

Quantitative Results on Algorithms for Zeros of Differences of Monotone Operators in Hilbert Space

Nicholas Pischke

*Department of Mathematics, Technische Universität Darmstadt, Germany
pischkenicholas@gmail.com*

Received: May 5, 2022

Accepted: August 8, 2022

We provide quantitative information in the form of a rate of metastability in the sense of T. Tao and (under a metric regularity assumption) a rate of convergence for an algorithm approximating zeros of differences of maximally monotone operators due to A. Moudafi by using techniques from ‘proof mining’, a subdiscipline of mathematical logic. For the rate of convergence, we provide an abstract and general result on the construction of rates of convergence for quasi-Fejér monotone sequences with metric regularity assumptions, generalizing previous results for Fejér monotone sequences due to U. Kohlenbach, G. López-Acedo and A. Nicolae.

Keywords: Maximally monotone operators, Zeros of set-valued operators, DC programming, Fejér monotonicity, proof mining.

2010 Mathematics Subject Classification: 47H05, 47J25, 03F10, 47H09.

1. Introduction

In [17], Moudafi introduced an algorithm for approximating critical points of differences of two maximally monotone operators by generalizing the method of proximity operators in DC programming (see below). Specifically, given two maximally monotone operators T, S on a real Hilbert space X , one wants to find points

$$x^* \in \Gamma := \{x^* \in X \mid T(x^*) \cap S(x^*) \neq \emptyset\}.$$

If T and S are both subdifferentials of convex functions f, g , respectively, this covers the prominent case of *DC programming*, i.e. mathematical programming for *differences of convex functions* (see e.g. [5, 26]) and the solution will be a critical point of $f - g$.

The algorithm given in [17] is motivated by noting that $\text{gra}T_\lambda \rightarrow \text{gra}T$ for $\lambda \rightarrow 0$ where $T_\lambda(x) = (x - J_\lambda^T x)/\lambda$ is the Yosida approximate of T , $J_\lambda^T = (Id + \lambda T)^{-1}$ is the resolvent of λT and $\text{gra}T$ is the graph of T , i.e. $\text{gra}T = \{(x, y) \mid y \in Tx\}$. This leads to regularizing the above problem to finding points x_λ with

$$T_\lambda(x_\lambda) \in S(x_\lambda)$$

for $\lambda \rightarrow 0$. This inclusion can be equivalently phrased as a fixed point problem

$$x_\lambda = J_\mu^S(x_\lambda + \mu T_\lambda x_\lambda)$$

with parameter $\mu > 0$.

This then leads to the iteration scheme

$$x_{n+1} = J_{\mu_n}^S(x_n + \mu_n T_{\lambda_n} x_n)$$

given initial data x_0 and parameters $\mu_n, \lambda_n > 0$ which, under suitable assumptions on the parameters, can be shown to be convergent:

Theorem 1.1. ([17]) *Let T, S be two maximally monotone operators on a finite dimensional Hilbert space X such that $\Gamma \neq \emptyset$, T is bounded on bounded sets and $\text{dom} S \subseteq \text{dom} T$ as well as*

- (1) $\lim_{n \rightarrow \infty} \lambda_n = 0$,
- (2) $\sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n} < \infty$,
- (3) $\lim_{n \rightarrow \infty} \|x_n - x_{n+1}\| / \mu_n = 0$,

for $\lambda_n, \mu_n > 0$. Then (x_n) converges to a point $x^* \in \Gamma$.

Moudafi's results does not give any quantitative information on the convergence. However, by inspection of the proof given in [17], it becomes apparent that it relies on a standard argument via establishing (quasi-)Fejér monotonicity (see [3]) and then inferring convergence from that.¹ In terms of quantitative results, this opens the door for applying the recent results of Kohlenbach, Leuştean and Nicolae [10] as well as of Kohlenbach, López-Acedo and Nicolae [12] on the finitary content of convergence of (quasi-)Fejér monotone sequences, which has been successfully applied in many other contexts of nonlinear analysis, in particular for the asymptotic regularity of compositions of two mappings [11], the proximal point algorithm in uniformly convex Banach spaces [9] and subgradient methods for equilibrium problems [22].

The results of the papers [10, 12] were obtained via the general methodological approach of 'proof mining', a subdiscipline of mathematical logic which aims at the extraction of quantitative information from *prima facie* nonconstructive proofs by logical transformations (see [7] for a book treatment and [8] for a recent survey). This approach has also been instrumental for obtaining the present results.

In terms of quantitative information, even for computable Fejér monotone sequences of real numbers, in general, there exists no computable rate of convergence which follows from fundamental results in recursion theory (see [10, 18]).²

However, in very general situations, one can extract effective rates of so-called metastability from non-effective proofs of convergence which are, moreover, highly uniform. This notion of metastability originates from a (noneffectively) equivalent reformulation of the Cauchy property in some metric space (X, d)

$$\forall k \in \mathbb{N} \exists n \in \mathbb{N} \forall i, j \geq n \left(d(x_i, x_j) < \frac{1}{k+1} \right)$$

$$\text{into} \quad \forall k \in \mathbb{N} \forall g \in \mathbb{N}^{\mathbb{N}} \exists n \in \mathbb{N} \forall i, j \in [n; n + g(n)] \left(d(x_i, x_j) < \frac{1}{k+1} \right)$$

where $[n; n + m] := \{n + i \mid i \in \mathbb{N} \wedge 0 \leq i \leq m\}$. From a logical perspective, this reformulation can be recognized as the so-called Herbrand normal form of (a slightly

¹ Although Fejér monotonicity is not mentioned explicitly in [17].

² These results extend the phenomena of 'arbitrary slow convergence' from optimization.

tweaked version of) the above Cauchy property which is in particular of the general form $\forall\exists$ (considering the leading two universal quantifiers as one and disregarding the last universal quantifier as it is bounded) and for statements of the above form, the logical metatheorems of proof mining guarantee the extractability of a rate of metastability, that is a (highly uniform and effective) bound on ‘ $\exists n \in \mathbb{N}$ ’ in the above reformulation (see [7]). This notion has also been recognized as an important finitary version of the Cauchy property from a non-logical perspective by Tao (see e.g. [24, 25]) who actually coined the term metastability.

In this paper, we provide a further case study to illustrate how the abstract approach from [10, 12] can be used in a particular scenario by obtaining an explicit quantitative version of Theorem 1.1 in the form of a fully effective and highly uniform rate of metastability. Moreover, in the latter parts of the paper, we even give a rate of convergence, modulo an additional metric regularity assumption in the sense of [12]. For this, we extend the main result from [12] to the case of quasi-Fejér monotone sequences. The quantitative analysis in particular relies on a mild form of uniform continuity of a set-valued operator T similar to that introduced in [14] and we expect that this quantitative notion and its use, together with the whole approach, detailed here will provide a guideline for future analyses of other convergence results from monotone operator theory, especially works dealing with differences of monotone operators like [1, 16, 19, 23]. In fact, the analysis presented here immediately generalizes to the extensions for Moudafi’s result considered in [23] where one similarly obtains a simple rate of metastability and even a rate of convergence under a metric regularity assumption but we omit any details regarding this.

2. Quasi-Fejér monotonicity, uniform continuity and rates of metastability

As mentioned in the introduction, the proof of Theorem 1.1 given in [17], and with that the following quantitative analysis, relies on the notion of quasi-Fejér monotonicity (see again [3]) which we want to briefly recall. For this, we actually rely on the following generalized version introduced in [10].

Definition 2.1. ([10]) Let $G : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be functions where

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

for any sequence (a_n) from $\mathbb{R}_+$, let (X, d) be a metric space and $F \subseteq X$ be nonempty. A sequence (x_n) in X is called *quasi- (G, H) -Fejér monotone with respect to F* , if

$$\forall n, m \in \mathbb{N} \forall p \in F \left(H(d(x_{n+m}, p)) \leq G(d(x_n, p)) + \sum_{i=n}^{n+m-1} \varepsilon_i \right),$$

where $(\varepsilon_i) \subseteq \mathbb{R}_+$ is such that $\sum_i \varepsilon_i < \infty$. □

For formulating the quantitative results, we pass from quasi-Fejér monotonicity to *uniform* quasi-Fejér monotonicity with an accompanying modulus in the sense of [10]. To do this, we assume a corresponding stratification of F by sets AF_k s.t.

$$AF_k \supseteq AF_{k+1} \text{ and } F = \bigcap_{k \in \mathbb{N}} AF_k.$$

Intuitively, the sets AF_k are meant to represent the set of k -good approximations of the set F and can take many forms in an actual application.

Definition 2.2. ([10]) Let G, H be as before. Then (x_n) is called *uniformly quasi- (G, H) -Fejér monotone with respect to F (and (AF_k))* if for all $r, n, m \in \mathbb{N}$:

$$\exists k \in \mathbb{N} \forall p \in AF_k \forall l \leq m \left(H(d(x_{n+l}, p)) < G(d(x_n, p)) + \sum_{i=n}^{n+l-1} \varepsilon_i + \frac{1}{r+1} \right).$$

Any function $\chi(n, m, r)$ producing an upper bound on such a $k \in \mathbb{N}$ is called a *modulus of uniform quasi- (G, H) -Fejér monotonicity for (x_n)* . $\square$

As a second ingredient, we need quantitative information on how the sequence (x_n) approaches the set F w.r.t. the stratification AF_k , in the sense of the following definition:

Definition 2.3. ([10]) (x_n) has the *lim inf-property w.r.t. F (and (AF_k))* if $\forall k, n \in \mathbb{N} \exists N \geq n (x_N \in AF_k)$. A bound $\Phi(k, n)$ on N , which is monotone in k and n , is called a *lim inf-bound for (x_n)* . $\square$

Using the proof mining macro established in [10], one can then combine these moduli, together with some further (minor) quantitative assumptions on the surrounding data, to a rate of metastability of the sequence. As discussed in [10], these moduli are guaranteed to exist (in very general situation) by logical metatheorems and can often be obtained by a separate application of corresponding logical bound extraction results of proof mining and it is the extraction of these moduli from the proof given in [17] which we detail (without any reference to logic) in the following.

For that, we will in particular rely on a certain notion of uniform continuity for a set-valued operator T which generalizes the usual notion of uniform continuity (as, e.g., stipulated in [15]) for a set-valued operator T

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in \text{dom} T (\|x - y\| \leq \delta \rightarrow H(Tx, Ty) \leq \varepsilon)$$

where H is the Hausdorff-metric.³ Motivated by logical considerations (see Remark 2.13), [14] introduced an ‘approximate version’ of the Hausdorff metric in the form of a *Hausdorff-like predicate H^** defined via

$$H^*[P, Q, \varepsilon] := \forall p \in P \exists q \in Q (\|p - q\| \leq \varepsilon).$$

One can then stipulate uniform continuity w.r.t. that predicate H^* by requiring

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in X (\|x - y\| \leq \delta \rightarrow H^*[Tx, Ty, \varepsilon]).$$

We say that $\varpi : \mathbb{N} \rightarrow \mathbb{N}$ is a *modulus of uniform continuity for T w.r.t. H^** if

$$\forall k \in \mathbb{N} \forall x, y \in X \left(\|x - y\| \leq \frac{1}{\varpi(k) + 1} \rightarrow H^*[Tx, Ty, \frac{1}{k+1}] \right)$$

Our analysis will in the following rely on such a modulus of uniform continuity for T w.r.t. H^* (see again Remark 2.13). For convenience, we will assume that ϖ is monotone increasing.

³ We have used H for another object before but the context will make it clear whether the Hausdorff-metric is meant.

Now, assume that $\text{dom}S \subseteq \text{dom}T$ and let $L \geq \text{diam}\{x_n \mid n \in \mathbb{N}\}$ for a concrete sequence (x_n) of the algorithm. Throughout, we will actually work over the compact space $X_0 = \overline{B}(x_0; L) \cap \overline{\text{dom}S}$ and all sets and moduli are to be understood as being relativized to this set. We first define appropriate instantiations Γ_k of the abstract approximations AF_k discussed before in the context of Moudafi's algorithm by setting

$$\Gamma_k := \left\{ x^* \in X_0 \left| \begin{array}{l} \exists y^* \left(\|y^*\| - \|T^\circ x^*\| \leq \frac{1}{k+1} \wedge H^* \left[y^*, Tx^*, \frac{1}{k+1} \right] \right) \\ \wedge \forall i \leq k \left(\|x^* - J_{\mu_i}^S(x^* + \mu_i y^*)\| \leq \frac{1}{k+1} \right) \end{array} \right. \right\}$$

where H^* is the previously discussed Hausdorff-like predicate (where we write y^* for the singleton $\{y^*\}$) and $T^\circ x = P_{Tx}0$ is the element of minimal norm in Tx (see, e.g., [2]). We write $\Gamma_k(x^*)$ for the set of all such y^* realizing the existential quantifier in the above definition with parameter x^* .

The motivation for this particular stratification of the set of solutions Γ very much follows the reasoning given by Moudafi in [17] for the algorithm as discussed in the introduction: $T(x^*) \cap S(x^*)$ is regularized to $T_\lambda(x_\lambda) \in Sx_\lambda$ for $\lambda \rightarrow 0$, which is equivalent to solving the fixed-point equation $x_\lambda = J_\mu^S(x_\lambda + \mu T_\lambda x_\lambda)$. To formalize what it means for x^* to be a k -good approximation of a solution in Γ , we replace $T_\lambda(x_\lambda)$, itself an approximation for a point in the intersection, by a generic point y^* which is supposed to be a k -good approximation of $T_\lambda(x^*)$ for suitable λ , formalized via being close to Tx^* in the sense of the Hausdorff-like predicate. In the same way, the fixed point condition $x_\lambda = J_\mu^S(x_\lambda + \mu T_\lambda x_\lambda)$ is relativized to an approximate fixed point condition where the parameter μ is replaced by the sequence μ_i from the algorithm. In that way, the fixed-point perspective of this algorithm is essential for the following quantitative analysis as it lends itself to useful approximate versions.

We begin by showing that the Γ_k are appropriate approximate versions of Γ . For that, we actually show two things. First, we have $\bigcap_k \Gamma_k \subseteq \Gamma$, i.e. an arbitrarily good approximation in the sense of the Γ_k is actually a solution. Second, the Γ_k are good approximate sets in the sense that $\bigcap_k \Gamma_k$ is uniformly closed w.r.t. the Γ_k in the following sense:

Definition 2.4. ([10]) F is called *uniformly closed* (w.r.t. AF_k) with moduli $\delta_F, \omega_F : \mathbb{N} \rightarrow \mathbb{N}$ if

$$\forall k \in \mathbb{N} \forall p, q \in X \left(q \in AF_{\delta_F(k)} \wedge d(p, q) \leq \frac{1}{\omega_F(k) + 1} \rightarrow p \in AF_k \right). \quad \square$$

Before we can state the corresponding result on uniform closedness, we need the following lemma which derives a modulus of continuity for T° from the modulus of uniform continuity ϖ for T .

Lemma 2.5. (1) For all $k \in \mathbb{N}$, $x, z \in X$, if $z \in Tx$ and $\langle T^\circ x - z, -z \rangle \leq \frac{1}{(k+1)^2}$,

$$\text{then } \|T^\circ x - z\| \leq \frac{1}{k+1}.$$

(2) For all $k \in \mathbb{N}$ and $x, z \in X$, if $z \in Tx$ and $\|z\|^2 - \|T^\circ x\|^2 \leq \frac{1}{(k+1)^2}$, then

$$\|T^\circ x - z\| \leq \frac{1}{k+1}.$$

- (3) Let $x, x' \in \text{dom}T$ and $B \geq \|T^\circ x'\|$ with $B \in \mathbb{N}^*$ and let ϖ be a modulus of uniform continuity for T . Then, $\varpi'(k) = \varpi(Bk^2 + 2Bk + B - 1)$ satisfies

$$\|x - x'\| \leq \frac{1}{\varpi'(k) + 1} \rightarrow \|T^\circ x - T^\circ x'\| \leq \frac{1}{k + 1} \quad \text{for all } x. \quad (\dagger)$$

Proof. Note that as $T^\circ x = P_{Tx}(0)$, Theorem 3.16 of [2] yields $\langle y - T^\circ x, -T^\circ x \rangle \leq 0$ for all x, y with $y \in Tx$. (1) Let $z \in Tx$ and $\langle T^\circ x - z, -z \rangle \leq \frac{1}{(k+1)^2}$. By $(\dagger)$, we have $\langle z - T^\circ x, -T^\circ x \rangle \leq 0$. Thus

$$\frac{1}{(k+1)^2} \geq \langle T^\circ x - z, -z \rangle + \langle z - T^\circ x, -T^\circ x \rangle = \|T^\circ x - z\|^2$$

which implies $\|T^\circ x - z\| \leq \frac{1}{k+1}$.

- (2) We skip the proof as it is an easy consequence of (1).

- (3) Let $\|x - x'\| \leq \frac{1}{\varpi'(k)+1}$. By definition of ϖ' , we have

$$\exists y' \in Tx' \left(\|T^\circ x - y'\| \leq \frac{1}{Bk^2 + 2Bk + B} \right).$$

Now, using $(\dagger)$ and the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} \langle T^\circ x - T^\circ x', -T^\circ x' \rangle &\leq \langle T^\circ x - y', -T^\circ x' \rangle + \langle y' - T^\circ x', -T^\circ x' \rangle \\ &= \langle T^\circ x - y', -T^\circ x' \rangle \leq \|T^\circ x - y'\| \|T^\circ x'\| \leq \frac{1}{(k+1)^2}. \end{aligned}$$

Item (1) now yields $\|T^\circ x - T^\circ x'\| \leq \frac{1}{k+1}$. □

Lemma 2.6. Let ϖ be a modulus of uniform continuity for T w.r.t. H^* and let $M \in \mathbb{N}^*$ satisfy $M \geq \|T^\circ x^*\|$ for all $x^* \in X_0$. Then $\bigcap_k \Gamma_k$ is uniformly closed w.r.t. Γ_k with moduli $\delta(k) = 2k + 1$ and $\omega(k) = \max\{4k + 3, \varpi(4M(k+1)^2 - 1)\}$ and, further,

$$\Gamma \supseteq \bigcap_k \Gamma_k.$$

Proof. Suppose $q \in \Gamma_{\delta(k)}$ and let $a \in \Gamma_k(q)$, i.e., for all $i \leq \delta(k)$,

$$|\|a\| - \|T^\circ q\|| \leq \frac{1}{\delta(k)+1}, \quad H^* \left[a, Tq, \frac{1}{\delta(k)+1} \right], \quad \text{and} \quad \|q - J_{\mu_i}^S(q + \mu_i a)\| \leq \frac{1}{\delta(k)+1}.$$

Let $p \in X_0$ be such that $\|p - q\| \leq \frac{1}{\omega(k)+1}$. Then in particular $\omega(k) \leq \varpi(2k+1)$ as ϖ is monotone increasing and therefore

$$\|p - q\| \leq \frac{1}{\omega(k) + 1} \leq \frac{1}{\varpi(2k+1) + 1}$$

and therefore $H^* \left[a, Tp, \frac{1}{\delta(k)+1} + \frac{1}{2(k+1)} \right]$, and so $H^* \left[a, Tp, \frac{1}{k+1} \right]$.

Further, as $\delta(k) \geq k$, we have

$$\begin{aligned} \|p - J_{\mu_i}^S(p + \mu_i a)\| &\leq \|p - q\| + \|q - J_{\mu_i}^S(q + \mu_i a)\| + \|J_{\mu_i}^S(q + \mu_i a) - J_{\mu_i}^S(p + \mu_i a)\| \\ &\leq 2\|p - q\| + \|q - J_{\mu_i}^S(q + \mu_i a)\| \leq \frac{2}{\omega(k)+1} + \frac{1}{\delta(k)+1} \leq \frac{2}{4(k+1)} + \frac{1}{2(k+1)} = \frac{1}{k+1} \end{aligned}$$

for all $i \leq k$ using nonexpansivity of $J_{\mu_i}^S$.

Finally we have

$$\begin{aligned} |\|a\| - \|T^\circ p\|| &\leq |\|a\| - \|T^\circ q\|| + |\|T^\circ p\| - \|T^\circ q\|| \\ &\leq \frac{1}{\delta(k)+1} + \|T^\circ p - T^\circ q\| \leq \frac{1}{2(k+1)} + \frac{1}{2(k+1)} = \frac{1}{k+1} \end{aligned}$$

using the modulus of uniform continuity for T° derived from ϖ in Proposition 2.5 (where one has to note, in particular, the definition of M). Combined, we have $a \in \Gamma_k(p)$ and thus $p \in \Gamma_k$ and Γ is therefore uniformly closed w.r.t. Γ_k (over X_0).

For the second claim, let $x^* \in \bigcap_k \Gamma_k$, i.e. for any k there are $y_k^* \in \Gamma_k(x^*)$ such that $|\|y_k^*\| - \|T^\circ x^*\|| \leq \frac{1}{k+1}$, $H^*[y_k^*, Tx^*, 1/(k+1)]$ and $\|x^* - J_{\mu_i}^S(x^* + \mu_i y_k^*)\| \leq \frac{1}{k+1}$ for all $i \leq k$.

Now, item (1) yields $\|y_k^*\| \rightarrow \|T^\circ x^*\|$ for $k \rightarrow \infty$. Item (2) yields

$$\forall k \exists z_k^* \in Tx^* \left(\|y_k^* - z_k^*\| \leq \frac{1}{k+1} \right)$$

and thus in particular $|\|z_k^*\| - \|y_k^*\|| \leq \frac{1}{k+1}$. Thus $\|z_k^*\| \rightarrow \|T^\circ x^*\|$ for $k \rightarrow \infty$. Uniqueness of $T^\circ x^*$ as the element of minimal norm, Lemma 2.5, (2), now yields $z_k^* \rightarrow T^\circ x^*$ for $k \rightarrow \infty$ (actually in a quantitative way). Thus, in particular also $y_k^* \rightarrow T^\circ x^*$ for $k \rightarrow \infty$. This yields

$$\|x^* - J_{\mu_0}^S(x^* + \mu_0 y_k^*)\| \rightarrow \|x^* - J_{\mu_0}^S(x^* + \mu_0 T^\circ x^*)\|$$

for $k \rightarrow \infty$ while item (3) yields $\|x^* - J_{\mu_0}^S(x^* + \mu_0 y_k^*)\| \rightarrow 0$. Thus

$$\|x^* - J_{\mu_0}^S(x^* + \mu_0 T^\circ x^*)\| = 0$$

and therefore $x^* = J_{\mu_0}^S(x^* + \mu_0 T^\circ x^*)$, i.e.

$$T^\circ x^* = \mu_0^{-1}(x^* + \mu_0 T^\circ x^* - x^*) \in Sx^*.$$

Now, as also $T^\circ x^* \in Tx^*$, we have $Sx^* \cap Tx^* \neq \emptyset$, i.e. $x^* \in \Gamma$. □

The next lemma gives a preliminary result for the extraction of a modulus of uniform quasi-Fejér monotonicity, obtained by an extraction from the proof of quasi-Fejér monotonicity given by Moudafi in [17].

Lemma 2.7. Let $n, r \in \mathbb{N}$ and $l \in \mathbb{N}^*$ be given and let $x^* \in X_0$ and y^* be such that

$$\left\| x^* - J_{\mu_{n+k}}^S(x^* + \mu_{n+k}y^*) \right\| \leq \frac{1}{r+1}$$

for all $k \in [0; l-1]$. Then

$$\begin{aligned} \|x_{n+l} - x^*\| &\leq \prod_{k=0}^{l-1} \left(1 + \frac{\mu_{n+k}}{\lambda_{n+k}}\right) \|x_n - x^*\| + (\|T^\circ x^*\| + \|y^*\|) \sum_{k=0}^{l-1} \mu_{n+k} \prod_{j=k+1}^{l-1} \left(1 + \frac{\mu_{n+j}}{\lambda_{n+j}}\right) \\ &\quad + \sum_{k=1}^l \frac{1}{r+1} \prod_{j=k}^{l-1} \left(1 + \frac{\mu_{n+j}}{\lambda_{n+j}}\right). \end{aligned}$$

Proof. Given n and r as well as x^* and y^* , we get

$$\begin{aligned} \|x_{n+1} - x^*\| &= \|J_{\mu_n}^S(x_n + \mu_n T_{\lambda_n} x_n) - x^*\| \\ &\leq \|J_{\mu_n}^S(x_n + \mu_n T_{\lambda_n} x_n) - J_{\mu_n}^S(x^* + \mu_n y^*)\| + \|J_{\mu_n}^S(x^* + \mu_n y^*) - x^*\| \\ &\leq \|x_n + \mu_n T_{\lambda_n} x_n - x^* - \mu_n y^*\| + \frac{1}{r+1} \\ &\leq \|x_n - x^*\| + \mu_n \|T_{\lambda_n} x_n - T_{\lambda_n} x^*\| + \mu_n \|T_{\lambda_n} x^* - y^*\| + \frac{1}{r+1} \\ &\leq \|x_n - x^*\| + \frac{\mu_n}{\lambda_n} \|x_n - x^*\| + \mu_n (\|T^\circ x^*\| + \|y^*\|) + \frac{1}{r+1} \\ &= \left(1 + \frac{\mu_n}{\lambda_n}\right) \|x_n - x^*\| + \mu_n (\|T^\circ x^*\| + \|y^*\|) + \frac{1}{r+1}. \end{aligned}$$

where we have used λ_n^{-1} -Lipschitz continuity of T_{λ_n} (Corollary 23.11 in [2]). This generalizes to the claim by induction on l . $\square$

This immediately gives a modulus of uniform quasi-Fejér monotonicity when we assume certain bounds on the objects involved. For this, we define a *bounded subtraction* $\div$ by $n \div m := \max\{0, n - m\}$.

Lemma 2.8. Let $M \geq \|T^\circ x^*\|$ for any $x^* \in X_0$. Further, let $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$ and assume that $\lambda_n \leq B$ for all n . Then $\sum_{n=0}^{\infty} \mu_n < \infty$ and (x_n) is uniformly quasi- $(e^A \text{id}_{\mathbb{R}_+}, \text{id}_{\mathbb{R}_+})$ -Fejér monotone w.r.t. Γ_k with modulus χ , that is for all $r, n, m \in \mathbb{N}$:

$$\forall x^* \in \Gamma_k \forall l \leq m \left(\|x_{n+l} - x^*\| < e^A \|x_n - x^*\| + (2M + 1)e^A \sum_{i=n}^{n+l-1} \mu_i + \frac{1}{r+1} \right)$$

where $k = \chi(r, n, m) := \max\{n + m \div 1, \lceil (r+1) \cdot m \cdot e^A \rceil\}$.

Proof. Assume $m \geq 1$ without loss of generality.

First, $\lambda_n \leq B$ together with $\sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n} \leq A < \infty$ gives $\sum_{n=0}^{\infty} \mu_n < \infty$.

Let $x^* \in \Gamma_k$ be arbitrary and $y^* \in \Gamma_k(x^*)$. Then

$$\left\| x^* - J_{\mu_i}^S(x^* + \mu_i y^*) \right\| \leq \frac{1}{k+1} \leq \frac{1}{\lceil (r+1) \cdot m \cdot e^A \rceil + 1}$$

for all $i \leq n + m - 1$, since $k \geq n + m \div 1$ and $m \geq 1$, and thus in particular for all $i \in [n; n + l - 1]$ and any $l \leq m$.

By Lemma 2.7, using $|\|y^*\| - \|T^\circ x^*\|| \leq 1$, i.e. $\|y^*\| \leq \|T^\circ x^*\| + 1 \leq M + 1$, we get

$$\begin{aligned} \|x_{n+l} - x^*\| &\leq e^A \|x_n - x^*\| + (2M + 1)e^A \sum_{i=n}^{n+l-1} \mu_i + e^A \sum_{k=1}^l \frac{1}{\lceil (r+1) \cdot m \cdot e^A \rceil + 1} \\ &< e^A \|x_n - x^*\| + (2M + 1)e^A \sum_{i=n}^{n+l-1} \mu_i + e^A \frac{l}{(r+1) \cdot m \cdot e^A} \\ &\leq e^A \|x_n - x^*\| + (2M + 1)e^A \sum_{i=n}^{n+l-1} \mu_i + \frac{1}{r+1}. \end{aligned} \quad \square$$

We move on to the lim-inf-property. For that, we first show a general inequality in the spirit of the proximal point algorithm (see [2]) which requires the following result:

Lemma 2.9. ([2], Proposition 23.31, (i)) *Let A be maximally monotone, $\gamma, \lambda > 0$ and x a point. Then $J_\gamma^A x = J_{\lambda\gamma}^A(\lambda x + (1 - \lambda)J_\gamma^A x)$.*

Lemma 2.10. *For any $n, i \in \mathbb{N}$, we have*

$$\|x_n - J_{\mu_i}^S(x_n + \mu_i T_{\lambda_n} x_n)\| \leq \|x_n - x_{n+1}\| + |\mu_n - \mu_i| \frac{\|x_n - x_{n+1}\|}{\mu_n}.$$

Proof. First, we have

$$\begin{aligned} &\|x_n - J_{\mu_i}^S(x_n + \mu_i T_{\lambda_n} x_n)\| \\ &\leq \|x_n - x_{n+1}\| + \|J_{\mu_n}^S(x_n + \mu_n T_{\lambda_n} x_n) - J_{\mu_i}^S(x_n + \mu_i T_{\lambda_n} x_n)\| \\ &\leq \|x_n - x_{n+1}\| + \|J_{\mu_n}^S x_n^n - J_{\mu_i}^S x_n^i\| \end{aligned}$$

where we write $x_n^j = x_n + \mu_j T_{\lambda_n} x_n$ for any j for simplicity. Then, we have

$$\begin{aligned} &\|J_{\mu_n}^S x_n^n - J_{\mu_i}^S x_n^i\| \\ &= \left\| J_{\mu_i}^S \left(\frac{\mu_i}{\mu_n} x_n^n + \left(1 - \frac{\mu_i}{\mu_n}\right) J_{\mu_n}^S x_n^n \right) - J_{\mu_i}^S x_n^i \right\| \quad (\text{by Lemma 2.9}) \\ &\leq \left\| \frac{\mu_i}{\mu_n} x_n^n + \left(1 - \frac{\mu_i}{\mu_n}\right) J_{\mu_n}^S x_n^n - x_n^i \right\| \quad (\text{nonexpansivity of } J_{\mu_i}^S) \\ &= \left\| \frac{\mu_i}{\mu_n} (x_n + \mu_n T_{\lambda_n} x_n) + \left(1 - \frac{\mu_i}{\mu_n}\right) x_{n+1} - (x_n + \mu_i T_{\lambda_n} x_n) \right\| \\ &= \left\| \frac{\mu_i}{\mu_n} x_n - x_n + \left(1 - \frac{\mu_i}{\mu_n}\right) x_{n+1} \right\| = |\mu_n - \mu_i| \frac{\|x_n - x_{n+1}\|}{\mu_n}. \end{aligned}$$

Combining the arguments above we obtain

$$\|x_n - J_{\mu_i}^S(x_n + \mu_i T_{\lambda_n} x_n)\| \leq \|x_n - x_{n+1}\| + |\mu_n - \mu_i| \frac{\|x_n - x_{n+1}\|}{\mu_n}. \quad \square$$

Under suitable quantitative reformulations of the assumptions, we get the following result translating a quantitative version of the assumption (3) of Theorem 1.1 into a lim inf-bound.

Lemma 2.11. *Let $C \geq 1$ be an upper bound on both $\text{diam}(\mu_n)$ and (μ_n) and let $M \in \mathbb{N}^*$ be such that $M \geq \|T^\circ x^*\|$ for any $x^* \in X_0$. Further, let ϕ be s.t.*

$$\forall k, n \exists N \in [n; \phi(k, n)] \left(\|x_N - x_{N+1}\| / \mu_N < \frac{1}{k+1} \right)$$

and such that it is monotone w.r.t. k and n . Let θ be a rate of convergence for $\lambda_n \rightarrow 0$, i.e.

$$\forall k \forall n \geq \theta(k) \left(\lambda_n \leq \frac{1}{k+1} \right).$$

Further, let ϖ be a modulus of uniform continuity for T w.r.t. H^ . Then the function*

$$\Phi(k, n) = \phi(\lceil 2C(k+1) \rceil - 1, \max\{\theta(M\varpi(k) + M - 1), n\})$$

is a lim inf-bound for x_n w.r.t. Γ_k .

Proof. By assumption on ϕ there exists $N \in [\max\{\theta(M\varpi(k) + M - 1), n\}; \Phi(k, n)]$ with

$$\frac{\|x_N - x_{N+1}\|}{\mu_N} < \frac{1}{\lceil 2C(k+1) \rceil - 1 + 1} \leq \frac{1}{2C(k+1)}.$$

Thus, we have

$$\|x_N - x_{N+1}\| = \mu_N \frac{\|x_N - x_{N+1}\|}{\mu_N} < C \frac{1}{2C(k+1)} = \frac{1}{2(k+1)}.$$

Thus, we get

$$\begin{aligned} \|x_N - J_{\mu_i}^S(x_N + \mu_i T_{\lambda_N} x_N)\| &\leq \|x_N - x_{N+1}\| + |\mu_N - \mu_i| \frac{\|x_N - x_{N+1}\|}{\mu_N} \\ &< \frac{1}{2(k+1)} + C \frac{1}{2C(k+1)} = \frac{1}{k+1} \end{aligned}$$

for all $i \leq k$. Now, we have $H^*[T J_{\lambda_N}^T x_N, T x_N, 1/(k+1)]$: using accretivity of the operator T , it is easy to see that

$$\|x_N - J_{\lambda_N}^T x_N\| \leq \lambda_N \|T^\circ x_N\| \leq \lambda_N M$$

and as $N \geq \theta(M\varpi(k) + M - 1)$, we have

$$\lambda_N \leq \frac{1}{M(\varpi(k) + 1)}$$

and thus $\|x_N - J_{\lambda_N}^T x_N\| \leq \frac{1}{\varpi(k)+1}$. By assumption on ϖ , we have

$$H^*[T J_{\lambda_N}^T x_N, T x_N, 1/(k+1)].$$

This implies $H^*[T_{\lambda_N} x_N, T x_N, 1/(k+1)]$ since we have $H^*[T_{\lambda_N} x_N, T J_{\lambda_N}^T x_N, 0]$ as $T_{\lambda_N} x_N \in T J_{\lambda_N}^T x_N$.

Therefore, in particular we have $\|T_{\lambda_N}x_N - z\| \leq \frac{1}{k+1}$ for some $z \in Tx_N$ and therefore

$$\begin{aligned} \frac{1}{k+1} &\geq \|T_{\lambda_N}x_N - z\| \geq |\|T_{\lambda_N}x_N\| - \|z\|| \\ &= \|z\| - \|T_{\lambda_N}x_N\| \geq \|T^\circ x_N\| - \|T_{\lambda_N}x_N\| = |\|T^\circ x_N\| - \|T_{\lambda_N}x_N\|| \end{aligned}$$

since $z \in Tx_N$, and thus $\|T_{\lambda_N}x_N\| \leq \|T^\circ x_N\| \leq \|z\|$. We have therefore shown that $x_N \in \Gamma_k$. $\square$

Theorem 2.12. *Let T, S be two maximally monotone operators on a finite dimensional Hilbert space X such that $\text{dom}S \subseteq \text{dom}T$. Let $M \in \mathbb{N}^*$ be such that $M \geq \|T^\circ x^*\|$ for any $x^* \in X_0$ and let ϖ be a modulus of uniform continuity for T w.r.t. H^* . Further, let $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$ and assume that $\lim_{n \rightarrow \infty} \lambda_n = 0$ with a rate of convergence θ . Let $C \geq 1$ be an upper bound on both $\text{diam}(\mu_n)$ and (μ_n) . Further, let ϕ be s.t.*

$$\forall k, n \exists N \in [n; \phi(k, n)] \|x_N - x_{N+1}\| / \mu_N < \frac{1}{k+1}$$

and such that it is monotone w.r.t. k and n . Let $L \geq \text{diam}(x_n)$ and let ξ be a Cauchy rate for $\sum_n \mu_n < \infty$. Then (x_n) is Cauchy and, moreover, for any $k \in \mathbb{N}$ and any $g : \mathbb{N} \rightarrow \mathbb{N}$:

$$\exists N \leq \Psi(k, g, \Phi, \chi, \tilde{\xi}, A, L) \forall i, j \in [N; N + g(N)] \left(\|x_i - x_j\| \leq \frac{1}{k+1} \right)$$

where $\Psi(k, g, \Phi, \chi, \tilde{\xi}, A, L) = \Psi_0(P, k, g, \Phi, \chi, \tilde{\xi})$ is defined by recursion with

$$\begin{cases} \Psi_0(0, k, g, \Phi, \chi, \tilde{\xi}) = 0 \\ \Psi_0(n+1, k, g, \Phi, \chi, \tilde{\xi}) = \Phi(\chi_g^M(\Psi_0(n, k, g, \Phi, \chi, \tilde{\xi}), 8k+7), \tilde{\xi}(8k+7)) \end{cases}$$

with $P = \lceil 2[8e^A(k+1)]\sqrt{d}L \rceil^d + 1$ where d is the dimension of X ,

$$\tilde{\xi}(n) = \xi(\lceil (2M+1)e^A(n+1) \rceil - 1)$$

and $\Phi(k, n) = \phi(\lceil 2C(k+1) \rceil - 1, \max\{\theta(M\varpi(k) + M - 1), n\})$

as well as $\chi(r, n, m) = \max\{n + m - 1, \lceil (r+1) \cdot m \cdot e^A \rceil\}$,

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

Further, for any k and any g as above:

$$\exists N \leq \Psi'(k, g, \Phi, \chi, \tilde{\xi}, A, L) \forall i, j \in [N; N + g(N)] \left(\|x_i - x_j\| \leq \frac{1}{k+1} \text{ and } x_i \in \Gamma_k \right),$$

where $\Psi'(k, g, \Phi, \chi, \tilde{\xi}, A, L) = \Psi(k_0, g, \Phi, \chi_k, \tilde{\xi}, A, L)$ with Ψ as before and with

$$k_0 := \max \left\{ k, \left\lceil \frac{\omega(k) - 1}{2} \right\rceil \right\}$$

where $\omega(k) := \max\{\varpi(2k+1), 4k+3, \varpi(4M((k+1)^2 - 1))\}$

with $\chi_k(r, n, m) := \max\{\delta(k), \chi(r, n, m)\}$

where $\delta(k) := 2k+1$ and χ as before.

Proof. The theorem arises from a direct application of Theorem 6.4 from [10] with $X := X_0$, $F := \Gamma \cap X_0$ and $AF_k := \Gamma_k$ with $G := e^A \text{id}_{\mathbb{R}_+}$ and $H := \text{id}_{\mathbb{R}_+}$. Note for this that if ξ is a Cauchy modulus for $\sum_n \mu_n < \infty$, then $n \mapsto \xi(\lceil (2M+1)e^A(n+1) \rceil - 1)$ is a Cauchy modulus for

$$(2M+1)e^A \sum_n \mu_n < \infty$$

as we have

$$\sum_{i=\xi(\lceil (2M+1)e^A(n+1) \rceil - 1)}^{\infty} (2M+1)e^A \mu_i < (2M+1)e^A \frac{1}{\lceil (2M+1)e^A(n+1) \rceil} \leq \frac{1}{n+1}.$$

Further, by Example 2.8 in [10], P is correctly defined since $\lceil 2(k+1)\sqrt{d}L \rceil^d$ is a modulus of total boundedness of $\overline{B}(0; L)$ and thus also of $\overline{B}(x_0; L) = \overline{B}(0; L) + x_0$ as moduli of total boundedness are easily seen to be translation invariant over normed spaces. This clearly makes it a modulus of total boundedness for X_0 as well. Lemma 2.11 gives that Φ is a lim inf-bound and Lemma 2.8 gives that χ is a modulus of uniform quasi-Fejér monotonicity.

The second claim can be concluded from the first claim in the same way that Theorem 5.3 in [10] is proved: χ_k is still a modulus of uniform quasi-Fejér monotonicity and thus

$$\exists N \leq \Psi' \forall i, j \in [N; N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k_0 + 1} \leq \frac{1}{k + 1} \right)$$

by the first result where actually, by inspecting the proof of Theorem 6.4 given in [10], there is an index n such that

- (1) $x_n \in \Gamma_{(\chi_k)_g(N, m)}$ for some m ,
- (2) $\forall i \in [N; N + g(N)] \left(d(x_i, x_n) \leq \frac{1}{2k_0 + 2} \leq \frac{1}{\omega(k) + 1} \right)$.

As $(\chi_k)_g(N, m) = \chi_k(N, g(N), m) \geq \delta(k)$, we get $x_n \in \Gamma_{\delta(k)}$. By Lemma 2.6, as δ and ω are moduli of uniform closedness, we get $x_i \in \Gamma_k$ for all $i \in [N; N + g(N)]$. $\square$

The above theorem is a finitization of Theorem 1.1 under the additional assumption that T is uniformly continuous w.r.t H^* (which is suggested by the logical metatheorems used to obtain this analysis, see Remark 2.13 for a further discussion of this): Assume that we are in the situation of the conclusion of the above theorem. The metastability of (x_n) trivially (but non-effectively) implies that (x_n) is Cauchy and thus convergent to some x .

By Lemma 2.6, we obtain that $\bigcap_k \Gamma_k$ is uniformly closed w.r.t. Γ_k . Further, Lemma 2.11 implies that the above sequence has the lim inf-property with respect to the sequence (Γ_k) .

By Lemma 3.6 of [10], we get that $x \in \bigcap_k \Gamma_k \subseteq \Gamma$ and x is therefore a solution to the original problem.

There are two further interesting notes to make here:

- (1) The assumption that $\lim_{n \rightarrow \infty} \|x_n - x_{n+1}\| / \mu_n = 0$ was weakened to

$$\liminf_{n \rightarrow \infty} \|x_n - x_{n+1}\| / \mu_n = 0.$$

- (2) The assumption $\Gamma \neq \emptyset$ got weakened to the existence of a bound $L \geq \text{diam}(x_n)$. This is indeed a weakening as if $\Gamma \neq \emptyset$, let $p \in \Gamma$ and $q \in Tp \cap Sp$. Then we get $\|x_n\| \leq \|x_n - p\| + \|p\|$ and the former term $\|x_n - p\|$ can be bounded using quasi-Fejér monotonicity as established in Lemma 2.7 by

$$\|x_n - p\| \leq e^A \|x_0 - p\| + (\|T^\circ p\| + \|q\|)e^A \sum_k \mu_k < \infty$$

with A as in the above theorem and thus $\|x_n\|$ is even bounded in that case.

Remark 2.13. (For logicians) The analysis of Moudafi's result as presented above can be explained by, and was obtained using, the general logical metatheorems for the extraction of uniform bounds from noneffective proofs involving set-valued operators developed in [20]. Further, the main analytical tools used in the proof of Theorem 1.1 given in [17] have interesting connections to new proof theoretic notions introduced in [20] (where this case study was instrumental in uncovering these connections). We expect that these connections will influence future approaches to quantitative results in monotone operator theory and we thus want to detail them in the following, motivated by a discussion of the logical aspects of this case study of Moudafi's algorithm. In particular, we want to focus on

- (1) the operator $T^\circ(x) = P_{Tx}(0)$,
- (2) the use of the closure of the graph of both T and S ,
- (3) the assumption that T is bounded on bounded sets.

At first, the analysis presented in the previous parts of the paper can be formalized in (extensions of) the system $\mathcal{T}^\omega$ introduced in [20]. These extensions amount to the treatment of two monotone operators together with the treatment of the uniform continuity of T w.r.t. H^* and the operator T° (see the discussion in [20]).

Now, the operator T° can be treated in the context of the logical metatheorems by a suitable additional constant of type $X(X)$ where X is an additional abstract type for the respective Hilbert space (see [4, 6]) together with characterizing axioms (see [20])

- (i) $\forall x^X (x \in \text{dom}T \rightarrow T^\circ x \in Tx)$,
- (ii) $\forall x^X, y^X (y \in Tx \rightarrow \langle y - T^\circ x, -T^\circ x \rangle \leq 0)$.

For that, the functional T° (extended to the whole space by $T^\circ x := 0$ for $x \notin \text{dom}T$) needs to be majorizable in the sense of [4, 6] (see also [7] for various perspectives on this), i.e. there needs to exist a function $f : \mathbb{N} \rightarrow \mathbb{N}$ such that

$$f \text{ is nondecreasing and } \forall x \in X, n \in \mathbb{N} (\|x\| \leq n \rightarrow \|T^\circ x\| \leq fn).$$

This turns out to connect intimately with a notion of majorizability for the set-valued operator T introduced in [20]: Call a set-valued operator T *majorizable* if there exists a selection function $t : X \rightarrow X$ such that

$$tx \in Tx \text{ for any } x \in \text{dom}T$$

and such that t is majorizable in the sense of the above. Then T is majorizable if and only if T° is majorizable.

Even further, the notion of T being bounded on bounded sets can be recognized as a uniform majorizability assumption (see also [20]): T is bounded on bounded sets if and only if it is uniformly majorizable in the sense that there exists some $f : \mathbb{N} \rightarrow \mathbb{N}$ such that any $t : X \rightarrow X$ with $tx \in Tx$ for $x \in \text{dom}T$ and $tx = 0$ otherwise is majorized by f .

So, majorizability of T (or, equivalently, T°) is already guaranteed by the much stronger assumption of uniform majorizability in Theorem 1.1. However, as a consequence of the proof-theoretic analysis, this assumption of T being bounded on bounded sets can be weakened to plain majorizability of T . Note that this was essentially also observed in [23] from an analytic perspective in the context of extensions of Moudafi's result.

In the above analysis, this majorant is represented by the bound M : The bound is defined via the property $M \geq \|T^\circ x^*\|$ for any $x^* \in X_0 \subseteq \overline{B}_L(x_0)$. Given a majorant f of T/T° , this is (a bound on) the value of f on (a bound on) $L + \|x_0\|$. So, we see that in this concrete situation, even only local information on the majorant is required. Note again that this was also observed in [23] from an analytical perspective.

Further discussions regarding this new notion of (uniform) majorizability of set-valued operators as well as its connections to other notions from monotone operator theory via a proof-theoretic perspective will be given in [13].

Now, item (2), i.e. the closure of the graph of a maximally monotone operator, turns out to be equivalent to the extensionality principle

$$\forall x, y \in X (x = y \rightarrow Tx = Ty)$$

of the set-valued operator T over the system $\mathcal{T}^\omega$ as shown in [20]. It is well known that extensionality can not be provable in systems which allow for the extraction of (uniform) bounds from proofs like, e.g., $\mathcal{T}^\omega$ and its extensions (see [7] for various general discussions of this) and thus a uniform quantitative version of extensionality, namely some uniform continuity principle, has to be added.

As already discussed in [14], there are certain problems with formulating one of the most widely known version of uniform continuity of a set-valued operator defined via the Hausdorff-metric H (see [15])

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in \text{dom}T (\|x - y\| \leq \delta \rightarrow H(Tx, Ty) \leq \varepsilon).$$

Motivated by this, the weaker notion of uniform continuity w.r.t. H^* as discussed before is introduced in [14]. This uniform continuity w.r.t. H^* can be added as an axiom to the system $\mathcal{T}^\omega$ such that one still obtains a bound extraction result (see [20]). Even further, this was recognized in [20] to be the uniform quantitative version of the following weak approximate extensionality principle

$$\forall x, y \in X \left(x = y \rightarrow \forall k \in \mathbb{N} \left(H^* \left[Tx, Ty, \frac{1}{k+1} \right] \right) \right).$$

This principle can be used in place of the full extensionality principle in some situations, for example whenever the rest of the proof following the application of extensionality is extensional in the variables.

Now, in Moudafi's proof, the application of extensionality of T in form of the closure of the graph of T can actually be recognized as just an application of this approximate extensionality principle: extensionality of T is used to conclude $y \in Tx$ given convergent subsequences $T_{\lambda_{n_k}} x_{n_k} \rightarrow y$ and $J_{\lambda_{n_k}}^T x_{n_k} \rightarrow x$ and using that $T_{\lambda_{n_k}} x_{n_k} \in T(J_{\lambda_{n_k}}^T x_{n_k})$. The rest of the proof is extensional as well as continuous in y and this thus reduces to an application of the above approximate extensionality principle. The metatheorems then immediately upgrade T to being uniformly continuous w.r.t. H^* and a modulus for this crucially features in the analysis presented above.

Further, in Moudafi's proof, extensionality of S can actually be completely avoided by instead using the resolvent and the fact that the resolvent is itself provably extensional (see [20]). $\square$

3. Moduli of regularity and rates of convergence

3.1. General theorems on rates of convergence

As mentioned in the introduction, Fejér monotone sequences, in general, do not have a computable rate of convergence. However, the existence of such can be guaranteed in some situations where additional quantitative assumptions are present. Choices for such were extensively studied in [12] under the very general notion of *moduli of regularity* (generalizing moduli of uniqueness and other regularity notions known from optimization like error bounds, weak sharp minima and metric subregularity, see the discussion in [12]) and based on a proof theoretic perspective, [12] presents theorems converting such moduli of regularity for Fejér monotone sequences, modulo some additional minor quantitative assumptions, into rates of convergence for the sequence. To apply these results in our context, we first extend the main quantitative result from [12] to the case of quasi-Fejér monotone sequences.

For that, we follow the setup and notation from [12] (which is conflicting with the notation used in the previous section which was derived from [10], but the context will make it clear which meaning is intended): let (X, d) be a metric space, $F : X \rightarrow \mathbb{R}$ with $\mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}$ be a mapping and assume $\text{zer}F \neq \emptyset$ where $\text{zer}F$ is the set of zeros of F .

Definition 3.1. ([12]) Let $z \in \text{zer}F$ and $r > 0$. A function $\phi : (0, \infty) \rightarrow (0, \infty)$ is a *modulus of regularity for F w.r.t. $\text{zer}F$* and $\overline{B}(z; r)$ if for all $\varepsilon > 0$ and $x \in \overline{B}(z; r)$:

$$|F(x)| < \phi(\varepsilon) \text{ implies } D(x, \text{zer}F) < \varepsilon$$

where D is the distance function between points and sets. It is a *modulus of regularity for F w.r.t. $\text{zer}F$* if there is a z such that it is a modulus of regularity for F w.r.t. $\text{zer}F$ and $\overline{B}(z; r)$ for any $r > 0$. $\square$

Adapting [10], we introduce G - and H -moduli for the functions G, H in the generalized notion of quasi-Fejér monotonicity: a function $\alpha_G : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is a *G -modulus for G* if

$$\forall \varepsilon > 0 \forall a \in \mathbb{R}_+^* (a \leq \alpha_G(\varepsilon) \text{ implies } G(a) \leq \varepsilon),$$

and $\beta_H : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ is a *H -modulus for H* if

$$\forall \varepsilon > 0 \forall a \in \mathbb{R}_+^* (H(a) \leq \beta_H(\varepsilon) \text{ implies } a \leq \varepsilon).$$

For convenience, we assume that a Cauchy rate for a sequence $\sum_i \varepsilon_i < \infty$ is now a mapping taking real values, i.e. a Cauchy rate will now be a function $\xi : \mathbb{R}_+^* \rightarrow \mathbb{N}$ which fulfills for any $\delta \in \mathbb{R}_+^*$:

$$\sum_{i=\xi(\delta)}^{\infty} \varepsilon_i < \delta.$$

We then can generalize the main result of [12], i.e. Theorem 4.1, to quasi- (G, H) -Fejér monotone sequences:

Theorem 3.2. *Let (X, d) be a metric space and $F : X \rightarrow \overline{\mathbb{R}}$ with $\text{zer}F \neq \emptyset$. Let (x_n) be quasi- (G, H) -Fejér monotone w.r.t. $\text{zer}F$. Let α_G be a G -modulus for G , β_H be an H -modulus for H and let β'_H be such that*

$$H(x) \leq a \text{ implies } x \leq \beta'_H(a).$$

for any $x, a \in \mathbb{R}_+$. Let $b \geq G(d(x_0, z))$ for some $z \in \text{zer}F$, $e \geq \sum_n \varepsilon_n$ and suppose there is a τ such that

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

Let ϕ be a modulus of regularity for F w.r.t. $\text{zer}F$ and $\overline{B}(z; \beta'_H(b + e))$ and let ξ be a Cauchy rate for $\sum_i \varepsilon_i < \infty$. Then (x_n) is Cauchy with Cauchy modulus θ :

$$\forall \delta > 0 \forall n, m \geq \theta(\delta) := \tau \left(\phi \left(\alpha_G \left(\frac{\beta_H(\delta/2)}{2} \right) \right), \xi \left(\frac{\beta_H(\delta/2)}{2} \right) \right) (d(x_n, x_m) < \delta).$$

Proof. The proof is a straightforward modification of that of Theorem 4.1 from [12]: Let $\delta > 0$ be given. By quasi- (G, H) -Fejér monotonicity, we have

$$H(d(x_n, z)) \leq G(d(x_0, z)) + \sum \varepsilon_i \leq b + e,$$

i.e. $(x_n) \subseteq \overline{B}(z; \beta'_H(b + e))$. By assumption we have

$$\exists N \in \left[\xi \left(\frac{\beta_H(\delta/2)}{2} \right); \theta(\delta) \right] \left(|F(x_N)| < \phi \left(\alpha_G \left(\frac{\beta_H(\delta/2)}{2} \right) \right) \right).$$

As ϕ is a corresponding modulus of regularity, we get

$$D(x_N, \text{zer}F) < \alpha_G \left(\frac{\beta_H(\delta/2)}{2} \right)$$

and therefore, there exists a $y \in \text{zer}F$ with $d(x_N, y) < \alpha_G \left(\frac{\beta_H(\delta/2)}{2} \right)$. This yields $G(d(x_N, y)) \leq \frac{\beta_H(\delta/2)}{2}$. Thus for any

$$n \geq N \geq \xi \left(\frac{\beta_H(\delta/2)}{2} \right),$$

by quasi- (G, H) -monotonicity, we obtain

$$H(d(x_n, y)) \leq G(d(x_N, y)) + \sum_{i=N}^{\infty} \varepsilon_i \leq \beta_H(\delta/2)/2 + \beta_H(\delta/2)/2 \leq \beta_H(\delta/2)$$

and thus $d(x_n, y) \leq \delta/2$ for any such n , i.e. in particular $d(x_n, x_m) \leq \delta$ for any $n, m \geq \theta(\delta)$. $\square$

Remark 3.3. β'_H is a quantitative version of the property (H1) from [10] (see Lemma 4.2 there). As apparent from the statement and proof, however, we don't need the full function but only its value at $b + e$. In fact, even an upper bound $B \geq \beta'_H(b + e)$ is sufficient as long as ϕ is a modulus of regularity for F w.r.t. $\text{zer}F$ and $\bar{B}(z; B)$.

3.2. An application to Moudafi's algorithm

In applications to problems involving zeros of set-valued operators, [12] describes the following approach for phrasing the corresponding problems in terms of the setup introduced above: Define $F_A(x) := D(0, A(x))$ for a set-valued operator $A: X \rightarrow \mathcal{P}(X)$. If we have that

$$D(0, A(x)) = 0 \text{ implies } x \in \text{zer}A \quad (\dagger)$$

for all $x \in X$, then we also get $\text{zer}F_A = \text{zer}A^4$ which makes F_A a suitable instantiation of the abstract function F from before regarding zeros of A . This always applies to maximally monotone operators A if X is a Hilbert space as $A(x)$ is then closed for any x . For differences $A = T - S$, even with T, S maximally monotone, A may not be maximally monotone anymore but $(T - S)(x)$ is still closed for any x if $T(x), S(x)$ are and thus $(\dagger)$ holds in that case. Thus, we may choose F_{T-S} if we are interested in zeros of differences of two maximally monotone operators.

Now, as exhibited in the above analysis of Moudafi's result, the sequence generated by the algorithm actually converges towards some x such that $T^\circ x \in Sx$ under the above additional quantitative assumptions.

In fact, it is easy to see that $\bigcap_k \Gamma_k = \{x^* \mid T^\circ x^* \in Sx^*\}$ for the previously used approximations Γ_k : if $T^\circ x^* \in Sx^*$, then immediately $x^* \in \Gamma_k$ for any k by setting $y^* = T^\circ x^*$. Conversely, if $x^* \in \Gamma_k$ for any k , we can obtain $T^\circ x^* \in Sx^*$ as in the proof of Lemma 2.6.

The above F_{T-S} is therefore not really faithful regarding the previous analysis and we in turn consider

$$F_1(x) := \|x - J_{\mu_0}^S(x + \mu_0 T^\circ x)\|$$

as $x = J_{\mu_0}^S(x + \mu_0 T^\circ x)$ if and only if $T^\circ x \in Sx$. Indeed, this is in particular supported by the following lemma which establishes a relation between the approximations Γ_k and the property $F_1(x) \leq \frac{1}{k+1}$.

Lemma 3.4. *Let $M \in \mathbb{N}^*$ be such that $M \geq \|T^\circ x^*\|$ for all $x \in X_0$ and let $B \in \mathbb{N}^*$ with $B \geq \mu_0$. If $x \in \Gamma_{\kappa(k)}$, then $F_1(x) \leq \frac{1}{k+1}$ where*

$$\kappa(k) := 4(M + 1)(B(4k + 4) - 1)^2 - 1.$$

Proof. Let $x \in \Gamma_{\kappa(k)}$, i.e. there exists y s.t.

$$\|y\| - \|T^\circ x\| \leq \frac{1}{\kappa(k) + 1}, \quad H^*\left[y, Tx, \frac{1}{(\kappa(k) + 1)}\right],$$

and

$$\|x - J_{\mu_i}^S(x + \mu_i y)\| \leq \frac{1}{\kappa(k) + 1} \text{ for all } i \leq \kappa(k).$$

⁴ Here, we write $\text{zer}A = \{x \mid 0 \in Ax\}$ for the set of zeros of A .

Then by item (2), there exists a $w \in Tx$ s.t. $\|y - w\| \leq \frac{1}{\kappa(k)+1}$, i.e.

$$|\|w\| - \|T^\circ x\|| \leq |\|y\| - \|T^\circ x\|| + |\|y\| - \|w\|| \leq \frac{2}{\kappa(k)+1}.$$

Therefore

$$\begin{aligned} |\|w\|^2 - \|T^\circ x\|^2| &\leq |\|w\| + \|T^\circ x\|| |\|w\| - \|T^\circ x\|| \\ &\leq (2M+2) |\|w\| - \|T^\circ x\|| \\ &\leq (2M+2) \frac{2}{\kappa(k)+1} \leq \frac{1}{(B(4k+4)-1)^2}. \end{aligned}$$

By the modulus of uniqueness for T° , Lemma 2.5, (2), we get

$$\|w - T^\circ x\| \leq \frac{1}{B(4k+4)}.$$

Thus, as $\kappa(k)+1 \geq B(4k+4)$:

$$\begin{aligned} \|y - T^\circ x\| &\leq \|y - w\| + \|w - T^\circ x\| \\ &\leq \frac{1}{\kappa(k)+1} + \frac{1}{B(4k+4)} \leq \frac{1}{B(2k+2)}. \end{aligned}$$

Now using $\|x - J_{\mu_0}^S(x + \mu_0 y)\| \leq \frac{1}{\kappa(k)+1}$, we get (using nonexpansivity of $J_{\mu_0}^S$)

$$\begin{aligned} \|x - J_{\mu_0}^S(x + \mu_0 T^\circ x)\| &\leq \|x - J_{\mu_0}^S(x + \mu_0 y)\| + \|J_{\mu_0}^S(x + \mu_0 T^\circ x) - J_{\mu_0}^S(x + \mu_0 y)\| \\ &\leq \frac{1}{\kappa(k)+1} + \mu_0 \|T^\circ x - y\| \leq \frac{1}{\kappa(k)+1} + \frac{1}{2k+2} \leq \frac{1}{k+1} \end{aligned}$$

as $\kappa(k) \geq 2k+1$. □

Having in mind that $\bigcap_k \Gamma_k = \{x^* \mid T^\circ x^* \in Sx^*\}$ as discussed before, another natural version for F is given by

$$F_2(x) := D(T^\circ x, Sx).$$

Note also that F_2 relates to F_{T-S} in the following way: for any x (in X_0), we have $F_{T-S}(x) \leq F_2(x)$. Indeed, this function F_2 can be used if we assume a further modulus of uniform continuity for S w.r.t. H^* as the following lemma shows:

Lemma 3.5. *Let $M \in \mathbb{N}^*$ be such that $M \geq \|T^\circ x^*\|$ for all $x \in X_0$, let $B \in \mathbb{N}^*$ and $B' \in \mathbb{N}$ with $B \geq \mu_0 \geq 2^{-B'}$, and let $\widehat{\omega}$ be a modulus of uniform continuity for S w.r.t. H^* . If $x \in \Gamma_{\widehat{\kappa}(k)}$, then $F_2(x) \leq \frac{1}{k+1}$ where*

$$\widehat{\kappa}(k) := \kappa \left(\max \left\{ \widehat{\omega}(2k+1), 2^{B'+1}(k+1) \div 1 \right\} \right)$$

with κ from the previous lemma.

Proof. By the previous lemma, we get

$$\|x - J_{\mu_0}^S(x + \mu_0 T^\circ x)\| \leq \frac{1}{\max\{\widehat{\varpi}(2k+1), 2^{B'+1}(k+1) \div 1\} + 1}.$$

Thus we get $H^* \left[SJ_{\mu_0}^S(x + \mu_0 T^\circ x), Sx, \frac{1}{2k+2} \right]$.

Thus, there exists $z \in Sx$ such that

$$\left\| \frac{1}{\mu_0}(x + \mu_0 T^\circ x - J_{\mu_0}^S(x + \mu_0 T^\circ x)) - z \right\| \leq \frac{1}{2k+2}.$$

This entails

$$\begin{aligned} \|z - T^\circ x\| &\leq \left\| \frac{1}{\mu_0}(x + \mu_0 T^\circ x - J_{\mu_0}^S(x + \mu_0 T^\circ x)) - z \right\| + \frac{1}{\mu_0} \|x - J_{\mu_0}^S(x + \mu_0 T^\circ x)\| \\ &\leq \frac{1}{2k+2} + \frac{1}{\mu_0} \frac{1}{2^{B'+1}(k+1)} \leq \frac{1}{2k+2} + \frac{1}{2k+2} \leq \frac{1}{k+1}. \end{aligned}$$

Thus, as $z \in Sx$, we have $D(T^\circ x, Sx) \leq \frac{1}{k+1}$. $\square$

As an immediate consequence of Theorem 3.2 together with the above lemmas, we get the following result for converting the quantitative assumptions into a rate of convergence for the sequence given by Moudafi's algorithm.

Theorem 3.6. *Let T, S be two maximally monotone operators on a Hilbert space X such that $\text{dom}S \subseteq \text{dom}T$ and $\text{zer}(F) \neq \emptyset$ and let $L \geq \text{diam}(x_n)$. Let $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$. Let $C \geq 1$ be an upper bound on both $\text{diam}(\mu_n)$ and (μ_n) , $B \in \mathbb{N}^*$ with $B \geq \mu_0$ and let $M \in \mathbb{N}^*$ be such that $M \geq \|T^\circ x^*\|$ for any $x^* \in X_0$. Further, let Φ be s.t.*

$$\forall k, n \exists N \in [n; \Phi(k, n)] \|x_N - x_{N+1}\| / \mu_N < \frac{1}{k+1}$$

and such that it is monotone w.r.t. k and n , ξ be a Cauchy rate (with input $k \in \mathbb{N}$) for $\sum_n \mu_n \leq d < \infty$ and $b \geq \|x_0 - z\|$ for some $z \in \text{zer}(T - S)$. Let θ be a rate of convergence for $\lambda_n \rightarrow 0$, i.e.

$$\forall k \forall n \geq \theta(k) \left(\lambda_n \leq \frac{1}{k+1} \right).$$

Further, let ϖ be a modulus of uniform continuity for T w.r.t. H^ .*

Let ϕ be a modulus of regularity for F_1 w.r.t. $\text{zer}(F_1)$ and $\overline{B}(z; e^A b + d)$. Then (x_n) is Cauchy with Cauchy modulus

$$\forall \varepsilon > 0 \forall n, m \geq \theta(\varepsilon) := \widehat{\Phi} \left(\kappa \left(\left\lceil \frac{1}{\phi \left(\frac{\varepsilon}{4e^A} \right)} \right\rceil \right), \tilde{\xi} \left(\frac{\varepsilon}{4} \right) \right) (d(x_n, x_m) < \varepsilon)$$

where

$$\tilde{\xi}(\varepsilon) = \xi \left(\left\lceil (2M+1)e^A \left(\left\lceil \frac{1}{\varepsilon} \right\rceil + 1 \right) \right\rceil - 1 \right)$$

and $\widehat{\Phi}(k, n) := \Phi(\lceil 2C(k+1) \rceil - 1, \max\{\theta(M\varpi(k) + M - 1), n\})$.

as well as

$$\kappa(k) := 4(M+1)(B(4k+4) - 1)^2 - 1.$$

Moreover, if we assume $\widehat{\omega}$ to be a modulus of uniform continuity for S w.r.t. H^* and $B' \in \mathbb{N}$ such that $\mu_0 \geq 2^{-B'}$, then the above claim holds for ϕ being a modulus of regularity for F_2 (or F_{T-S}) w.r.t. $\text{zer}(F_2)$ (or $\text{zer}(F_{T-S})$) and $\overline{B}(z; e^A b + d)$ and with κ replaced by $\widehat{\kappa}$ defined by

$$\widehat{\kappa}(k) = \kappa \left(\max \left\{ \widehat{\omega}(2k+1), 2^{B'+1}(k+1) \div 1 \right\} \right).$$

Proof. The proof is a direct application of Theorem 3.2. We want to note a few things, however. Again, we work over X_0 . Then Lemma 2.8 actually established quasi- $(e^A \text{id}_{\mathbb{R}_+}, \text{id}_{\mathbb{R}_+})$ -Fejér monotonicity w.r.t. $\text{zer}(F)$. Naturally, $\delta \mapsto \delta/e^A$ is a G -modulus for $e^A \text{id}_{\mathbb{R}_+}$ and $\delta \mapsto \delta$ is a H -modulus for $\text{id}_{\mathbb{R}_+}$ and β'_H can just be set to be the identity. Further, the function

$$(\varepsilon, n) \mapsto \widehat{\Phi} \left(\kappa \left(\left\lceil \frac{1}{\varepsilon} \right\rceil \right), n \right)$$

is a $\liminf$ -modulus for F_1 by Lemma 2.11 and Lemma 3.4. $\square$

Note that it is no longer necessary for X to be finite dimensional in this case. This in particular relies on the fact that the previous moduli do not rely on finite dimensionality either.

Acknowledgments. This paper is a revised version of parts of my master thesis [21] written under the supervision of Prof. Dr. Ulrich Kohlenbach at TU Darmstadt. I want to thank Prof. Kohlenbach. His suggestion to work on Moudafi's algorithm proved to be a very interesting case study in the context of proof mining and I have immensely enjoyed and benefited from our various discussions on the subject matter.

References

- [1] S. Adly, W. Oettli: *Solvability of generalized nonlinear symmetric variational inequalities*, J. Austral. Math. Soc. Ser. B, Appl. Math. 40/3 (1999) 289–300.
- [2] H. Bauschke, P. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, Cham (2017).
- [3] P. L. Combettes: *Quasi-Fejérian analysis of some optimization algorithms*, in: *Inherently Parallel Algorithms for Feasibility and Optimization*, D. Butnariu, Y. Censor, S. Reich (eds.), Elsevier, Amsterdam (2001) 115–152.
- [4] P. Gerhardy, U. Kohlenbach: *General logical metatheorems for functional analysis*, Trans. Amer. Math. Soc. 360 (2008) 2615–2660.
- [5] R. Horst, N. V. Thoai: *DC programming: Overview*, J. Optimization Theory Appl. 103 (1999) 1–43.
- [6] U. Kohlenbach: *Some logical metatheorems with applications in functional analysis*, Trans. Amer. Math. Soc. 357/1 (2005) 89–128.
- [7] U. Kohlenbach: *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*, Springer Monographs in Mathematics, Springer, Berlin (2008).
- [8] U. Kohlenbach: *Proof-theoretic methods in nonlinear analysis*, in: Proc. Int. Congress Math. 2018, Vol. 2, B. Sirakov, P. Ney de Souza and M. Viana (eds.), World Scientific, Singapore (2019) 61–82.

- [9] U. Kohlenbach: *Quantitative results on the proximal point algorithm in uniformly convex Banach spaces*, J. Convex Analysis 28/1 (2021) 11–18.
- [10] U. Kohlenbach, L. Leuştean, A. Nicolae: *Quantitative Results on Fejér Monotone Sequences*. Commun. Contemp. Math. 20/2 (2018), art. no. 1750015.
- [11] U. Kohlenbach, G. López-Acedo, A. Nicolae: *Quantitative asymptotic regularity results for the composition of two mappings*, Optimization 66 (2017) 1291–1299.
- [12] U. Kohlenbach, G. López-Acedo, A. Nicolae: *Moduli of regularity and rates of convergence for Fejér monotone sequences*, Israel J. Math. 232 (2019) 261–297.
- [13] U. Kohlenbach, N. Pischke: *Proof theory and nonsmooth analysis*, Phil. Trans. Royal Society A (2023), to appear.
- [14] U. Kohlenbach, T. Powell: *Rates of convergence for iterative solutions of equations involving set-valued accretive operators*, Comp. Math. Appl. 80 (2020) 490–503.
- [15] C. Moore, B. V. C. Nnoli: *Iterative solution of nonlinear equations involving set-valued uniformly accretive operators*, Comp. Math. Appl. 42/1-2 (2001) 131–140.
- [16] A. Moudafi: *On the difference of two maximal monotone operators: Regularization and algorithmic approaches*, Appl. Math. Comput. 202 (2008) 446–452.
- [17] A. Moudafi: *On critical points of the difference of two maximal monotone operators*, Afrika Matematika 26 (2015) 457–463.
- [18] E. Neumann: *Computational problems in metric fixed point theory and their Weihrauch degrees*, Logical Methods Comp. Sci. 11/4 (2015), art. no. 20.
- [19] M. A. Noor, K. I. Noor, A. Hamdi, E. H. El-Shemas: *On difference of two monotone operators*, Optim. Letters 3 (2009) 329–335.
- [20] N. Pischke: *Logical metatheorems for accretive and (generalized) monotone set-valued operators*, arXiv: 2205.01788 (2022).
- [21] N. Pischke: *Logical metatheorems for set-valued operators and their use in the analysis of Moudafi’s algorithm for the difference of two monotone operators in Hilbert space*, Master’s Thesis, Technische Universität Darmstadt (2022).
- [22] N. Pischke, U. Kohlenbach: *Quantitative analysis of a subgradient-type method for equilibrium problems*, Numer. Algorithms 90/1 (2022) 197–219.
- [23] M. Ramazannejad, M. Alimohammady, C. Cattani: *On algorithms for difference of monotone operators*, in: *Computational Mathematics and Variational Analysis*, N. J. Daras and T. M. Rassias (eds.), Springer Series on Optimization and its Applications 159, Springer, Berlin (2020) 381–401.
- [24] T. Tao: *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dyn. Systems 28 (2008) 657–688.
- [25] 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 (2008).
- [26] H. A. L. Thi, T. P. Dinh: *DC programming and DCA: Thirty years of developments*, Math. Programming 169 (2018) 5–68.