

# Fejér monotone sequences revisited

Ulrich Kohlenbach and Pedro Pinto

Department of Mathematics

Technische Universität Darmstadt

Schlossgartenstraße 7, 64289 Darmstadt, Germany

kohlenbach@mathematik.tu-darmstadt.de, pinto@mathematik.tu-darmstadt.de

September 19, 2023

## Abstract

In this paper we introduce a localized and relativized generalization of the usual concept of Fejér monotonicity together with uniform and quantitative versions thereof and show that the main quantitative results obtained by the 1st author together with Nicolae and Leuştean in 2017 and with López-Acedo and Nicolae in 2019 respectively, extend to this generalization. Our framework, in particular, covers the sequence generated by the Dykstra algorithm while the latter is not Fejér-monotone in the ordinary sense. This gives a theoretical explanation why under a metric regularity assumption one obtains an explicit rate of convergence for Dykstra's algorithm which was proved recently by the 2nd author.

## 1 Introduction

Let  $X$  be a real Banach space,  $C \subseteq X$  a nonempty subset and  $(x_n)$  a sequence of points in  $X$ . An important feature enjoyed by many iterative methods in convex optimization is that of Fejér monotonicity:  $(x_n)$  is Fejér monotone w.r.t.  $C$  if

$$\forall p \in C \forall n \in \mathbb{N} (\|x_{n+1} - p\| \leq \|x_n - p\|).$$

Fejér monotone algorithms are frequently favored due to their good asymptotic behavior but, in general are only weakly convergent to some point in the set  $C$ . As most applications are naturally restricted to a finite dimensional setting, weak convergence gets then upgraded into strong convergence. Nevertheless, the general consensus appears to be that Fejér monotone methods are weakly convergent, and that strong convergence will prevent Fejér monotonicity since the sequence must eventually exhibit a preference towards some point in  $C$ .

In [2, 3], in a metric setting, the property of Fejér monotonicity was extensively studied through the lenses of proof mining techniques. In [2], it was shown that under a compactness assumption one can construct a rate of metastability in the sense of Tao for  $(x_n)$  if a quantitative version of Fejér monotonicity holds. In [3], a notion of modulus of regularity was introduced. It was shown that the availability of a modulus of regularity for a Fejér monotone sequence will entail the construction of a

rate of convergence without any compactness assumption and using just the usual non-quantitative form of Fejér monotonicity. This points to the nontriviality of some metric regularity assumptions in the literature, as it is known that already for fairly simple algorithms no computable rate of convergence exists. That notwithstanding, the metric regularity assumption pertains to the specifics of the underlying set  $C$  and in instances where the set  $C$  enjoys nice geometry properties, a modulus of regularity (and – correspondingly – a rate of convergence) turns out to be available.

Recently [5], the second author provided a proof-theoretical analysis of the strong convergence of Dykstra’s algorithm of cyclic projections. Strikingly, even here without the sequence being Fejér monotone, the existence of a modulus of regularity allowed for rates of convergence. Since the iteration fails to be Fejér monotone, this result escapes the reach of the general results in [3].

In this paper, we introduce generalized notions of *locally relativized Fejér monotonicity* and show that the main results in [2, 3] extend to these notions. In particular, this provides a theoretical justification for [5, Section 4].

## 2 Locally $S$ -relativized Fejér monotone sequences

In the following,  $(X, d)$  is a metric space and  $F \subseteq X$  a nonempty subset. As in [2], whose notations we follow, we assume that

$$F = \bigcap_{k \in \mathbb{N}} \tilde{F}_k,$$

where  $\tilde{F}_k \subseteq X$  for every  $k \in \mathbb{N}$  and we say that the family  $(\tilde{F}_k)$  is a *representation* of  $F$ . It is clear that  $F$  has a trivial representation, by letting  $\tilde{F}_k := F$  for all  $k$  but the intended interpretation is that

$$AF_k := \bigcap_{l \leq k} \tilde{F}_l$$

is some weakened approximate form of  $F$ . A point  $p \in AF_k$  is said to be a  *$k$ -approximate  $F$ -point*.

In the following we always view  $F$  not just as a set but we suppose it is equipped with a representation  $(\tilde{F}_k)$  to which we refer implicitly in many of the notations introduced below.

Let  $(x_n)$  be a sequence in  $X$  and  $S(n, k)$  be an arbitrary property about  $(n, k) \in \mathbb{N}^2$ . We think of  $k$  as an error  $\delta > 0$  via  $\delta = \frac{1}{k+1}$ . Throughout this paper we assume that  $S(n, k)$  is monotone in  $\delta$  in the sense of

$$\forall n, k_1, k_2 \in \mathbb{N} (k_1 \leq k_2 \wedge S(n, k_2) \rightarrow S(n, k_1)).$$

Note that if  $S$  does not satisfy this we may replace it by  $\tilde{S}(n, k) := \bigwedge_{i=0}^k S(n, i)$ .

**Definition 2.1.**  $(x_n)$  has approximate  $F/S$ -points if  $\forall k \in \mathbb{N} \exists n \in \mathbb{N} (x_n \in AF_k \wedge S(n, k))$ . A function  $\Phi : \mathbb{N} \rightarrow \mathbb{N}$  is called an approximate  $F/S$ -bound if

$$\forall k \in \mathbb{N} \exists n \leq \Phi(k) (x_n \in AF_k \wedge S(n, k)).$$

We assume w.l.o.g. that  $\Phi$  is nondecreasing since, otherwise, we may take  $\Phi^M(k) := \max\{\Phi(i) : i \leq k\}$ .

**Definition 2.2** ([2]). We say that  $F$  is explicitly closed (w.r.t. the representation  $(\tilde{F}_k)$ ) if

$$\forall p \in X \left( \forall N, M \in \mathbb{N} \left( AF_M \cap \overline{B} \left( p, \frac{1}{N+1} \right) \neq \emptyset \right) \rightarrow p \in F \right).$$

As in [2, Section 4], let  $G, H : (0, \infty) \rightarrow (0, \infty)$  be functions which possess moduli  $\alpha_G, \beta_H : \mathbb{N} \rightarrow \mathbb{N}$  such that

$$\forall k \in \mathbb{N} \forall a \in (0, \infty) \left( a \leq \frac{1}{\alpha_G(k) + 1} \rightarrow G(a) \leq \frac{1}{k + 1} \right)$$

and

$$\forall k \in \mathbb{N} \forall a \in (0, \infty) \left( H(a) \leq \frac{1}{\beta_H(k) + 1} \rightarrow a \leq \frac{1}{k + 1} \right).$$

Let  $(x_n)$  be a sequence in the metric space  $(X, d)$  and  $\emptyset \neq F \subseteq X$ .

The next definition generalizes the concept of  $(G, H)$ -Fejér monotonicity from [2, Definition 4.1]:

**Definition 2.3.**  $(x_n)$  is locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F$  if

$$(*) \left\{ \begin{array}{l} \forall p \in F \forall r \in \mathbb{N} \exists m \in \mathbb{N} \forall n \in \mathbb{N} \\ \left( d(x_n, p) < \frac{1}{m+1} \wedge S(n, m) \rightarrow \forall l \in \mathbb{N} (H(d(x_{n+l}, p)) \leq G(d(x_n, p)) + \frac{1}{r+1}) \right) \end{array} \right\}.$$

**Remark 2.4.** If  $(x_n)$  is  $(G, H)$ -Fejér monotone w.r.t.  $F$  in the sense of [2, Definition 4.1], then it, in particular, is locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F$  for any property  $S(n, m)$ .

The next theorem generalizes [2, Proposition 4.3] to locally  $S$ -relativized  $(G, H)$ -Fejér monotone sequences:

**Theorem 2.5.** Let  $X$  be a compact metric space and  $F$  be explicitly closed. Assume that  $(x_n)$  is locally  $S$ -relativized  $(G, H)$ -Fejér monotone with respect to  $F$  and that  $(x_n)$  has approximate  $F/S$ -points. Then  $(x_n)$  converges to a point  $x \in F$ .

*Proof.* For each  $k \in \mathbb{N}$  let  $n_k \in \mathbb{N}$  be such that

$$x_{n_k} \in AF_k \wedge S(n_k, k)$$

and define  $y_k := x_{n_k}$ . Since  $X$  is compact,  $(y_k)$  has a convergent subsequence  $(y_{k_l})$ . Let  $x := \lim y_{k_l}$ . Similarly to [2] one shows that

- (i)  $x$  is an adherent point of  $S_k := \{x_n : n \in \mathbb{N} \wedge S(n, k)\}$  for all  $k \in \mathbb{N}$  and
- (ii) (using that  $F$  is explicitly closed)  $x \in F$ .

**Proof of (i):** Let  $k, m \in \mathbb{N}$ . We want to show that

$$\exists y \in S_k \left( d(x, y) \leq \frac{1}{m+1} \right).$$

Let  $l \geq k$  be so large that  $d(y_{k_l}, x) \leq \frac{1}{m+1}$ . By construction

$$y_{k_l} = x_{n_{k_l}} \in AF_{k_l} \wedge S(n_{k_l}, k_l).$$

Since  $(y_{k_l})$  is a subsequence of  $(y_k)$  we have that  $k_l \geq l \geq k$  and so by the monotonicity properties of  $AF_K$  and  $S(n, K)$  in  $K \in \mathbb{N}$

$$y_{k_l} \in AF_k \wedge S(n_{k_l}, k).$$

Hence we may take  $y := y_{k_i}$ . ■

**Proof of (ii):** By the proof of (i), the premise of the property of  $F$  being ‘explicitly closed’ is satisfied for  $p := x$  and so  $x \in F$ . ■

We now show using (i) and (ii) that  $\lim x_n = x$ : let  $r \in \mathbb{N}$  be arbitrary and define  $r' := 2\beta_H(r) + 1$ . Let for this  $r'$  and  $p := x \in F$  be  $m$  as in Definition 2.3 and also satisfying that

$$m \geq \max \{ \alpha_G(2\beta_H(r) + 1), 2\beta_H(r) + 1 \}$$

(this is possible to achieve by the monotonicity of  $S(n, m)$  in  $m$ ). By (i),

$$\exists n \in \mathbb{N} \left( d(x_n, x) < \frac{1}{m+1} \wedge S(n, m) \right).$$

Hence by Definition 2.3 applied to this  $n$ :

$$\forall l \in \mathbb{N} \left( H(d(x_{n+l}, x)) \leq G(d(x_n, x)) + \frac{1}{2\beta_H(r) + 2} \stackrel{\alpha_G\text{-def.}}{\leq} \frac{1}{\beta_H(r) + 1} \right)$$

and so by  $\beta_H$ -definition

$$\forall l \in \mathbb{N} \left( d(x_{n+l}, x) \leq \frac{1}{r+1} \right). \quad \square$$

**Application 2.6.** Let  $C_1, \dots, C_N \subseteq H$  be closed and convex subsets of a (real) Hilbert space with

$$C := \bigcap_{i=1}^N C_i \neq \emptyset$$

and  $P_i$  be the metric projection onto  $C_i$  for  $i = 1, \dots, N$ . Let  $(x_n), (q_n)$  be the sequences in  $H$  generated by Dykstra’s algorithm with  $x_0$  as starting point. Let  $z \in C$  and  $\mathbb{N} \ni b > 0$  with  $\|z - x_0\| \leq b$ . Then, by [5, Lemma 3.4],  $(x_n) \subset \overline{B}(z, b)$ . Define

$$f(x) := \max_{i=1, \dots, N} \|x - P_i x\|.$$

Note that  $C = \bigcap_{k=0}^{\infty} AF_k := \{x \in X : |f(x)| \leq \frac{1}{k+1}\} = \text{zer } f$  is explicitly closed.

We also define  $G(a) := H(a) := a^2$  and may take as in [2, Lemma 7.10]

$$\alpha_G(k) := \lceil \sqrt{k} \rceil, \quad \beta_H(k) := k^2.$$

Now define (where for negative indices  $k$  we take  $x_k$  as arbitrary points in  $H$  while  $q_k := 0$ )

$$S(n, r) := \left( \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle < \frac{1}{r+1} \right).$$

Since  $(x_n)$  is locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $C$  and possesses approximate  $C/S$ -points (these facts are proven in stronger form in Application 3.6 below), Theorem 2.5 (applied to  $X := C \cap \overline{B}(z, b)$ ) gives a simple proof of the convergence of  $(x_n)$  towards a point in  $C$  if  $H$  is finite dimensional (compared to the much more complicated strong convergence proof known in the general case, see e.g. [1]).

**Application 2.7.** Let  $(X, d)$  be compact metric space and  $F$  be explicitly closed. For each  $r \in \mathbb{N}$ , consider the subset  $F_r \subseteq F$  of points  $p$  in  $F$  which satisfy

$$\exists m \in \mathbb{N} \forall n \in \mathbb{N} \left( d(x_n, p) < \frac{1}{m+1} \wedge S(n, m) \rightarrow \forall l \in \mathbb{N} (H(d(x_{n+l}, p)) \leq G(d(x_n, p)) + \frac{1}{r+1}) \right).$$

Assume that

$$\forall r \in \mathbb{N} \forall q \in F \setminus F_r \exists N \in \mathbb{N} \forall n \in \mathbb{N} \left( S(n, N) \rightarrow d(x_n, q) \geq \frac{1}{N+1} \right).$$

If  $(x_n)$  has approximate  $F/S$ -points, then  $(x_n)$  strongly converges to a point in  $F$ : it suffices to show that  $(x_n)$  is locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F$ . Let  $r \in \mathbb{N}$  and  $p \in F$ . If  $p \in F_r$ , then  $(*)$  holds by definition of  $F_r$ . If  $p \notin F_r$ , then  $(*)$  holds for  $m := N$  since  $S(n, m)$  implies that  $d(x_n, p) \geq \frac{1}{m+1}$ .

By Theorem 2.5 a sequence  $(x_n)$  strongly converges to a point  $x \in F$  (for explicitly closed  $F$ ) if for some condition  $S$ , it is locally  $S$ -relativized  $(G, H)$ -Fejér monotone and has approximate  $F/S$ -points. We now show that also the converse holds (with  $G := H := Id$ ) if  $F$  satisfies the following condition (which in most applications is trivially satisfied, e.g. when  $F := \text{zer } f$  for some continuous  $f : X \rightarrow \mathbb{R}$  and  $AF_k := \{x \in X : |f(x)| \leq \frac{1}{k+1}\}$ ).

**Definition 2.8.** We say that  $F$  is properly approximated by  $(AF_k)$  if

$$\forall x \in F \forall k \in \mathbb{N} \exists m \in \mathbb{N} \forall y \in X \left( d(x, y) \leq \frac{1}{m+1} \rightarrow y \in AF_k \right).$$

**Proposition 2.9.** If  $F$  is properly approximated by  $(AF_k)$  and  $(x_n)$  converges to  $x \in F$ , then for a suitable condition  $S$ , the sequence  $(x_n)$  is locally  $S$ -relativized  $(Id, Id)$ -Fejér monotone and has approximate  $F/S$ -points.

*Proof.* Let  $x \in F$  and assume that  $(x_n)$  converges to  $x$ . Take

$$S(n, r) := \forall k \geq n \left( d(x_k, x) < \frac{1}{2r+2} \right).$$

Then clearly for all  $p \in X, n, r \in \mathbb{N}$

$$S(n, r) \rightarrow \forall l \in \mathbb{N} \left( d(x_{n+l}, p) \leq d(x_{n+l}, x_n) + d(x_n, p) \leq d(x_n, p) + \frac{1}{r+1} \right)$$

and so  $(x_n)$  satisfies  $(*)$  with  $m := r$ . Moreover,  $(x_n)$  has approximate  $F/S$ -points: given  $r \in \mathbb{N}$ , let  $m$  be so large that  $d(x, y) \leq \frac{1}{2m+2}$  implies  $y \in AF_r$ . Take  $n \in \mathbb{N}$  be so large that  $S(n, \max\{m, r\})$ . Then  $S(n, r)$  and  $x_n \in AF_r$ .  $\square$

The relevance of Theorem 2.5, of course, stems from the fact that in applications  $S$  will be such that it is much easier to construct an approximate  $F/S$ -point bound than a rate of convergence for  $(x_n)$  (which is needed for the specific  $S$  used in the proof above).

### 3 Uniform locally $S$ -relativized $(G, H)$ -Fejér monotone sequences w.r.t. $F$ -points

We now strengthen the Definition 2.3 by demanding that ‘ $\exists m \in \mathbb{N}$ ’ is uniform w.r.t.  $p \in F$ .

**Definition 3.1.** *A sequence  $(x_n)$  in  $X$  is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F$  with modulus  $\rho : \mathbb{N} \rightarrow \mathbb{N}$  if*

$$\forall p \in F \forall r, n \in \mathbb{N} \left( d(x_n, p) < \frac{1}{\rho(r)+1} \wedge S(n, \rho(r)) \rightarrow \forall l \in \mathbb{N} (H(d(x_{n+l}, p)) \leq G(d(x_n, p)) + \frac{1}{r+1}) \right).$$

**Remark 3.2.** Even with this strengthened version, any ordinary  $(G, H)$ -Fejér monotone sequence in the sense of [2, Definition 4.1] (and hence by taking  $G := H := Id$  any Fejér monotone sequence) satisfies the above definition for any choice of  $S(n, r)$  and any  $\rho$ .

Let  $f : X \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$  an arbitrary function with  $\text{zer } f := \{x \in X : f(x) = 0\} \neq \emptyset$ . We recall the following definition from [3] (but written in  $1/(k+1)$ -form rather than with  $\varepsilon > 0$ ):

**Definition 3.3** ([3, Definition 3.1]). *Let  $z \in \text{zer } f$  and  $b > 0$ . Then  $\mu : \mathbb{N} \rightarrow \mathbb{N}$  is a modulus of regularity for  $f$  w.r.t.  $\text{zer } f$  and  $\overline{B}(z, b)$  if for all  $k \in \mathbb{N}$  and  $x \in \overline{B}(z, b)$  we have the following implication*

$$|f(x)| < \frac{1}{\mu(k)+1} \rightarrow \text{dist}(x, \text{zer } f) < \frac{1}{k+1}.$$

Let  $b > 0$  and  $(x_n)$  be a sequence in  $\overline{B}(z, b)$  for some  $z \in \text{zer } f := \{x \in X : f(x) = 0\}$  and  $\mu$  be a modulus of regularity w.r.t.  $\text{zer } f$  and  $\overline{B}(z, b)$ .

Let  $\Phi : \mathbb{N} \rightarrow \mathbb{N}$  be an approximate  $F/S$ -bound for  $(x_n)$  with  $F := \text{zer } f = \bigcap_{k=0}^{\infty} AF_k$ , where  $AF_k := \{x \in X : |f(x)| \leq \frac{1}{k+1}\}$ , more precisely

$$\forall k \in \mathbb{N} \exists n \leq \Phi(k) \left( |f(x_n)| < \frac{1}{k+1} \wedge S(n, k) \right).$$

Recall that we assume  $S(n, m)$  to be monotone in  $m$ .

**Theorem 3.4.** *Let  $\mu$  be a modulus of regularity w.r.t.  $F := \text{zer } f$  and  $\overline{B}(z, b)$ . If  $(x_n)$  is a sequence in  $\overline{B}(z, b)$  which is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F$  with modulus  $\rho$  and if  $\Phi$  is an approximate  $F/S$ -bound for  $(x_n)$ , then  $(x_n)$  is a Cauchy sequence with rate  $\Psi(2k+1)$  for*

$$\Psi(k) := \Phi(\max\{\rho(2\beta_H(k)+1), \mu(\max\{\alpha_G(2\beta_H(k)+1), \rho(2\beta_H(k)+1)\})\}), \text{ i.e.}$$

$$\forall k, m, \tilde{m} \in \mathbb{N} \left( m, \tilde{m} \geq \Psi(2k+1) \rightarrow \left( \|x_m - x_{\tilde{m}}\| \leq \frac{1}{k+1} \right) \right)$$

and

$$\forall k \in \mathbb{N} \forall n \geq \Psi(k) \left( \text{dist}(x_n, \text{zer } f) \leq \frac{1}{k+1} \right).$$

Moreover, if  $X$  is complete and  $\text{zer } f$  is closed, then  $(x_n)$  converges to a zero of  $f$  with rate of convergence  $\Psi(2k+1)$ .

*Proof.* Let  $k \in \mathbb{N}$  be given. By the definition of  $\Phi$  there exists an  $n \leq \Psi(k)$  s.t.

$$(1) S(n, \rho(2\beta_H(k) + 1)) \wedge |f(x_n)| < \frac{1}{\mu(\max\{\alpha_G(2\beta_H(k) + 1), \rho(2\beta_H(k) + 1)\}) + 1}.$$

Hence by  $\mu$ -definition

$$(2) \exists p \in \text{zer } f \left( d(x_n, p) < \frac{1}{\max\{\alpha_G(2\beta_H(k) + 1), \rho(2\beta_H(k) + 1)\} + 1} \right).$$

By the definition of  $\rho$  applied to  $p$ , (1) and (2) imply that

$$(3) \forall l \in \mathbb{N} \left( H(d(x_{n+l}, p)) \leq G(d(x_n, p)) + \frac{1}{2\beta_H(k) + 2} \stackrel{(2), \alpha_G\text{-def}}{\leq} \frac{1}{\beta_H(k) + 1} \right).$$

Hence by the definition of  $\beta_H$

$$(4) \forall l \in \mathbb{N} \left( d(x_{n+l}, p) \leq \frac{1}{k + 1} \right)$$

and so

$$(5) \forall m, \tilde{m} \geq n \left( d(x_m, x_{\tilde{m}}) \leq d(x_m, p) + d(x_{\tilde{m}}, p) \leq \frac{2}{k + 1} \right).$$

(4) and (5) establish the first two claims.

For the remaining claim we argue as in the proof of [3, Theorem 4.1]: let  $X$  be complete and  $\text{zer } f$  be closed. Then by the above  $z' := \lim x_n$  exists and by the Cauchy rate (5) we get that

$$(6) \forall m \geq \Psi(2k + 1) \left( d(x_m, z') \leq \frac{1}{k + 1} \right).$$

Together with (4) this yields

$$(7) \text{dist}(z', \text{zer } f) \leq \text{dist}(x_m, \text{zer } f) + d(z', x_m) \leq \frac{1}{2(k + 1)} + \frac{1}{k + 1} < \frac{2}{k + 1}.$$

As  $k$  was arbitrary, we obtain that  $\text{dist}(z', \text{zer } f) = 0$  which implies that  $z' \in \text{zer } f$  as  $\text{zer } f$  is closed.  $\square$

**Remark 3.5.** Note that in Theorem 3.4,  $X$  is not required to be compact.

**Application 3.6.** We now show that the rate of convergence for the Dykstra cyclic projection algorithm, recently obtained in [5] under a metric regularity assumption in the case of an arbitrary real Hilbert space  $H$ , is covered by Theorem 3.4 (but not by Theorem 4.1 in [2] as this algorithm is not  $(G, H)$ -Fejér monotone but only uniformly locally  $S$ -relativized so for a suitable  $S$ ): we use the notations from Application 2.6. In [5, Section 4] it is assumed that  $C_1, \dots, C_N$  are metrically regular with a modulus  $\mu : \mathbb{N}^2 \rightarrow \mathbb{N}$  in the sense of [3, Definition 4.6] (we switch here from the  $\varepsilon/\delta$ -formulation used in [3] to  $\frac{1}{k+1}$  instead of  $\varepsilon$  etc. to fit the notations used in [2] and above), i.e.

$$\forall k, r \in \mathbb{N} \forall x \in \overline{B}(z, b) \left( \bigwedge_{i=1}^N \|x - P_i x\| < \frac{1}{\mu_r(k) + 1} \rightarrow \exists p \in C \left( \|x - p\| < \frac{1}{k + 1} \right) \right).$$

As shown in the proof of [3, Corollary 4.8] (and the sentence after that proof), such a  $\mu_r$  is a modulus of regularity for zer  $f$  and  $\overline{B}(z, b)$ , where

$$f(x) := \max_{i=1, \dots, N} \|x - P_i x\|.$$

Note that  $C = \text{zer } f$ .

We also define  $G(a) := H(a) := a^2$  and may take as in [2, Lemma 7.10]

$$\alpha_G(k) := \lceil \sqrt{k} \rceil, \quad \beta_H(k) := k^2.$$

Now define

$$A(n, r) := \left( \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle < \frac{1}{r+1} \right).$$

Implicitly in the proof of [5, Theorem 4.2], an  $C/S$ -bound  $\Phi$  for  $(x_n)$  is constructed.<sup>1</sup>

Define  $\rho(r) := 8b(r+1) - 1$ . It remains to show that  $(x_n)$  is locally  $S$ -relativized  $(G, H)$ -Fejér monotone (with  $G, H, S$  as above) w.r.t.  $C$  with modulus  $\rho$ : let  $p \in C, r \in \mathbb{N}$  and  $n \in \mathbb{N}$  be such that

$$\|x_n - p\| < \frac{1}{\rho(r) + 1} \wedge S(n, \rho(r)).$$

By (7) in [5] one has for all  $l \geq n$

$$\begin{aligned} \|x_l - p\|^2 &\leq \|x_n - p\|^2 + 2 \underbrace{\sum_{k=n-N+1}^n \langle x_k - p, q_k \rangle}_{\stackrel{!}{\leq} 1/(2r+2)} - 2 \underbrace{\sum_{k=l-N+1}^l \langle x_k - p, q_k \rangle}_{\geq 0, \text{ by [5, L.3.1(iii)]}} \\ &\leq \|x_n - p\|^2 + \frac{1}{r+1}. \end{aligned}$$

Here ‘!’ follows from  $\|x_n - p\| < \frac{1}{8b(r+1)}$  and  $\sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle < \frac{1}{4(r+1)}$  by reasoning as in [5, p.17]:

$$\begin{aligned} \sum_{k=n-N+1}^n \langle x_k - p, q_k \rangle &= \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle + \underbrace{\langle x_n - p, \sum_{k=n-N+1}^n q_k \rangle}_{= \langle x_n - p, x_0 - x_n \rangle, \text{ by [5, L.3.1(ii)]}} \\ &\leq \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle + \|x_n - p\| \cdot \|x_0 - x_n\| \\ &\leq \frac{1}{4(r+1)} + 2b \cdot \frac{1}{8b(r+1)} = \frac{1}{2(r+1)}. \end{aligned}$$

<sup>1</sup>In the notation of [5] this bound can be taken as  $\alpha(b, N, \varepsilon, \Phi(b, N, \varepsilon, \cdot)) + \Phi(b, N, \varepsilon, \alpha(\dots))$  with  $\varepsilon := 1/(r+1)$ .



We now discuss a simple argument which in some sense gives a reversal to Theorem 3.4. Fix a function  $I : \mathbb{N} \times X \rightarrow X$  standing for some iterative process employed to approximate zeros of  $f$ . Consider the following natural condition on  $I$  (see remark 3.8 below):

$$\forall n \in \mathbb{N} \forall p \in \text{zer } f (I(n, p) = p), \quad (+)$$

stating that the iterative process leaves the points in  $\text{zer } f$  unchanged (such iterations are called retractive in [4]). Consider a sequence  $(x_n)$  recursively defined by the function  $I$ , i.e.  $x_0 \in X$  and  $x_{n+1} := I(n, x_n)$ . Therefore, the assumption (+) on  $I$  entails,

$$\forall x_0 \in X (x_0 \in \text{zer } f \rightarrow \forall n \in \mathbb{N} (x_n = x_0)). \quad (++)$$

**Definition 3.7.** We say that a function  $\tau : \mathbb{N}^2 \rightarrow \mathbb{N}$  is a modulus of coincidence for  $I$  if it satisfies

$$\forall n \in \mathbb{N} \forall k \in \mathbb{N} \forall x_0 \in X \left( |f(x_0)| < \frac{1}{\tau(k, n) + 1} \rightarrow d(x_n, x_0) < \frac{1}{k + 1} \right),$$

where  $(x_n)$  is initiated at  $x_0$  and recursively defined by  $I$ .

**Remark 3.8.** Most iterative methods in optimization satisfy the condition (+). For example, in a normed linear space with  $f(x) = \|x - T(x)\|$  for a nonexpansive map  $T$ , if  $I(n, x)$  is defined as

- (i)  $T(x)$  – Picard iteration,
- (ii)  $\alpha_n x + (1 - \alpha_n)T(x)$ , for  $(\alpha_n) \subseteq [0, 1]$  – Krasnoselski-Mann iteration,
- (iii)  $\alpha_n x + (1 - \alpha_n)T(\beta_n x + (1 - \beta_n)T(x))$ , for  $(\alpha_n), (\beta_n) \subseteq [0, 1]$  – Ishikawa iteration,

then the condition (+) is satisfied. Another well-known iterative method is the Halpern iteration,  $I(n, x) = \alpha_n u + (1 - \alpha_n)T(x)$ . Despite not satisfying (+), it still satisfies (++) whenever  $u = x_0$ . One easily checks that for the Ishikawa iteration  $\tau(k, n) := 2n(k + 1)$  gives a modulus of coincidence, and for all the other cases (including the Halpern iteration) one can take  $\tau(k, n) := n(k + 1)$ . Naturally, if one imposes further conditions on the parameter sequences  $(\alpha_n), (\beta_n)$ , then the definition of the function  $\tau$  may be improved.

We have the following easy result.

**Proposition 3.9.** Consider a function  $I$  as above, and let  $\tau$  be a modulus of coincidence for  $I$ . Given  $z \in \text{zer } f$  and  $b > 0$ , assume that for any  $x_0 \in \overline{B}(z, b)$  the sequence initiated at  $x_0$  and recursively defined by  $I$  converges to a point  $\ell_I(x_0) \in \text{zer } f$  with a common rate of convergence  $\Psi : \mathbb{N} \rightarrow \mathbb{N}$ , i.e.

$$d(x_0, z) \leq b \rightarrow \forall k \in \mathbb{N} \forall n \geq \Psi(k) \left( d(x_n, \ell_I(x_0)) \leq \frac{1}{k + 1} \right).$$

Then, the function  $\mu(k) := \tau(2k + 1, \Psi(2k + 1))$  is a modulus of regularity for  $f$  w.r.t.  $\text{zer } f$  and  $\overline{B}(z, b)$ .

*Proof.* Let  $k \in \mathbb{N}$  and  $x \in \overline{B}(z, b)$  be given, and assume that  $|f(x)| < \frac{1}{\mu(k) + 1}$ . Consider  $(x_n)$  to be the sequence recursively defined by  $I$  with initial point  $x_0 = x$ . By the hypothesis of  $\Psi$ , we have

$d(x_{\Psi(2k+1)}, \ell_I(x)) \leq \frac{1}{2(k+1)}$ . On the other hand, the assumption on  $\tau$  entails  $d(x_{\Psi(2k+1)}, x) < \frac{1}{2(k+1)}$ . We conclude that

$$\text{dist}(x, \text{zer } f) \leq d(x, \ell_I(x)) \leq d(x, x_{\Psi(2k+1)}) + d(x_{\Psi(2k+1)}, \ell_I(x)) < \frac{1}{k+1},$$

and hence  $\mu$  is a modulus of regularity for  $f$  w.r.t.  $\text{zer } f$  and  $\overline{B}(z, b)$ .  $\square$

**Remark 3.10.** Theorem 3.4 states that the assumption of a modulus of regularity in the case of a sequence which is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t.  $F = \text{zer } f$ , and has approximate  $F/S$ -points, will converge (when  $X$  is complete and  $\text{zer } f$  is closed) to a point in  $\text{zer } f$  with a uniform rate of convergence. Proposition 3.9 above states that, in most iterative methods, if such uniform rate of convergence is to exist, then the assumption of metric regularity is indeed necessary. This argument was used in [3, Proposition 4.4] in the context of the Picard iteration. It was also used recently in [5, Proposition 4.7], in the context of Dykstra's method for solving the convex feasibility problem. In this iterative method, the computation of the function  $\tau$  is more convoluted and was given in [5, Propositions 4.5 and 4.6].

## 4 Uniform locally $S$ -relativized $(G, H)$ -Fejér monotone sequences w.r.t. approximate $F$ -points

We now generalize the concept of 'uniform  $(G, H)$ -Fejér monotonicity' introduced in [2, Definition 4.6]:

**Definition 4.1.** We say that  $(x_n)$  is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t. approximate  $F$ -points if for all  $r, n, m \in \mathbb{N}$  with moduli  $\chi : \mathbb{N}^3 \rightarrow \mathbb{N}, \rho : \mathbb{N} \rightarrow \mathbb{N}$  if

$$\forall r, n, m \in \mathbb{N} \forall p \in X \left( p \in AF_{\chi(n, m, r)} \wedge d(x_n, p) < \frac{1}{\rho(r)+1} \wedge S(n, \rho(r)) \rightarrow \forall l \leq m \left( H(d(x_{n+l}, p)) \leq G(d(x_n, p)) + \frac{1}{r+1} \right) \right).$$

**Remark 4.2.** If  $(x_n)$  is uniformly  $(G, H)$ -Fejér monotone w.r.t.  $F$  with modulus  $\chi$  in the sense of [2, Definition 4.6], then  $(x_n)$  is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t. approximate  $F$ -points with moduli  $\chi$  and an arbitrary  $\rho : \mathbb{N} \rightarrow \mathbb{N}$  for any property  $S$ .

**Example 4.3.** In the setting of Application 2.6 for the case of Dykstra's algorithm, we can obtain moduli  $\chi, \rho$  in the sense of Definition 4.1. Indeed, for each  $r, n, m \in \mathbb{N}$  define

$$\rho(r) := 4(2b+1)(r+1) - 1 \text{ and } \chi(n, m, r) := 8b(n+m)(r+1) \div 1,$$

where  $\mathbb{N} \ni b \geq \|z - x_0\|$  for some  $z \in C$ , as before. Assume that  $\|x_n - p\| \leq \frac{1}{\rho(r)+1}$ ,

$$p \in AF_{\chi(n, m, r)}, \text{ i.e. } \max_{i=1, \dots, N} \|p - P_i(p)\| \leq \frac{1}{\chi(n, m, r) + 1}, \text{ and } \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle \leq \frac{1}{\rho(r) + 1}$$

for given  $r, m, n \in \mathbb{N}$  and  $p \in X$ .<sup>2</sup> Then, similarly to the proof of [5, Theorem 3.11], we have for  $G(a) = H(a) = a^2$  and any  $\ell \leq m$ ,

$$H(\|x_{n+\ell} - p\|) \leq G(\|x_n - p\|) + 2 \left( \underbrace{\sum_{k=n-N+1}^n \langle x_k - p, q_k \rangle}_{=:t_1} + \underbrace{\sum_{k=n+\ell-N+1}^{n+\ell} \langle p - x_k, q_k \rangle}_{=:t_2} \right).$$

Then,

$$\begin{aligned} t_1 &= \sum_{k=n-N+1}^n \langle x_k - x_n, q_k \rangle + \sum_{k=n-N+1}^n \langle x_n - p, q_k \rangle \\ &\stackrel{[5, \text{L.3.1(ii),L.3.4}]}{\leq} \frac{1}{\rho(r) + 1} + 2b\|x_n - p\| \leq \frac{2b + 1}{\rho(r) + 1} = \frac{1}{4(r + 1)} \end{aligned}$$

and (with  $P_k(p)$  being an arbitrary point for negative  $k$ )

$$\begin{aligned} t_2 &= \sum_{k=n+\ell-N+1}^{n+\ell} \langle p - P_k(p), q_k \rangle + \sum_{k=n+\ell-N+1}^{n+\ell} \underbrace{\langle P_k(p) - x_k, q_k \rangle}_{\leq 0, \text{ by [5, L.3.1(iii)]}} \\ &\leq \sum_{k=n+\ell-N+1}^{n+\ell} \|p - P_k(p)\| \cdot \|q_k\| \stackrel{[5, \text{L.3.2}]}{\leq} \frac{1}{\chi(n, m, r) + 1} \sum_{k=0}^{n+\ell-1} \|x_k - x_{k+1}\| \\ &\leq \frac{2b(n + \ell)}{\chi(n, m, r) + 1} \leq \frac{1}{4(\ell + 1)}. \end{aligned}$$

Therefore  $2(t_1 + t_2) \leq \frac{1}{r+1}$ , which concludes the proof.

The next theorem generalizes [2, Theorem 5.1] to sequences which are uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t. approximate  $F$ -points:

**Theorem 4.4.** *Let  $(X, d)$  be a totally bounded metric space and  $\gamma$  be a modulus of total boundedness for  $X$  in the sense of [2, Definition 2.2]. Assume that*

- (i)  $(x_n)$  is uniformly locally  $S$ -relativized  $(G, H)$ -Fejér monotone w.r.t. approximate  $F$ -points, with moduli  $\chi, \rho$ ;
- (ii)  $(x_n)$  has approximate  $F/S$ -points, with  $\Phi$  being a nondecreasing approximate  $F/S$ -point bound.

Then  $(x_n)$  is Cauchy and, moreover, for all  $k \in \mathbb{N}$  and all  $g : \mathbb{N} \rightarrow \mathbb{N}$

$$\exists N \leq \Psi(k, g, \Phi, \chi, \rho, \alpha_G, \beta_H, \gamma) \forall i, j \in [N, N + g(N)] \left( d(x_i, x_j) \leq \frac{1}{k + 1} \right),$$

where  $\Psi(k, g, \Phi, \chi, \rho, \alpha_G, \beta_H, \gamma) := \Psi_0(P - 1) := \Psi_0(P - 1, k, g, \Phi, \chi, \rho, \beta_H)$ , with

<sup>2</sup>Recall that the last conjunct stands for the property  $S(n, \rho(r))$ .

$P := \gamma(\max\{\alpha_G(2\beta_H(2k+1)+1), \rho(2\beta_H(2k+1)+1)\})$  and

$$\begin{aligned}\Psi_0(0) &:= \Phi(\rho(2\beta_H(2k+1)+1)), \\ \Psi_0(n+1) &:= \Phi(\max\{\chi_g^M(\Psi_0(n), 2\beta_H(2k+1)+1), \rho(2\beta_H(2k+1)+1)\})\end{aligned}$$

with

$$\chi_g(n, k) := \chi(n, g(n), k), \quad \chi_g^M(n, k) := \max\{\chi_g(i, k) \mid i \leq n\}.$$

*Proof.* We follow closely the proof of [2, Theorem 5.1] with some decisive changes though. Let  $k \in \mathbb{N}$  and  $g : \mathbb{N} \rightarrow \mathbb{N}$ . Define

$$\varphi(k) := \min m (x_m \in AF_k \wedge S(m, k)).$$

Both  $\varphi$  and  $\Phi$  are nondecreasing and  $\Phi$  is a pointwise bound on  $\varphi$ . By induction we readily prove that  $\Psi_0(n, k, g, \varphi, \chi, \rho, \beta_H) \leq \Psi_0(n+1, k, g, \varphi, \chi, \rho, \beta_H)$ ,  $\Psi_0(n, k, g, \Phi, \chi, \rho, \beta_H) \leq \Psi_0(n+1, k, g, \Phi, \chi, \rho, \beta_H)$  and  $\Psi_0(n, k, g, \varphi, \chi, \rho, \beta_H) \leq \Psi_0(n, k, g, \Phi, \chi, \rho, \beta_H)$  for all  $n \in \mathbb{N}$ .

Define for every  $i \in \mathbb{N}$

$$n_i := \Psi_0(i, k, g, \varphi, \chi, \rho, \beta_H).$$

**Claim 1:** For all  $j \geq 1$  and all  $0 \leq i < j$ ,  $x_{n_j}$  is a  $\chi_g(n_i, 2\beta_H(2k+1)+1)$ -approximate  $F$ -point.

**Proof of claim:** As  $j \geq 1$  and

$$\begin{aligned}n_j &= \Psi_0(j, k, g, \varphi, \chi, \rho, \beta_H) \geq \varphi(\chi_g^M(\Psi_0(j-1, k, g, \varphi, \chi, \rho, \beta_H), 2\beta_H(2k+1)+1)) \\ &= \varphi(\chi_g^M(n_{j-1}, 2\beta_H(2k+1)+1)),\end{aligned}$$

$x_{n_j}$  is a  $\chi_g^M(n_{j-1}, 2\beta_H(2k+1)+1)$ -approximate  $F$ -point. Since  $0 \leq i < j-1$ , we have that  $n_i \leq n_{j-1}$ . Apply now the fact that  $\chi_g^M$  is nondecreasing in the first argument to get that

$$\begin{aligned}\chi_g(n_i, 2\beta_H(2k+1)+1) &\leq \chi_g^M(n_i, 2\beta_H(2k+1)+1) \\ &\leq \chi_g^M(n_{j-1}, 2\beta_H(2k+1)+1)\end{aligned}$$

which establishes Claim 1. ■

**Claim 2:** There exist  $0 \leq I < J \leq P$  satisfying

$$\forall l \in [n_I, n_I + g(n_I)] \left( d(x_l, x_{n_J}) \leq \frac{1}{2k+2} \right).$$

**Proof of claim:** Utilizing that  $\gamma$  is modulus of total boundedness for  $X$  we get that there exist  $0 \leq I < J \leq P$  with

$$(1) \quad d(x_{n_I}, x_{n_J}) \leq \min \left\{ \frac{1}{\alpha_G(2\beta_H(2k+1)+1)+1}, \frac{1}{\rho(2\beta_H(2k+1)+1)+1} \right\}.$$

Since  $\alpha_G$  is a  $G$ -modulus this, in particular, implies

$$(2) \quad G(d(x_{n_I}, x_{n_J})) \leq \frac{1}{2\beta_H(2k+1)+2}.$$

By the first claim, we have that  $x_{n_J}$  is a  $\chi_g(n_I, 2\beta_H(2k+1)+1)$ -approximate  $F$ -point. By construction,  $n_I$  satisfies  $S(n_I, \rho(2\beta_H(2k+1)+1))$ . Applying now the uniform locally  $S$ -relativized  $(G, H)$ -Féjer monotonicity of  $(x_n)$  w.r.t. approximate  $F$ -points with  $r := 2\beta_H(2k+1)+1$ ,  $n := n_I$ ,  $m := g(n_I)$  and  $p := x_{n_J}$ , we get – using (1) again – that for all  $l \leq g(n_I)$ ,

$$H(d(x_{n_I+l}, x_{n_J})) \leq G(d(x_{n_I}, x_{n_J})) + \frac{1}{2\beta_H(2k+1)+2} \stackrel{(2)}{\leq} \frac{1}{\beta_H(2k+1)+1}.$$

Since  $\beta_H$  is an  $H$ -modulus,

$$\forall l \leq g(n_I) \left( d(x_{n_I+l}, x_{n_J}) \leq \frac{1}{2k+2} \right).$$

which establishes Claim 2. ■

From Claim 2 it is immediate that

$$\forall k, l \in [n_I, n_I + g(n_I)] \left( d(x_k, x_l) \leq \frac{1}{k+1} \right).$$

As  $n_I = \Psi_0(I, k, g, \varphi, \chi, \rho, \beta_H) \leq \Psi_0(I, k, g, \Phi, \chi, \rho, \beta_H)$  and  $I \leq P-1$ , we get that

$$n_I \leq \Psi_0(P-1, k, g, \Phi, \chi, \rho, \beta_H) = \Psi(k, g, \Phi, \chi, \rho, \alpha_G, \beta_H, \gamma).$$

The theorem holds with  $N := n_I$ . □

**Corollary to the proof:** One of the numbers  $n_0, \dots, n_{P-1}$  is a point of metastability.

**Application 4.5.** As discussed already at the end of Application 2.6, this relativized notion of Féjer monotonicity allows for a simple convergence proof of Dykstra’s algorithm in the finite dimensional case. In light of the moduli  $\chi$  and  $\rho$  in Example 4.3 (and also the  $C/S$ -bound  $\Phi$  implicit in [5, Theorem 4.2], cf. footnote 1), by Theorem 4.4 we moreover obtain rates of metastability for Dykstra’s iteration.

#### Acknowledgments:

The authors were supported by the German Science Foundation (DFG Project KO 1737/6-2).

## References

- [1] H.H. Bauschke and P.L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, New York, 2nd edn., 2017.
- [2] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative Results on Féjer Monotone Sequences. *Communications in Contemporary Mathematics*, 20(2), 42pp., 2018.
- [3] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Féjer monotone sequences. *Israel Journal of Mathematics*, 232:261-297, 2019.
- [4] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11, 44pp, 2015.
- [5] P. Pinto. On the finitary content of Dykstra’s cyclic projections algorithm. arXiv:2306.09791, 2023.