded Hölder regularity of the individual operators T_j and β is the constant of bounded Hölder regularity of the collection $\{\operatorname{Fix} T_j\}_{j\in J}$, respectively, on an appropriate compact set. Consequently, these expressions, appropriately specialized, also hold for all the subsequent corollaries of Theorem 3.1.

Remark 3.3. Theorem 3.1 generalizes [10, Thm. 6.1], which considers the special case in which the Hölder exponents are independent of the bounded set K and are specified by $\gamma_{1j} = \gamma_2 = 1$, $j = 1, \ldots, m$.

A slight refinement of Theorem 3.1 which allows for extrapolations as well as different averaging constants is possible. More precisely, an *extrapolation* of the operator T (in the sense of [8]) is a (nonconvex) combination of the form $wT + (1-w)I$, where the weight w may take values larger than 1. Recall that for a finite set Ω, we use $|\Omega|$ to denote the number of elements of Ω.

COROLLARY 3.4 (extrapolated quasi-cyclic algorithm). *Let* $J := \{1, 2, \ldots, m\}$, *let* $\{T_j\}_{j\in J}$ *be a finite family of* α_j*-averaged operators on a Hilbert space* H *with* $\cap_{j\in J}\operatorname{Fix} T_j \neq \emptyset$ *and* $\alpha_j \in (0,1)$, *and let* $\alpha > 0$ *be such that* $\max_{j\in J}\alpha_j \leq \alpha$. *For each* $t \in \mathbb{N}$, *let* $w_{0,t} \in \mathbb{R}$ *and* $w_{j,t} \in [0, \frac{\alpha}{\alpha_j}]$ $j \in J$, *be such that*

$$\sum_{j\in J} w_{j,t} + w_{0,t} = 1 \text{ and } \sup_{t\in\mathbb{N}}\left\{\sum_{j\in J}\alpha_j w_{j,t}\right\} < \alpha.$$

Let $x^0 \in H$ *and consider the extrapolated quasi-cyclic algorithm generated by*

$$x^{t+1} = \sum_{j\in J} w_{j,t} T_j(x^t) + w_{0,t}x^t \quad \forall t \in \mathbb{N}. \tag{3.18}$$

Assume the following hypotheses:

(a) *For each* $j \in J$, *the operator* T_j *is bounded Hölder regular.*
(b) $\{\operatorname{Fix} T_j\}_{j\in J}$ *has a boundedly Hölder regular intersection.*
(c) $\sigma := \inf_{t\in\mathbb{N}}\inf_{j\in J_+(t)}\{w_{j,t}\} > 0$, *where* $J_+(t) = \{j \in J : w_{j,t} > 0\}$ *for each* $t \in \mathbb{N}$, *and there exists an* $s \in \mathbb{N}$ *such that*

$$J_+(t) \cup J_+(t+1) \cup \cdots \cup J_+(t+s-1) = J \quad \forall t \in \mathbb{N}.$$

Then $x^t \to \bar{x} \in \cap_{j\in J}\operatorname{Fix} T_j \neq \emptyset$ *at least with a sublinear rate* $O(t^{-\rho})$ *for some* $\rho > 0$.

In particular, if we assume (a′), (b′), *and* (c) *hold where* (a′), (b′) *are given by*

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

then there exist constants $M_1, M_2 \geq 0$ *and* $r \in [0,1)$ *such that*

$$\|x^t - \bar{x}\| \leq \begin{cases} M_1 t^{-\frac{\gamma}{2(1-\gamma)}}, & \gamma \in (0,1), \\ M_2\, r^t, & \gamma = 1, \end{cases}$$

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

Proof. For each $j \in J$, by Definition 2.3(2.3), the operator $\overline{T}_j$ is α-averaged where

$$\overline{T}_j := \frac{\alpha}{\alpha_j} T_j - \left(\frac{\alpha}{\alpha_j} - 1\right) I.$$

Let $\overline{w}_{j,t} := \frac{w_{j,t}\alpha_j}{\alpha}$, $j \in J$. Then, $\overline{w}_{j,t} \geq 0$ and

$$\sum_{j\in J} \overline{w}_{j,t} = \sum_{j\in J} \frac{w_{j,t}\alpha_j}{\alpha} = \frac{1}{\alpha}\sum_{j\in J} w_{j,t}\alpha_j < 1.$$

Let $\overline{w}_{0,t} := 1 - \sum_{j\in J} \overline{w}_{j,t}$, $\overline{T}_0(x) := x$ for all $x \in H$, and $\overline{J} = J \cup \{0\}$. Then, for all $t \in \mathbb{N}$, $\sum_{j\in\overline{J}} \overline{w}_{j,t} = 1$ with $w_{j,t} \geq 0$, $j \in \overline{J}$, and

$$\begin{aligned} x^{t+1} &= \sum_{j\in J} w_{j,t} T_j\left(x^t\right) + w_{0,t}x^t \\ &= \sum_{j\in J} \overline{w}_{j,t}\frac{\alpha}{\alpha_j} T_j(x^t) - \left(\sum_{j\in J} \overline{w}_{j,t}\frac{\alpha}{\alpha_j} - 1\right) x^t \\ &= \sum_{j\in J} \overline{w}_{j,t}\left[\frac{\alpha}{\alpha_j} T_j(x^t) - \left(\frac{\alpha}{\alpha_j} - 1\right) x^t\right] + \left(1 - \sum_{j\in J}\overline{w}_{j,t}\right) x^t \\ &= \sum_{j\in\overline{J}} \overline{w}_{j,t}\overline{T}_j(x^t). \end{aligned}$$

Since $\mathrm{Fix}\ \overline{T}_j = \mathrm{Fix}\ T_j$ for all $j \in J$, $\{\mathrm{Fix}\ \overline{T}_j\}_{j\in J}$ is bounded Hölder regular (with uniform exponent γ_2) if and only if $\{\mathrm{Fix}\ T_j\}_{j\in J}$ is bounded Hölder regular (with uniform exponent γ_2). This together with $\mathrm{Fix}\ \overline{T}_0 = H$ implies that $\{\mathrm{Fix}\ \overline{T}_j\}_{j\in\overline{J}}$ also has a boundedly Hölder regular intersection. For all $j \in J$, $\|x - \overline{T}_j x\| = \frac{\alpha}{\alpha_j}\|T_j x - x\|$ and so $\overline{T}_j$ is bounded Hölder regular (with uniform exponent $\gamma_{1,k}$) if and only if T_j is bounded Hölder regular (with uniform exponent $\gamma_{1,k}$). Clearly, $\overline{T}_{l+1}$ is bounded Hölder regular with a uniform exponent 1. Thus, assumptions (a) and (b) of Theorem 3.1 hold for $\{\overline{w}_{j,t}\}_{j\in\overline{J}}$. Moreover, noting that by assumption $\sup_{t\in\mathbb{N}}\{\sum_{j\in J} w_{j,t}\alpha_j\} < \alpha$, we have

$$\inf_{t\in\mathbb{N}} \overline{w}_{0,t} = \inf_{t\in\mathbb{N}}\left\{1 - \sum_{j\in J}\overline{w}_{j,t}\right\} = \inf_{t\in\mathbb{N}}\left\{1 - \sum_{j\in J}\frac{w_{j,t}\alpha_j}{\alpha}\right\} = \frac{\alpha - \sup_{t\in\mathbb{N}}\{\sum_{j\in J} w_{j,t}\alpha_j\}}{\alpha} > 0.$$

This together with assumption (c) of this corollary implies that assumption (c) of Theorem 3.1 holds for $\{\overline{w}_{j,t}\}_{j\in\overline{J}}$. Therefore, the claimed result now follows from Theorem 3.1. ∎

Remark 3.5 (common fixed points). Throughout this work we assume the collection of operators $\{T_j\}_{j\in J}$ (J a finite index set) to have a common fixed point. In this setting, with appropriate nonexpansivity properties, one has that the fixed point set of convex combinations or compositions of the operators $\{T_j\}_{j\in J}$ is precisely the set of their common fixed points. While there do exist several instances where such an assumption does not hold (e.g., regularization schemes such as [44]), this does not preclude their analysis using the theory presented here (see Proposition 4.2). Indeed, for such cases, the fixed point set of an appropriate convex combination or composition of operators is nonempty, and this aggregated operator is thus amenable to our results (rather than the individual operators themselves). The question of usefully characterizing the fixed point set of this aggregated operator must then be addressed, a matter significantly more subtle in the absence of a common fixed point.

We next provide four important specializations of Theorem 3.1. The first result is concerned with a simple fixed point iteration, the second with a Kransnoselskii–Mann scheme, the third with the method of cyclic projections, and the fourth with forward-backward splitting for variational inequalities.

COROLLARY 3.6 (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 $x^t \to \bar{x} \in \operatorname{Fix} T \neq \emptyset$ at least with a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. In particular, if T is bounded Hölder regular with uniform exponent $\gamma \in (0,1]$, then 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 3.1. ∎

COROLLARY 3.7 (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 $x^t \to \bar{x} \in \operatorname{Fix} T \neq \emptyset$ at least with a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. In particular, if T is bounded Hölder regular with uniform 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 uniform 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 3.1. ∎

The following result includes [17, Thm. 4.4] and [6, Thm. 3.12] as special cases.

COROLLARY 3.8 (cyclic projection algorithm with Hölder regularity). *Let $J = \{1,2,\dots,m\}$ and let $\{C_j\}_{j\in J}$ be a collection of closed convex subsets of a Hilbert space H with nonempty 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 $x^t \to \bar{x} \in \cap_{j\in J} C_j \neq \emptyset$ at least with a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. In particular, if the collection $\{C_j\}_{j\in J}$ is bounded Hölder regular with uniform 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 uniform exponent 1. The result follows from Theorem 3.1. ∎

We now turn our attention to *variational inequalities* [7, Chap. 25.5]. Let $f : H \to (-\infty, +\infty]$ be a proper lower semicontinuous (l.s.c.) convex function and let $F : H \mapsto H$ be β-*cocoercive*, that is,

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

The *generalized variational inequality problem*, denoted $\mathrm{VIP}(F, f)$, is

$$\text{Find } x^* \in H \text{ such that } f(x) - f(x^*) + \langle F(x^*), x - x^* \rangle \geq 0, \tag{3.19}$$

and the set of *solutions* to (3.19) is denoted $\mathrm{Sol}(\mathrm{VIP}(F, f))$.

The *forward-backward splitting method* is often employed to solve $VIP(F, f)$ (see, for instance, [7, Prop. 25.18]) and generates a sequence $\{x^t\}$ according to

$$x^{t+1} = x^t + \lambda_t \big(R_{\gamma \partial f}(x^t - \gamma F(x^t)) - x^t \big) \quad \forall t \in \mathbb{N}, \tag{3.20}$$

where $R_{T_0} := (I + T_0)^{-1}$ denotes the *resolvent* of an operator T_0. Here, we note that ∂f is a maximal monotone operator and so its resolvent is single-valued.

COROLLARY 3.9 (forward-backward splitting method for variational inequality problem). *Let* $f : H \to (-\infty, +\infty]$ *be a proper l.s.c. convex function, let* $F : H \to H$ *be a* β*-cocoercive operator with* $\beta > 0$*, let* $\gamma \in (0, 2\beta)$*, and let* $(\lambda_t)_{t\in\mathbb{N}}$ *be a sequence of real numbers with* $\sigma_0 := \inf_{t\in\mathbb{N}} \{\lambda_t(1 - \lambda_t)\} \in (0, 1)$*. Suppose* $\mathrm{Sol}(\mathrm{VIP}(F, f)) \neq \emptyset$*. Let* $x^0 \in H$ *and set*

$$x^{t+1} = (1 - \lambda_t)\, x^t + \lambda_t R_{\gamma \partial f}(x^t - \gamma F(x^t)) \quad \forall t \in \mathbb{N}. \tag{3.21}$$

If $T_2 := R_{\gamma \partial f}(I - \gamma F)$ *is bounded Hölder regular, then* $x^t \to \bar{x} \in \mathrm{Sol}(\mathrm{VIP}(F, f))$ *at least with a sublinear rate* $O(t^{-\rho})$ *for some* $\rho > 0$*. In particular, if* T *is bounded Hölder regular with uniform 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. Since F is β-cocoercive, [7, Prop. 4.33] shows that $I - \gamma F$ is $\gamma/(2\beta)$-averaged. The operator $R_{\gamma \partial f}$ is $\frac{1}{2}$-averaged, as the resolvent of a maximally monotone operator. By [7, Prop. 4.32], T_2 is $\frac{2}{3}$-averaged. Applying Corollary 3.7 with $T = T_2$, the claimed convergence rate to a point $\bar{x} \in \operatorname{Fix} T_2$ thus follows.

It remains to show that $\operatorname{Fix} T_2 = \mathrm{Sol}(\mathrm{VIP}(F, f))$. Indeed, $x \in \operatorname{Fix} T_2$ if and only if

$$x = (I + \gamma \partial f)^{-1} \big((I - \gamma F)(x) \big) \iff x - \gamma F(x) \in x + \gamma \partial f(x) \iff 0 \in (\partial f + F)(x).$$

The claimed result follows. ∎

Remark 3.10 (semialgebraic forward-backward splitting). Corollary 3.9 applies, in particular, when f and F are semialgebraic functions and $H = \mathbb{R}^n$. To see this, first note that ∂f is semialgebraic as the subdifferential of the semialgebraic function [26, p. 6]. The graph of the resolvent to $\gamma\partial f$ may then be expressed as

$$\text{gph } R_{\gamma\partial f} = \{\big(z, (I+\gamma\partial f)^{-1}(z)\big) : z \in \mathbb{R}^n\} = \{((I+\gamma\partial f)(z), z) : z \in \mathbb{R}^n\},$$

which is a semialgebraic set since $(I + \gamma\partial f)$ is a semialgebraic function by Fact 2.4, and hence the resolvent is a semialgebraic mapping. A further application of Fact 2.4 shows that T_2 is semialgebraic, as the composition of semialgebraic maps.

We may therefore define two continuous semialgebraic functions $\phi(x) := \text{dist}(x, \text{Fix}T_2)$ and $\varphi(x) := \|x - Tx\|$. Since $\phi^{-1}(0) = \text{Fix}T_2 = \varphi^{-1}(0)$, (P6) of Fact 2.4 yields, for any compact semialgebraic set K, the existence of constants $c > 0$ and $\tau \in (0, 1]$ such that

$$\|x - T_2x\| \leq c\,\text{dist}^\tau(x, \text{Fix}T_2) \quad \forall x \in K,$$

or, in other words, the bounded Hölder regularity of T_2.

4. The rate of convergence of DR algorithms for convex feasibility problems. We now specialize our convergence results to the classical DR algorithm and its variants in the setting of convex feasibility problems. In doing so, a convergence rate is obtained under the Hölder regularity condition. Recall that the basic DR algorithm for two set feasibility problems can be stated as in Algorithm 1.

Direct verification shows that the relationship between consecutive terms in the sequence (x^t) of (4.1) can be described in terms of the firmly nonexpansive *(two-set) DR operator* which is of the form

$$T_{C,D} = \frac{1}{2}\left(I + R_D R_C\right), \tag{4.2}$$

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

We shall also consider the abstraction given by Algortihm 2, which chooses two constraint sets from some finite collection at each iteration. Note that iterations (4.1) and (4.3) have the same structure.

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

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

Algorithm 1. Basic DR algorithm.

Data: Two closed and convex sets $C, D \subseteq H$

Choose an initial point $x^0 \in H$ **for** $t = 0, 1, 2, 3, \dots$ **do**

Set:

$$\begin{cases} y^{t+1} := \mathrm{P}_C(x^t), \\ z^{t+1} := \mathrm{P}_D(2y^{t+1} - x^t), \\ x^{t+1} := x^t + (z^{t+1} - y^{t+1}). \end{cases} \tag{4.1}$$

end

Algorithm 2. A multiple-sets DR 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 $\Omega_1, \ldots, \Omega_p \in \{(i_1, i_2) : i_1, i_2 = 1, 2, \ldots, m \text{ and } i_1 \neq i_2\}$ with $\cup_{j=1}^{p} \Omega_j = \{1, \ldots, m\}$ Define $\Omega_0 = \Omega_m$ Choose an initial point $x^0 \in H$ **for** $t = 0, 1, 2, 3, \ldots$ **do**

Set the indices $(i_1, i_2) := \Omega_{t'}$ where $t' = t + 1 \bmod s$ Set

$$\begin{cases} y^{t+1} := \mathrm{P}_{C_{i_1}}(x^t), \\ z^{t+1} := \mathrm{P}_{C_{i_2}}(2y^{t+1} - x^t), \\ x^{t+1} := x^t + (z^{t+1} - y^{t+1}). \end{cases} \tag{4.3}$$

end

which corresponds to Algorithm 2 with $u^t = x^{mt}$, $t \in \mathbb{N}$, $p = m$ and $\Omega_j = (j, j+1)$, $j = 1, \ldots, m-1$, and $\Omega_m = (m, 1)$. The second scheme the *cyclically anchored DR algorithm* and is generated according to

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

which corresponds to Algorithm 2 with $u^t = x^{(m-1)t}$, $t \in \mathbb{N}$, $p = m - 1$ and $\Omega_j = (1, j+1)$, $j = 1, \ldots, m-1$. The following lemma shows that underlying operators of both these methods are also averaged.

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

Proof. The two-set *DR operator* of (4.2) is firmly nonexpansive and hence 1/2-averaged. The result follows by [7, Prop. 4.32]. □

Corollary 4.1 (convergence rate for the multiple-sets DR algorithm). *Let $C_1, C_2, \ldots, C_m$ be closed convex sets in a Hilbert space H with nonempty intersection. Let $\{\Omega_j\}_{j=1}^{p}$ and $\{(y^t, z^t, x^t)\}$ be generated by the multiple-sets DR algorithm* (4.1). *Suppose the follwoing:*

(a) *For each $j \in \{1, \ldots, p\}$, the operator T_{Ω_j} is bounded Hölder regular.*
(b) *The collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{p}$ has a bounded Hölder regular intersection.*

Then $x^t \to \bar{x} \in \cap_{j=1}^{p} \mathrm{Fix}\, T_{\Omega_j}$ with at least a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. In particular, suppose we assume the stronger assumptions:

(a′) *For each $j \in \{1, \ldots, p\}$, the operator T_{Ω_j} is bounded Hölder regular with uniform exponent $\gamma_{1,j}$.*
(b′) *The collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{p}$ has a bounded Hölder regular intersection with uniform 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 all $j \in J$, set $T_j = T_{\Omega_j}$ and

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

Since T_{Ω_j} is firmly nonexpansive (that is, 1/2-averaged), the conclusion follows immediately from Theorem 3.1. ∎

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

PROPOSITION 4.1 (semialgebraicity implies Hölder regularity and sublinear convergence). *Let $C_1, C_2, \ldots, C_m$ be convex, semialgebraic sets in $\mathbb{R}^n$ with nonempty intersection which can be described by polynomials (in the sense of* (2.1)*) on $\mathbb{R}^n$ having degree d. Let $\{\Omega_j\}_{j=1}^p$ be a list of 2-tuples with $\cup_{j=1}^p \Omega_j = \{1, \ldots, m\}$. Then the following hold:*

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

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

In particular, let $\{(y^t, z^t, x^t)\}$ be generated by the multiple-sets DR algorithm (4.3)*. Then there exists $\rho > 0$ such that $x^t \to \bar{x} \in \cap_{j=1}^p \mathrm{Fix}\, T_{\Omega_j}$ with at least a sublinear rate $O(t^{-\rho})$.*

Proof. Fix any $j \in \{1, \ldots, p\}$. 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 *DR 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 semialgebraic set $C \subseteq \mathbb{R}^n$, the projection mapping $x \mapsto \mathrm{P}_C(x)$ is a semialgebraic 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 semialgebraic mappings. Since the composition of semialgebraic maps remains semialgebraic ((P5) of Fact 2.4), we deduce that $f \colon x \mapsto \|T_{C_1,C_2}x - x\|^2$ is a continuous semialgebraic function. By (P4) of Fact 2.4, $\mathrm{Fix}\, T_{C_1,C_2} = \{x \mid f(x) = 0\}$ which is therefore a semialgebraic set. By (P2) of Fact 2.4, the function $\mathrm{dist}(\cdot, \mathrm{Fix}\, T_{C_1,C_2})$ is semialgebraic, and clearly $\mathrm{dist}(\cdot, \mathrm{Fix}\, T_{C_1,C_2})^{-1}(0) = f^{-1}(0)$.

By the Łojasiewicz inequality for semialgebraic functions ((P6) of Fact 2.4), we see that for every $\rho > 0$, one can find $\mu > 0$ and $\gamma \in (0, 1]$ such that

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

So, the *DR 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 [47], 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 [49] implies that for all $a \in \mathbb{R}^n$, there exist $\mu > 0$, $\epsilon > 0$ such that

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

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

Next, we assert that the collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^p$ has a bounded Hölder regular intersection. To see this, as in the proof of part (a), we can show that for each $j = 1, \ldots, p$, $\mathrm{Fix}\, T_{\Omega_j}$ is a semialgebraic set. Then, their intersection $\cap_{j=1}^p \mathrm{Fix}\, T_{\Omega_j}$ is also a semialgebraic set. Thus, $\psi(x) = \mathrm{dist}(x, \cap_{j=1}^{p} \mathrm{Fix}\, T_{\Omega_j})$ and $\phi(x) = \max_{1\le j\le p}$ $\mathrm{dist}(x, \mathrm{Fix}\, T_{\Omega_j})$ are semialgebraic functions. It is easy to see that $\phi^{-1}(0) = \psi^{-1}(0)$ and hence the Łojasiewicz inequality for semialgebraic functions ((P6) of Fact 2.4) implies that the collection $\{\mathrm{Fix}\, T_{\Omega_j}\}_{j=1}^{p}$ has a bounded Hölder regular intersection.

The final conclusion follows by Theorem 3.1. ∎

Next, we establish the convergence rate for DR algorithm assuming bounded Hölder regularity of the *DR operator* $T_{C,D}$.

COROLLARY 4.2 (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* DR operator*. Let $\{(y^t, z^t, x^t)\}$ be generated by the DR algorithm* (4.1)*. Suppose that $T_{C,D}$ is bounded Hölder regular. Then $x^t \to \bar{x} \in \mathrm{Fix}\, T_{C,D}$ with at least a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. In particular, if $T_{C,D}$ is bounded Hölder regular with uniform exponent $\gamma \in (0, 1]$, then there exist $M > 0$ and $r \in (0, 1)$ such that*

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

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 3.1. ∎

Similar to Proposition 4.1, if C and D are basic convex semialgebraic sets, then the DR algorithm exhibits at least a sublinear convergence rate.

Remark 4.3 (linear convergence of the DR algorithm for convex feasibility problems). We note that if $H = \mathbb{R}^n$ and C, D are convex sets with $\mathrm{ri}C \cap \mathrm{riD} \neq \emptyset$, then $T_{C,D}$ is bounded linear regular, that is, bounded Hölder regular with uniform exponent 1. In this case, the DR algorithm converges linearly as was previously established in [10]. Further, 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 established in [11]. It should be noted that [11] deduced the stronger result that the linear convergence rate is exactly the cosine of the Friedrichs angle. We also remark that linear convergence of the DR algorithm under the regularity condition $\mathrm{ri}C \cap \mathrm{riD} \neq \emptyset$ may alternatively be deduced as a consequence of the recently established local linear convergence of nonconvex DR algorithms [40, 45] by specializing C and D to be convex sets.

To conclude this section, we consider a regularization of the DR algorithm [44] which converges even when the target intersection is empty where the sequence is generated by $x^{t+1} = T_R(x^t)$ where $T_R := \beta \mathrm{P_C} + (1-\beta) T_{C,D}$ and $\beta \in (0, 1)$. When the target intersection $C \cap D$ is empty, this gives an example of a useful algorithm in which the two operators of interest, P_C and $T_{C,D}$, have no common fixed point but can still be analyzed within our framework.

PROPOSITION 4.2 (convergence rate of regularized DR algorithm). *Let C, D be two basic convex semialgebraic sets and let $T_{C,D}$ be the* DR operator*. Let $\{x^t\}$ be generated by $x^{t+1} := T_R(x^t)$, where $T_R = \beta \mathrm{P_C} + (1-\beta) T_{C,D}$ and $\beta \in (0, 1)$. Then $\mathrm{P}_D(x^t) \to \bar{x}_1 \in D$ and $\mathrm{P}_C(\mathrm{P}_D(x^t)) \to \bar{x}_2 \in C$ both with at least a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$ and $\|\bar{x}_1 - \bar{x}_2\| = \mathrm{dist}(C, D)$. In particular, if C, D are polyhedral, then there exist $M > 0$ and $r \in (0, 1)$ such that*

$$\max\{\|\mathrm{P}_D(x^t) - \bar{x}_1\|, \|\mathrm{P}_C(\mathrm{P}_D(x^t)) - \bar{x}_2\|\} \le M\, r^t.$$

Proof. As C, D are two basic convex semialgebraic sets, $D-C$ is a closed set [17, Lem. 4.7]. Let $g = P_{D-C}(0)$, $E = C \cap (D-g)$ and $F = (C+g) \cap D$. Then, [44, Lem. 2.1] shows that $\mathrm{Fix}T_R = F - \frac{\beta}{1-\beta}\, g$, $P_D(\mathrm{Fix}T_R) \subseteq F$ and $P_C(P_D(\mathrm{Fix}T_R)) \subseteq E$. We first show that T_R is bounded Hölder regular, and T_R is bounded Hölder regular with uniform exponent 1 if C, D are polyhedral. To see this, we use the same argument as in Proposition 4.1 and it suffices to establish that $T_R - I$ is a semialgebraic (resp., piecewise affine) map when C and D are semialgebraic (resp., polyhedral). For the sake of avoiding repetition, we only show the latter. Observe that

$$T_R - I = \beta(\mathrm{P_C} - I) + (1-\beta)(\mathrm{P_D}\mathrm{R_C} - \mathrm{P_C}).$$

Thus $T_R - I$ can be represented in terms of linear combinations and compositions of continuous semialgebraic (resp., continuous piecewise affine) operators, more precisely, the projectors onto C and D. Since projectors of convex semialgebraic (resp., polyhedral) sets are semialgebraic (resp., piecewise affine) and continuous, it follows that $T_R - I$ is semialgebraic (resp., piecewise affine) and continuous. It then follows from Theorem 3.1 that $x^t \to \bar{x} \in \mathrm{Fix}T_R$ with at least a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. From the nonexpansive property of projection mapping, we see that $\mathrm{P}_D(x^t) \to \bar{x}_1 \in P_D(\mathrm{Fix}T_R) \subseteq F$ and $\mathrm{P}_C(\mathrm{P}_D(x^t)) \to \bar{x}_2 = P_C(\bar{x}_1) \subseteq E$ both with at least a sublinear rate $O(t^{-\rho})$. From the definitions of E, F, and g, it follows that $\|\bar{x}_1 - \bar{x}_2\| = \mathrm{dist}(C, D)$. The assertion for the linear convergence in the case where C and D are polyhedral also follows from Theorem 3.1. ∎

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 DR 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

$$\mathrm{P}_A^\eta(x) = \mathrm{prox}_{\eta\mathrm{dist}_A^2}(x) = \operatorname*{arg\,min}_{y\in H}\left\{\mathrm{dist}_A^2(y) + \frac{1}{2\eta}\|y-x\|^2\right\} \quad \forall x \in H, \tag{5.1}$$

where prox_f denotes the *proximity operator* of the function f. The damped variant can be stated as in Algorithm 3.

Remark 5.1. In the more general setting in which P_C^η and P_D^η are, respectively, replaced by $\mathrm{prox}_{\eta f}$ and $\mathrm{prox}_{\eta g}$ for f and g proper l.s.c convex functions, Algorithm 3 can be found, for instance, in [7, Cor. 27.4], [44, Alg. 1.8], and [23]. Convergence of this algorithm (without an explicit estimate of the convergence rate) has been established in [44, Cor. 1.11] and [23, Cor. 5.2]. We also note that a similar relaxation of the DR algorithm for lattice cone constraints has been proposed and analyzed in [18].

While it is possible to analyze the damped DR algorithm within the quasi-cyclic framework, we learn more by proving the following result directly.

Algorithm 3. Damped DR algorithm.

Data: Two closed sets $C, D \subseteq H$

Choose $\eta > 0$ and $\lambda \in (0, 2]$;

Choose an initial point $x^0 \in H$ **for** $t = 0, 1, 2, 3, \dots$ **do**

Choose $\lambda_t \in (0, 2]$ with $\lambda_t \geq \lambda$ and set:

$$\left\{\begin{array}{l} y^{t+1} := \mathrm{P}_C^{\eta}(x^t), \\ z^{t+1} := \mathrm{P}_D^{\eta}(2y^{t+1} - x^t), \\ x^{t+1} := x^t + \lambda_t(z^{t+1} - y^{t+1}). \end{array}\right. \tag{5.2}$$

end

THEOREM 5.2 (convergence rate for the damped DR 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 DR algorithm* (5.2)*. Suppose that the pair of sets $\{C, D\}$ has a bounded Hölder regular intersection. Then $x^t \to \bar{x} \in C \cap D$ with at least a sublinear rate $O(t^{-\rho})$ for some $\rho > 0$. Furthermore, if the pair $\{C, D\}$ has a bounded Hölder regular intersection with uniform exponent $\gamma \in (0, 1]$, then there exist $M > 0$ and $r \in (0, 1)$ such that*

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

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

$$2\eta\lambda\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \leq \|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2. \tag{5.3}$$

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

$$\begin{aligned}
&2\eta\lambda\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \\
&\leq 2\eta\lambda_t\big(\mathrm{dist}^2(y^{t+1}, C) + \mathrm{dist}^2(z^{t+1}, D)\big) \\
&= 2\eta\lambda_t\big(\mathrm{dist}^2(y^{t+1}, C) - \mathrm{dist}^2(x^*, C) + \mathrm{dist}^2(z^{t+1}, D) - \mathrm{dist}^2(x^*, D)\big) \\
&\leq 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - x^*\rangle\big) \\
&= 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}), y^{t+1} - x^*\rangle + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - y^{t+1}\rangle \\
&\quad + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle\big) \\
&= 4\eta\lambda_t\big(\langle y^{t+1} - \mathrm{P_C}(y^{t+1}) + z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle \\
&\quad + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), z^{t+1} - y^{t+1}\rangle) \\
&= 4\eta\big(\lambda_t\langle y^{t+1} - \mathrm{P_C}(y^{t+1}) + z^{t+1} - \mathrm{P}_D(z^{t+1}), y^{t+1} - x^*\rangle \\
&\quad + \langle z^{t+1} - \mathrm{P}_D(z^{t+1}), x^{t+1} - x^t\rangle\big),
\end{aligned} \tag{5.4}$$

where the last equality follows from the last relation in (5.2). Now using (5.1), we see that

$$0 = \nabla\bigg(\mathrm{dist}^2(\cdot, C) + \frac{1}{2\eta}\|\cdot - x^t\|^2\bigg)(y^{t+1}) = 2\left(y^{t+1} - \mathrm{P_C}(y^{t+1})\right) + \frac{1}{\eta}(y^{t+1} - x^t),$$

and similarly

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

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 (5.4) gives

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

Step 2 *(establishing a recurrence for* $\operatorname{dist}^2(x^t, C\cap D)$*).* 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,

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

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 nonexpansiveness of $\mathrm{dist}(\cdot, D)$, we have

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

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 (5.3), we have

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

Note that

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

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

$$\begin{aligned}\|x^t - x^*\|^2 - \|x^{t+1} - x^*\|^2 &\ge 2\eta\left(\frac{1}{2\eta+1}\right)^2 c^{-1}\max\{\mathrm{dist}^2(x^t, C), \mathrm{dist}^2(x^t, D)\} \\ &= 2\eta\left(\frac{1}{2\eta+1}\right)^2 c^{-1}(\max\{\mathrm{dist}(x^t, C), \mathrm{dist}(x^t, D)\})^2. \qquad (5.5)\end{aligned}$$

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

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

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

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

where $\theta = \frac{1}{\gamma} \in [1, \infty)$. So, Fact 2.1 implies that $P_{C\cap D}(x^t) \to \bar{x}$ for some $\bar{x} \in C \cap D$. Setting $x^* = P_{C\cap D}(x^t)$ in (5.5) we therefore obtain

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

Now, the conclusion follows by applying Proposition 3.1 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 (5.4) 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 [6]). The preceding proposition therefore implies that the damped DR 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 convex basic semialgebraic sets.

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

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

where g_i, h_j, $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$, 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 DR algorithm (5.2). Then, $x^t \to \bar{x} \in C \cap D$. Moreover, there exist $M > 0$ and $r \in (0, 1)$ such that

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

where $\gamma = [\min\{\frac{(2d-1)^n+1}{2}, B(n-1)d^n\}]^{-1}$ 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 2.1 with $\theta = 1$, we see that for any compact set K, there exists $c > 0$ such that for all $x \in K$,

$$\operatorname{dist}(x, C \cap D) \leq c\,(\operatorname{dist}(x, C) + \operatorname{dist}(x, D))^\gamma \leq 2^\gamma c \max\{\operatorname{dist}(x, C), \operatorname{dist}(x, D)\}^\gamma,$$

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

Remark 5.6. Let C, D be two basic convex semialgebraic sets in $\mathbb{R}^n$ with $C \cap D \neq \emptyset$ and consider the associated convex feasibility problem: find $x^* \in C \cap D$. As an easy consequence of Theorem 5.5, we see that a solution with ϵ-tolerance of the convex feasibility problem, the number of iterations needed of the damped DR algorithm is at worst $O(\frac{1}{\sqrt[\rho]{\epsilon}})$, where $\rho := \frac{\gamma}{2(1-\gamma)}$ and γ is a constant given in Theorem 5.5 that can be explicitly determined.

6. Two examples. In this section we fully examine two concrete problems which illustrate the difficulty of establishing optimal rates. In addition to illustrating our approach, these examples also give some further insight into the sharpness of our derived qualitative behavior. 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 [12, Ex. 5.4].

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

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

where $d > 0$ is an even number. Clearly, $C \cap D = \{0_{\mathbb{R}^{n+1}}\}$ and $\mathrm{ri}C \cap \mathrm{ri}D = \emptyset$. 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 $r_k := \frac{1}{k^d}$,

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

Let $T_{C,D}$ be the *DR 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 4.1, the sequence (x^t, r^t) generated by the DR 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

First, on the route to showing bounded Hölder regularity, it can be verified (see also [9, 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

$$a = \frac{1}{1 + d\|a\|^{2d-2} + d\|a\|^{d-2}|r|}x. \tag{6.1}$$

So,

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

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 $x \neq 0_{\mathbb{R}^n}$

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

where the equality follows from (6.1). 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 $(x,r) \in K \backslash \{0_{\mathbb{R}^{n+1}}\}$,

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

where the equality follows from (6.1). 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 DR 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 [12] (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 for this example is not tight in the case $n = 1$. On the other hand, as noted in [12], their analysis is largely limited to the two-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 semialgebraic sets in $\mathbb{R}^2$:

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

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 [9, Cor. 3.9], the fixed point set is given by

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

We therefore have that

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

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

$$T_{C,D}x := \left(\frac{I + R_D R_C}{2}\right) x = (\alpha - 1 - \alpha x_1, \alpha x_2)\,, \tag{6.2}$$

and thus

$$\|x - T_{C,D}x\|^2 = ((1-\alpha) + x_1(1+\alpha))^2 + (x_2(1-\alpha))^2\,. \tag{6.3}$$

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

$$\frac{\sqrt{m^2 + 2m}}{2} \ge M \frac{m^2}{1+m}. \tag{6.4}$$

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 (6.3) 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

$$\|x - T_{C,D}x\| \ge \frac{m^2}{2(1+m)}\|x\|^3 = \frac{m^2}{2(1+m)}\operatorname{dist}^3(x, T_{C,D}x). \tag{6.5}$$

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 $\alpha \geq m$. So, the elementary inequality $\sqrt{a^2+b^2} \geq (a+b)/2$ for all $a, b \geq 0$ implies that

$$\|x - T_{C,D}x\| \geq \frac{\sqrt{m^2+2m}}{2}x_1 + \frac{1-\alpha}{2}\|x\|. \tag{6.6}$$

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

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

Then, by combining with (6.6), we deduce

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

The claimed equation (6.5) now follows from (6.4).

Subcase II ($x_1 \leq 0$). In this case, $\operatorname{dist}(x, \operatorname{Fix} T_{C,D}) = |x_2|$ 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, (6.5) also follows in this subcase.

Combining the two cases we have

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

That is, $T_{C,D}$ is bounded Hölder regular with exponent $\gamma = 1/3$. Therefore, for this example, Corollary 4.2 implies that the DR algorithm generated a sequence $\{x^t\}$ which converges to $\bar{x} \in \operatorname{Fix} T_{C,D} = (-\infty, 0] \times \{0\}$ at least with a sublinear convergence rate of $O(\frac{1}{\sqrt[4]{t}})$. 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 (6.2), 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.

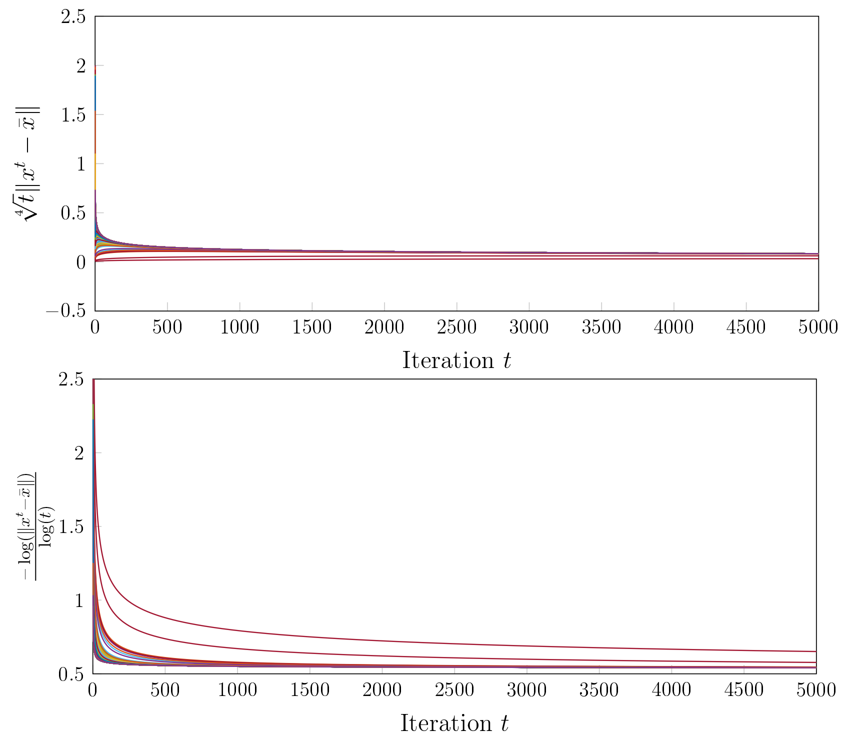

FIG. 1. *Numerical simulation results: (top) the successive change* $\sqrt[4]{t}\,\|x^t - \bar{x}\|$ *and (bottom) the ratio* $\frac{-\log(\|x^t-\bar{x}\|)}{\log(t)}$ *as a function of the number of iterations, t.*

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 of the choice of the initial points.

We now illustrate the sublinear convergence rate by numerical simulation. To do this, we first randomly generated an initial point in $[-100, 100]^2$. We then ran the DR algorithm for this example (starting with the corresponding random starting point) while 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 are plotted in Figure 1.

From the first graph, we see that the value of $\sqrt[4]{t}\,\|x^t - \bar{x}\|$ quickly decreases with increasing t. This supports the result 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 (6.2) 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 DR sequence $\{x^t\}$ is contained in $\mathbb{R}_{--} \times \mathbb{R}_{++}$. By Example 6.2, we have that $x^t = (u_t, v_t) \to (0,0)$. Below we verify that the sequence converges with exact sublinear order $O(1/\sqrt{t})$.

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

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

Taking square roots and inverting both sides we obtain

$$\limsup_{t\to\infty} \frac{v_t}{1/\sqrt{t}} \leq \sqrt{2}. \tag{6.7}$$

Now, recall that

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

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

$$0 > u_{t+1} \geq -v_t^2. \tag{6.8}$$

Combining (6.7) and (6.8), 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. Combined with (6.8), 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 have been established. The framework was then specialized to various fixed point algorithms including Krasnoselskii–Mann iterations, the cyclic projection algorithm, and the DR feasibility algorithm along with some variants. In the case where the underlying sets are convex semialgebraic, in a finite dimensional space, the results apply without any further regularity assumptions.

In particular, for our damped DR 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 emphasize that, unlike for the damped DR algorithm, we were not able to provide an explicit estimate of the sublinear convergence rate for the classical DR 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 [13]. As shown in [20, Cor. 3.1], if only two sets are involved and the initial point is chosen in a specific way, the cyclic DR 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 4.1 that the basic/cyclic DR method enjoys a sublinear convergence rate if the underlying sets are convex semialgebraic 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 semialgebraic convex sets applies only 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 [30].

Acknowledgments. The authors wish to thank Neal Hermer, Victor Isaac Kolobov, Simeon Reich, Rafał Zalas, and the three anonymous referees for their insightful comments.

REFERENCES

[1] F. J. A. 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. A. 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] J.-B. Baillon, R. E. Bruck, and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces,* Houston J. Math. 4 (1978), pp. 1–9.
[5] 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.
[6] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.
[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* Springer-Verlag, New York, 2011.
[8] H. H. Bauschke, P. L. Combettes, and S. G. Kruk, *Extrapolation algorithm for affine-convex feasibility problems,* Numer. Algorithms, 41 (2006), pp. 239–274.
[9] 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.
[10] 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.
[11] 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.
[12] 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., 23 (2016), pp. 237–261.
[13] 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.
[14] J. Bochnak, M. Coste, and M. F. Roy, *Real Algebraic Geometry,* Springer-Verlag, Berlin, 1998.
[15] J. M. Borwein, *Maximum entropy and feasibility methods for convex and nonconvex inverse problems,* Optimization, 61 (2012), pp. 1–33.
[16] 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.
[17] 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.
[18] J. M. Borwein, B. Sims, and M. K. Tam, *Norm convergence of realistic projection and reflection methods,* Optimization, 64 (2015), pp. 161–178.
[19] 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.
[20] J. M. Borwein and M. K. Tam, *A cyclic Douglas–Rachford iteration scheme,* J. Optim. Theory Appl., 160 (2014), pp. 1–29.
[21] Y. Censor and A. Cegielski, *Projection methods: an annotated bibliography of books and reviews,* Optimization, 64 (2015), pp. 2343–2358.
[22] P. L. Combettes, *The convex feasibility problem in image recovery,* Adv. Imaging Electron Phys., 95 (1996), pp. 155–270.
[23] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), no. 5-6, pp. 475–504.
[24] D. Davis, *Convergence rate analysis of the forward-Douglas–Rachford splitting scheme,* SIAM J. Optim., 25 (2015), pp. 1760–1786.
[25] D. Davis and W. Yin, *Convergence Rate Analysis of Several Splitting Schemes*, arXiv:1406.4834v3, 2015.
[26] D. Drusvyatskiy and A. S. Lewis, *Semi-algebraic functions have small subdifferentials*, Math. Program., 140 (2013), pp. 5–29.
[27] R. Escalante and M. Raydan, *Alternating Projection Methods,* SIAM, Philadelphia, 2011.
[28] P. Giselsson, *Tight Global Linear Convergence Rate Bounds for Douglas–Rachford Splitting*, arXiv:1506.01556, 2015.
[29] 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. Phys., 7 (1967), pp. 1–24.
[30] J. M. Gutiérrez, J. A. Jaramillo, and J. G. Llavona, *Polynomials and geometry of Banach spaces,* Extracta Math., 10 (1995), pp. 77–114.
[31] A. Y. Kruger. *Error bounds and metric subregularity,* Optimization, 64 (2015), pp. 49–79.
[32] A. Y. Kruger and H. T. Nguygen, *About [q]-regularity properties of collections of sets,* J. Math. Anal. Appl., 416 (2014), pp. 471–496.
[33] G. Li, *On the asymptotic well behaved functions and global error bound for convex polynomials,* SIAM J. Optim., 20 (2010), pp. 1923–1943.

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

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

[36] G. Li and T. K. Pong, *Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems,* Math. Program., 159 (2016), pp. 371–401, doi:10.1007/s10107-015-0963-5.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

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

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

[44] D. R. Luke, *Finding best approximation pairs relative to a convex and a prox-regular set in a Hilbert space,* SIAM J. Optim, 19(2008), pp. 714–739.

[45] H. Phan, *Linear convergence of the Douglas–Rachford method for two closed sets,* Optimization, 65 (2016), pp. 369–385, doi:10.1080/02331934.2015.1051532.

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

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

[48] 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, 72 (2016), 297–323.

[49] S. M. Robinson, *Some continuity properties of polyhedral multifunctions,* in Mathematical Programming at Oberwolfach, Math. Program. Stud. 14, Springer, Berlin, 1981, pp. 206–214.