

GENERALIZED FEJÉR MONOTONE SEQUENCES AND THEIR FINITARY CONTENT

NICHOLAS PISCHKE

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,
E-mail: pischke@mathematik.tu-darmstadt.de

ABSTRACT. We provide quantitative and abstract strong convergence results for sequences from a compact metric space satisfying a certain form of *generalized Fejér monotonicity* where (1) the metric can be replaced by a much more general type of function measuring distances (including, in particular, certain Bregman distances), (2) full Fejér monotonicity is relaxed to a partial variant and (3) the distance functions are allowed to vary along the iteration. For such sequences, the paper provides explicit and effective rates of metastability and even rates of convergence, the latter under a regularity assumption that generalizes the notion of metric regularity introduced by Kohlenbach, López-Acedo and Nicolae, itself an abstract generalization of many regularity notions from the literature.

In the second part of the paper, we apply the abstract quantitative results established in the first part to two algorithms: one algorithm for approximating zeros of maximally monotone and maximally ρ -comonotone operators in Hilbert spaces (in the sense of Combettes and Pennanen as well as Bauschke, Moursi and Wang) that incorporates inertia terms *every other* term and another algorithm for approximating zeros of monotone operators in Banach spaces (in the sense of Browder) that is only Fejér monotone w.r.t. a certain Bregman distance.

Keywords: Fejér monotonicity; metastability; rates of convergence; proof mining; Bregman distance
MSC2020 Classification: 47J25, 47H05, 03F10, 47H09

1. INTRODUCTION

This paper is about quantitative information on convergence results from analysis and optimization where the sequence in question satisfies a particular form of monotonicity, generalizing the well-known Fejér monotonicity (see [5, 21, 76] and see the next section for precise definitions).¹ As discussed e.g. in [21, 76], many sequences studied in nonlinear analysis and optimization are Fejér or quasi-Fejér monotone (in the sense of [20]) and very general convergence theorems are available for these classes which classify additional properties required for strong or weak convergence (see [11] for such results already implicitly and in particular see [5, 21]). Consequently, in the last decade(s), Fejér monotonicity has become a unifying concept for a wide range of convergence results in those areas. Particular examples include the generic example of Picard iterations for nonexpansive mappings but also the proximal point algorithm (stemming from [55, 67]), Mann iterations [54] as well as Ishikawa iterations [32] for generalized nonexpansive mappings (see [28]) and range to involved examples like, e.g., algorithms approximating equilibrium points [31] or algorithms for solving problems in DC programming [58], and beyond.

Even with this breadth of notions and schemes encompassed by the concept of Fejér monotonicity, there are many influential iteration schemes which, although related to the concept in principle, are not truly Fejér monotone and thus can not be covered by the previous study thereof.

For one, there are iterations which are only partially Fejér in the sense that only a subsequence is properly Fejér monotone and the other elements connect to this subsequence just via weaker monotonicity requirements. Such a situation in particular arises in the context of “alternating methods”, i.e. algorithms which alternate between two (or more) “schemas”, e.g. like in recent modifications of the proximal point algorithm [33, 59] which incorporate inertia terms *every other* term.

For another, iterations can be Fejér monotone w.r.t. functions measuring the distance of points other than metrics. The most prominent and instructive example for this case are Bregman distances as derived from Bregman’s seminal work [10] or even generalized Bregman distances in the sense of Burachik, Dao and Lindstrom [14]. From that perspective, the theory of sequences which are Fejér monotone w.r.t. Bregman distances has been recognized, in a similar way as ordinary Fejér monotone sequences which are defined w.r.t. a norm or metric, as a powerful framework for unifying many convergence results in nonlinear analysis (see in particular

Date: August 12, 2024.

¹Fejér monotonicity was introduced in [57], named after the use made of a similar concept by L. Fejér [26].

the seminal work by Bauschke, Borwein and Combettes [4]) and applications of this framework include many prominent iterative schemes in the context of Bregman distances (see in particular the references in [4]).

Lastly, in particular in the recent years, methods have been studied where instead of solving subproblems by e.g. minimizing some sub-objective function relative to a fixed distance, these distances are allowed to vary along the iteration so that these subproblems can be tweaked adaptively. Particular examples for such iterations are the recent works on variable-metric forward-backward type methods [9, 19, 23] and in particular similar methods that allow for varying Bregman distances [13, 61], among many more. The treatment of all these methods relies on establishing some form of Fejér monotonicity where the distances in question are allowed to vary along the iteration and for the particular case of varying metrics induced by bilinear forms, an abstract treatment of such types of methods has been given in [22].

In this paper, we will study a generalized form of Fejér monotonicity which unifies all of these variants. Concretely, our particular emphasis will be on sequences where the property of Fejér monotonicity is relaxed in the following three ways:

- (1) We allow for a very general class of functions (in particular including metrics as well as large classes of Bregman distances) to measure distances between points in the (quasi-)Fejér monotonicity property.
- (2) We simultaneously allow that the sequence only fulfills a partial form of Fejér monotonicity where one subsequence retains a generalized form of (quasi-)Fejér monotonicity while the other sequence is only required to fulfill a very relaxed form of monotonicity which generalizes the conditions on the complementary subsequence currently in place in the literature on such methods.
- (3) We allow the general distances to vary along the (partial) Fejér monotonicity property, capturing the multiple notions of varying distances currently discussed in the literature.

To that end, we give abstract and unified strong convergence results for such sequences over compact (or relatively compact and thus in particular finite-dimensional) spaces which are moreover quantitative in the sense that, depending only on few quantitative reformulations of the key properties of the sequence, they provide a construction for an effective and highly uniform rate of metastability in the sense of T. Tao [74, 75] and even a rate of convergence under an additional regularity assumption introduced here² that generalizes the notion of metric regularity introduced in [45].³ These abstract and quantitative results arise as generalizations of previous constructions of rates of metastability and rates of convergence for Fejér and quasi-Fejér monotone sequences obtained in [43, 45] (see also [63]). Moreover, by “reversing” the process and forgetting about the quantitative aspects, we are then able to derive a new and very general “non-quantitative” convergence result for these generalized Fejér monotone sequences in the previous three senses which in particular generalizes the classical results on Fejér and quasi-Fejér monotone sequences as well as the general convergence results on Bregman monotone sequences obtained in [4] and on variable metric Fejér monotone sequences obtained in [22] (at least in relatively compact spaces). Not only does this treatment therefore unify all of these different types of Fejér monotonicity but it also can accommodate new methods that would not have been covered by either of these various strands of notions of Fejér monotonicity individually (as also commented on further below).

Similar to the circumstances of the quantitative results on Fejér monotone sequences from [43, 45], the present work has been obtained through the general methodological approach of proof mining, a program in mathematical logic conceptually originating from Kreisel’s unwinding of proofs [49, 50] and brought to maturity by U. Kohlenbach and his collaborators, which aims to classify and extract the computational content of *prima facie* “non-computational” proofs (see [38] for a comprehensive monograph on the subject, [46] for an early survey and [39] for a survey of recent results up to 2019).

Proof mining has been extensively applied in the context of nonlinear (functional) analysis and optimization to derive quantitative results in numerous circumstances. Particularly relevant for the present paper are, e.g., the quantitative treatments of equilibrium problems [65], zeros of differences of monotone operators [63], the proximal point algorithm in various settings [40, 41, 42, 51, 52] and the composition of two firmly nonexpansive mappings in nonlinear spaces [44] as well as Korpelevich’s extragradient method [62] as those applications were obtained using the previously mentioned abstract results on Fejér monotone sequences [43, 45].

²The dependence on such an additional assumption is indeed necessary if one wants to obtain computable rates of convergence. This follows from adaptations (see [60]) of a classical result in computable analysis by Specker [72] which, in general, rule out effective rates of convergence for many common iteration schemes, even in the context where all parameters of the sequence like potential mappings, etc., are effective (see the discussion in [60]).

³This notion of regularity introduced in [45] provides a unified approach to several regularity notions known from optimization such as error bounds and metric subregularity, among others (see the discussion in [45]).

However, as common in the proof mining program, we want to emphasize that while this logic-based approach was crucial in the course of obtaining the present results, there is no explicit use of logical tools in the derivation of our results presented here and all relevant logical particularities are discussed, if at all, only in separate remarks.

We expect that the abstract results established here will exhibit a similar range of applications in the future. To that end, the second half of the paper is devoted to two particular exemplary case studies. Here, we provide quantitative results on the previously mentioned alternating inertia proximal point algorithm [59] and its generalization given in [33] to averaged mappings (or, respectively, comonotone operators in the sense of [8]) as well as on a modification of the well-known proximal point algorithm due to [36] (see also the related work [56]) to monotone operators in Banach spaces (in the sense of Browder, see in particular [12]) which also modifies the iteration to be of a style similar to that of Mann's iteration (see also [77] for similar modifications in Hilbert spaces).

As a further indication of the applicability of the results presented here, we however want to additionally mention that beyond the previously discussed applications, the methods seem to be immediately further applicable in the context of the recent works on proximal gradient algorithms [34], on a modification of Korpelevich's extragradient algorithm for variational inequalities (see [48]) to incorporate alternating inertia [71], on extragradient methods for equilibrium problems [69], on algorithms for the split feasibility problem [53, 70] or on modifications of the Krasnosel'skii-Mann iteration to incorporate alternating inertia [24] as well as on algorithms using the Bregman distance in the proximal point style [17, 25], on Landweber's iteration in Banach spaces [68] and potential extensions to Bregman distances, on iterations of Bregman retractions [6], on iterations of Bregman projections [1], on algorithms for solving the variational inequality problem in Banach spaces using the Bregman distance [15, 16] or on algorithms which are only quasi-Fejér w.r.t. Bregman distances like iterations of inexact orbits [18]. Further, also the previously mentioned works on variable-metric forward-backward type methods [9, 19, 23] and on varying Bregman distances [13, 61] seem to immediately fit in the framework proposed here and could thus be quantitatively treated.

Lastly, we want to note that the general approach developed here is crucially used in the recent work [64] to study a new notion of Bregman distance in metric spaces where, in that context, the generality of the results presented here was crucial for obtaining corresponding convergence results of sequences which are Fejér monotone w.r.t. this metric variant of the Bregman distance as it was neither covered by previous work on Bregman monotone iterations, being set in normed spaces, nor on Fejér monotone sequences in metric spaces, as this new Bregman distance in metric spaces is not a metric itself.

2. PRELIMINARIES: CONVERGENCE THEOREMS FOR (QUASI-)FEJÉR MONOTONE SEQUENCES

Throughout, if not stated otherwise, we consider an underlying metric space (X, d) . In this section, we now recap the main results on the usual (quasi-)Fejér monotone sequences underlying this work. In that way, we mainly exposit the work [43], being closest in spirit to this paper. Therefore, we mostly follow the notational setup of [43] if not stated otherwise. In particular, we write \mathbb{R}_+ for the interval $[0, \infty)$ and we write \mathbb{R}_+^* for the interval $(0, \infty)$. Further, we use $\overline{B}(p, r)$ to denote the closed ball around p with radius r in (X, d) . Also, we write $n \dot{-} m = \max\{0, n - m\}$ for $n, m \in \mathbb{N}$.

The central notion investigated by the authors in [43] is the following form of Fejér monotonicity relative to the metric:

Definition 2.1 ([43]). Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and a non-empty subset $F \subseteq X$, a sequence $(x_n) \subseteq X$ is called (G, H) -Fejér monotone w.r.t. F if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(d(p, x_{n+m})) \leq G(d(p, x_n)).$$

Following [43], the functions G and H are later assumed to satisfy the following continuity properties: we assume that G satisfies that

$$a_n \rightarrow 0 \text{ implies } G(a_n) \rightarrow 0$$

and that H satisfies that

$$H(a_n) \rightarrow 0 \text{ implies } a_n \rightarrow 0,$$

both for any $(a_n) \subseteq \mathbb{R}_+$. Those properties are equivalent, respectively, to the existence of corresponding moduli $\alpha_G, \beta_H : \mathbb{N} \rightarrow \mathbb{N}$ which satisfy

$$a \leq \frac{1}{\alpha_G(k) + 1} \rightarrow G(a) \leq \frac{1}{k + 1}$$

as well as

$$H(a) \leq \frac{1}{\beta_H(k) + 1} \rightarrow a \leq \frac{1}{k + 1}$$

for any $k \in \mathbb{N}$ and any $a \in \mathbb{R}_+$. Following [43], we call α_G a G -modulus for G and β_H an H -modulus for H .

For the set F , we assume (following [43]) a representation via

$$F = \bigcap_{k \in \mathbb{N}} AF_k$$

where the sequence of the sets $AF_k \subseteq X$ is monotone decreasing, i.e. $AF_k \supseteq AF_{k+1}$. The elements of AF_k are to be intuitively understood as being k -good approximations of points in F but can take arbitrary forms in our general setting.

In that context, we will need to assume that the set F is not only closed but closed in a special sense relative to the approximations AF_k :

Definition 2.2 ([43]). The set F is *explicitly closed* w.r.t. AF_k if for all $p \in X$:

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

As is well known from the context of usual Fejér monotone sequences (see [43]), obtaining a convergence result for such sequences requires an additional assumption in the form of an ‘‘approximation property’’, which here takes the form of the following definition:

Definition 2.3 ([43]). We say that a sequence $(x_n) \subseteq X$ has *approximate F -points* (w.r.t. (AF_k)) if

$$\forall k \in \mathbb{N} \exists N \in \mathbb{N} (x_N \in AF_k).$$

The main result on Fejér monotone sequences in compact spaces is now the following proposition (taken in this form from [43]):

Proposition 2.4. *Let (X, d) be compact and $F \subseteq X$ be explicitly closed. If $(x_n) \subseteq X$ is (G, H) -Fejér monotone w.r.t. F and has approximate F -points (w.r.t. (AF_k)), then (x_n) converges to some $x^* \in F$.*

In many situations, Fejér monotonicity is considered in the context of an additional sequence of summable errors in the sense of the following definition, which is then called quasi-Fejér monotonicity:

Definition 2.5 ([43]). Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$ and a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$, a sequence $(x_n) \subseteq X$ is called (G, H) -*quasi-Fejér monotone* w.r.t. F and (ε_n) if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(d(p, x_{n+m})) \leq G(d(p, x_n)) + \sum_{i=n}^{n+m-1} \varepsilon_i.$$

In the context of this extended notion, the assumption of the sequence having approximate F -points has to be upgraded to a lim inf-like property in the sense of the following definition:

Definition 2.6 ([43]). We say that a sequence $(x_n) \subseteq X$ has the *lim inf-property* w.r.t. F (and (AF_k)) if

$$\forall k, n \in \mathbb{N} \exists N \in \mathbb{N} (N \geq n \text{ and } x_N \in AF_k).$$

Then one obtains the following main convergence result for quasi-Fejér monotone sequences in compact spaces (again taken in this form from [43]):

Proposition 2.7. *Let (X, d) be compact and $F \subseteq X$ be explicitly closed. If $(x_n) \subseteq X$ is (G, H) -quasi-Fejér monotone w.r.t. F and (ε_n) and has the lim inf-property w.r.t. F (and (AF_k)), then (x_n) converges to some $x^* \in F$.*

For these convergence results, the work [43] then provided quantitative variants in the form of an explicit construction of a rate of metastability depending only on moduli witnessing uniform quantitative reformulations of the core components of the above result (as will be discussed in more detail later on). Further, the work [45] similarly provided explicit constructions even for full rates of convergence in the context of Proposition 2.4, subject to an additional dependence on a modulus of regularity (as will also be discussed later on). This construction of rates of convergence for Fejér monotone sequences was extended to quasi-Fejér monotone sequences in [63].

The main aim of this paper now is to generalize these quantitative results from [43, 45] to much more general circumstances, uniformly covering all the Fejér-like sequences discussed in the introduction. While in [43, 45] the main aim was at providing quantitative variants of already established convergence results, we here not only aim at quantitative convergence results for these generalized Fejér monotone sequences but, as discussed before, also use to these quantitative results to obtain new “non-quantitative” convergence results. The following section now formally introduces the desired generalizations.

3. GENERALIZED (QUASI-)FEJÉR MONOTONE SEQUENCES

As discussed before, the general situation which we want to consider in this paper simultaneously treats the following three extensions of the usual metric notion of Fejér monotonicity: we want to allow

- (1) more general distance functions $\phi : X \times X \rightarrow \mathbb{R}_+$ to uniformly capture notions like metrics, Bregman distances and beyond,
- (2) a partial version of Fejér monotonicity where only a subsequence is properly Fejér monotone,
- (3) the distance function to be able to vary along the sequence in the Fejér-type property.

The following subsections now discuss these different extensions. We will always define these notions already in their “quasi-type” variants, incorporating errors.

3.1. (Weakly) triangular mappings. The most general class of distance functions $\phi : X \times X \rightarrow \mathbb{R}_+$ which we want to permit is only characterized by a particular weak property which we will call *weakly triangular* which derives as a uniform quantitative version of the property

$$\phi(y, x) = \phi(y, z) = 0 \rightarrow \phi(x, z) = 0,$$

which is a particular special case of the triangle inequality for ϕ . Concretely, we consider the following notion:

Definition 3.1. Let $a \in X$. A function $\phi : X \times X \rightarrow \mathbb{R}_+$ is called *weakly triangular* (w.r.t. a) with modulus $\theta : \mathbb{N}^2 \rightarrow \mathbb{N}$ if for any $k, b \in \mathbb{N}$ and any $x, y, z \in X$ such that $d(a, x), d(a, y), d(a, z) \leq b$, it holds that

$$\phi(y, x), \phi(y, z) \leq \frac{1}{\theta(k, b) + 1} \rightarrow \phi(x, z) \leq \frac{1}{k + 1}.$$

We call ϕ *weakly triangular* if there exists an accompanying modulus and we say that a modulus is uniform if it does not depend on b .⁴

We also consider a strengthening of this notion which connects a function ϕ with the underlying metric d in a certain way.

Definition 3.2. Let $a \in X$. A function $\phi : X \times X \rightarrow \mathbb{R}_+$ is called *triangular* (w.r.t. a) with modulus $\theta : \mathbb{N}^2 \rightarrow \mathbb{N}$ if for any $k, b \in \mathbb{N}$ and any $x, y, z \in X$ such that $d(a, x), d(a, y), d(a, z) \leq b$, it holds that

$$\phi(y, x), \phi(y, z) \leq \frac{1}{\theta(k, b) + 1} \rightarrow d(x, z) \leq \frac{1}{k + 1}.$$

We call ϕ *triangular* if there exists an accompanying modulus and we say that a modulus is uniform if it does not depend on b .⁵

We also introduce a notion of comparison for two distance functions: for $\phi, \psi : X \times X \rightarrow \mathbb{R}_+$ and $A : \mathbb{N} \rightarrow \mathbb{N}$, we write $\phi \leq^A \psi$ if

$$\psi(x, y) \leq \frac{1}{A(k) + 1} \rightarrow \phi(x, y) \leq \frac{1}{k + 1}$$

for any $x, y \in X$.

As for a choice of naming, we want to note that being triangular can be “conceived of” as being stronger than weak triangularity as in many cases, it can be inferred that $\phi \leq^A d$ for a suitable function A (e.g. if ϕ is uniformly continuous) in which case a modulus for triangularity can be converted into a modulus of weak triangularity (see in particular the discussions in Section 5 later on).

In the following, at the most basic level only this property of being (weakly) triangular will be required from ϕ and we in particular allow ϕ to not be symmetric and to not satisfy the full triangle inequality.

⁴In B -bounded metric spaces, such a uniform modulus can of course always be derived from a usual modulus $\theta(k, b)$ by considering $\theta'(k) = \theta(k, B)$.

⁵As before, such a uniform modulus can be derived from a usual modulus bounded metric spaces.

Remark 3.3. The notion of triangular distances is related to the notion of w -distances introduced by Kada, Suzuki and Takahashi in [35] (see also [73]). Concretely, a map $\rho : X \times X \rightarrow \mathbb{R}_+$ is called a w -distance if

- (1) $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ for all $x, y, z \in X$,
- (2) $\rho(x, \cdot)$ is lower semicontinuous for any $x \in X$,
- (3) for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x, y, z \in X$:

$$\rho(z, x), \rho(z, y) \leq \delta \rightarrow d(x, y) \leq \varepsilon.$$

Condition (3) is equivalent to ρ being triangular. However being triangular is clearly more general than being a w -distance and, as mentioned before, we in the following in particular do not require the triangle inequality, i.e. condition (1) above, which allows us later to treat Bregman distances as these generally do not enjoy that property.

Particular examples of such distances and their moduli will be discussed later on but we already want to give some easy instances here.

Lemma 3.4 (essentially [73]). *The following functions are weakly triangular and even triangular:⁶*

- (1) Any metric is uniformly triangular and uniformly weakly triangular with the common modulus $\theta(k) = 2k + 1$.
- (2) Assume X is a normed space with norm $\|\cdot\|$. Then $\phi(x, y) = \|x\| + \|y\|$ and $\phi(x, y) = \|y\|$ are both uniformly triangular and uniformly weakly triangular with the common modulus $\theta(k) = 2k + 1$.
- (3) Given $c > 0$, define $\phi(x, y) = c$ for any $x, y \in X$. Then ϕ is uniformly triangular and uniformly weakly triangular with the common modulus $\theta(k) = \bar{c} + 1$ where $\bar{c} \in \mathbb{N}$ satisfies $c \geq 1/(\bar{c} + 1)$.
- (4) Let $T : X \rightarrow X$ be any mapping. Then $\phi(x, y) = \max\{d(Tx, y), d(Tx, Ty)\}$ is uniformly triangular and uniformly weakly triangular with the common modulus $\theta(k) = 2k + 1$.
- (5) Let F be bounded and $c > 0$ be a bound on its diameter. Define

$$\phi(x, y) = \begin{cases} d(x, y) & \text{if } x, y \in F, \\ c & \text{if } x \notin F \text{ or } y \notin F. \end{cases}$$

Then ϕ is uniformly triangular and uniformly weakly triangular with the common modulus $\theta(k) = 2n_0(k + 1) \div 1$ where $n_0 = \lceil 1/c(k + 1) \rceil + 1$.

Further, the maximum and linear combination of finitely many (weakly) triangular mappings is again (weakly) triangular and moduli for the combinations can be computed from the moduli of the constituents. Lastly, if $f : X \rightarrow \mathbb{R}_+$ is any function and ϕ is a triangular mapping, then $\phi'(x, y) = \max\{f(x), \phi(x, y)\}$ is also triangular with the same modulus.

The proofs are immediate, so we omit them.

Beyond these examples, we now just note two things: For one, note that the reference point a used for “bounding” the elements x, y, z is not crucial for the existence of a modulus since if θ is such a modulus of (weak) triangularity w.r.t. a and a' is such that $d(a, a') \leq r$, then $\theta'(k, b) = \theta(k, b + r)$ is a modulus of (weak) triangularity w.r.t. a' .

Also, we note for another that if ϕ is continuous in both arguments, then a simple compactness argument shows that over compact spaces X the property

$$\phi(y, x) = \phi(y, z) = 0 \rightarrow \phi(x, z) = 0$$

already implies that ϕ is uniformly weakly triangular and that the property

$$\phi(y, x) = \phi(y, z) = 0 \rightarrow d(x, z) = 0$$

already implies that ϕ is uniformly triangular.

Remark 3.5 (For logicians). The underlying logical methods from proof mining (see in particular [30, 37] and the treatment in [38]) actually guarantee that corresponding moduli of (weak) triangularity can be extracted from a large class of proofs of the above respective properties

$$\phi(y, x) = \phi(y, z) = 0 \rightarrow \phi(x, z) = 0$$

and

$$\phi(y, x) = \phi(y, z) = 0 \rightarrow d(x, z) = 0,$$

⁶The examples are adapted from [73] and as such many of these distances are w -distances, or can be turned into such with a few additional conditions, in the sense of the previous Remark 3.3.

even potentially in the absence of any compactness assumptions.

Relating to such distance functions, we will in the following consider an associated notion of Fejér monotonicity (derived by generalizing the previous definition):

Definition 3.6. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any function. Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$ and a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$, a sequence $(x_n) \subseteq X$ is called ϕ - (G, H) -quasi-Fejér monotone w.r.t. F and (ε_n) if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(\phi(p, x_{n+m})) \leq G(\phi(p, x_n)) + \sum_{i=n}^{n+m-1} \varepsilon_i.$$

3.2. Partial Fejér monotonicity. Now, regarding the “partial Fejér monotonicity”, we consider the following more concrete version of the previously only loosely defined notion. By partial Fejér monotonicity, we want to understand that only a subsequence is properly Fejér monotone while the other elements “connect” to this subsequence just by some very weak form of monotonicity which is some form of “slowed-down” Fejér monotonicity.

Concretely, we will (in some sense without loss of generality) identify this designated subsequence with the elements of the sequence at an even index and thus require that for a given sequence (x_n) , only (x_{2n}) is Fejér monotone w.r.t. F . For the odd elements, we require only a weak f -monotonicity property that relates the odd elements $x_{2(n+m)+1}$ to previous even elements $x_{2f(n)}$ where $f : \mathbb{N} \rightarrow \mathbb{N}$ is some number-theoretic function which is nondecreasing but otherwise can be mostly arbitrary (and in particular can be essentially arbitrarily slow increasing). Additionally incorporating the previous generalizations introduced in the context of Fejér monotonicity like the functions G, H as well as using a general function ϕ for measuring distances, we arrive at the following definition:

Definition 3.7. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any function. Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$, a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$ and given a function $f : \mathbb{N} \rightarrow \mathbb{N}$, a sequence $(x_n) \subseteq X$ is called ϕ - (G, H) -quasi- f -monotone w.r.t. F and (ε_n) if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(\phi(p, x_{2(n+m)+1})) \leq G(\phi(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+m} \varepsilon_i.$$

3.3. Variable distances. At last, to capture the various methods employing changing metrics or Bregman distances, we want to consider the above notions relativized to a sequence of distance functions (ϕ_n) .

Definition 3.8. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$ and a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$, a sequence $(x_n) \subseteq X$ is called (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F and (ε_n) if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(\phi_{n+m}(p, x_{n+m})) \leq G(\phi_n(p, x_n)) + \sum_{i=n}^{n+m-1} \varepsilon_i.$$

Similarly, we also consider a version of the accompanying notion of f -monotonicity, relativized to a sequence (ϕ_n) :

Definition 3.9. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$, a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$ and given a function $f : \mathbb{N} \rightarrow \mathbb{N}$, a sequence $(x_n) \subseteq X$ is called (ϕ_n) - (G, H) -quasi- f -monotone w.r.t. F and (ε_n) if for all $n, m \in \mathbb{N}$ and all $p \in F$:

$$H(\phi_{2(n+m)+1}(p, x_{2(n+m)+1})) \leq G(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+m} \varepsilon_i.$$

Concretely, this constitutes our final notion of generalized Fejér monotonicity:

Definition 3.10. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Given two functions $G, H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a non-empty subset $F \subseteq X$, a sequence $(\varepsilon_n) \subseteq \mathbb{R}_+$ with $\sum \varepsilon_n < \infty$ and given a function $f : \mathbb{N} \rightarrow \mathbb{N}$, we call a sequence $(x_n) \subseteq X$ *partially* (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) if (x_{2n}) is (ϕ_{2n}) - (G, H) -quasi-Fejér monotone w.r.t. F and (ε_n) and (x_n) is (ϕ_n) - (G, H) -quasi- f -monotone w.r.t. F and (ε_n) .

The exemplary applications given later, in particular also in combination with the work [64], then illustrate the versatility and applicability of this generalized Fejér monotonicity.

4. RATES OF METASTABILITY FOR GENERALIZED FEJÉR MONOTONE SEQUENCES

One of the main results of the first part of this paper will be theorems that extend the usual convergence result for quasi-Fejér monotone sequences presented in Proposition 2.7 to these generalized Fejér monotone sequences. This generalization will be obtained through a preceding quantitative result by “forgetting” the quantitative information and moving to a simple convergence conclusion.

These quantitative versions, which we therefore shall establish first, form the other main cluster of results of this first part. They themselves are obtained by generalizing the quantitative versions of Proposition 2.7 due to [43, 45]. To that end, we generalize the moduli introduced in [43, 45] that witness uniform quantitative versions of the central assumptions of Proposition 2.7 (which are equivalent to the original assumptions in the context of compact spaces).

Remark 4.1 (For logicians). Similar to [43, 45], these moduli are introduced as guided by the underlying proof-theoretic methodology of proof mining. In that way, these moduli all have the essential property that the so-called logical metatheorems used in the underlying proof mining program (recall the previously mentioned references) guarantee the existence of a quantitative variant of Proposition 2.7 similar to the ones presented in [43, 45] which depends only on such moduli. Moreover, in many cases, the moduli in question can be extracted from proofs of the underlying non-quantitative properties at hand, again as guided by the same logical methodology.

In our case, the main generalizations will be obtained by extending the previous moduli from [43] to be applicable in the context of a general mapping ϕ . Concretely, the main assumptions of Proposition 2.7 are translated to the following ϕ -relative variants and associated uniform quantitative versions in similarity to [43] which will prominently feature in the quantitative results later on:

total ϕ -boundedness	moduli of total ϕ -boundedness γ
explicit ϕ -closedness	moduli of uniform ϕ -closedness ω, δ
approximate F -points	approximate F -point bounds Φ
lim inf-property w.r.t. F	lim inf-bounds Φ w.r.t. F
(ϕ_n) - (G, H) -quasi-Fejér monotonicity	moduli of uniform (ϕ_n) - (G, H) -quasi-Fejér monotonicity χ
(ϕ_n) - (G, H) -quasi- f -monotonicity	moduli of uniform (ϕ_n) - (G, H) -quasi- f -monotonicity ζ

The main new quantitative notion considered here is the last modulus of uniform (ϕ_n) - (G, H) -quasi- f -monotonicity whose details and motivation we will discuss later in a more precise way. The other notions, while here considered in variants relative to the distances (ϕ_n) , arise as straightforward modifications of the metric-based notions introduced in [43]. We nevertheless, both due to the alteration by the functions (ϕ_n) as well as for reasons of self-containedness, discuss also those moduli briefly.

While these moduli and the corresponding convergence results proved later will be presented in an abstract fashion, we refer to [43], the applications derived thereof as well as to the applications presented later on in this paper and those developed in [64] for many versatile examples of concrete instantiations of these moduli.

4.1. Moduli of total boundedness. Compactness features as an essential assumption already in the strong convergence results formulated in Propositions 2.4 and 2.7. As such, also in this paper, usual compactness as well as a version relativized to general distances will be (at least at first) feature crucially in the context of the upcoming strong convergence results. The formulation of compactness that is most crucial here is that obtained via the notion of total boundedness: a set $A \subseteq X$ is called *totally bounded* if there exists an ε -net covering it for any $\varepsilon > 0$. An alternative characterization of this notion commonly used in the context of proof mining was introduced in [29]:

Definition 4.2 ([29]). For $A \neq \emptyset$, $\gamma : \mathbb{N} \rightarrow \mathbb{N}$ is called a *modulus of total boundedness for A* if for all $k \in \mathbb{N}$ and any $(x_n) \subseteq A$:

$$\exists 0 \leq i < j \leq \gamma(k) \left(d(x_i, x_j) \leq \frac{1}{k+1} \right).$$

Indeed, A is totally bounded if, and only if A has a modulus of total boundedness (see [43], Proposition 2.4).⁷

As mentioned before, we will require this notion in the context of general functions $\phi : X \times X \rightarrow \mathbb{R}_+$.

⁷Notice that the notion of a modulus of total boundedness from [29] is discussed in [43] under the name of a II-modulus of total boundedness.

Definition 4.3. We say that a set $A \subseteq X$ is totally ϕ -bounded if for any $k \in \mathbb{N}$ there are $a_0, \dots, a_{j_k} \in X$ such that

$$\forall x \in A \exists 0 \leq i \leq j_k \left(\phi(a_i, x) \leq \frac{1}{k+1} \right).$$

This gives rise to the following analogous notion of modulus:

Definition 4.4. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any function. For $A \neq \emptyset$, $\gamma : \mathbb{N} \rightarrow \mathbb{N}$ is called a *modulus of total ϕ -boundedness* for A if for all $k \in \mathbb{N}$ and any $(x_n) \subseteq A$:

$$\exists 0 \leq i < j \leq \gamma(k) \left(\phi(x_j, x_i) \leq \frac{1}{k+1} \right).$$

The existence of such a modulus is in fact equivalent to being totally ϕ -bounded, at least if ϕ is weakly triangular, as the following proposition shows (which is constructed in analogy to Proposition 2.4 from [43]):

Proposition 4.5. *Let ϕ be uniformly weakly triangular. If a set $A \subseteq X$ is totally ϕ -bounded, then it has a modulus of total ϕ -boundedness. Conversely, if ϕ additionally satisfies $\phi(x, x) = 0$ for all $x \in X$, then a set $A \subseteq X$ that has a modulus of total ϕ -boundedness is also totally ϕ -bounded.*

Proof. Throughout, let θ be a modulus of uniform weak triangularity of ϕ .

Note that a set A is clearly totally ϕ -bounded if, and only if, there exists a modulus $\alpha : \mathbb{N} \rightarrow \mathbb{N}$ (derived from the notion of I-modulus from [43]) satisfying that for any $k \in \mathbb{N}$ there are $a_0, \dots, a_{\alpha(k)} \in X$ such that

$$\forall x \in A \exists 0 \leq i \leq \alpha(k) \left(\phi(a_i, x) \leq \frac{1}{k+1} \right).$$

We show that such a modulus α and a modulus of total ϕ -boundedness can be translated into each other.

Suppose now first that α is a modulus for a set A as above and define $\gamma(k) = \alpha(\theta(k)) + 1$. We show that γ is a modulus of total ϕ -boundedness of A . For this, let

$$a_0, \dots, a_{\alpha(\theta(k))} \in X$$

be such that

$$\forall x \in A \exists 0 \leq i \leq \alpha(\theta(k)) \left(\phi(a_i, x) \leq \frac{1}{\theta(k)+1} \right).$$

Let $x_0, \dots, x_{\alpha(\theta(k))+1}$ be any sequence from A . By the pigeonhole principle, there are $i < j$ such that for some $l \leq \alpha(\theta(k))$:

$$\phi(a_l, x_i), \phi(a_l, x_j) \leq \frac{1}{\theta(k)+1}.$$

Thus we in particular have

$$\phi(x_j, x_i) \leq \frac{1}{k+1}$$

as θ is a modulus of uniform triangularity of ϕ .

Conversely, suppose that γ is a modulus of total ϕ -boundedness for a set A and define $\alpha(k) = \gamma(\theta(k)) - 1$. Note first that, by the defining property of γ , we have $\gamma(k) \geq 1$ for any k as γ is a modulus so that α is well-defined. We show that α witnesses that A is totally ϕ -bounded as above. Suppose not, i.e. there exists a $k \in \mathbb{N}$ such that

$$(*) \quad \forall a_0, \dots, a_{\gamma(\theta(k))-1} \in X \exists x \in A \forall 0 \leq i \leq \gamma(\theta(k)) - 1 \left(\phi(a_i, x) > \frac{1}{k+1} \right).$$

By induction we show that for all $l \leq \gamma(\theta(k))$:

$$\exists \beta_0, \dots, \beta_l \in A \forall 0 \leq i < j \leq l \left(\phi(\beta_j, \beta_i) > \frac{1}{\theta(k)+1} \right)$$

which, for $l = \gamma(\theta(k))$, in particular contradicts that γ is a modulus of total ϕ -boundedness. For $l = 0$, pick $\beta_0 \in A$ arbitrary (where we assume w.l.o.g. $A \neq \emptyset$). For the step from l to $l+1$, let β_0, \dots, β_l be chosen from the induction hypothesis. Now apply (*) to

$$a_i = \begin{cases} \beta_i & \text{if } i \leq l, \\ \beta_l & \text{if } l \leq i \leq \gamma(\theta(k)) - 1, \end{cases}$$

to get an $x \in A$ such that

$$\phi(a_i, x) > \frac{1}{k+1} \text{ for all } i \leq l,$$

i.e. in particular $\phi(\beta_i, x) > 1/(k+1)$ for all $i \leq l$. Now suppose that

$$\phi(x, \beta_i) \leq \frac{1}{\theta(k)+1}.$$

Then as $\phi(x, x) = 0$, we get

$$\phi(\beta_i, x) \leq \frac{1}{k+1}$$

which is a contradiction to before. Thus we have $\phi(x, \beta_i) > \frac{1}{\theta(k)+1}$ and thus $\beta_0, \dots, \beta_l, \beta_{l+1}$ with $\beta_{l+1} = x$ satisfies the desired claim for $l+1$. \square

4.2. Moduli of uniform closedness. In [43], the explicit closedness of F w.r.t. AF_k is upgraded to the existence of moduli of uniform closedness in the following sense:

Definition 4.6 ([43]). The set F is called *uniformly closed with moduli* $\delta_F, \omega_F : \mathbb{N} \rightarrow \mathbb{N}$ (w.r.t. (AF_k)) if for all $k \in \mathbb{N}$ and all $p, q \in X$:

$$q \in AF_{\delta_F(k)} \text{ and } d(q, p) \leq \frac{1}{\omega_F(k)+1} \text{ implies } p \in AF_k.$$

Following [43], this quantitative reformulation is motivated by rewriting the definition of explicit closedness via

$$\begin{aligned} \forall p \in X \left(\forall N, M \in \mathbb{N} \left(AF_M \cap \overline{B} \left(p, \frac{1}{N+1} \right) \neq \emptyset \right) \rightarrow p \in F \right) \\ \equiv \forall p \in X \forall k \in \mathbb{N} \left(\forall N, M \in \mathbb{N} \left(AF_M \cap \overline{B} \left(p, \frac{1}{N+1} \right) \neq \emptyset \right) \rightarrow p \in AF_k \right) \\ \equiv \forall p \in X \forall k \in \mathbb{N} \exists N, M \in \mathbb{N} \forall q \in X \left(q \in AF_M \text{ and } d(q, p) \leq \frac{1}{N+1} \rightarrow p \in F \right) \end{aligned}$$

and then requiring that corresponding moduli exist which bound (and thus witness) the quantifiers over N, M uniformly in k , i.e. without dependence on p .

As with moduli of total ϕ -boundedness, we also require this notion to be relativized to general functions $\phi : X \times X \rightarrow \mathbb{R}_+$.

Definition 4.7. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any function. The set F is *explicitly ϕ -closed* w.r.t. AF_k if for all $p \in X$:

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

where we write $\overline{B}_\phi(p, r) = \{x \in X \mid \phi(x, p) \leq r\}$.

In a similar vein as the moduli of uniform closedness arose above from a suitable rewriting of the property of explicit closedness, we now define moduli of uniform ϕ -closedness:

Definition 4.8. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any function. The set F is called *uniformly ϕ -closed with moduli* $\delta_F, \omega_F : \mathbb{N} \rightarrow \mathbb{N}$ (w.r.t. (AF_k)) if for all $k \in \mathbb{N}$ and all $p, q \in X$:

$$q \in AF_{\delta_F(k)} \text{ and } \phi(q, p) \leq \frac{1}{\omega_F(k)+1} \text{ implies } p \in AF_k.$$

Remark 4.9 (For logicians). Similar to before, also these moduli of uniform closedness can be extracted by separate applications of the logical metatheorems of proof mining to the respective proofs of the explicit closedness property (see also the discussion in [43]).

4.3. Approximate F -point bounds. The approximate F -point property will be quantitatively witnessed by an accompanying bound on the index in the sense of the following definition:

Definition 4.10 ([43]). A bound $\Phi(k)$ on the quantifier “ $\exists N \in \mathbb{N}$ ” in the definition of approximate F -points detailed before, i.e. a bound such that

$$\forall k \in \mathbb{N} \exists N \leq \Phi(k) (x_N \in AF_k),$$

is called an *approximate F -point bound for (x_n) (w.r.t. (AF_k))*.

Throughout, we will always assume that any approximate F -point bound Φ is monotone nondecreasing (note that this is without loss of generality as given an arbitrary approximate F -point bound Φ , the function $\Phi^M(k) := \max\{\Phi(j) \mid j \leq k\}$ is also an approximate F -point bound and monotone nondecreasing, see also [43]).

Remark 4.11 (For logicians). The extractability of these bounds from proofs of the underlying property of approximate F -points via the logical metatheorems of proof mining can in particular be guaranteed if the relation $x_N \in AF_k$ can be expressed by an existential formula in the language used by these metatheorems (see also in particular the discussion in [43]).

4.4. \liminf -bounds w.r.t. F . The \liminf -property that extends the approximate F -points as an assumption in the context of quasi-Fejér monotone sequences will be quantitatively witnessed by a so-called \liminf -bound in the sense of [43]:

Definition 4.12 ([43]). A bound $\Phi(k, n)$ is a *\liminf -bound for (x_n) w.r.t. F (and (AF_k))* if

$$\forall k, n \in \mathbb{N} \exists N \in [n; \Phi(k, n)] (x_N \in AF_k).$$

Also here, we will throughout assume that any \liminf -bound Φ is monotone nondecreasing in both arguments (which is, by a similar argument as with the approximate F -point bounds, again without loss of generality).

Remark 4.13 (For logicians). As before with the approximate F -point bounds, the extractability of such bounds from proofs of the \liminf -property can in particular be guaranteed if $x_N \in AF_k$ can be expressed by an existential formula.

4.5. Moduli of uniform Fejér monotonicity. In the case of Fejér monotone sequences, we move to the notion of uniform Fejér monotonicity in the sense of [43] with accompanying modulus:

Definition 4.14 ([43]). A sequence $(x_n) \subseteq X$ is called *uniformly (G, H) -Fejér monotone w.r.t. F (and (AF_k))* if for all $r, n, m \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ such that for all $p \in AF_k$:

$$\forall l \leq m \left(H(d(p, x_{n+l})) < G(d(p, x_n)) + \frac{1}{r+1} \right).$$

A modulus of uniform (G, H) -Fejér monotonicity w.r.t. F (and (AF_k)) is a function $\chi(n, m, r)$ bounding (and thus witnessing) such a k in terms of the r, n, m .

Now, for our generalized notion of Fejér monotonicity involving sequences of distance functions (ϕ_n) , we consider the following natural extension of this notion of modulus (also incorporating the error sequence featuring in quasi-Fejér monotone sequences, as already discussed in [43]):

Definition 4.15. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. A sequence $(x_n) \subseteq X$ is called *uniformly (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F (and (AF_k)) with errors (ε_n)* if for all $r, n, m \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ such that for all $p \in AF_k$:

$$\forall l \leq m \left(H(\phi_{n+l}(p, x_{n+l})) < G(\phi_n(p, x_n)) + \sum_{i=n}^{n+l-1} \varepsilon_i + \frac{1}{r+1} \right).$$

A modulus of uniform (ϕ_n) - (G, H) -quasi-Fejér monotonicity w.r.t. F (and (AF_k)) is a function $\chi(n, m, r)$ bounding (and thus witnessing) such a k in terms of the r, n, m .

If all ϕ_n coincide with one function ϕ , then we call such a modulus just a *modulus of uniform ϕ - (G, H) -quasi-Fejér monotonicity* and we omit the ϕ if it coincides with the metric d .

A simple compactness argument shows that if F is explicitly closed w.r.t. AF_k , all ϕ_n are continuous in their left argument and G, H are continuous, then normal (ϕ_n) - (G, H) -quasi-Fejér monotonicity already implies uniform (ϕ_n) - (G, H) -quasi-Fejér monotonicity (i.e. the existence of such a modulus).

Remark 4.16 (For logicians). For a discussion on the extractability of such moduli from proofs of the underlying properties of the various forms of Fejér monotonicity (which in essence can be guaranteed if the property $p \in AF_k$ can be expressed by a universal formula in the underlying language of the metatheorems), we again refer to the discussion given in [43, Section 4.1].

4.6. Moduli of uniform f -monotonicity. We now want to discuss the new quantitative notion necessary for the following results on rates of metastability for generalized Fejér monotone sequences. To this end, we want to give a more detailed account on the reasoning leading up to this notion (which is similar to that presented in [43] for the moduli of uniform Fejér monotonicity): Consider the (ϕ_n) - (G, H) -quasi- f -monotonicity property, i.e.

$$H(\phi_{2(n+m)+1}(p, x_{2(n+m)+1})) \leq G(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+m} \varepsilon_i$$

for all $n, m \in \mathbb{N}$ and all $p \in F$. This property can be rewritten into

$$\begin{aligned} \forall n, m \in \mathbb{N} \forall p \in X \left(\forall k \in \mathbb{N} (p \in AF_k) \right. \\ \left. \rightarrow \forall r \in \mathbb{N} \forall l \leq m \left(H(\phi_{2(n+l)+1}(p, x_{2(n+l)+1})) < G(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+l} \varepsilon_i + \frac{1}{r+1} \right) \right) \end{aligned}$$

which in turn is equivalent to

$$\begin{aligned} \forall r, n, m \in \mathbb{N} \forall p \in X \exists k \in \mathbb{N} \left(p \in AF_k \right. \\ \left. \rightarrow \forall l \leq m \left(H(\phi_{2(n+l)+1}(p, x_{2(n+l)+1})) < G(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+l} \varepsilon_i + \frac{1}{r+1} \right) \right). \end{aligned}$$

Like with uniform Fejér monotonicity, we now call sequence uniformly f -monotone if we can bound (and thus witness) such a k uniformly in p , i.e. only depending on $r, n, m \in \mathbb{N}$. This lead us to the following definition:

Definition 4.17. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. A sequence (x_n) is called *uniformly (ϕ_n) - (G, H) -quasi- f -monotone w.r.t. F (and (AF_k)) with errors (ε_n)* if for all $r, n, m \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ such that for all $p \in AF_k$:

$$\forall l \leq m \left(H(\phi_{2(n+l)+1}(p, x_{2(n+l)+1})) < G(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+l} \varepsilon_i + \frac{1}{r+1} \right).$$

A modulus of *uniform (ϕ_n) - (G, H) -quasi- f -monotonicity w.r.t. F (and (AF_k)) with errors (ε_n)* is an upper bound $\zeta(n, m, r)$ (and consequently a realizer) for k in terms of r, n, m .

Again, a simple compactness argument shows that the notions of normal and uniform (ϕ_n) - (G, H) -quasi- f -monotonicity coincide in compact spaces if F is explicitly closed and all ϕ_n as well as G, H are suitably continuous.

If all ϕ_n coincide with one function ϕ , then we call such a modulus just a *modulus of uniform ϕ - (G, H) -quasi- f -monotonicity* and we omit the ϕ as before if it coincides with the metric d .

Remark 4.18 (For logicians). As the above motivating discussion shows, similar to the case of the modulus of uniform Fejér monotonicity (see again the discussion in [43]), such a ζ can be extracted from a proof of the corresponding property, provided that $p \in AF_k$ can be written in a purely universal way.

4.7. A quantitative convergence result. The previously detailed moduli can now be combined to a first quantitative convergence result, generalizing the quantitative versions of the convergence of (quasi-)Fejér monotone sequences from Propositions 2.4 and Proposition 2.7, that is Theorems 5.1, 5.3 and 6.4 of [43]. The proof of the following theorem is, in that way, an extension of the proofs for these theorems presented in [43].

For this, we just need to discuss some last minor assumptions and their quantitative versions. The first such assumption is the summability of the error terms which is quantitatively witnessed via a Cauchy rate ξ :

Definition 4.19. A function $\xi : \mathbb{N} \rightarrow \mathbb{N}$ is a Cauchy rate for $\sum \varepsilon_n < \infty$ if

$$\forall k \in \mathbb{N} \left(\sum_{i=\xi(k)}^{\infty} \varepsilon_i \leq \frac{1}{k+1} \right).$$

Further, inspired by the conditions on the varying distances exhibited in e.g. [13, 22, 23, 61], we will require that the sequence of distance functions (ϕ_n) behaves suitably regular in the sense that the functions converge from below and are bounded from below in a certain sense. For the former, we will concretely assume that the sequence (ϕ_n) approximates another function ϕ from below with a rate π in the following sense:

Definition 4.20. A function $\pi : \mathbb{N} \rightarrow \mathbb{N}$ is a rate of convergence for $\phi_n \nearrow \phi$ uniformly if

$$\forall k \in \mathbb{N} \forall n \geq \pi(k) \forall x, y \in X \left(\phi_n(x, y) \leq \phi(x, y) + \frac{1}{k+1} \right).$$

For the latter, using the previous notion of comparison \leq^A of two distance functions via a given function A , we will assume that there exist functions $\psi : X \times X \rightarrow \mathbb{R}_+$ and $A : \mathbb{N} \rightarrow \mathbb{N}$ such that $\psi \leq^A \phi_n$ for all n (together with some further properties of ψ). This in particular encompasses the common assumption from the literature that $\alpha\psi(x, y) \leq \phi_n(x, y)$ for all $x, y \in X$ and some $\alpha > 0$, as in that case, if $\phi_n(x, y) \leq 1/(A(k) + 1)$ for

$$A(k) = (a+1)(k+1) \div 1$$

with $\alpha \geq 1/(a+1)$ for some $a \in \mathbb{N}$, then

$$\frac{1}{a+1} \psi(x, y) \leq \alpha\psi(x, y) \leq \phi_n(x, y) \leq \frac{1}{A(k)+1} \leq \frac{1}{(a+1)(k+1)}$$

and thus $\psi(x, y) \leq 1/(k+1)$ which yields $\psi \leq^A \phi_n$ for the specified A .

In the context of the varying distances ϕ_n , we also need to consider the further requirement on f that $\lim f(n) = \infty$. For this, we rely on the following quantitative notion of a rate of divergence:

Definition 4.21. We say that κ is a rate of divergence for f if for any $L \in \mathbb{N}$, we get

$$\forall n \geq \kappa(L) (f(n) \geq L).$$

Note that if f is nondecreasing, then κ is already a rate of divergence if $f(\kappa(L)) \geq L$.

Theorem 4.22. Let (X, d) be a metric space and (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Assume $\phi_n \nearrow \phi$ uniformly with a rate of convergence π and that $\psi \leq^A \phi_n$ for some $A : \mathbb{N} \rightarrow \mathbb{N}$ where ψ is uniformly weakly triangular with modulus θ . Assume that X has a modulus of total ϕ -boundedness γ , that α_G is a G -modulus for G and β_H is a H -modulus for H and that f is nondecreasing and $f(n) \leq n$. Further, let $(x_n) \subseteq X$ be a sequence and

- (1) χ be a modulus for uniform (ϕ_{2n}) - (G, H) -quasi-Fejér monotonicity w.r.t. F and (AF_k) for (x_{2n}) with errors (ε_n) ,
- (2) ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$,
- (3) Φ be a lim inf-bound w.r.t. F and (AF_k) for $(x_{2f(n)})$,
- (4) κ be a rate of divergence for f ,
- (5) ζ be a modulus of uniform (ϕ_n) - (G, H) -quasi- f -monotonicity w.r.t. F and (AF_k) for (x_n) with errors (ε_n) .

Then (x_n) is ψ -Cauchy with a rate of metastability $\Psi(k, g)$ (to be defined below), i.e. for all $k \in \mathbb{N}, g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \Psi(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k+1} \right).$$

If ψ is uniformly triangular with modulus θ , then (x_n) is Cauchy with the same rate of metastability $\Psi(k, g)$.

Here: $\Psi(k, g) = 2\Psi_0(P, k, g)$ with

$$P = \gamma(\max\{2\alpha_G(4\beta_H(A(\theta(k)))) + 3, 1, 2\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1, 1\})$$

and Ψ_0 defined by

$$\begin{cases} \Psi_0(0, k, g) = 0, \\ \Psi_0(n+1, k, g) = \Phi(\eta^M(\Psi_0(n, k, g), 4\beta_H(A(\theta(k)))) + 3, \hat{k}), \end{cases}$$

where

$$\hat{k} = \max\{\kappa(\xi(4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 3), \xi(2\beta_H(A(\theta(k))) + 1), \kappa(\xi(2\beta_H(A(\theta(k))) + 1)), \\ \kappa(\pi(\max\{2\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1 + 1, 2\alpha_G(4\beta_H(A(\theta(k))) + 3) + 1\})\})$$

and

$$\eta(n, r) = \max\{\chi_g(n, r), \zeta_g(n, r), \chi(f(n), n - f(n), 4\beta_H(\alpha_G(r)) + 3)\}$$

with

$$\chi_g(n, r) = \chi\left(n, \left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right), \\ \zeta_g(n, r) = \zeta\left(n, \left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right).$$

Proof. Set $\varphi(k, n) = \min\{m \geq n \mid x_{2f(m)} \in AF_k\}$ which is well-defined as $(x_{2f(n)})_n$ has the lim inf-property w.r.t. F . Let $(n_i)_i$ defined by $n_0 = 0$ and

$$n_{i+1} = \varphi(\eta^M(n_i, 4\beta_H(A(\theta(k)))) + 3, \hat{k}).$$

Then for any $j \geq 1$ and any $0 \leq i < j$, $x_{2f(n_j)}$ is an $\eta^M(n_i, 4\beta_H(A(\theta(k)))) + 3$ -approximate F -point as by definition, $x_{2f(x_j)}$ is an $\eta^M(n_{j-1}, 4\beta_H(A(\theta(k)))) + 3$ -approximate- F -point and $n_i \leq n_{j-1}$ yields $\eta^M(n_{j-1}, 4\beta_H(A(\theta(k)))) + 3 \geq \eta^M(n_i, 4\beta_H(A(\theta(k)))) + 3$.

Further, by the definition of P and as X is totally ϕ -bounded with modulus γ , we have that $\exists 0 < I < J \leq P$ (with > 0 through the additional $+1$ in the definition of P) such that

$$\phi(x_{2f(n_j)}, x_{2f(n_I)}) \leq \frac{1}{2\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1} + 2.$$

As we have

$$n_I \geq \hat{k} \geq \kappa(\pi(\max\{2\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1 + 1, 2\alpha_G(4\beta_H(A(\theta(k))) + 3) + 1\}))$$

this yields

$$\phi_{2f(n_I)}(x_{2f(n_j)}, x_{2f(n_I)}) \leq \frac{1}{\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1} + 1$$

and thus we have

$$G(\phi_{2f(n_I)}(x_{2f(n_j)}, x_{2f(n_I)})) \leq \frac{1}{2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 2}.$$

By the above first result, we have that $x_{2f(n_j)}$ is an $\eta^M(n_I, 4\beta_H(A(\theta(k)))) + 3$ -approximate F -point. Therefore, for

$$l \leq n_I - f(n_I),$$

we get

$$H(\phi_{2f(n_I)+2l}(x_{2f(n_j)}, x_{2f(n_I)+2l})) \leq \frac{1}{\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1}.$$

To see that, note that

$$\eta^M(n_I, 4\beta_H(A(\theta(k)))) + 3 \geq \chi(f(n_I), n_I - f(n_I), 4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 3$$

and as $x_{2f(n_j)}$ is an $\eta^M(n_I, 4\beta_H(A(\theta(k)))) + 3$ -approximate F -point and $f(n_I) \geq \xi(4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 3$ since $n_I \geq \kappa(\xi(4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 3)$, we get that

$$H(\phi_{2f(n_I)+2l}(x_{2f(n_j)}, x_{2f(n_I)+2l})) \\ = H(\phi_{2(f(n_I)+l)}(x_{2f(n_j)}, x_{2(f(n_I)+l)})) \\ \leq G(\phi_{2f(n_I)}(x_{2f(n_j)}, x_{2f(n_I)})) + \sum_{i=f(n_I)}^{\infty} \varepsilon_i + \frac{1}{4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 4} \\ \leq \frac{1}{\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1}.$$

Now, as $f(n) \leq n$, we in particular have that $2f(n_I) \leq 2n_I$ and thus $2n_I = 2f(n_I) + 2l$ for $l = n_I - f(n_I)$. Thus we in particular have

$$H(\phi_{2n_I}(x_{2f(n_j)}, x_{2n_I})) \leq \frac{1}{\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1}.$$

Thus, in particular

$$\phi_{2n_I}(x_{2f(n_J)}, x_{2n_I}) \leq \frac{1}{\alpha_G(4\beta_H(A(\theta(k))) + 3) + 1}$$

as β_H is an H -modulus and as α_G is a G -modulus, we get

$$G(\phi_{2n_I}(x_{2f(n_J)}, x_{2n_I})) \leq \frac{1}{4\beta_H(A(\theta(k))) + 4}.$$

Now, also by definition of P , note that this same pair $(x_{2f(n_J)}, x_{2f(n_I)})$ satisfies

$$\phi(x_{2f(n_J)}, x_{2f(n_I)}) \leq \frac{1}{2\alpha_G(4\beta_H(A(\theta(k))) + 3) + 2}.$$

As we have

$$n_I \geq \hat{k} \geq \kappa(\pi(\max\{2\alpha_G(2\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 1, 2\alpha_G(4\beta_H(A(\theta(k))) + 3) + 1\})),$$

this yields

$$\phi_{2f(n_I)}(x_{2f(n_J)}, x_{2f(n_I)}) \leq \frac{1}{\alpha_G(4\beta_H(A(\theta(k))) + 3) + 1}$$

and thus

$$G(\phi_{2f(n_I)}(x_{2f(n_J)}, x_{2f(n_I)})) \leq \frac{1}{4\beta_H(A(\theta(k))) + 4}.$$

We can then show that for any $l \leq g(2n_I)$

$$H(\phi_{2n_I+l}(x_{2f(n_J)}, x_{2n_I+l})) \leq \frac{1}{\beta_H(A(\theta(k))) + 1}.$$

To see that, assume first that l is even, i.e. $l = 2l'$. Since

$$\eta^M(n_I, r) \geq \chi\left(n_I, \left\lfloor \frac{g(2n_I)}{2} \right\rfloor, r\right)$$

and as $x_{2f(n_J)}$ is an $\eta^M(n_I, 4\beta_H(A(\theta(k))) + 3)$ -approximate F -point and since

$$l \leq g(2n_I) \text{ implies } l' \leq \left\lfloor \frac{g(2n_I)}{2} \right\rfloor,$$

we get that

$$\begin{aligned} H(\phi_{2n_I+l}(x_{2f(n_J)}, x_{2n_I+l})) &= H(\phi_{2(n_I+l')}(x_{2f(n_J)}, x_{2(n_I+l')})) \\ &\leq G(\phi_{2n_I}(x_{2f(n_J)}, x_{2n_I})) + \sum_{i=n_I}^{\infty} \varepsilon_i + \frac{1}{4\beta_H(A(\theta(k))) + 4} \\ &\leq \frac{1}{\beta_H(A(\theta(k))) + 1} \end{aligned}$$

where we have used $n_I \geq \xi(2\beta_H(A(\theta(k))) + 1)$.

Now, assume that l is odd, i.e. $l = 2l' + 1$. Then, since

$$\eta^M(n_I, r) \geq \zeta\left(n_I, \left\lfloor \frac{g(n_I)}{2} \right\rfloor, r\right)$$

and as $x_{2f(n_J)}$ is an $\eta^M(n_I, 4\beta_H(A(\theta(k))) + 3)$ -approximate F -point and since

$$l \leq g(2n_I) \text{ implies } l' \leq \left\lfloor \frac{g(2n_I)}{2} \right\rfloor$$

as before, we get that

$$\begin{aligned} H(\phi_{2n_I+l}(x_{2f(n_J)}, x_{2n_I+l})) &= H(\phi_{2(n_I+l')+1}(x_{2f(n_J)}, x_{2(n_I+l')+1})) \\ &\leq G(\phi_{2f(n_I)}(x_{2f(n_J)}, x_{2f(n_I)})) + \sum_{i=f(n_I)}^{\infty} \varepsilon_i + \frac{1}{4\beta_H(A(\theta(k))) + 4} \\ &\leq \frac{1}{\beta_H(A(\theta(k))) + 1} \end{aligned}$$

where we have used $n_I \geq \kappa(\xi(2\beta_H(A(\theta(k))) + 1))$ which implies $f(n_I) \geq \xi(2\beta_H(A(\theta(k))) + 1)$.

As β_H is a H -modulus, we get

$$\phi_{2n_I+l}(x_{2f(n_J)}, x_{2n_I+l}) \leq \frac{1}{A(\theta(k)) + 1}$$

for all $l \leq g(2n_I)$ and thus

$$\psi(x_{2f(n_J)}, x_{2n_I+l}) \leq \frac{1}{\theta(k) + 1}$$

for all $l \leq g(2n_I)$. Thus, for $m, l \in [2n_I, 2n_I + g(2n_I)]$, we get

$$\psi(x_m, x_l) \leq \frac{1}{k + 1}$$

which yields the result for $N = 2n_I$. If ψ is uniformly triangular with modulus θ , then it similarly follows that for $m, l \in [2n_I, 2n_I + g(2n_I)]$:

$$d(x_m, x_l) \leq \frac{1}{k + 1}.$$

Note that we in any way have $n_i \leq \Psi_0(i, k, g)$ for all i (using the monotonicity of Φ) and thus $N \leq 2\Psi_0(I, k, g) \leq 2\Psi_0(P, k, g)$. \square

This quantitative result can now be extended to include the moduli of uniform closedness for a stronger conclusion (similar to Theorem 5.3 of [43]) that actually guarantees that the resulting limit of the sequence belongs to the solution set F .

Theorem 4.23. *In addition to the assumptions of Theorem 4.22, assume that F is uniformly ψ -closed with moduli δ and ω . Then for all $k \in \mathbb{N}$ and all $g \in \mathbb{N}^{\mathbb{N}}$:*

$$\exists N \leq \tilde{\Psi}(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k + 1} \text{ and } x_i \in AF_k \right),$$

where $\tilde{\Psi}(k, g) = 2\tilde{\Psi}_0(P, k, g)$ and where $\tilde{\Psi}_0$ is defined similar to Ψ_0 in Theorem 4.22 with

$$\tilde{\eta}_{k, \delta}(n, r) = \max\{\delta(k), \eta(n, r)\}$$

instead of η and with $\tilde{\theta}(k) = \max\{\theta(k), \omega(k)\}$ instead of θ .

If ψ is instead uniformly triangular with a modulus θ , then for all $k \in \mathbb{N}$ and all $g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \tilde{\Psi}(k, g) \forall i, j \in [N; N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k + 1} \text{ and } x_i \in AF_k \right)$$

with the same $\tilde{\Psi}$.

Proof. Note that $\tilde{\theta}$ is also a modulus of uniform weak triangularity for ψ . Thus, as in the proof of Theorem 4.22, we obtain

$$\exists N \leq \tilde{\Psi}(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k + 1} \right)$$

and, more precisely, that there exists $0 < I < J \leq P$ such that $N = 2n_I$ and

$$\psi(x_{2f(n_J)}, x_{2n_I+l}) \leq \frac{1}{\tilde{\theta}(k) + 1} \leq \frac{1}{\omega(k) + 1}$$

for all $l \leq g(2n_I)$ and that $x_{2f(n_J)}$ is an $\tilde{\eta}_{k, \delta}^M(n_I, 4\beta_H(A(\theta(k))) + 3)$ -approximate F -point. As $\tilde{\eta}_{k, \delta}^M(n_I, 4\beta_H(A(\theta(k))) + 3) \geq \delta(k)$, we have that $x_{2f(n_J)}$ is a $\delta(k)$ -approximate F -point. Thus, we get $x_{2n_I+l} \in AF_k$ for any $l \leq g(n_I)$.

If θ is a modulus of uniform triangularity, then $\tilde{\theta}$ is also such a modulus and the resulting claim follows likewise. \square

With a similar reasoning as in [43], in both results above it is actually sufficient to have a rate of metastability for the summability of the error sequence (ε_n) instead of a full Cauchy rate. Similarly, a rate of metastability for the uniform convergence $\phi_n \nearrow \phi$ would be sufficient but for simplicity, we omit this here. Note also that we only for simplicity assume that the error sequences for the quasi-Fejér monotonicity and the quasi- f -monotonicity coincide as in the case of different error sequences (ε_n) and (ε'_n) , constructing $(\varepsilon_n + \varepsilon'_n)$ provides a common error sequence for which a Cauchy rate can immediately be constructed by combining the two separate Cauchy rates.

Already from Theorem 4.23, we would now be able to infer a “plain” convergence result for generalized Fejér monotone sequences in the spirit of Proposition 2.7. However, this result would only take place over

totally ϕ -bounded spaces and we would have to require that all mappings are ϕ -continuous. As both notions could be rather hard to verify in practice, we refrain from spelling this out here and consider the main “plain” convergence result to be that contained in Theorem 5.7 which infers the normal convergence for the sequence under an additional mild assumption connecting ϕ and the usual metric d .

Remark 4.24. In the context of (ϕ_n) -Fejér monotone sequences w.r.t. a set F , it holds that

$$\phi_{n+1}(z, x_{n+1}) \leq \phi_n(z, x_n) \leq \cdots \leq \phi_0(z, x_0)$$

for $z \in F$ and any n . Now, if $\phi_n \rightarrow \phi$ holds uniformly, i.e.

$$\forall k \in \mathbb{N} \exists N \in \mathbb{N} \forall n \geq N \forall x, y \in X \left(|\phi_n(x, y) - \phi(x, y)| \leq \frac{1}{k+1} \right),$$

then for large enough n we get $\phi(z, x_n) \leq \phi_0(z, x_0) + 1$ and thus we get that x_n is ϕ -bounded. This naturally extends to the various extensions like including G, H or errors as well as partiality.

In this case where the sequence (x_n) is contained in a ϕ -bounded set, i.e.

$$(x_n) \subseteq \{y \in X \mid \phi(z, y) \leq \alpha\}$$

for some $z \in X$ and $\alpha > 0$, the above results naturally extend to spaces which are ϕ -boundedly compact, i.e. where for any $z \in X$ and $\alpha > 0$, the set

$$\{y \in X \mid \phi(z, y) \leq \alpha\}$$

is ϕ -compact.

4.8. Some special cases. The first special case that we want to discuss is when all ϕ_n coincide with a single ϕ . In that case, we can not only simplify the resulting bounds but, in the absence of errors in the Fejér monotonicity, we can even weaken the assumption of the lim inf-property to that of just approximate F -points. This would not have been possible before as, beyond the treatment of errors, the lim inf-property was needed to deal with the assumption that $\phi_n \nearrow \phi$.

Concretely, instantiating the previous results, we get the following:

Theorem 4.25. *Let (X, d) be a metric space and ϕ be a function $\phi : X \times X \rightarrow \mathbb{R}_+$. Assume that $\psi \leq^A \phi$ for some $A : \mathbb{N} \rightarrow \mathbb{N}$ where ψ is uniformly weakly triangular with modulus θ . Assume that X has a modulus of total ϕ -boundedness γ , that α_G is a G -modulus for G and β_H is a H -modulus for H and that f is nondecreasing and $f(n) \leq n$. Further, let $(x_n) \subseteq X$ be a sequence and*

- (1) χ be a modulus for uniform ϕ -(G, H)-quasi-Fejér monotonicity w.r.t. F and (AF_k) for (x_{2n}) with errors (ε_n) ,
- (2) ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$,
- (3) Φ be a lim inf-bound w.r.t. F and (AF_k) for $(x_{2f(n)})$,
- (4) κ be a rate of divergence for f ,
- (5) ζ be a modulus of uniform ϕ -(G, H)-quasi- f -monotonicity w.r.t. F and (AF_k) for (x_n) with errors (ε_n) .

Then (x_n) is ψ -Cauchy with a rate of metastability $\Psi(k, g)$ (to be defined below), i.e. for all $k \in \mathbb{N}, g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \Psi(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k+1} \right).$$

If ψ is uniformly triangular with modulus θ , then (x_n) is Cauchy with the same rate of metastability $\Psi(k, g)$.

Here: $\Psi(k, g) = 2\Psi_0(P, k, g)$ with Ψ_0 defined as in Theorem 4.22, now with

$$\hat{k} = \max\{\kappa(\xi(4\beta_H(\alpha_G(4\beta_H(A(\theta(k)))) + 3)) + 3), \xi(2\beta_H(A(\theta(k))) + 1), \kappa(\xi(2\beta_H(A(\theta(k))) + 1))\}.$$

If in addition to the above assumptions, F is uniformly ψ -closed with modulus δ and ω , then for all $k \in \mathbb{N}$ and all $g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \tilde{\Psi}(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k+1} \text{ and } x_i \in AF_k \right),$$

where $\tilde{\Psi}(k, g) = 2\tilde{\Psi}_0(P, k, g)$ and where $\tilde{\Psi}_0$ is defined similar to Ψ_0 before but with

$$\tilde{\eta}_{k, \delta}(n, r) = \max\{\delta(k), \eta(n, r)\}$$

instead of η and with $\tilde{\theta}(k) = \max\{\theta(k), \omega(k)\}$ instead of θ . As before, if ψ is actually uniformly triangular with a modulus θ , then for all $k \in \mathbb{N}$ and all $g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \tilde{\Psi}(k, g) \forall i, j \in [N; N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k+1} \text{ and } x_i \in AF_k \right),$$

Lastly, if in any of the cases, instead of (1) and (5) we even have

- (1)' χ is a modulus for uniform ϕ -(G, H)-Fejér monotonicity w.r.t. F and (AF_k) for (x_{2n}) ,
(5)' ζ is a modulus of uniform ϕ -(G, H)- f -monotonicity w.r.t. F and (AF_k) for (x_n) ,

then ξ and κ from assumptions (2) and (4) can be omitted and it suffices to assume that

- (3)' Φ is an approximate F -point bound w.r.t. (AF_k) for $(x_{2f(n)})$,

and we get the same conclusions, respectively, with a rate $\hat{\Psi}(k, g) = 2\hat{\Psi}_0(P, k, g)$ and where $\hat{\Psi}_0$ is defined as

$$\begin{cases} \hat{\Psi}_0(0, k, g) = 0, \\ \hat{\Psi}_0(n+1, k, g) = \Phi(\eta^M(\hat{\Psi}_0(n, k, g), 4\beta_H(A(\theta(k)))) + 3), \end{cases}$$

with the other constants defined as before.

The proof is immediate and we thus omit it.

Further, we now want to indicate how this result encompasses the previous quantitative results for plain Fejér monotone sequences. For that, we first consider how a modulus of uniform Fejér monotonicity for the full sequence actually can be used to construct the other relevant moduli for the uniform partial Fejér monotonicity.

Proposition 4.26. *Suppose (x_n) is uniformly (ϕ_n) -(G, H)-quasi-Fejér monotone w.r.t. F and (AF_k) with errors (ε_n) and a modulus χ . Define $\tilde{\varepsilon}_n = \varepsilon_{2n} + \varepsilon_{2n+1}$. Then (x_{2n}) is uniformly (ϕ_{2n}) -(G, H)-quasi-Fejér monotone w.r.t. F and (AF_k) with errors $(\tilde{\varepsilon}_n)$ and a modulus*

$$\chi'(n, m, r) = \chi(2n, 2m, r)$$

and (x_n) is uniformly (ϕ_n) -(G, H)-quasi-id-monotone w.r.t. F and (AF_k) with errors $(\tilde{\varepsilon}_n)$ and a modulus

$$\eta(n, m, r) = \chi(2n, 2m+1, r).$$

We omit the proof as it is rather immediate.

Up to this point, we assumed the functions G and H to be the same in both the Fejér and the f -monotonicity property. This can be immediately extended to the case where the respective functions do not match which we will establish through the following results:

Proposition 4.27. *Let $G_{1,2}, H_{1,2}$ be given. Set $H = \min\{H_1, H_2\}$ and $G = \max\{G_1, G_2\}$. Let α_{G_i} be G -moduli for G_i and β_{H_i} be H -moduli for H_i . Then*

$$\alpha_G(k) = \max\{\alpha_{G_1}(k), \alpha_{G_2}(k)\}$$

is a G -modulus for G and

$$\beta_H(k) = \max\{\beta_{H_1}(k), \beta_{H_2}(k)\}$$

is a H -modulus for H .

Proof. Let $a \leq \frac{1}{\alpha_G(k)+1} \leq \frac{1}{\alpha_{G_i}(k)+1}$ for $i = 1, 2$. Thus $G_i(a) \leq \frac{1}{k+1}$ for $i = 1, 2$ and thus $G(a) \leq \frac{1}{k+1}$.

Similarly, let $H(a) \leq \frac{1}{\beta_H(k)+1} \leq \frac{1}{\beta_{H_i}(k)+1}$ for $i = 1, 2$. Then $H_i(a) \leq \frac{1}{\beta_{H_i}(k)+1}$ for some i . Thus $a \leq \frac{1}{k+1}$. \square

Proposition 4.28. *Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Let $G_{1,2}, H_{1,2}$ be given. Set $H = \min\{H_1, H_2\}$ and $G = \max\{G_1, G_2\}$. Then:*

- (1) *If χ is a modulus for uniform (ϕ_n) -(G_1, H_1)-quasi-Fejér monotonicity of (x_n) w.r.t. errors (ε_n) , then it is also a modulus for uniform (ϕ_n) -(G, H)-quasi-Fejér monotonicity of (x_n) w.r.t. (ε_n) .*
- (2) *If ζ is a modulus of uniform (ϕ_n) -(G_2, H_2)-quasi- f -monotonicity of (x_n) w.r.t. (ε_n) , then it is also a modulus of uniform (ϕ_n) -(G, H)-quasi- f -monotonicity of (x_n) w.r.t. (ε_n) .*

We again omit the proof as it is rather immediate.

Proposition 4.29. *Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Let $G_{1,2}, H_{1,2}$ be given and assume that $G_2(x) \leq H_1(x)$ for all x . Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be nondecreasing and $f(n) \leq n$.*

Let χ be a modulus of uniform (ϕ_{2n}) -(G_1, H_1)-quasi-Fejér monotonicity for (x_{2n}) w.r.t. F and (AF_k) with errors (ε_n) and let $\zeta : \mathbb{N}^2 \rightarrow \mathbb{N}$ be nondecreasing in the left argument such that for all $n, r \in \mathbb{N}$, there exists a $k \leq \zeta(n, r)$ such that

$$\forall p \in AF_k \left(H_2(\phi_{2n+1}(p, x_{2n+1})) \leq G_2(\phi_{2f(n)}(p, x_{2f(n)})) + \sum_{i=f(n)}^n \varepsilon_i + \frac{1}{r+1} \right).$$

Then $\hat{\zeta}(n, m, r) = \max\{\chi(f(n), f(n+m) - f(n), 2r+1), \zeta(n+m, 2r+1)\}$ is a modulus of uniform (ϕ_n) - (G_1, H_2) -quasi- f -monotonicity for (x_n) w.r.t. F and (AF_k) with errors (ε_n) .

Proof. Let $k = \hat{\zeta}(n, m, r)$ and $p \in AF_k$. Then $k \geq \zeta(n+m, 2r+1) \geq \zeta(n+l, 2r+1)$ and thus

$$H_2(\phi(p, x_{2(n+l)+1})) \leq G_2(\phi(p, x_{2f(n+l)})) + \sum_{i=f(n+l)}^{n+l} \varepsilon_i + \frac{1}{2r+1}.$$

Using $G_2 \leq H_1$ and $k \geq \chi(n, m, 2r+1)$, the above then yields

$$\begin{aligned} H_2(\phi(p, x_{2(n+l)+1})) &\leq H_1(\phi(p, x_{2f(n+l)})) + \sum_{i=f(n+l)}^{n+l} \varepsilon_i + \frac{1}{2r+1} \\ &\leq G_1(\phi(p, x_{2f(n)})) + \sum_{i=f(n+l)}^{n+l} \varepsilon_i + \sum_{i=f(n)}^{f(n+l)-1} \varepsilon_i + \frac{1}{r+1} \\ &= G_1(\phi(p, x_{2f(n)})) + \sum_{i=f(n)}^{n+l} \varepsilon_i + \frac{1}{r+1} \end{aligned}$$

as $f(n) \leq f(n+l)$ since f is nondecreasing. \square

Lastly, we want to consider a special situation of the f -monotonicity property. Assume that we are in a situation where

$$H(\phi_{2(n+m)+s}(p, x_{2(n+m)+s})) \leq G(\phi_{2n}(p, x_{2n}))$$

for any $n, m \in \mathbb{N}$ and any $p \in F$ where $s = 2s' + 1$ is a given odd step length. Such a situation can be converted to the above setting in a straightforward way:

Proposition 4.30. *Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$ and let χ be a modulus of uniform (ϕ_{2n}) - (G, H) -quasi-Fejér monotonicity for (x_{2n}) w.r.t. F and (AF_k) with errors (ε_n) . Assume $s = 2s' + 1$ and let $\zeta : \mathbb{N}^3 \rightarrow \mathbb{N}$ be such that for all $n, m, r \in \mathbb{N}$, there exists a $k \leq \zeta(n, m, r)$ such that for all $p \in AF_k$:*

$$\forall l \leq m \left(H(\phi_{2(n+l)+s}(p, x_{2(n+l)+s})) \leq G(\phi_{2n}(p, x_{2n})) + \sum_{i=f(n)}^{n+l+s'} \varepsilon_i + \frac{1}{r+1} \right).$$

Then, for $f(n) = n \dot{-} s'$, $\hat{\zeta}(n, m, r) = \max\{\zeta(n, m, r), \chi(0, s', r)\}$ is a modulus of uniform $(\hat{\phi}_n)$ - (G, H) -quasi- f -monotonicity for (\hat{x}_n) w.r.t. F and (AF_k) with errors (ε_n) where for any sequence (a_n) of objects, we define

$$\hat{a}_n = \begin{cases} a_n & \text{for } n \geq s, \\ a_n & \text{for even } n < s, \\ a_{n-1} & \text{for odd } n < s. \end{cases}$$

So, we in particular see that the above situation is covered already via affine linear f . In that case, note that a modulus of uniform Fejér monotonicity for the sequence (x_{2n}) is also a modulus of uniform Fejér monotonicity for (\hat{x}_{2n}) . Similarly, we get the following result:

Proposition 4.31. *Assume $s = 2s' + 1$ and let Φ be an approximate F -point bound for (x_{2n}) . Then $\Phi'(k) = \Phi(k) + s'$ is an approximate F -point bound for $(x_{2f(n)})$ where $f(n) = n \dot{-} s'$.*

5. RELATION TO THE METRIC CASE AND TO BREGMAN DISTANCES

5.1. Consistent distances and their properties. We now want to discuss the relation of the previously introduced relativized notions to their original metric counterparts if the distance ϕ in question is suitably related to the metric. Concretely, in many cases⁸ one works with distances ϕ which satisfy

$$\phi(x, y) = 0 \leftrightarrow d(x, y) = 0.$$

In the following, we will consider distances that fulfill the above property quantitatively in the sense of the following definition:

⁸In fact, in the important special case of Bregman distances (which will be later discussed) this property was singled out as one of the central properties of the abstract generalized distances considered by Bregman already in the seminal paper [10] (see Condition I on p. 200 therein) which formed the basis for the concrete Bregman distances now prevalent in the literature.

Definition 5.1. We call ϕ consistent if there exist two functions $\lambda, \Lambda : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, called moduli of consistency, that satisfy

$$\begin{aligned} \phi(x, y) \leq \frac{1}{\lambda(k, b) + 1} &\rightarrow d(x, y) \leq \frac{1}{k + 1}, \\ d(x, y) \leq \frac{1}{\Lambda(k, b) + 1} &\rightarrow \phi(x, y) \leq \frac{1}{k + 1}, \end{aligned}$$

for any $k, b \in \mathbb{N}$ and for any $x, y \in X$ such that $d(a, x), d(a, y) \leq b$ holds w.r.t. some reference point $a \in X$.⁹ We call such functions moduli of uniform consistency (and ϕ uniformly consistent) if they do not depend on b .

Remark 5.2. Note that given $\lambda, \Lambda : \mathbb{N} \rightarrow \mathbb{N}$, these functions are moduli of uniform consistency for ϕ if, and only if $d \leq^\lambda \phi$ and $\phi \leq^\Lambda d$.

We will later discuss an example of a consistent distance beyond the metric.

In the context of such moduli of consistency, we can immediately derive that ϕ is weakly triangular as well as triangular and derive corresponding moduli from λ, Λ : Define $\theta(k, b) = \lambda(2k + 1, b)$. Then for x, y, z such that $d(a, x), d(a, y), d(a, z) \leq b$, assume that

$$\phi(y, x), \phi(y, z) \leq \frac{1}{\theta(k, b) + 1}.$$

Then in particular

$$d(y, x), d(y, z) \leq \frac{1}{2k + 2}$$

by assumption on λ and thus $d(x, z) \leq 1/(k + 1)$. Thus θ is a modulus of triangularity. Clearly, it also follows immediately from this that $\theta(k, b) = \lambda(2\Lambda(k, b) + 1, b)$ is a modulus of weak triangularity.

Further, a simple compactness argument shows again that if the space is compact and ϕ is continuous, then the above property

$$\phi(x, y) = 0 \leftrightarrow d(x, y) = 0$$

already implies the existence of moduli of uniform consistency for ϕ .

Remark 5.3 (For logicians). As in the context of the moduli of (weak) triangularity, also here the underlying logical methods from proof mining actually guarantee that corresponding moduli λ, Λ can be extracted from a large class of proofs of the property

$$\phi(x, y) = 0 \leftrightarrow d(x, y) = 0,$$

even in the absence of any compactness assumptions.

Even further, using such moduli, we can translate corresponding moduli of total ϕ -boundedness and uniform ϕ -closedness into moduli for their metric counterparts, and vice versa. For that, we again rely on these uniform versions of the moduli λ, Λ which in particular can be derived from the usual moduli in any bounded metric space. As the proof is rather straightforward, we omit it.

Proposition 5.4. *Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be a uniformly consistent mapping with moduli $\lambda, \Lambda : \mathbb{N} \rightarrow \mathbb{N}$. Then:*

- (1) *If γ is a modulus of total boundedness for some set $A \subseteq X$, then $\gamma'(k) = \gamma(\Lambda(k))$ is a modulus of total ϕ -boundedness.*
- (2) *If ω, δ are moduli of uniform closedness for a given F and a sequence (AF_k) , then $\omega'(k) = \lambda(\omega(k))$ and $\delta'(k) = \delta(k)$ are respective moduli of uniform ϕ -closedness.*

Derived from these conversions, we can now obtain the following modified general quantitative convergence result where the assumptions are phrased in terms of the metric (and thus are witnessed by moduli already commonly used in the literature).

Theorem 5.5. *Let (X, d) be a metric space and (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Assume that $\phi : X \times X \rightarrow \mathbb{R}_+$ is uniformly consistent with moduli $\lambda, \Lambda : \mathbb{N} \rightarrow \mathbb{N}$ and satisfies $\phi_n \nearrow \phi$ uniformly with a rate of convergence π . Further assume that $\psi \leq^A \phi_n$ for some $A : \mathbb{N} \rightarrow \mathbb{N}$ where ψ is uniformly weakly triangular with a modulus θ . Assume that X is totally bounded with a modulus of total boundedness γ , α_G be a G -modulus for G , β_H be a H -modulus for H and $f : \mathbb{N} \rightarrow \mathbb{N}$ be nondecreasing and $f(n) \leq n$. Further, let $(x_n) \subseteq X$ be a sequence and*

⁹As before with the modulus θ , the existence of such moduli does not depend on the concrete reference point.

- (1) χ be a modulus for uniform (ϕ_{2n}) - (G, H) -quasi-Fejér monotonicity w.r.t. F and (AF_k) for (x_{2n}) with errors (ε_n) ,
- (2) ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$,
- (3) Φ be a lim inf-bound w.r.t. F and (AF_k) for $(x_{2f(n)})$,
- (4) κ be a rate of divergence for f ,
- (5) ζ be a modulus of uniform (ϕ_n) - (G, H) -quasi- f -monotonicity w.r.t. F and (AF_k) for (x_n) with errors (ε_n) .

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

$$\exists N \leq \Psi'(k, g) \forall i, j \in [N; N + g(N)] \left(\psi(x_i, x_j) \leq \frac{1}{k+1} \right),$$

where $\Psi'(k, g)$ is defined similar to Ψ from Theorem 4.22, now using $\gamma'(k) = \gamma(\Lambda(k))$. As before, if ψ is uniformly triangular with a modulus θ , then the sequence is Cauchy with the same rate of metastability.

If in addition to the above assumptions, F is uniformly ψ -closed with moduli δ and ω , then for all $k \in \mathbb{N}$ and all $g \in \mathbb{N}^{\mathbb{N}}$:

$$\exists N \leq \tilde{\Psi}'(k, g) \forall i, j \in [N; N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k+1} \text{ and } x_i \in AF_k \right),$$

where $\tilde{\Psi}'(k, g)$ is defined similar to Ψ' from above but with

$$\tilde{\eta}_{k, \delta}(n, r) = \max\{\delta(k), \eta(n, r)\}$$

instead of η and with $\tilde{\theta}(k) = \max\{\theta(k), \omega(k)\}$ instead of θ .

If ψ is also uniformly consistent with moduli $\lambda', \Lambda' : \mathbb{N} \rightarrow \mathbb{N}$, then it suffices to assume that F is uniformly closed with moduli δ and ω and the above rates apply with $\delta'(k) = \delta(k)$ and $\omega'(k) = \lambda(\omega(k))$ used instead.

Lastly, if in any of the cases, we have $\phi_n = \phi$ for all n and instead of (1) and (5) we even have

- (1)' χ is a modulus for uniform ϕ - (G, H) -Fejér monotonicity w.r.t. F and (AF_k) for (x_{2n}) ,
- (5)' ζ is a modulus of uniform ϕ - (G, H) - f -monotonicity w.r.t. F and (AF_k) for (x_n) ,

then ξ and κ from assumptions (2) and (4) can be omitted and it suffices to assume that

- (3)' Φ is an approximate F -point bound w.r.t. (AF_k) for $(x_{2f(n)})$,

and we get the same conclusions, respectively, with a rate $\hat{\Psi}(k, g) = 2\hat{\Psi}_0(P, k, g)$ where $\hat{\Psi}_0$ is defined as

$$\begin{cases} \hat{\Psi}_0(0, k, g) = 0, \\ \hat{\Psi}_0(n+1, k, g) = \Phi(\eta^M(\hat{\Psi}_0(n, k, g), 4\beta_H(A(\theta(k)))) + 3), \end{cases}$$

with the other constants defined as before.

Again, the proof is immediate and we thus omit it.

Remark 5.6. Similar to Remark 4.24, also these results hold in the more general case of boundedly compact spaces if the sequence in question is bounded. However, in contrast to Fejér monotonicity w.r.t. the metric d , the property of a sequence (x_n) being Fejér monotone w.r.t. (ϕ_n) even for $\phi_n \rightarrow \phi$ does not suffice to infer that said sequence is bounded but only that it is ϕ -bounded as discussed before in Remark 4.24. This is effectively due to the fact that we know almost nothing about ϕ . In fact, the property that (ϕ_n) -Fejér monotonicity implies boundedness of the sequence is not even guaranteed for the special case of Bregman monotone sequences, i.e. where $\phi_n = D_f$ for any n . In that context, it is thus a common requirement in the literature (being contained in various notions of a ‘‘Bregman function’’, see e.g. [25], and in the stronger form of requiring compactness already being mentioned in Bregman’s work [10] for his general class of distances D) to require that the level sets

$$\begin{aligned} L_1(y, \alpha) &= \{x \in X \mid D_f(x, y) \leq \alpha\}, \\ L_2(x, \alpha) &= \{y \in X \mid D_f(x, y) \leq \alpha\}, \end{aligned}$$

are bounded for every $\alpha > 0$ and $x, y \in X$.

In similarity, we thus may additionally require that our function ϕ satisfies the following boundedness condition:

$$\{y \in X \mid \phi(z, y) \leq \alpha\} \text{ is bounded for any } z \in X, \alpha > 0.$$

Then the ϕ -boundedness translates into the boundedness of (x_n) and thus the above results holds in boundedly compact spaces in that context.

This result now gives rise to the following general convergence result by “forgetting” about the quantitative data in the assumptions and conclusions. Concretely we obtain the following generalization of Proposition 2.7.

Theorem 5.7. *Let (X, d) be compact and $F \subseteq X$ be explicitly closed. Let (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$. Assume $\phi_n \nearrow \phi$ uniformly and that $\psi \leq^A \phi_n$ where $A : \mathbb{N} \rightarrow \mathbb{N}$ and where $\psi, \phi : X \times X \rightarrow \mathbb{R}_+$ are uniformly consistent as well as continuous in their left arguments. If $(x_n) \subseteq X$ is such that there exists a nondecreasing function $f : \mathbb{N} \rightarrow \mathbb{N}$ with $f(n) \leq n$ and $f(n) \rightarrow \infty$ for $n \rightarrow \infty$ and where (x_n) is partially (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) and $(x_{2f(n)})$ has the \liminf -property w.r.t. F and (AF_k) , then (x_n) converges to some $x^* \in F$.*

In particular, the above holds if ψ and ϕ , instead of being uniformly consistent, are continuous also in their right arguments and it additionally holds that

$$\psi(x, y) = 0 \leftrightarrow d(x, y) = 0 \leftrightarrow \phi(x, y) = 0$$

for all $x, y \in X$.

If $\phi_n = \phi$ and $\varepsilon_n = 0$ for all n , then it suffices that $(x_{2f(n)})$ has approximate F -points w.r.t. (AF_k) .

Proof. The result follows from Theorem 5.5, noting that all the relevant uniform premises follow from the nonuniform variants by simple compactness arguments as mentioned before and by noting that metastability implies convergence. \square

This result not only generalizes the previous Proposition 2.7 but also encompasses (special cases of) other results on convergence for “Fejér-like monotone” sequences. As a representative example, besides those discussed in subsequent papers, we in particular want to mention the work of [4] where a convergence result (see Theorem 4.11 therein) was obtained for sequences which are Fejér monotone w.r.t. Bregman-distances

$$D_f : X \times X \rightarrow [0, +\infty], \quad (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle & \text{if } y \in \text{intdom} f, \\ +\infty & \text{otherwise,} \end{cases}$$

for $f : X \rightarrow (0, +\infty]$ a lower semicontinuous convex function which is Gâteaux-differentiable on $\text{intdom} f \neq \emptyset$ over a reflexive Banach space X . In particular, in the context of the strong convergence result presented there, the additional requirement (called Condition 4.4 in [4]) that

$$D_f(x_n, y_n) \rightarrow 0 \text{ implies } x_n - y_n \rightarrow 0$$

for all bounded sequences (x_n) and (y_n) in $\text{intdom} f$ is discussed. If f is real-valued (i.e. $f(x) \neq +\infty$ for all x) then the above clearly yields that $D_f(x, y) = 0$ implies $\|x - y\| = 0$. If f is strictly convex, the Bregman distance D_f also naturally satisfies that $\|x - y\| = 0$ implies $D_f(x, y) = 0$ and further, if f is even Fréchet-differentiable, then D_f is continuous in both arguments. In that vein, our above convergence results covers these partial cases of the result obtained in [4] if the sequence further belongs to a compact subset (as is e.g. in particular the case if the space is finite dimensional as the sequences are naturally bounded).

5.2. A more involved example of a uniformly consistent distance. We are now concerned with a more involved particular example of a concrete Bregman distances where moduli of consistency can be naturally constructed. Now, for general Bregman distances D_f , the situation is in general quite complicated as not all distances are immediately consistent. We refer to [66] for further discussions on (the use of) moduli of consistency in this general case and in this subsection only survey the special case D_f of the Bregman distance for $f = \|\cdot\|^2/2$ in the context of uniformly convex and uniformly smooth Banach spaces where such moduli can be readily obtained (although this is non-trivial). Instead of working with D_f directly, we follow the setup of e.g. the works [2, 36, 47] and rather work with the function ϕ defined by¹⁰

$$\phi(x, y) = \|x\|^2 - 2\langle x, Jy \rangle + \|y\|^2$$

where J is the (single-valued) duality map

$$J(x) = \{x^* \in X^* \mid \|x\|^2 = \|x^*\|^2 = \langle x, x^* \rangle\}.$$

Basic considerations on ϕ immediately yield that

$$0 \leq (\|x\| - \|y\|)^2 \leq \phi(x, y).$$

Central for obtaining moduli of consistency for this distance ϕ is the following result due to Alber:

¹⁰It can be easily seen via the use of the Fenchel conjugate of f for the above f that ϕ is nothing else than $2D_f$.

Proposition 5.8 (Alber [2]). *Let X be a uniformly convex and uniformly smooth Banach space with a modulus of convexity δ_X , i.e.*

$$\delta_X(\varepsilon) = \inf\{1 - \|x + y\|/2 \mid x, y \in X, \|x\|, \|y\| \leq 1, \|x - y\| \geq \varepsilon\},$$

and a modulus of uniform smoothness ρ_X , i.e.

$$\rho_X(\tau) = \sup\{1/2(\|x + y\| + \|x - y\| - 2) \mid x, y \in X, \|x\| = 1, \|y\| \leq \tau\}.$$

Then for all $x, y \in X$:

$$8C^2\delta_X(\|x - y\|/4C) \leq \phi(x, y) \leq 4C^2\rho_X(4\|x - y\|/C)$$

where

$$C = \sqrt{(\|x\|^2 + \|y\|^2)/2}.$$

In particular, utilizing the Figiel inequalities [27], for $\|x\|, \|y\| \leq b$:

$$2L^{-1}b^2\delta_X(\|x - y\|/4b) \leq \phi(x, y) \leq 4L^{-1}b^2\rho_X(4\|x - y\|/b)$$

where $1 < L < 3.18$ is the constant of the Figiel inequalities [27].

From this, suitable moduli λ, Λ can be immediately derived.

Lemma 5.9. *Let X be a uniformly convex and uniformly smooth Banach space and let $\eta : (0, 2] \rightarrow (0, 1]$ be a nondecreasing modulus of uniform convexity for X , i.e.*

$$\forall \varepsilon \in (0, 2] \forall x, y \in X \left(\|x\|, \|y\| \leq 1 \wedge \|x - y\| \geq \varepsilon \rightarrow \left\| \frac{x + y}{2} \right\| \leq 1 - \eta(\varepsilon) \right).$$

Then, the functions

$$\begin{aligned} \tilde{\Lambda}(\varepsilon, b) &= \frac{\varepsilon}{16L^{-1}b}, \\ \tilde{\lambda}(\varepsilon, b) &= \frac{b^2}{L} \eta\left(\frac{\varepsilon}{4b}\right), \end{aligned}$$

satisfy

$$\begin{aligned} \|x - y\| \leq \tilde{\Lambda}(\varepsilon, b) &\rightarrow \phi(x, y) \leq \varepsilon, \\ \phi(x, y) \leq \tilde{\lambda}(\varepsilon, b) &\rightarrow \|x - y\| \leq \varepsilon, \end{aligned}$$

for all $x, y \in X$ with $\|x\|, \|y\| \leq b$. In particular, the functions

$$\begin{aligned} \Lambda(k, b) &= \left\lceil \frac{k+1}{16L^{-1}b} \right\rceil \div 1, \\ \lambda(k, b) &= \left\lfloor \frac{L}{b^2} \left(\eta\left(\frac{1}{4b(k+1)}\right) \right)^{-1} \right\rfloor \div 1, \end{aligned}$$

are moduli of consistency for ϕ .

Proof. Note that $\eta(\varepsilon) \leq \delta_X(\varepsilon)$ and that $\rho_X(\varepsilon) \leq \varepsilon$. Therefore, for one, if

$$\|x - y\| \leq \tilde{\Lambda}(\varepsilon, b) = \frac{\varepsilon}{16L^{-1}b},$$

then we get

$$\begin{aligned} \phi(x, y) &\leq 4L^{-1}b^2\rho_X(4\|x - y\|/b) \\ &\leq 4L^{-1}b4\|x - y\| \\ &= 16L^{-1}b\|x - y\| \\ &\leq \varepsilon. \end{aligned}$$

For another, if

$$\phi(x, y) \leq \tilde{\lambda}(\varepsilon, b) = \frac{b^2}{L} \eta\left(\frac{\varepsilon}{4b}\right),$$

then we get

$$\begin{aligned} 2L^{-1}b^2\eta(\|x - y\|/4b) &\leq 2L^{-1}b^2\delta_X(\|x - y\|/4b) \\ &\leq \phi(x, y) \\ &< \frac{2b^2}{L}\eta\left(\frac{\varepsilon}{4b}\right). \end{aligned}$$

Thus, in particular, we have $\eta(\|x - y\|/4b) < \eta(\varepsilon/4b)$ and thus $\|x - y\|/4b < \varepsilon/4b$, i.e. $\|x - y\| \leq \varepsilon$, as η is nondecreasing. \square

As discussed before, from λ and Λ we can immediately derive

$$\theta(k, b) = \left[\frac{L}{b^2} \left(\eta \left(\frac{1}{8b(k+1)} \right) \right)^{-1} \right] \div 1$$

as a modulus for the triangularity and

$$\tilde{\theta}(k, b) = \left[\frac{L}{b^2} \left(\eta \left(\frac{1}{8b(\lceil \frac{k+1}{16L^{-1}b} \rceil + 1)} \right) \right)^{-1} \right] \div 1$$

as a modulus for the weak triangularity of ϕ .

6. MODULI OF REGULARITY AND RATES OF CONVERGENCE

To establish the main results for rates of convergence under a metric regularity assumption, we now follow the setup established in [45]. Between this and the previous, there are some notational conflicts but the context will make it clear to which objects we refer.

The first main difference is that while the previous setup operated on approximations AF_k of a solution set F governed by natural numbers k , the setup from [45] works with a general function $F : X \rightarrow \overline{\mathbb{R}}$, mapping into the extended reals, for measuring the ‘‘approximateness’’ of any point x to the solution set now represented by $\text{zer}F$, i.e. instead of finding points in a set F we are now interested in finding zeros of the function F and $|F(x)| < \varepsilon$ is to be understood as x being an ε -approximate solution.

The whole setup of Fejér monotonicity is naturally formulated in this framework with Fejér monotonicity now understood over $\text{zer}F$ and with approximate solutions understood in the previously explained sense instead of the sets AF_k .

The main notion on which we rely in the following is that of a modulus of regularity from [45]:

Definition 6.1 (essentially [45]). Let $F : X \rightarrow \overline{\mathbb{R}}$ be a given function. For a set $S \subseteq X$, a function $\rho : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is called a modulus of regularity for F w.r.t. $\text{zer}F$ and S if for all $\varepsilon > 0$ and all $x \in S$:

$$|F(x)| < \rho(\varepsilon) \rightarrow \text{dist}(x, \text{zer}F) < \varepsilon.$$

In that way, a modulus of regularity, given an error $\varepsilon > 0$, allows for the construction of an error $\rho(\varepsilon)$ such that being a $\rho(\varepsilon)$ -approximate solution implies that one is ε -close to an actual solution. We refer to [45] for a discussion on how various notions of regularity known from optimization and nonlinear analysis fit into this framework and do not discuss this here any further.

Remark 6.2. Note that in [43], the above notion of a modulus of regularity is only introduced relative to the specific choice of $S = \overline{B}(z, r)$ for a given $z \in \text{zer}F$ and $r > 0$ but we here allow general sets S (mainly out of simplicity so that the comparisons to moduli of regularity relative to a distance function ϕ introduced later become simpler and do not require tedious translations between metric balls and balls relative to ϕ). In particular, we want to note that in the context of metrically bounded sets S , the uniformity of the modulus ρ regarding the elements $x \in S$ is in particular suggested by the logical metatheorems of proof mining. Although this is not the case for general sets S , we here nevertheless included this uniformity in the above definition of the moduli of regularity. Note that in all case studies presented later, the sets are naturally bounded anyhow.

In the context of the mapping ϕ , we here consider the following relativized notion:

Definition 6.3. Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be any mapping. For a set $S \subseteq X$, a function $\rho : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is called a modulus of ϕ -regularity for F w.r.t. $\text{zer}F$ and S if for all $\varepsilon > 0$ and all $x \in S$:

$$|F(x)| < \rho(\varepsilon) \rightarrow \text{dist}_\phi(x, \text{zer}F) < \varepsilon.$$

Here, dist_ϕ is the ϕ -distance function defined by

$$\text{dist}_\phi(x, \text{zer}F) = \inf_{y \in \text{zer}F} \phi(y, x).$$

We can then obtain the following generalization of the main result of [45] on rates of convergence derived from moduli of regularity. We state the result already in its full generality. To that end, the sequence is only assumed to be partially (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) in the sense of before. In that way, it also generalizes the result given in [63] which extended Theorem 4.1 of [45] by error terms to treat quasi-Fejér monotone sequences.

In this setup of working with real ε 's as representations for errors, the previous moduli α_G , β_H as well as π and θ then need to be real-valued in the sense that now:

- (1) A modulus of uniform weak triangularity or of uniform triangularity is now a function $\theta : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ that satisfies

$$\phi(y, x), \phi(y, z) \leq \theta(\varepsilon) \rightarrow \phi(x, z) \leq \varepsilon \text{ or } \phi(y, x), \phi(y, z) \leq \theta(\varepsilon) \rightarrow d(x, z) \leq \varepsilon$$

for any $\varepsilon \in \mathbb{R}_+^*$ and any $x, y, z \in X$, respectively.

- (2) G - and H -moduli are now functions $\alpha_G, \beta_H : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ that satisfy

$$a \leq \alpha_G(\varepsilon) \rightarrow G(a) \leq \varepsilon$$

as well as

$$H(a) \leq \beta_H(\varepsilon) \rightarrow a \leq \varepsilon$$

for any $\varepsilon \in \mathbb{R}_+^*$ and any $a \in \mathbb{R}_+$, respectively.

- (3) A rate for $\phi_n \nearrow \phi$ uniformly is now a function $\pi : \mathbb{R}_+^* \rightarrow \mathbb{N}$ that satisfies

$$\forall n \geq \pi(\varepsilon) \forall x, y \in X (\phi_n(x, y) \leq \phi(x, y) + \varepsilon)$$

for all $\varepsilon \in \mathbb{R}_+^*$.

Also, we now understand an A witnessing $\psi \leq^A \phi$ to be a real-valued function $A : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ with

$$\psi \leq^A \phi \equiv \forall \varepsilon > 0 \forall x, y \in X (\phi(x, y) \leq A(\varepsilon) \rightarrow \psi(x, y) \leq \varepsilon).$$

Theorem 6.4. Let (X, d) be a metric space and (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$ such that $\psi \leq^A \phi_n$ for $A : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ and where $\psi : X \times X \rightarrow \mathbb{R}_+$ is uniformly weakly triangular with modulus θ and such that $\phi_n \nearrow \phi$ uniformly with a rate π for a distance $\phi : X \times X \rightarrow \mathbb{R}_+$. Let $F : X \rightarrow \overline{\mathbb{R}}$ be given with $\text{zer}F \neq \emptyset$. Let f be a nondecreasing function with $f(n) \leq n$ and let κ be a rate of divergence for f . Assume that (x_n) is partially (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) . Let α_G be a G -modulus for G and β_H be a H -modulus for H . Suppose there is a $\tau : \mathbb{R}_+^* \times \mathbb{N} \rightarrow \mathbb{N}$ such that

$$\forall \delta > 0 \forall n \in \mathbb{N} \exists N \in [n; \tau(\delta, n)] (|F(x_{2f(N)})| < \delta).$$

Let ρ be a modulus of ϕ -regularity for F w.r.t. $\text{zer}F$ and S such that $(x_n) \subseteq S$ and let ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$.

Then (x_n) is ψ -Cauchy and

$$\forall \delta > 0 \forall n, m \geq 2\mu(\delta) (\psi(x_n, x_m) < \delta)$$

with

$$\mu(\delta) = \tau \left(\rho \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right), \max \left\{ \kappa \left(\xi \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) \right), \kappa \left(\pi \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right) \right\} \right).$$

If ψ is uniformly triangular with a modulus θ , then (x_n) is Cauchy with the same rate.

Proof. Let $\delta > 0$ be given. By assumption we have

$$\exists N \in \left[\max \left\{ \kappa \left(\xi \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) \right), \kappa \left(\pi \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right) \right\}; \mu(\delta) \right] \\ \left(|F(x_{2f(N)})| < \rho \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right).$$

As ρ is a corresponding modulus of regularity and since $(x_n) \subseteq S$, we get

$$\text{dist}_\phi(x_{2f(N)}, \text{zer}F) < \alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2$$

and therefore, there exists a $y \in \text{zer}F$ with $\phi(y, x_{2f(N)}) < \alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2$. As

$$N \geq \kappa \left(\pi \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right)$$

we obtain

$$f(N) \geq \pi \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right)$$

and thus we get $\phi_{2f(N)}(y, x_{2f(N)}) \leq \alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right)$ and thus $G(\phi_{2f(N)}(y, x_{2f(N)})) \leq \frac{\beta_H(A(\theta(\delta)))}{2}$. As further

$$N \geq \kappa \left(\xi \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) \right),$$

we obtain

$$N \geq f(N) \geq \xi \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right)$$

and thus for $n \geq 2N$, by (ϕ_{2n}) - (G, H) -quasi-Fejér monotonicity, we obtain

$$\begin{aligned} H(\phi_n(y, x_n)) &\leq G(\phi_{2f(N)}(y, x_{2f(N)})) + \sum_{i=f(N)}^{\infty} \varepsilon_i \\ &\leq \beta_H(A(\theta(\delta))) / 2 + \beta_H(A(\theta(\delta))) / 2 \\ &\leq \beta_H(A(\theta(\delta))) \end{aligned}$$

for $n \geq 2N$ even and by (ϕ_n) - (G, H) -quasi- f -monotonicity, we obtain similarly

$$H(\phi_n(y, x_n)) \leq G(\phi_{2f(N)}(y, x_{2f(N)})) + \sum_{i=f(N)}^{\infty} \varepsilon_i \leq \beta_H(A(\theta(\delta)))$$

for $n \geq 2N$ odd. Thus $\phi_n(y, x_n) \leq A(\theta(\delta))$ for any $n \geq 2N$. This yields $\psi(y, x_n) \leq \theta(\delta)$ for any $n \geq 2N$ (and so for any $n \geq 2\mu(\delta)$ since $N \leq \mu(\delta)$). If ψ is uniformly weakly triangular with modulus θ , we in particular derive from this that $\psi(x_n, x_m) \leq \delta$ for any $n, m \geq 2\mu(\delta)$. If ψ is even uniformly triangular with modulus θ , then we derive from this that $d(x_n, x_m) \leq \delta$ for any $n, m \geq 2\mu(\delta)$. \square

Remark 6.5. Note that in the context of the above Theorem 6.6, we in particular get that the sequence in question strongly converges without any additional compactness assumption.

We again explicitly list the case where all ϕ_n coincide with a single function ϕ :

Theorem 6.6. *Let (X, d) be a metric space and $\phi : X \times X \rightarrow \mathbb{R}_+$ be a function such that $\psi \leq^A \phi$ for $A : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ and for $\psi : X \times X \rightarrow \mathbb{R}_+$ being uniformly weakly triangular with modulus θ . Let $F : X \rightarrow \overline{\mathbb{R}}$ be given with $\text{zer}F \neq \emptyset$. Let f be a nondecreasing function with $f(n) \leq n$ and let κ be a rate of divergence for f . Assume that (x_n) is partially (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) . Let α_G be a G -modulus for G , β_H be an H -modulus for H . Suppose there is a $\tau : \mathbb{R}_+^* \times \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\forall \delta > 0 \forall n \in \mathbb{N} \exists N \in [n; \tau(\delta, n)] (|F(x_{2f(N)})| < \delta).$$

Let ρ be a modulus of ϕ -regularity for F w.r.t. $\text{zer}F$ and S such that $(x_n) \subseteq S$ and let ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$. Then (x_n) is ψ -Cauchy and

$$\forall \delta > 0 \forall n, m \geq 2\mu(\delta) (\psi(x_n, x_m) < \delta)$$

with

$$\mu(\delta) = \tau \left(\rho \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right), \kappa \left(\xi \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) \right) \right).$$

If ψ is uniformly triangular with a modulus θ , then (x_n) is Cauchy with the same rate.

Lastly, if $\varepsilon_n = 0$ for any $n \in \mathbb{N}$, then the assumption that $f(n) \rightarrow \infty$ for $n \rightarrow \infty$ can be omitted and it suffices to assume that $\tau : \mathbb{R}_+^ \rightarrow \mathbb{N}$ is such that*

$$\forall \delta > 0 \exists N \leq \tau(\delta) (|F(x_{2f(N)})| < \delta),$$

to derive that (x_n) is ψ -Cauchy or Cauchy, depending on the assumption on ψ , with a rate $2\mu'(\delta)$ where

$$\mu'(\delta) = \tau \left(\rho \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right)$$

and where the other constants are as before.

We omit the proof as it is rather immediate.

Again, if we have moduli relating the distance function ϕ to the usual metric distance in the sense that we have moduli $\lambda, \Lambda : \mathbb{R}_+^* \times \mathbb{N} \rightarrow \mathbb{R}_+^*$ such that

$$\begin{aligned} \phi(x, y) \leq \lambda(\varepsilon, b) &\rightarrow d(x, y) \leq \varepsilon, \\ d(x, y) \leq \Lambda(\varepsilon, b) &\rightarrow \phi(x, y) \leq \varepsilon, \end{aligned}$$

for any $\varepsilon > 0$ and any $x, y \in X$ such that $d(a, x), d(a, y) \leq b$, then we can modify the above result to only require the usual standard metric versions of the assumptions. Similarly, we also in this case call ϕ consistent with moduli $\lambda, \Lambda : \mathbb{R}_+^* \times \mathbb{N} \rightarrow \mathbb{R}_+^*$ if these satisfy the above inequalities and we call ϕ uniformly consistent with moduli $\lambda, \Lambda : \mathbb{R}_+^* \rightarrow \mathbb{N}$ if these satisfy the above inequalities without any reference to b .

As before, we can again immediately derive a corresponding moduli θ and $\tilde{\theta}$ from the moduli λ, Λ for the triangularity and weak triangularity of ϕ , respectively: Define

$$\theta(\varepsilon, b) = \lambda \left(\frac{\varepsilon}{2}, b \right).$$

Then for $x, y, z \in X$ such that $d(a, x), d(a, y), d(a, z) \leq b$, assume that we have $\phi(y, x), \phi(y, z) \leq \theta(\varepsilon, b)$. Then in particular $d(y, x), d(y, z) \leq \varepsilon/2$ by assumption on λ and thus $d(x, z) \leq \varepsilon$. Thus θ is a modulus for the triangularity of θ . It is then clear that

$$\tilde{\theta}(\varepsilon, b) = \lambda \left(\frac{\Lambda(\varepsilon, b)}{2}, b \right)$$

is a corresponding modulus for the weak triangularity.

Similarly, we can provide a translation for the main notion of a modulus of ϕ -regularity to its metric counterpart.

Lemma 6.7. *Let $\phi : X \times X \rightarrow \mathbb{R}_+$ be a uniformly consistent mapping with moduli $\lambda, \Lambda : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$. Let further $z \in \text{zer}F$, $r > 0$ and $\rho : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ be a modulus of regularity for F w.r.t. $\text{zer}F$ and S . Then $\rho'(\varepsilon) = \rho(\Lambda(\varepsilon/2))$ is a modulus of ϕ -regularity for F w.r.t. $\text{zer}F$ and S .*

Proof. Let ε be given and let $x \in S$. Then suppose that $|F(x)| < \rho'(\varepsilon) = \rho(\Lambda(\varepsilon/2))$, then by assumption on ρ , we have $\text{dist}(x, \text{zer}F) < \Lambda(\varepsilon/2)$ and thus

$$\exists y \in \text{zer}F \ (d(y, x) < \Lambda(\varepsilon/2)).$$

Thus for this y , we have $\phi(y, x) \leq \varepsilon/2 < \varepsilon$ and thus it follows that $\text{dist}_\phi(x, \text{zer}F) < \varepsilon$. \square

In the following theorem, we now just collect the resulting rates of convergence that can be constructed in the context of a consistent distance:

Theorem 6.8. *Let (X, d) be a metric space and (ϕ_n) be a sequence of functions $\phi_n : X \times X \rightarrow \mathbb{R}_+$ such that $\psi \leq^A \phi_n$ for $A : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ and where $\psi : X \times X \rightarrow \mathbb{R}_+$ is uniformly weakly triangular with a modulus θ . Further assume that $\phi_n \nearrow \phi$ uniformly with a rate π for a given function $\phi : X \times X \rightarrow \mathbb{R}_+$ which is uniformly consistent with moduli $\lambda, \Lambda : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$. Let $F : X \rightarrow \bar{\mathbb{R}}$ be given with $\text{zer}F \neq \emptyset$. Let f be a nondecreasing function with $f(n) \leq n$ and let κ be a rate of divergence for f . Assume that (x_n) is partially (ϕ_n) - (G, H) -quasi-Fejér monotone w.r.t. F , f and (ε_n) . Let α_G be a G -modulus for G , β_H be an H -modulus for H . Suppose there is a $\tau : \mathbb{R}_+^* \times \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\forall \delta > 0 \forall n \in \mathbb{N} \exists N \in [n; \tau(\delta, n)] (|F(x_{2f(N)})| < \delta).$$

Let ρ be a modulus of regularity for F w.r.t. $\text{zer}F$ and S such that $(x_n) \subseteq S$ and let ξ be a Cauchy rate for $\sum \varepsilon_n < \infty$.

Then (x_n) is ψ -Cauchy and

$$\forall \delta > 0 \forall n, m \geq 2\mu'(\delta) \ (\psi(x_n, x_m) < \delta)$$

with μ' defined as μ in Theorem 6.6 but with $\rho'(\varepsilon) = \rho(\Lambda(\varepsilon/2))$ instead of ρ . If ψ is even uniformly triangular with a modulus θ , then the sequence is Cauchy with the same rate.

If $\phi_n = \phi$ and $\varepsilon_n = 0$ for all n , then the assumption that $f(n) \rightarrow \infty$ for $n \rightarrow \infty$ can be omitted and we may assume that $\tau : \mathbb{R}_+^* \rightarrow \mathbb{N}$ is such that

$$\forall \delta > 0 \exists N \leq \tau(\delta) (|F(x_{2f(N)})| < \delta),$$

to derive that (x_n) is ψ -Cauchy or Cauchy, depending on whether ψ is uniformly weakly triangular or uniformly triangular with modulus θ as before, with a rate $2\mu'(\delta)$ where

$$\mu'(\delta) = \tau \left(\rho' \left(\alpha_G \left(\frac{\beta_H(A(\theta(\delta)))}{2} \right) / 2 \right) \right)$$

and where the other constants are as before.

Again, we omit the proof as it is rather immediate.

7. APPLICATIONS TO A METHOD WITH ALTERNATING INERTIA

Let X be a finite dimensional Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$ and dimension d . We consider the algorithm given in [33] to approximate fixed points of α -averaged mappings T , i.e. $T : X \rightarrow X$ satisfies

$$(1 - \alpha) \|(Id - T)x - (Id - T)y\|^2 \leq \alpha \left(\|x - y\|^2 - \|Tx - Ty\|^2 \right).$$

Concretely, given x^0 , $\lambda > 0$ and $0 \leq \alpha_k \leq (1 - \alpha)/\alpha$ for all k , we consider the iteration defined by

$$x^{k+1} = T\bar{x}^k \text{ with } \bar{x}^k = \begin{cases} x^k & \text{if } k \text{ is even,} \\ x^k + \alpha_k(x^k - x^{k-1}) & \text{if } k \text{ is odd.} \end{cases}$$

By modifying a result from [59] for the case of $\alpha = 1/2$, the authors then obtain the following convergence result:

Theorem 7.1 ([33], Lemma 3.3). *Let $0 \leq \alpha_k \leq (1 - \alpha)/\alpha$ for all k . Then (x^k) converges to a fixed point of T .*

The proof of this result proceeds by establishing partial Fejér monotonicity (w.r.t. the usual metric distance) and thus fits into the above general result which we will apply in the following to derive quantitative information.

Note before that, that using the correspondence between averaged mappings and monotone operators (see, e.g., [8] as well as [7]) this immediately covers the case when T is given as the resolvent $J_\lambda^A = (Id + \gamma A)^{-1}$ for $\gamma > 0$ for a maximal monotone or maximally ρ -comonotone operator $A : X \rightarrow 2^X$ as J_λ^A is firmly nonexpansive, i.e. $1/2$ -averaged, in the former case and α -averaged for $\alpha = \frac{1}{2(\rho/\lambda + 1)}$ in the latter case (see [8]).

In that case, by $\text{Fix} J_\lambda^A = A^{-1}(0)$, approximating fixed points of $T = J_\lambda^A$ means approximating zeros $\hat{x} \in A^{-1}(0)$.

In that formulation, as mentioned before, the algorithm considered above is exactly the algorithm considered in [59] for maximally monotone operators and the following quantitative analysis thus also applies in that case by setting $\alpha = 1/2$.

Now, for a quantitative approach in the spirit of the previous sections, take \hat{x} with $\hat{x} = T\hat{x}$ fixed, suppose $M \geq \|x^k - \hat{x}\|, \|\bar{x}^k - \hat{x}\|$ for all k . Define $X_0 = \overline{B}(\hat{x}, M)$ as well as

$$F = \{x \in X_0 \mid x = Tx\}$$

as well as

$$AF_k = \left\{ x \in X_0 \mid \|x - Tx\| \leq \frac{1}{k+1} \right\}.$$

Lemma 7.2. *If $\sum_i a_i \leq A$, then*

$$\min_{0 \leq i \leq k-1} a_i \leq \frac{A}{k} \text{ for all } k \geq 1.$$

Proof. Suppose not. Then there is a $k \geq 1$ such that for any $0 \leq i \leq k-1$: $a_i > A/k$. Then

$$\sum_i a_i \geq \sum_{i=0}^{k-1} a_i > A$$

which is a contradiction. □

Lemma 7.3 ([33]). *For any k :*

$$\|x^{k+2} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|Tx^k + \alpha_{k+1}(Tx^k - x^k) - x^{k+2}\|^2.$$

Lemma 7.4.

$$\left(\frac{1-\alpha}{\alpha}\right) \sum_i \|x^{2i+2} - \bar{x}^{2i+1}\|^2 = \left(\frac{1-\alpha}{\alpha}\right) \sum_i \|Tx^{2i} - \alpha_{2i+1}(Tx^{2i} - x^{2i}) - x^{2i+2}\|^2 \leq \|x^0 - \hat{x}\|^2.$$

Thus in particular

$$\min_{0 \leq i \leq k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2 \leq \left(\frac{\alpha}{1-\alpha}\right) \frac{\|x^0 - \hat{x}\|^2}{k}$$

for any $k \geq 1$.

We omit the proof as it is immediate.

Lemma 7.5. *For any $k \geq 0$: $\|x^{2k+3} - x^{2k+2}\| \leq \|x^{2k+2} - \bar{x}^{2k+1}\|$.*

Proof. As T is α -averaged:

$$(1-\alpha) \|x^{2k+2} - x^{2k+3} - (\bar{x}^{2k+1} - x^{2k+2})\|^2 \leq \alpha \left(\|x^{2k+2} - \bar{x}^{2k+1}\|^2 - \|x^{2k+3} - x^{2k+2}\|^2 \right).$$

Thus, in particular

$$\begin{aligned} \|x^{2k+3} - x^{2k+2}\|^2 &\leq \|x^{2k+2} - \bar{x}^{2k+1}\|^2 - \left(\frac{1-\alpha}{\alpha}\right) \|x^{2k+2} - x^{2k+3} - (\bar{x}^{2k+1} - x^{2k+2})\|^2 \\ &\leq \|x^{2k+2} - \bar{x}^{2k+1}\|^2. \end{aligned}$$

□

Lemma 7.6. *Let $b \geq \|x^0 - \hat{x}\|$. Then $\Phi(k) = 2 \max \left\{ 1, \left\lceil \frac{\alpha}{1-\alpha} b^2 (k+1)^2 \right\rceil \right\}$ is an approximate F -point bound for (x^{2n}) .*

Proof. Write k' for $\max \left\{ 1, \left\lceil \frac{\alpha}{1-\alpha} b^2 (k+1)^2 \right\rceil \right\}$. As $k' \geq 1$, by Lemma 7.4, there exists an $i \in [0; k'-1]$ such that

$$\|x^{2i+2} - \bar{x}^{2i+2}\|^2 \leq \left(\frac{\alpha}{1-\alpha}\right) \frac{\|x^0 - \hat{x}\|^2}{k'} \leq \left(\frac{\alpha}{1-\alpha}\right) \frac{b^2}{k'}.$$

Using Lemma 7.5, we get

$$\|x^{2i+3} - x^{2i+2}\| \leq \left(\frac{\sqrt{\alpha}}{\sqrt{1-\alpha}}\right) \frac{b}{\sqrt{k'}} \leq \frac{1}{k'+1}.$$

As $2i+2$ is even, we have $x^{2i+3} = T\bar{x}^{2i+2} = Tx^{2i+2}$. Thus $x^{2i+2} \in AF_k$. Lastly, we have $2i+2 \leq 2(k'-1)+2 = 2k' = \Phi(k)$. □

Lemma 7.7. *Let $\|x^* - Tx^*\| \leq \varepsilon$ and $b \geq \|x^k - x^*\|, \|x^k - Tx^*\|, \|Tx^k - x^*\|, \|x^k - Tx^k\|$ for any k . Then for any k :*

$$\|x^{k+2} - x^*\|^2 \leq \|x^k - x^*\|^2 + \left(\frac{3-\alpha}{\alpha^2} + 2\right) b\varepsilon.$$

Proof. We have

$$\begin{aligned} \|x^{k+2} - x^*\| &= \langle x^{k+2} - Tx^*, x^{k+2} - x^* \rangle + \langle Tx^* - x^*, x^{k+2} - x^* \rangle \\ &= \langle x^{k+2} - Tx^*, x^{k+2} - Tx^* \rangle + \langle x^{k+2} - Tx^*, Tx^* - x^* \rangle + \langle Tx^* - x^*, x^{k+2} - x^* \rangle \\ &\leq \|x^{k+2} - Tx^*\|^2 + 2b\varepsilon \\ &\leq \|Tx^k + \alpha_{k+1}(Tx^k - x^k) - x^*\|^2 + 2b\varepsilon. \end{aligned}$$

Now

$$\begin{aligned} &\|Tx^k + \alpha_{k+1}(Tx^k - x^k) - x^*\|^2 \\ &= \|(1 + \alpha_{k+1})(Tx^k - x^*) - \alpha_{k+1}(x^k - x^*)\|^2 \\ &= (1 + \alpha_{k+1}) \|Tx^k - x^*\|^2 - \alpha_{k+1} \|x^k - x^*\|^2 + \alpha_{k+1}(1 + \alpha_{k+1}) \|Tx^k - x^k\|^2 \end{aligned}$$

using Corollary 2.15 from [7]. Using

$$\begin{aligned} \|Tx^k - x^*\|^2 &= \langle Tx^k - Tx^*, Tx^k - x^* \rangle + \langle Tx^* - x^*, Tx^k - x^* \rangle \\ &\leq \langle Tx^k - Tx^*, Tx^k - Tx^* \rangle + 2b\varepsilon \\ &\leq \|x^k - x^*\|^2 - \frac{1-\alpha}{\alpha} \|x^k - Tx^k - x^* + Tx^*\|^2 + 2b\varepsilon, \end{aligned}$$

we get

$$\begin{aligned} \|Tx^k + \alpha_{k+1}(Tx^k - x^k) - x^*\|^2 &\leq \|x^k - x^*\|^2 - (1 + \alpha_{k+1}) \frac{1-\alpha}{\alpha} \|x^k - Tx^k - x^* + Tx^*\|^2 \\ &\quad + \alpha_{k+1}(1 + \alpha_{k+1}) \|Tx^k - x^k\|^2 + (1 + \alpha_{k+1})2b\varepsilon. \end{aligned}$$

Using

$$\begin{aligned} \|x^k - Tx^k - x^* + Tx^*\|^2 &= \|x^k - Tx^k\|^2 + \langle x^k - Tx^k, Tx^* - x^* \rangle + \langle Tx^* - x^*, x^k - Tx^k - x^* + Tx^* \rangle \\ &\geq \|x^k - Tx^k\|^2 - b\varepsilon + \langle Tx^* - x^*, x^k - x^* \rangle + \langle Tx^* - x^*, Tx^* - Tx^k \rangle \\ &\geq \|x^k - Tx^k\|^2 - 3b\varepsilon \end{aligned}$$

this implies

$$\begin{aligned} \|Tx^k + \alpha_{k+1}(Tx^k - x^k) - x^*\|^2 &\leq \|x^k - x^*\|^2 - (1 + \alpha_{k+1}) \frac{1-\alpha}{\alpha} \left(\|x^k - Tx^k\|^2 - 3b\varepsilon \right) \\ &\quad + \alpha_{k+1}(1 + \alpha_{k+1}) \|Tx^k - x^k\|^2 + (1 + \alpha_{k+1})2b\varepsilon \\ &\leq \|x^k - x^*\|^2 + (1 + \alpha_{k+1}) \frac{1-\alpha}{\alpha} 3b\varepsilon + (1 + \alpha_{k+1})2b\varepsilon \\ &\leq \|x^k - x^*\|^2 + \left(1 + \frac{1-\alpha}{\alpha} \right) \frac{1-\alpha}{\alpha} 3b\varepsilon + \left(1 + \frac{1-\alpha}{\alpha} \right) 2b\varepsilon \\ &= \|x^k - x^*\|^2 + \left(\frac{3-\alpha}{\alpha^2} \right) b\varepsilon \end{aligned}$$

which yields

$$\|x^{k+2} - x^*\|^2 \leq \|x^k - x^*\|^2 \left(\frac{3-\alpha}{\alpha^2} + 2 \right) b\varepsilon.$$

□

Lemma 7.8. *Let $b \geq \|x^k - x^*\|$, $\|x^k - Tx^*\|$, $\|Tx^k - x^*\|$, $\|x^k - Tx^k\|$ for all k and $\|x^* - Tx^*\| \leq \varepsilon$. Then for any k :*

- (1) $\|x^{2k+2} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 + \left(\frac{3-\alpha}{\alpha^2} + 2 \right) b\varepsilon$,
- (2) $\|x^{2k+3} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 + \left(\frac{3-\alpha}{\alpha^2} + 4 \right) b\varepsilon$.

Proof. The first item follows immediately from Lemma 7.7 and for the latter, note that

$$\|x^{2k+3} - x^*\|^2 \leq \|x^{2k+2} - x^*\|^2 + 2b\varepsilon.$$

□

This immediately yields a modulus of uniform Fejér monotonicity and the respective uniform modulus for the odd elements:

Lemma 7.9. *For all r, n, m :*

- (1) $\forall x^* \in AF_{\chi(n,m,r)} \forall l \leq m \left(\|x^{2(n+l)} - x^*\| \leq \|x^{2n} - x^*\| + \frac{1}{r+1} \right)$,
- (2) $\forall x^* \in AF_{\zeta(n,m,r)} \forall l \leq m \left(\|x^{2(n+l)+3} - x^*\| \leq \|x^{2n} - x^*\| + \frac{1}{r+1} \right)$,

where

$$\chi(n, m, r) = 2m^2 \left[\frac{3-\alpha}{\alpha^2} + 2 \right] M(r+1)^2 \div 1 \text{ and } \zeta(n, m, r) = 2m^2 \left[\frac{3-\alpha}{\alpha^2} + 4 \right] M(r+1)^2 \div 1.$$

Proof. W.l.o.g. $l \geq 1$ in both cases. Let $x^* \in AF_{\chi(n,m,r)}$. Then, by Lemma 7.8, we get

$$\begin{aligned} \|x^{2(n+l)} - x^*\| &\leq \|x^{2n} - x^*\| + m\sqrt{2\left[\frac{3-\alpha}{\alpha^2} + 2\right]M\frac{1}{\chi(n,m,r)+1}} \\ &\leq \|x^{2n} - x^*\| + \frac{1}{r+1}. \end{aligned}$$

Similarly, we obtain for $x^* \in AF_{\zeta(n,m,r)}$:

$$\begin{aligned} \|x^{2(n+l)+3} - x^*\| &\leq \|x^{2n} - x^*\| + m\sqrt{2\left[\frac{3-\alpha}{\alpha^2} + 4\right]M\frac{1}{\zeta(n,m,r)+1}} \\ &\leq \|x^{2n} - x^*\| + \frac{1}{r+1}. \end{aligned}$$

□

Combined, we obtain a rate of metastability for the sequences by an application of Theorem 5.5.

Theorem 7.10. *The sequence (x_n) is Cauchy and for all $k \in \mathbb{N}, g \in \mathbb{N}^{\mathbb{N}}$:*

$$\exists N \leq \Psi(k, g) \forall i, j \in [N; N + g(N)] \left(\|x_i - x_j\| \leq \frac{1}{k+1} \text{ and } x_i \in AF_k \right)$$

where $\Psi(k, g) = 2\Psi_0(P, 2k+1, g') + 3$ with $P = \left[(64k+64)\sqrt{d}M \right]^d$, $g'(n) = g(n+3) + 3$ and Ψ_0 defined by

$$\begin{cases} \Psi_0(0, k, g) = 0, \\ \Psi_0(n+1, k, g) = \Phi^M(\eta_{g,k}^M(\Psi_0(n, k, g), 4k+3)), \end{cases}$$

where

$$\Phi(k) = 2 \max \left\{ 1, \left\lceil \frac{\alpha}{1-\alpha} M^2(k+1)^2 \right\rceil \right\} + 1$$

and $\Phi^M(k) = \max\{\Phi(j) \mid j \leq k\}$ as well as

$$\eta_{g,k}(n, r) = \max \left\{ 2k+1, 2 \left\lceil \frac{g(2n)}{2} \right\rceil^2 \left\lceil \frac{3-\alpha}{\alpha^2} + 4 \right\rceil M(2r+4)^2 \div 1 \right\}.$$

Proof. Note that T is nonexpansive and thus by Lemma 7.1 of [43], we get that F is uniformly closed with moduli $\omega_F(k) = 4k+3$ and $\delta_F(k) = 2k+1$. Further, by Example 2.8 of [43] we get that $\gamma(k) = \left[2(k+1)\sqrt{d}M \right]^d$ is a modulus of total boundedness of $\overline{B}(\hat{x}, M)$. Then, for the sequence (\tilde{x}_n) defined by

$$\tilde{x}_n = \begin{cases} x_n & \text{for } n \geq 3, \\ x_n & \text{for even } n < 3, \\ x_{n-1} & \text{for odd } n < 3, \end{cases}$$

we get from Theorem 5.5 by the use of Proposition 4.30, Proposition 4.31, Lemma 7.9, Lemma 7.6 and $f(n) := n \div 1$ that

$$\exists N \leq 2\Psi_0(P, 2k+1, g') \forall i, j \in [N; N + g(N+3) + 3] \left(\|\tilde{x}_i - \tilde{x}_j\| \leq \frac{1}{k+1} \text{ and } \tilde{x}_i \in AF_k \right)$$

after some obvious simplifications. Therefore we get, using $\tilde{x}_i = x_i$ for $i \geq 3$, that

$$\exists N \leq 2\Psi_0(P, 2k+1, g') \forall i, j \in [N+3; N+3 + g(N+3)] \left(\|x_i - x_j\| \leq \frac{1}{k+1} \text{ and } x_i \in AF_k \right)$$

which yields the result. □

Under the assumption of a modulus of regularity, we can also derive a rate of convergence using Theorem 6.8. For this, the central instantiation of the notion of the moduli of regularity is the following:

Definition 7.11. Let $T : X \rightarrow X$ be a map. For a set $S \subseteq X$, a function $\rho : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is called a modulus of regularity for T w.r.t. $\text{Fix}T$ and S if for all $\varepsilon > 0$ and all $x \in S$:

$$\|x - Tx\| < \rho(\varepsilon) \rightarrow \text{dist}(x, \text{Fix}T) < \varepsilon.$$

With such a modulus, we get the following result.

Theorem 7.12. *Let $x^* \in \text{zer}F$ and $b \geq \max\{\|x^0 - x^*\|, \|x^1 - x^*\|\}$. Let ρ be a modulus of regularity for T w.r.t. $\text{Fix}T$ and $\overline{B}(x^*, b)$. Then (x_n) is Cauchy with*

$$\forall \delta > 0 \forall n, m \geq \max \left\{ 3, 2 \left\lceil \frac{\alpha}{1-\alpha} b^2 \left(\left\lceil \frac{1}{\rho(\frac{\delta}{4})} \right\rceil + 1 \right)^2 \right\rceil \right\} (\|x_n - x_m\| < \delta).$$

Proof. For (\tilde{x}_n) as before we get from Theorem 6.8 (and using Proposition 4.30, Proposition 4.31, Lemma 7.9, Lemma 7.6 and $f(n) := n \div 1$) that

$$\forall \delta > 0 \forall n, m \geq 2 \left\lceil \frac{\alpha}{1-\alpha} b^2 \left(\left\lceil \frac{1}{\rho(\frac{\delta}{4})} \right\rceil + 1 \right)^2 \right\rceil (\|\tilde{x}_n - \tilde{x}_m\| < \delta).$$

The result then follows as for $n \geq 3$, we have $\tilde{x}_n = x_n$. \square

As already discussed in Remark 6.5, also here the strong convergence in particular holds without any additional compactness assumptions in the context of these moduli of regularity.

We refer to the extensive discussions in [45] for examples of mappings T where such a modulus can be readily constructed.

8. APPLICATIONS TO THE MANN-TYPE PROXIMAL POINT ALGORITHM IN BANACH SPACES

For a second application, we work over a Banach space X with norm $\|\cdot\|$ and dual space X^* . Since the seminal work by Rockafellar [67] and Martinet [55], the proximal point algorithm has been one of the most influential algorithmic approaches to many problems in optimization.

We are here concerned with the previously mentioned work [36] for a combination of Mann's iteration with the proximal point algorithm (see again [77] for this kind of iteration in Hilbert space and e.g. [56] for related iterations) for a monotone operator of the Banach space X in the sense of Browder [12]. Concretely, a set-valued operator $T \subseteq X \times X^*$ is called monotone if

$$\langle x - y, x^* - y^* \rangle \geq 0$$

for all $(x, x^*), (y, y^*) \in T$. For these monotone operators in Banach spaces, there are natural analogues for the resolvent. Over a smooth Banach space with a (single-valued) duality map J and given a maximally monotone operator T , one particular choice is the function $J_r : X \rightarrow X$ defined by

$$J_r x = (J + rT)^{-1} Jx$$

for $r > 0$ and $x \in X$ (see also [3] for this and other variants of the resolvent in this context). Then, by combining the aspects of Mann's iteration and the proximal point algorithm, Kamimura, Kohsaka and Takahashi obtained the following result:

Theorem 8.1 ([36]). *Let X be a uniformly smooth and uniformly convex Banach space whose duality map is weakly sequentially continuous. Let T be a maximally monotone operator and x_n defined by*

$$x_{n+1} = J^{-1} (\alpha_n Jx_n + (1 - \alpha_n) J J_{r_n} x_n)$$

for $x_0 \in X$ and where $(\alpha_n) \subseteq [0, 1]$, $(r_n) \subseteq \mathbb{R}_+^*$ satisfy

$$\limsup_{n \rightarrow \infty} \alpha_n < 1 \text{ and } \liminf_{n \rightarrow \infty} r_n > 0.$$

If $T^{-1}0 \neq \emptyset$, then the sequence (x_n) converges weakly to an element of $T^{-1}0$.

Note that in this context of a uniformly smooth space, the mapping J is uniformly continuous on bounded subsets.

Essential for the proof of weak convergence, and thus also for the convergence analysis, is establishing Fejér monotonicity of this algorithm w.r.t. the previously discussed distance

$$\phi(x, y) = \|x\|^2 - \langle x, Jy \rangle + \|y\|^2$$

for $x, y \in X$. In that way, this fits into the general quantitative result established before and we apply this result in the following to extract quantitative information on the above convergence from the proof. Recall for

this in particular the previously derived moduli λ, Λ and θ for the distance ϕ in a uniformly convex and smooth space with a non-decreasing modulus of uniform convexity η :

$$\begin{aligned}\Lambda(k, b) &= \left\lceil \frac{k+1}{16L^{-1}b} \right\rceil \div 1, \\ \lambda(k, b) &= \left\lceil \frac{L}{b^2} \left(\eta \left(\frac{1}{4b(k+1)} \right) \right)^{-1} \right\rceil \div 1, \\ \theta(k, b) &= \left\lceil \frac{L}{b^2} \left(\eta \left(\frac{1}{8b(k+1)} \right) \right)^{-1} \right\rceil \div 1.\end{aligned}$$

For the quantitative analysis, we again suppose $M \geq \|x_n\|$ for all n and define $X_0 = \overline{B}(0, M)$ as well as

$$F = \text{zer}T \cap X_0 = \text{Fix}J_r \cap X_0 \text{ (for any } r > 0\text{)}$$

as well as

$$AF_k = \left\{ x \in X_0 \mid \|x - J_{r_n}x\| \leq \frac{1}{k+1} \text{ for all } n \leq k \right\}.$$

The result given in [36] relies on a particular result on the relation between the distance and the resolvents of the operator from [47]:

Lemma 8.2 ([47]). *Let X be smooth, strictly convex and reflexive and let T be maximally monotone with $\text{zer}T \neq \emptyset$. Then*

$$\phi(x, J_r y) + \phi(J_r y, y) \leq \phi(x, y)$$

for any $x \in \text{zer}T$ and $y \in X$.

Based on the approximations AF_k , we get a quantitative version of this result. For this, we still need the following preliminary bound on the norm of the resolvent:

Lemma 8.3. *Let T be maximally monotone and let $d \in Tc$ with $C \geq \|c\|$, $D \geq \|d\|$. Then for any $r > 0$ and any $x \in X$:*

$$\|J_r x\| \leq \max\{(1+C)(b+RD) + C, 1\} =: \mu(R, b)$$

where $R \geq r$ and $b \geq \|x\|$.

Proof. By definition of the resolvent, we have for any $r > 0$ and any $x \in X$:

$$(J_r x, r^{-1}(Jx - JJ_r x)) \in T.$$

As $d \in Tc$, we have by monotonicity of T that

$$\langle J_r x - c, r^{-1}(Jx - JJ_r x) - d \rangle \geq 0.$$

By linearity, we obtain

$$\begin{aligned}0 &\leq \langle J_r x - c, r^{-1}(Jx - JJ_r x) - d \rangle \\ &= \langle J_r x, r^{-1}Jx - d \rangle - \langle J_r x, r^{-1}JJ_r x \rangle - \langle c, r^{-1}Jx - d \rangle + \langle c, r^{-1}JJ_r x \rangle\end{aligned}$$

and therefore we get

$$\begin{aligned}\|J_r x\|^2 &= \langle J_r x, JJ_r x \rangle \leq \langle J_r x, Jx - rd \rangle - \langle c, Jx - rd \rangle + \langle c, JJ_r x \rangle \\ &\leq \|J_r x\| (\|x\| + rD) + C(\|x\| + rD) + C \|J_r x\|.\end{aligned}$$

Suppose now that $\|J_r x\| > \max\{\|x\| + rD + C(\|x\| + rD) + C, 1\}$. Then by multiplying by $\|J_r x\|^{-1}$, we get

$$\begin{aligned}\|J_r x\| &\leq \|x\| + rD + \|J_r x\|^{-1} C(\|x\| + rD) + C \\ &\leq \|x\| + rD + C(\|x\| + rD) + C\end{aligned}$$

from the above which is a contradiction. This gives the claim. \square

Lemma 8.4. *Let X be uniformly smooth and uniformly convex and let T be maximally monotone. Let $d \in Tc$ with $C \geq \|c\|$, $D \geq \|d\|$. Let $\omega(\varepsilon, b) \leq \varepsilon$ be a modulus of uniform continuity of J on bounded subsets, i.e.*

$$\forall z, w \in X, b \in \mathbb{N}, \varepsilon > 0 (\|z\|, \|w\| \leq b \wedge \|z - w\| \leq \omega(\varepsilon, b) \rightarrow \|Jz - Jw\| \leq \varepsilon).$$

Then for any $\varepsilon > 0$, $x, y \in X$ and $r, s > 0$, if x and s are such that

$$\|x - J_s x\| \leq \omega\left(\frac{\varepsilon}{2E}, \max\{b, \mu(s, b)\}\right)$$

for $b \geq \|x\|, \|y\|$, we have

$$\phi(x, J_r y) + \phi(J_r y, y) \leq \phi(x, y) + \varepsilon$$

where

$$E \geq \max\{2(\mu(r, b) + b), 2rs^{-1}(\mu(r, b) + \mu(s, b))\}$$

with μ as in Lemma 8.3.

Proof. It is known that

$$\begin{aligned} \phi(x, y) &= \phi(x, J_r y) + \phi(J_r y, y) + 2\langle x - J_r y, J J_r y - J y \rangle \\ &= \phi(x, J_r y) + \phi(J_r y, y) + 2r\langle x - J_r y, -A_r y \rangle. \end{aligned}$$

Now, we also immediately have (using monotonicity of T):

$$\begin{aligned} \langle x - J_r y, -A_r y \rangle &= \langle x - J_s x, -A_r y \rangle + \langle J_s x - J_r y, -A_r y \rangle \\ &= \langle x - J_s x, -A_r y \rangle + \langle J_s x - J_r y, s^{-1}(Jx - J J_s x) - A_r y \rangle + \langle J_s x - J_r y, -s^{-1}(Jx - J J_s x) \rangle \\ &\geq \langle x - J_s x, -A_r y \rangle + s^{-1}\langle J_s x - J_r y, J J_s x - Jx \rangle \\ &\geq -\|x - J_s x\| \|A_r y\| - s^{-1} \|J_s x - J_r y\| \|J J_s x - Jx\| \\ &\geq -\|x - J_s x\| r^{-1}(\|J_r y\| + \|y\|) - s^{-1}(\|J_s x\| + \|J_r y\|) \|J J_s x - Jx\|. \end{aligned}$$

Thus, using the previous Lemma 8.3, we have

$$\begin{aligned} \phi(x, y) &\geq \phi(x, J_r y) + \phi(J_r y, y) - 2r \|x - J_s x\| r^{-1}(\|J_r y\| + \|y\|) - s^{-1}2r(\|J_s x\| + \|J_r y\|) \|J J_s x - Jx\| \\ &\geq \phi(x, J_r y) + \phi(J_r y, y) - \|x - J_s x\| 2(\mu(r, \|y\|) + \|y\|) \\ &\quad - 2rs^{-1}(\mu(r, \|y\|) + \mu(s, \|x\|)) \|J J_s x - Jx\| \\ &\geq \phi(x, J_r y) + \phi(J_r y, y) - E(\|x - J_s x\| + \|J J_s x - Jx\|) \end{aligned}$$

and therefore, for x such that

$$\|x - J_s x\| \leq \omega\left(\frac{\varepsilon}{2E}, \max\{b, \mu(s, b)\}\right),$$

we get that

$$\phi(x, y) \geq \phi(x, J_r y) + \phi(J_r y, y) - \varepsilon$$

which is the claim. \square

Remark 8.5. If $\omega : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ is a modulus of uniform continuity of J in the sense that

$$\|z - w\| \leq \frac{1}{\omega(k, b) + 1} \rightarrow \|Jz - Jw\| \leq \frac{1}{k + 1}$$

for all $k \in \mathbb{N}$ and all $z, w \in \overline{B}(0, b)$, then it holds that

$$\phi(x, J_r y) + \phi(J_r y, y) \leq \phi(x, y) + \frac{1}{k + 1}$$

for any $k, b \in \mathbb{N}$, $x, y \in X$ and $r, s > 0$ whenever

$$\|x - J_s x\| \leq \frac{1}{\omega(2E(k + 1) + 1, B + 1) + 1}$$

for $B \geq \|x\|$.

Using this, we can give a quantitative version (in the sense of a modulus of uniform ϕ -Fejér monotonicity as introduced before) of the ‘‘Fejér monotonicity’’-type result implicitly shown in the course of the proof of Theorem 3.1 in [36].

Lemma 8.6. *Let X be uniformly smooth and uniformly convex and let T be maximally monotone. Let $d \in Tc$ with $C \geq \|c\|$, $D \geq \|d\|$ and let $\omega : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ be a modulus of uniform continuity of J . Then for all $r, n, m \in \mathbb{N}$:*

$$\forall z \in AF_{\chi(n, m, r)} \forall l \leq m \left(\phi(z, x_{n+l}) \leq \phi(z, x_n) + \frac{1}{r + 1} \right),$$

where

$$\chi(n, m, r) = \max\{n, \omega(2E_{n, m}^0((r + 1)(m \div 1) \div 1 + 1) + 1, M + 1)\}$$

as well as

$$E_{n, m}^0 \geq \max\{2(\mu(\hat{r}_{n, m}, M) + M), 2\hat{r}_{n, m} r_n^{-1}(\mu(\hat{r}_{n, m}, M) + \mu(r_n, M))\},$$

where $\hat{r}_{n, m} = \max\{r_i \mid i \leq n + m \div 1\}$.

Proof. We get

$$\phi(z, x_{n+1}) \leq \alpha_n \phi(z, x_n) + (1 - \alpha_n) \phi(z, J_{r_n} x_n).$$

as in the proof of Theorem 3.1 in [36]. Now, since $z \in AF_{\chi(n,m,r)}$, we have

$$\|z - J_{r_n} z\| \leq \frac{1}{\omega(2E_{n,m}^0((r+1)(m \div 1) \div 1 + 1), M+1) + 1}$$

and by Lemma 8.4 (see also Remark 8.5), this implies

$$\phi(z, J_{r_n} x_n) \leq \phi(z, x_n) + \frac{1}{(r+1)(m \div 1) \div 1 + 1}$$

and by induction, combined with the above, we get

$$\begin{aligned} \phi(z, x_{n+l}) &\leq \phi(z, x_n) + \frac{1}{(r+1)(m \div 1)} \sum_{i=n}^{n+m \div 1} (1 - \alpha_n) \\ &\leq \phi(z, x_n) + \frac{m \div 1}{(r+1)(m \div 1)} \\ &\leq \phi(z, x_n) + \frac{1}{r+1} \end{aligned}$$

for any $l \leq m$ since $\|z\|, \|x_i\| \leq M$ for all i as well as

$$E_{n,m}^0 \geq \max\{2(\mu(\hat{r}_{n,m}, M) + M), 2\hat{r}_{n,m}r_n^{-1}(\mu(\hat{r}_{n,m}, M) + \mu(r_n, M))\}$$

where $\hat{r}_{n,m} = \max\{r_i \mid i \leq n + m \div 1\}$. □

Next, we will obtain an approximate F -point bound from the asymptotic regularity result

$$\phi(y_n, x_n) \rightarrow 0 \text{ for } n \rightarrow \infty$$

established in the proof of Theorem 3.1 from [36]. For this, we first collect a few preliminary results.

Lemma 8.7. *If $(a_n) \subseteq [0, b]$ and $a_{n+1} \leq a_n$ for all n , then*

$$\forall \varepsilon > 0 \exists n \leq 2 \left\lceil \frac{b}{\varepsilon} \right\rceil (a_{2n} - a_{2n+1} \leq \varepsilon).$$

Proof. Suppose not. Then

$$a_0 > \varepsilon + a_2 > \dots > 2 \left\lceil \frac{b}{\varepsilon} \right\rceil \varepsilon + a_{2 \lceil \frac{b}{\varepsilon} \rceil} \geq b$$

which is a contradiction. □

Lemma 8.8 ([36]). *Let $z \in \text{zer}T$. Then for any n : $\phi(z, x_{n+1}) \leq \phi(z, x_n)$ as well as*

$$\phi(y_n, x_n) \leq \frac{\phi(z, x_n) - \phi(z, x_{n+1})}{1 - \alpha_n}.$$

Lemma 8.9. *Let $z \in \text{zer}T$, let $\bar{\alpha}$ be such that $1 > \bar{\alpha} > \alpha_n$ for all n and let \bar{r} be such that $r_n \geq \bar{r} > 0$ for all n . Let $b \geq \phi(z, x_0)$. Then*

$$\forall \varepsilon \exists n \leq 2 \left\lceil \frac{b}{\varepsilon(1 - \bar{\alpha})} \right\rceil (\phi(J_{r_{2n}} x_{2n}, x_{2n}) \leq \varepsilon).$$

Proof. The claim follows from the previous lemmas. □

Lemma 8.10. *Let $z \in \text{zer}T$, let $\bar{\alpha}$ be such that $1 > \bar{\alpha} > \alpha_n$ for all n and let \bar{r} be such that $r_n \geq \bar{r} > 0$ for all n . Let $b \geq \phi(z, x_0)$. Then*

$$\forall k \in \mathbb{N} \exists n \leq \Phi(k) (x_{2n} \in AF_k)$$

with

$$\Phi(k) = 2 \left\lceil \frac{(\lambda(\omega(2E_k^1(\lambda(k)) + 1, M+1)) + 1)b}{1 - \bar{\alpha}} \right\rceil$$

as well as

$$E_k^1 \geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k(\bar{r}^{-1}\mu(\tilde{r}_k, M) + \max\{(1+C)(\bar{r}^{-1}M + D) + C, \bar{r}^{-1}\})\}$$

where $\tilde{r}_k = \max\{r_i \mid i \leq k\}$.

Proof. By Lemma 8.9 and the properties of λ , we have

$$\forall k \in \mathbb{N} \exists n \leq \Phi(k) \left(\|J_{r_{2n}} x_{2n} - x_{2n}\| \leq \frac{1}{\omega(2E_k^1(\lambda(k)) + 1, M + 1) + 1} \right).$$

By Lemma 8.4, we then in particular obtain that

$$\begin{aligned} \phi(x_{2n}, J_{r_i} x_{2n}) &\leq \phi(x_{2n}, x_{2n}) + \frac{1}{\lambda(k) + 1} \\ &= \frac{1}{\lambda(k) + 1} \end{aligned}$$

for all $i \leq k$. This in particular holds since

$$\begin{aligned} E_k^1 &\geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k(\bar{r}^{-1}\mu(\tilde{r}_k, M) + \max\{(1 + C)(\bar{r}^{-1}M + D) + C, \bar{r}^{-1}\})\} \\ &\geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k(\bar{r}^{-1}\mu(\tilde{r}_k, M) + r_{2n}^{-1} \max\{(1 + C)(M + r_{2n}D) + C, 1\})\} \\ &\geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k(\bar{r}^{-1}\mu(\tilde{r}_k, M) + r_{2n}^{-1}\mu(r_{2n}, M))\} \\ &\geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k r_{2n}^{-1}(\mu(\tilde{r}_k, M) + \mu(r_{2n}, M))\} \end{aligned}$$

where $\tilde{r}_k = \max\{r_i \mid i \leq k\}$. Thus $\|x_{2n} - J_{r_i} x_{2n}\| \leq 1/(k + 1)$ for all $i \leq k$. \square

The combination of the previous modulus of uniform Fejér monotonicity with the above approximate F -point bound then yields the following rate of metastability as an application of Theorem 4.25.

Theorem 8.11. *Let X be a uniformly convex and uniformly smooth Banach space and let $\eta : (0, 2] \rightarrow (0, 1]$ be a non-decreasing modulus of uniform convexity for X . Assume that X_0 as defined before is totally bounded with a modulus of total boundedness γ . Let $z \in \text{zer}T$, let $\bar{\alpha}$ be such that $1 > \bar{\alpha} > \alpha_n$ for all n and let \bar{r} be such that $r_n \geq \bar{r} > 0$ for all n . Let $b \geq \phi(z, x_0)$. Let $d \in Tc$ with $C \geq \|c\|$, $D \geq \|d\|$ and let $\omega : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ be a modulus of uniform continuity of J . Then (x_n) is Cauchy and for all $k \in \mathbb{N}$ and all $g : \mathbb{N} \rightarrow \mathbb{N}$:*

$$\exists N \leq \tilde{\Psi}(\lambda(k), g) \forall i, j \in [N; N + g(N)] \left(\|x_i - x_j\| \leq \frac{1}{k + 1} \text{ and } x_i \in AF_k \right)$$

with $\tilde{\Psi}(k, g) = \Psi(k_0, g)$ where

$$k_0 = \max\{k, \lambda(\omega_F(k))\}$$

where $\Psi(k, g) = 2\Psi_0(P, k, g)$ with $P = \gamma(\Lambda(16\theta(k) + 15))$ and Ψ_0 defined by

$$\begin{cases} \Psi_0(0, k, g) = 0, \\ \Psi_0(n + 1, k, g) = \Phi^M(\eta_k^M(\Psi_0(n, k, g), 2\theta(k) + 1)), \end{cases}$$

where

$$\begin{aligned} \Lambda(k) &= \left\lceil \frac{k + 1}{16L^{-1}(M + 1)} \right\rceil \div 1, \\ \lambda(k) &= \left\lceil \frac{L}{(M + 1)^2} \left(\eta \left(\frac{1}{4(M + 1)(k + 1)} \right) \right)^{-1} \right\rceil \div 1, \\ \theta(k) &= \left\lceil \frac{L}{(M + 1)^2} \left(\eta \left(\frac{1}{8(M + 1)(k + 1)} \right) \right)^{-1} \right\rceil \div 1, \end{aligned}$$

as well as

$$\Phi(k) = 2 \left\lceil \frac{(\lambda(\omega(2E_k^1(\lambda(k)) + 1, M + 1) + 1)b)}{1 - \bar{\alpha}} \right\rceil$$

and $\Phi^M(k) = \max\{\Phi(j) \mid j \leq k\}$ and where

$$\eta_k(n, r) = \max\{\chi'_{k, g}(n, r), \zeta_g(n, r), \delta(n, r)\}$$

with

$$\begin{aligned} \chi(n, m, r) &= \max\{n, \omega(2E_{n, m}^0((r + 1)(m \div 1) \div 1 + 1) + 1, M + 1)\}, \\ \chi_k(n, m, r) &= \max\{\delta_F(k), \chi(n, m, r)\}, \\ \chi'_k(n, m, r) &= \chi_k(2n, 2m, r), \\ \zeta(n, m, r) &= \chi(2n, 2m + 1, r), \end{aligned}$$

as well as

$$\begin{aligned}\delta(n, r) &= \chi(f(n), n - f(n), 4r + 3), \\ \chi'_{k,g}(n, r) &= \chi'_k\left(n, \left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right), \\ \zeta_g(n, r) &= \zeta\left(n, \left\lfloor \frac{g(2n)}{2} \right\rfloor, r\right).\end{aligned}$$

for

$$\begin{aligned}E_{n,m}^0 &\geq \max\{2(\mu(\hat{r}_{n,m}, M) + M), 2\hat{r}_{n,m}r_n^{-1}(\mu(\hat{r}_{n,m}, M) + \mu(r_n, M))\}, \\ E_k^1 &\geq \max\{2(\mu(\tilde{r}_k, M) + M), 2\tilde{r}_k(\bar{r}^{-1}\mu(\tilde{r}_k, M) + \max\{(1+C)(\bar{r}^{-1}M + D) + C, \bar{r}^{-1}\})\},\end{aligned}$$

where $\tilde{r}_k = \max\{r_i \mid i \leq k\}$ and $\hat{r}_{n,m} = \max\{r_i \mid i \leq n + m \div 1\}$ as well as

$$\mu(R, b) = \max\{(1+C)(b + RD) + C, 1\}.$$

If we have a modulus of regularity for the operator in the sense of the following definition, then an application of Theorem 6.6 even allows for the derivation of a rate of convergence.

Definition 8.12. Let $S \subseteq X$ be a set. A function $\rho : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is a *modulus of regularity for the (generalized) proximal point algorithm w.r.t. S* if for all $\varepsilon > 0$ and $x \in S$:

$$\|x - J_1x\| < \rho(\varepsilon) \text{ implies } \text{dist}(x, \text{Fix } J_1) < \varepsilon.$$

Theorem 8.13. Let X be a uniformly convex and uniformly smooth Banach space and let $\eta : (0, 2] \rightarrow (0, 1]$ be a modulus of uniform convexity for X . Let $\omega(\varepsilon, b) \leq \varepsilon$ be a modulus of uniform continuity of J on bounded subsets, i.e.

$$\forall z, w \in X, b \in \mathbb{N}, \varepsilon > 0 (\|z\|, \|w\| \leq b \wedge \|z - w\| \leq \omega(\varepsilon, b) \rightarrow \|Jz - Jw\| \leq \varepsilon).$$

Let X_0 be defined as before. Let $z \in \text{zer}T$ and let $b \geq \phi(z, x_0)$. Let $\bar{\alpha}$ be such that $1 > \bar{\alpha} > \alpha_n$ for all n and let \bar{r} be such that $r_n \geq \bar{r} > 0$ for all n . Let ρ be a modulus of regularity for the proximal point algorithm w.r.t. X_0 . Then (x_n) is Cauchy and

$$\forall \delta > 0 \forall n, m \geq \mu(\delta) (\|x_n - x_m\| < \delta)$$

with

$$\mu(\delta) = 2\tau\left(\rho'\left(\frac{\theta(\lambda(\delta))}{2}\right)\right)$$

and where

$$\begin{aligned}\tilde{\Lambda}(\varepsilon) &= \frac{\varepsilon}{16L^{-1}(M+1)}, \\ \tilde{\lambda}(\varepsilon) &= \frac{(M+1)^2}{L}\eta\left(\frac{\varepsilon}{4b(M+1)}\right), \\ \theta(\varepsilon) &= \tilde{\lambda}(\varepsilon/2), \\ \rho'(\varepsilon) &= \rho(\tilde{\Lambda}(\varepsilon/2)).\end{aligned}$$

and where

$$\tau(\varepsilon) = 2\left\lceil \frac{b}{\tilde{\lambda}\left(\omega\left(\frac{\tilde{\lambda}(\varepsilon)}{2E}, M\right)\right)(1-\bar{\alpha})} \right\rceil$$

with

$$E \geq \max\{2(\mu(1, M) + M), 2\bar{r}^{-1}\mu(1, M) + \max\{(1+C)(\bar{r}^{-1}M + D) + C, \bar{r}^{-1}\}\}$$

as well as $\mu(R, b) = \max\{(1+C)(b + RD) + C, 1\}$.

Proof. By Lemma 8.9, we have

$$\forall \varepsilon \exists n \leq 2\left\lceil \frac{b}{\varepsilon(1-\bar{\alpha})} \right\rceil (\phi(J_{r_{2n}}x_{2n}, x_{2n}) \leq \varepsilon).$$

Thus, we get that there exists an $n \leq \tau(\varepsilon)$ such that

$$\|x_{2n} - J_{r_{2n}}x_{2n}\| \leq \omega\left(\frac{\tilde{\lambda}(\varepsilon)}{2E}, M\right).$$

By Lemma 8.4, we have

$$\phi(x_{2n}, J_1 x_{2n}) \leq \phi(x_{2n}, x_{2n}) + \tilde{\lambda}(\varepsilon) = \tilde{\lambda}(\varepsilon)$$

and we thus have $\|x_{2n} - J_1 x_{2n}\| \leq \varepsilon$. For this, note that

$$\begin{aligned} E &\geq \max\{2(\mu(1, M) + M), 2(\bar{r}^{-1}\mu(1, M) + \max\{(1 + C)(\bar{r}^{-1}M + D) + C, \bar{r}^{-1}\})\} \\ &\geq \max\{2(\mu(1, M) + M), 2(\bar{r}^{-1}\mu(1, M) + r_{2n}^{-1} \max\{(1 + C)(M + r_{2n}D) + C, 1\})\} \\ &\geq \max\{2(\mu(1, M) + M), 2(\bar{r}^{-1}\mu(1, M) + r_{2n}^{-1}\mu(r_{2n}, M))\} \\ &\geq \max\{2(\mu(1, \|x_{2n}\|) + \|x_{2n}\|), 2r_{2n}^{-1}(\mu(1, \|x_{2n}\|) + \mu(r_{2n}, \|x_{2n}\|))\}. \end{aligned}$$

Thus τ confines to the properties of Theorem 6.6 and this result consequently implies the present theorem. \square

Similar to before, as already discussed in Remark 6.5, also here the strong convergence in particular holds without any additional compactness assumptions in the context of these moduli of regularity.

Again, we refer to [45] for the discussion of examples of concrete moduli of regularity for resolvents of monotone operators in Hilbert spaces which can be easily transformed into corresponding moduli of regularity for resolvents of monotone operators in Banach spaces in the above sense.

Acknowledgments I want to thank Ulrich Kohlenbach for suggesting the algorithms presented in [33, 59] as interesting case studies as well as for suggesting to consider potential abstract results as presented in the first part of this paper, generalizing the previous works [43, 45] to the setting of partial Fejér monotonicity. For another, I want to emphasize that Lemmas 7.7, 7.8 and 7.9 are generalizations of quantitative versions due to him of the analogous results from [59] that he communicated to me in personal correspondence. Further, I want to thank Thomas Powell and Morenikeji Neri for many remarks that led to simplifications in the assumptions of various theorems in this paper. Lastly, I want to thank the anonymous referees for the valuable comments that improved the paper at various places.

Funding and/or Conflicts of interests This work was supported by the ‘Deutsche Forschungsgemeinschaft’ Project DFG KO 1737/6-2. The author has no relevant financial or non-financial interests to disclose.

REFERENCES

- [1] Y. Alber and D. Butnariu. Convergence of Bregman Projection Methods for Solving Consistent Convex Feasibility Problems in Reflexive Banach Spaces. *Journal of Optimization Theory and Applications*, 92(1):33–61, 1997.
- [2] Y.I. Alber. Metric and generalized projections in banach spaces: properties and applications. In A.G. Kartsatos, editor, *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, pages 15–50. Marcel Dekker, New York, 1996.
- [3] V. Barbu. *Nonlinear semigroups and differential equations in Banach spaces*. Springer Netherlands, 1976.
- [4] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.
- [5] H.H. Bauschke and P.L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of Operations Research*, 26(2):248–264, 2001.
- [6] H.H. Bauschke and P.L. Combettes. Iterating Bregman retractions. *SIAM Journal on Optimization*, 13(4):1159–1173, 2003.
- [7] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, Cham, 2017.
- [8] H.H. Bauschke, W.M. Moursi, and X. Wang. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, 189:55–74, 2021.
- [9] S. Bonettini, I. Loris, F. Porta, and M. Prato. Variable metric inexact line-search-based methods for nonsmooth optimization. *SIAM Journal on Optimization*, 26(2):891–921, 2016.
- [10] L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.
- [11] F.E. Browder. Convergence theorems for sequences of nonlinear operators in Banach spaces. *Mathematische Zeitschrift*, 100:201–225, 1967.
- [12] F.E. Browder. Nonlinear maximal monotone operators in Banach space. *Mathematische Annalen*, 175(2):89–113, 1968.
- [13] M.N. Bùì and P.L. Combettes. Bregman forward-backward operator splitting. *Set-Valued and Variational Analysis*, 29:583–603, 2021.
- [14] R.S. Burachik, M.N. Dao, and S.B. Lindstrom. The generalized Bregman distance. *SIAM Journal on Optimization*, 31(1):404–424, 2021.
- [15] R.S. Burachik and A.N. Iusem. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM Journal on Optimization*, 8(1):197–216, 1998.
- [16] R.S. Burachik and S. Scheimberg. A proximal point method for the variational inequality problem in Banach spaces. *SIAM Journal on Control and Optimization*, 39(5):1633–1649, 2001.
- [17] D. Butnariu and A.N. Iusem. On a proximal point method for convex optimization in Banach spaces. *Numerical Functional Analysis and Optimization*, 18(7–8):723–744, 1997.

- [18] D. Butnariu, S. Reich, and A. Zaslavski. Convergence to fixed points of inexact orbits of Bregman-monotone and nonexpansive operators in Banach spaces. In H.F. Nathansky, B.G. de Buen, K. Goebel, W.A. Kirk, and B. Sims, editors, *Fixed Point Theory and Applications*, pages 11–32. Yokohama Publishers, Yokohama, 2006.
- [19] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function. *Journal of Optimization Theory and Applications*, 162(1):107–132, 2014.
- [20] P.L. Combettes. Quasi-fejérian analysis of some optimization algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms for Feasibility and Optimization*, pages 115–152. Elsevier, New York, 2001.
- [21] P.L. Combettes. Fejér monotonicity in convex optimization. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1016–1024. Springer, New York, 2009.
- [22] P.L. Combettes and B.C. Vũ. Variable metric quasi-Fejér monotonicity. *Nonlinear Analysis: Theory, Methods & Applications*, 78:17–31, 2013.
- [23] P.L. Combettes and B.C. Vũ. Variable metric forward–backward splitting with applications to monotone inclusions in duality. *Optimization*, 63(9):1289–1318, 2014.
- [24] Q.-L. Dong, Y.J. Cho, S. He, P.M. Pardalos, and T.M. Rassias. *The Krasnosel’skiĭ-Mann Iterative Method*. Springer Briefs in Optimization. Springer Cham, 2022.
- [25] J. Eckstein. Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.
- [26] L. Fejér. Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen. *Mathematische Annalen*, 85:41–48, 1922.
- [27] T. Figiel. On the moduli of convexity and smoothness. *Studia Mathematica*, 56(2):121–155, 1976.
- [28] J. García-Falset, E. Llorens-Fuster, and T. Suzuki. Fixed point theory for a class of generalized nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 375:185–195, 2011.
- [29] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49:431–446, 2008.
- [30] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360:2615–2660, 2008.
- [31] H. Iiduka and I. Yamada. A subgradient-type method for the equilibrium problem over the fixed point set and its applications. *Optimization*, 58(2):251–261, 2009.
- [32] S. Ishikawa. Fixed points by a new iteration methods. *Proceedings of the American Mathematical Society*, 44:147–150, 1974.
- [33] F. Iutzeler and J.M. Hendrickx. A generic online acceleration scheme for optimization algorithms via relaxation and inertia. *Optimization Methods and Software*, 34(2):383–405, 2019.
- [34] F. Iutzeler and J. Malick. On the Proximal Gradient Algorithm with Alternated Inertia. *Journal of Optimization Theory and Applications*, 176:688–710, 2018.
- [35] O. Kada, T. Suzuki, and W. Takahashi. Nonconvex minimization theorems and fixed point theorems in complete metric spaces. *Mathematica Japonica*, 44(2):381–391, 1996.
- [36] S. Kamimura, F. Kohsaka, and W. Takahashi. Weak and strong convergence theorems for maximal monotone operators in a Banach space. *Set-Valued Analysis*, 12:417–429, 2004.
- [37] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.
- [38] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
- [39] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proc. ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.
- [40] U. Kohlenbach. Quantitative analysis of a Halpern-type Proximal Point Algorithm for accretive operators in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 21(9):2125–2138, 2020.
- [41] U. Kohlenbach. Quantitative results on the Proximal Point Algorithm in uniformly convex Banach spaces. *Journal of Convex Analysis*, 28(1):11–18, 2021.
- [42] U. Kohlenbach. On the Proximal Point Algorithm and its Halpern-type variant for generalized monotone operators in Hilbert space. *Optimization Letters*, 16:611–621, 2022.
- [43] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative Results on Fejér Monotone Sequences. *Communications in Contemporary Mathematics*, 20(2), 2018.
- [44] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Quantitative Asymptotic Regularity Results for the Composition of Two Mappings. *Optimization*, 66:1291–1299, 2017.
- [45] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequences. *Israel Journal of Mathematics*, 232:261 – 297, 2019.
- [46] U. Kohlenbach and P. Oliva. Proof Mining: A Systematic Way of Analysing Proofs in Mathematics. *Proceedings of the Steklov Institute of Mathematics*, 242:136–164, 2003.
- [47] F. Kohsaka and W. Takahashi. Strong convergence of an iterative sequence for maximal monotone operators in a Banach space. *Abstract and Applied Analysis*, 3:239–249, 2004.
- [48] G.M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i matematicheskie metody*, 12(4):747–756, 1976.
- [49] G. Kreisel. On the Interpretation of Non-Finitist Proofs—Part I. *The Journal of Symbolic Logic*, 16(4):241–267, 1951.
- [50] G. Kreisel. On the Interpretation of Non-Finitist Proofs—Part II. Interpretation of Number Theory. Applications. *The Journal of Symbolic Logic*, 17(1):43–58, 1952.
- [51] L. Leuştean and A. Sipoş. An application of proof mining to the proximal point algorithm in CAT(0) spaces. In A. Bellow, C. Calude, and T. Zamfirescu, editors, *Mathematics almost everywhere. In memory of Solomon Marcus*, pages 153–167. World Scientific, Hackensack, NJ, 2018.

- [52] L. Leuştean and A. Sipoş. Effective strong convergence of the proximal point algorithm in CAT(0) spaces. *Journal of Nonlinear and Variational Analysis*, 2:219–228, 2018.
- [53] H. Li, Y. Wu, and F. Wang. Convergence analysis for solving equilibrium problems and split feasibility problems in Hilbert spaces. *Optimization*, 72(7):1863–1898, 2023.
- [54] W.R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.
- [55] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4:154–158, 1970.
- [56] S.Y. Matsushita and W. Takahashi. Weak and strong convergence theorems for relatively nonexpansive mappings in Banach spaces. *Fixed Point Theory and Applications*, 1:1–11, 2004.
- [57] T. Motzkin and I. Schoenberg. The Relaxation Method for Linear Inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.
- [58] A. Moudafi. On critical points of the difference of two maximal monotone operators. *Afrika Matematika*, 26:457–463, 2015.
- [59] Z. Mu and Y. Peng. A note on the inertial proximal point method. *Statistics, Optimization and Information Computing*, 3(3):241–248, 2015.
- [60] E. Neumann. Computational Problems in Metric Fixed Point Theory and their Weihrauch Degrees. *Logical Methods in Computer Science*, 11(4), 2015.
- [61] Q. Van Nguyen. Forward-backward splitting with Bregman distances. *Vietnam Journal of Mathematics*, 45(3):519–539, 2017.
- [62] N. Pischke. On Korpelevich's Extragradient Algorithm. *Zeitschrift für Analysis und ihre Anwendungen*, 42(1/2):117–129, 2023.
- [63] N. Pischke. Quantitative Results on Algorithms for Zeros of Differences of Monotone Operators in Hilbert Space. *Journal of Convex Analysis*, 30(1):295–315, 2023.
- [64] N. Pischke. Duality, Fréchet differentiability and Bregman distances in hyperbolic spaces. *Israel Journal of Mathematics*, 2024. to appear, manuscript available at <https://sites.google.com/view/nicholaspischke/notes-and-papers>.
- [65] N. Pischke and U. Kohlenbach. Quantitative Analysis of a Subgradient-Type Method for Equilibrium Problems. *Numerical Algorithms*, 90(1):197–219, 2022.
- [66] N. Pischke and U. Kohlenbach. Effective rates for iterations involving Bregman strongly nonexpansive operators. 2023. manuscript available at <https://www2.mathematik.tu-darmstadt.de/~kohlenbach/>.
- [67] T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal of Control and Optimization*, 14:877–898, 1976.
- [68] F. Schöpfer, A.K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse problems*, 22:311–329, 2006.
- [69] Y. Shehu, Q.-L. Dong, L. Liu, and J.-C. Yao. Alternated inertial subgradient extragradient method for equilibrium problems. *TOP*, 31:1–30, 2023.
- [70] Y. Shehu and A. Gibali. New inertial relaxed method for solving split feasibilities. *Optimization Letters*, 15:2109–2126, 2021.
- [71] Y. Shehu and O.S. Iyiola. Projection methods with alternating inertial steps for variational inequalities: Weak and linear convergence. *Applied Numerical Mathematics*, 157:315–337, 2020.
- [72] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *Journal of Symbolic Logic*, 14:145–158, 1949.
- [73] W. Takahashi. *Nonlinear Functional Analysis*. Yokohama Publishers, Yokohama, 2000.
- [74] T. Tao. Norm Convergence of Multiple Ergodic Averages for Commuting Transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.
- [75] T. Tao. *Structure and Randomness: Pages from Year One of a Mathematical Blog*, chapter Soft analysis, hard analysis, and the finite convergence principle. American Mathematical Society, Providence, RI, 2008.
- [76] V.V. Vasin and I.I. Eremin. *Operators and Iterative Processes of Fejér Type*. Springer Briefs in Optimization. Walter de Gruyter, Berlin, 2009.
- [77] H.K. Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66(1):240–256, 2002.