EFFECTIVE RATES FOR CONTINUOUS-TIME QUASI-FEJÉR MONOTONE DYNAMICAL SYSTEMS

ANTON FREUND^a AND NICHOLAS PISCHKE^b

^{a} Institute of Mathematics, University of Würzburg,
Emil-Fischer-Str. 40, 97074 Würzburg, Germany,
^{b} Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom.
E-mails: anton.freund@uni-wuerzburg.de, nnp39@bath.ac.uk

ABSTRACT. We provide quantitative convergence results for continuous-time dynamical systems in metric spaces that satisfy a continuous-time analog of quasi-Fejér monotonicity. More precisely, we provide a (strong) convergence result for such dynamical systems over compact metric spaces which is quantitatively outfitted with a continuous-time rate of metastability, which moreover can be explicitly and effectively constructed in a very uniform way, only depending on a few moduli representing quantitative witnesses to key properties of the dynamical system and a measure for the compactness of the space. We further show how this convergence result can be extended to non-compact spaces under a regularity assumption of the associated problem, where moreover rates of convergence can then be explicitly constructed which are similarly uniform. In both cases, already the associated "infinitary" convergence result is qualitatively novel in its present generality. Beyond this abstract quantitative theory for such dynamical systems, we motivate how the presently studied continuous-time variant of quasi-Fejér monotonicity naturally occurs as a unifying property of many dynamical systems and differential equations and inclusions, and in that way can be used to provide a comprehensive quantitative theory for many such dynamical systems. We illustrate this with three case studies for both classical first- and second-order dynamical systems in Hilbert spaces as well as (generalized) gradient flows and associated semigroups in nonlinear Hadamard spaces.

Keywords: dynamical systems, continuous Fejér monotonicity, rates of metastability, rates of convergence, proof mining
MSC2020 Classification: 34G25, 47J25, 90C25, 03F10

1. INTRODUCTION

Fejér monotonicity is one of the most fundamental concepts in modern (convex) programming. Traditionally formulated by

$$\forall n \in \mathbb{N} \ \forall z \in F \left(d(x_{n+1}, z) \leqslant d(x_n, z) \right)$$

for a sequence (x_n), say in some metric space (X, d), it expresses that the sequence as a whole descends towards a solution set $F \subseteq X$. This rather convenient property, or at least one of its many extensions, is enjoyed by many, if not most, iterative methods in optimization, and as such has considerably streamlined and unified the convergence analysis of such dynamical systems, where next to establishing Fejér monotonicity, all that is required to prove a method convergent is generally some mild approximation property of the sequence towards the solution set. As such, Fejér monotonicity over time became one of the most central organisational principles in the analysis of algorithms.

The notion itself has a rather long history already. The name Fejér monotonicity was coined in [72], referring to a use of a similar concept made by Fejér in [49]. Subsequently, it has

been heavily studied by the "Russian school" of convex programming, in particular Eremin [44, 45] as well as Ermol'ev [46, 47, 48]. The modern appreciation of Fejér monotonicity as a central unifying concept for many strands of convex programming is however centrally due to Combettes and Bauschke, in particular due to the surveys and expositions of Combettes [37, 38] and the central monograph [15].

This paper is devoted to a study of a continuous-time analog of Fejér monotonicity, as formulated by

$$\forall t \geqslant s \geqslant 0 \ \forall z \in F \left(d(x(t), z) \leqslant d(x(s), z)\right)$$

and an associated quasi-variant incorporating an error function e with $\lim_{t \geqslant s \to \infty} e(s, t) = 0$, that is

$$\forall t \geqslant s \geqslant 0 \ \forall z \in F \left(d(x(t), z) \leqslant d(x(s), z) + e(s, t)\right),$$

now for a continuous-time dynamical system $x : [0, \infty) \to X$, still over some metric space (X, d). This property, and its extensions by perturbations of the metric or inclusions of error terms, is intimately linked to differential inequalities for associated energy functionals as arising in the context of Lyapunov analyses of differential equations and inclusions, but in particular allows one to capture many further situations like flows and semigroups, also over nonlinear spaces (as discussed in greater detail in Section 2). Nevertheless, this property is often left implicit, only explicitly coming to the forefront in the proofs, if at all. A notable exception is the well-known survey [79] by Peypouquet and Sorin, where the property is prominently isolated as a link between continuous- and discrete-time settings (see in particular property (a3) therein). But even in [79], the investigation is not centered around the above property but rather constructed ad-hoc around the particular problem at hand, that is the first-order Cauchy problem for a maximal monotone operator.

As such, the present paper is intended to illustrate how the above, more "sequential", notion naturally allows for a uniform study of such dynamical systems arising through differential equations, and beyond. In particular, this more sequential character seems particularly suitable for quantitative considerations, which are the main object of this paper.

Concretely, the first part of the present paper presents highly general and abstract convergence theorems for dynamical systems subscribing to (a generalized version of) the above continuous-time Fejér monotonicity property. These results fall into two categories:

(1) We provide a strong convergence result under a (relative) compactness assumption on the space, utilizing only mild approximation assumptions on the dynamical systems.
(2) We provide a strong convergence result without any explicit compactness assumption, but while imposing a certain rather general type of regularity for the associated problem.

In the former case, we in particular provide rates of metastability for the dynamical system x, that is a function $\Delta(\varepsilon, g)$ such that

$$\forall \varepsilon > 0 \ \forall g : \mathbb{N} \to \mathbb{N} \ \exists n \leqslant \Delta(\varepsilon, g) \ \forall t, s \in [n, n + g(n)] \left(d(x(t), x(s)) \leqslant \varepsilon\right).$$

These rates thereby represent a continuous-time analog of the usual rates of metastability as coined by Tao [90, 91] for discrete-time sequences, and similar to that context present an alternative phrasing of the Cauchy property of a dynamical system which allows for effective quantitative information, contrary to usual rates of convergence which, due to results from computability theory (see [86] as well as [76]) are generally non-computable.

Nevertheless, in the latter case, these rates of metastability can generally be upgraded to full rates of convergence, that is we can even provide a function $\rho(\varepsilon)$ such that

$$\forall \varepsilon > 0 \ \forall t, s \geqslant \rho(\varepsilon) \left(d(x(t), x(s)) \leqslant \varepsilon\right).$$

In either case, these rates can be explicitly constructed from some simple data representing quantitative formulations of the main assumptions, and we provide these constructions in full detail. In particular, the resulting rates are highly uniform, depending only on these data and not on further auxiliary properties of the space or dynamical system.

The quantitative results pertaining to the first item in particular generalize related such results for discrete-time Fejér monotone sequences presented in [59], while the latter generalize similar such results given in [60] (see also [80]). In particular, our methods are quite similar, and result in rates of a similar style and complexity, thereby highlighting how one can give a unified treatment of continuous- and discrete-time systems, in particular for quantitative questions, by following the more discrete reasoning to a large degree in the continuous context. We regard this as another methodological contribution of the present paper, and regard our paper in a similar spirit to the survey [79], which already heavily highlights such similarities, also in various other parts of the analysis of continuous- and discrete-time systems.

Like the works mentioned above, also the present results are derived using methods from proof mining, a subfield of mathematical logic which aims at utilizing tools from the foundations of mathematics to provide additional computational information for ineffectively proven results in mathematics. As such, it has found wide success in particular in the context of (nonlinear) analysis and optimization. We refer to the seminal monograph [57] as well as the surveys [58, 61] for further background on proof mining. However, regardless of this logical background, the entire paper is formulated without explicit reference to logic, as typical for results arising in the proof mining program, and in that way can be comprehended without any knowledge of mathematical logic.

All these methodological aspects notwithstanding, we regard the actual applications of the abstract results presented here as the key exponents of our paper. To illustrate this, we provide three concrete case studies in this paper. The first presents a quantitative analysis of the work [23] by Boţ and Csetnek for a first-order dynamical system over a nonexpansive operator and the associated first-order continuous-time variant of the forward-backward scheme. The second presents a quantitative analysis of the work [22] by Boţ and Csetnek for a second-order dynamical system over a cocoercive operator, and its associated second-order continuous-time variant of the forward-backward scheme. Both case studies are chosen due to their rather abstract presentation and in particular their generality, allowing for a uniform treatment of various other preceding systems (as discussed in more detail in the respective introductions to the case studies later on, but we already want to refer to the survey [42]). The second case study in particular presents an example where our approach, even in the context of metric regularity assumptions, cannot provide rates of convergence in general, as we can only provide a rate of metastability for the respective error function in the quasi-Fejér monotonicity property, which presents a novel structural reason for why no rates of convergence have appeared for that particular scheme, at that level of generality. This should in particular be compared to more restrictive choices of parameter processes and operators, in particular covering the continuous-time forward-backward method in the context of strongly monotone operators, where fast effective rates can nevertheless be constructed as illustrated in [25], which we partially extend in the present paper to more general settings, again however following a quite different methodology motivated by the discrete-time approach. The last case study is concerned with nonlinear semigroups on nonpositively curved metric spaces, making use of the general metric setting of our results. In particular, we consider the perhaps canonical example of the gradient flow semigroup as studied in this context by Mayer [71] and later Bačák [16], and then further consider a related nonlinear semigroup generated by a nonexpansive mapping as studied by Stojkovic [87] and

later Bačák and Reich [18], which in particular allows for an abstract approach towards heat flows in singular spaces (as further discussed later on).

At last, we think of our paper as laying the groundwork for quantitative studies of other facets of continuous-time Fejér monotonicity, and can conceive of various natural successors to the present work. Indeed, discrete-time Fejér monotonicity encompasses many variants, and we think that also these variants can be quantitatively lifted to a continuous-time setting by following the present methods. Initial examples of such variants include adaptations to varying metrics [39] or to different distance functions like Bregman distances [14] or even broader general classes of distances on metric spaces as considered in [81] (see also [75]). However, these extensions would be rather routine. More interestingly, one could also generalize the present results to continuous-time variants of the recently introduced Fejér* monotonicity [7, 19, 20] as well as the rather general and extensive localized and relativized generalization of Fejér monotonicity [63] (which was also introduced through a proof mining perspective), where then the very interesting question arises whether natural dynamical systems fall under that regime, with a natural contender being continuous-time variants of the method presented in [19]. Last but not least, the present results naturally lend themselves to extensions into the stochastic setting. In particular, the present results on rates of convergence under regularity assumptions could be generalized in that way by combining the present approach with that of the recent works [75, 83] on rates of convergence for discrete-time stochastically quasi-Fejér monotone sequences under stochastic regularity assumptions, while the general constructions of rates of metastability under a compactness assumption could be extended by instead following the approach of the upcoming work [82]. The previously cited works rely on a quantitative theory for discrete-time supermartingales, and extensions of these results would then require a similar quantitative approach to continuous-time martingale theory, which would be interesting in its own right. In particular, such results could then unlock quantitative treatment of associated stochastic differential equations and inclusions, as to render them applicable to recent works in that direction such as [30, 31, 66, 67, 69, 70].

Notation. Throughout, if not stated otherwise, we let (X, d) be a given metric space. We denote closed balls in such spaces by

$$\overline{B}_r(x) := \{y \in X \mid d(x, y) \leqslant r\}.$$

Natural numbers $\mathbb{N}$ are throughout assumed to contain 0. The function $\lceil \cdot \rceil$ is defined as usual, that is by $\lceil x \rceil := \min\{z \in \mathbb{Z} \mid z \geqslant x\}$. Further, we define bounded subtraction on $\mathbb{N}$ via $n \dot{-} m := \max\{n - m, 0\}$. If we are given a function $f : \mathbb{N} \to \mathbb{N}$, we denote iteration of that function by $f^{(i)}$, defined recursively by $f^{(0)} := \text{Id}$ and $f^{(i+1)} := f \circ f^{(i)}$.

2. Continuous-time Fejér and quasi-Fejér monotonicity

2.1. **Definition and motivation.** We begin with our notion of continuous-time quasi-Fejér monotonicity, the central notion studied in the present paper. This notion, like all other variants of Fejér monotonicity, refers to an underlying solution set which we in this paper take to be given as the set of zeros

$$\text{zer} F := \{x \in X \mid F(x) = 0\}$$

for a given solution function $F : X \to [0, +\infty]$.

Definition 2.1 (Quasi-Fejér monotone dynamical system). Let $F : X \to [0, +\infty]$ be a given solution function and let $G, H : [0, \infty) \to [0, \infty)$ be functions with $a \to 0$ whenever $H(a) \to 0$ as well as $G(a) \to 0$ whenever $a \to 0$.

A dynamical system $x : [0, +\infty) \to X$ is called (G, H)-quasi-Fejér monotone w.r.t. F if

$$H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s, t)$$

for all $t \geqslant s$ and all $y \in \text{zer} F$, where e is some error function with $\lim_{t \geqslant s \to \infty} e(s, t) = 0$. If $e \equiv 0$, then x is called (G, H)-Fejér monotone w.r.t. F.

Motivated by their use in [59], the general perturbation functions G and H above allow for quite a great variety of different notions of Fejér monotonicity familiar from the discrete-time world to be encompassed by the above continuous notion.

As such, beyond the more "ordinary" quasi-Fejér monotonicity property

$$d(x(t), y) \leqslant d(x(s), y) + e(s, t) \text{ for all } t \geqslant s$$

derived by $G = H = \text{Id}$, which presents a continuous-time analog of quasi-Fejér monotonicity of type I as considered by Combettes [37], they in particular encompass functions such as $G = H = (\cdot)^p$ for $p \in (0, +\infty)$, where the associated monotonicity property

$$d^p(x(t), y) \leqslant d^p(x(s), y) + e(s, t) \text{ for all } t \geqslant s$$

in particular presents a continuous-time analog of quasi-Fejér monotonicity of order p as discussed in [48]. In the case of $p = 2$, this in particular presents a continuous-time analog of quasi-Fejér monotonicity of type II as considered in [37].

A more or less immediate lift of the discrete-time notion of quasi-Fejér monotonicity, the above notion nevertheless abstracts and unifies various scattered formulations and conditions often associated with "continuous-time quasi-Fejér monotonicity". These, while often distinct in their use and analysis, commonly take the form of differential inequalities

$$\frac{d}{dt} \mathcal{E}_y(t) \leqslant \overline{e}(t)$$

for a given (family of) function(s) $\mathcal{E}_y(t) = \rho(d(x(t), y))$ over $y \in \text{zer} F$, where $x(t)$ is a given dynamical system, $\overline{e} \in L^1$ is an error function and ρ again serves as a perturbation (commonly taking the forms $\rho = \text{Id}$ or $\rho = (\cdot)^2$). Indeed, it is often already the stricter variant

$$\frac{d}{dt} \mathcal{E}_y(t) \leqslant 0,$$

formulated without an error function, which occurs. The perhaps canonical case for this that we have in mind is the first-order Cauchy problem over a maximally monotone operator in Hilbert spaces (already due to [33, 41, 64], see also [34, 40] and e.g. the survey [79]), that is

$$\begin{cases} -\dot{x}(t) \in A(x(t)), \\ x(0) = x_0, \end{cases}$$

where A is a maximally monotone operator on a Hilbert space X and $x_0 \in \text{dom}(A)$, in particular generalizing the gradient flow equation $\dot{x} = \nabla f(x)$ to general inclusions. Now, any solution $x : [0, \infty) \to X$ to the above problem (that is x is an absolutely continuous function satisfying $x(0) = x_0$ and $-\dot{x}(t) \in A(x(t))$ almost everywhere; in fact, there is a unique such solution, see e.g. [79]) satisfies the differential inequality

$$\frac{d}{dt} \left(\frac{1}{2} \|x(t) - y\|^2 \right) = \langle \dot{x}(t), x(t) - y \rangle \leqslant 0$$

almost everywhere, where $y \in \text{zer}(A)$ is any zero of A, i.e. $0 \in A(y)$. Beyond this rather immediate example, we further refer to [9, 13, 23, 28, 29, 43] for other instances of differential inequalities such as the above, just to name a few.

In any case, in the context of the above differential inequality, integration of that property yields

$$\rho(d(x(t), y)) \leqslant \rho(d(x(s), y)) + \int_s^t \overline{e}(\tau)\, d\tau \text{ for all } t \geqslant s$$

for all $y \in \mathrm{zer}F$, so that the above family of differential inequalities results in the dynamical system x being (ρ, ρ)-quasi-Fejér monotone in the sense of Definition 2.1, with errors $e(s, t) := \int_s^t \overline{e}(\tau)\, d\tau$. In the special case of $\overline{e} \equiv 0$ already mentioned above, we in particular get that x is (ρ, ρ)-Fejér monotone.

Such a presentation of the errors via an integral, that is

$$e(s, t) = \int_s^t \overline{e}(\tau)\, d\tau$$

for $\overline{e} \in L^1$, is in a way the standard form of an error that we want to accommodate in the above Definition 2.1. In this error form, the (G, H)-quasi-Fejér monotonicity of a dynamical system x can be even more directly recognised as a direct lift of the usual discrete-time (G, H)-quasi-Fejér monotonicity of a sequence (x_n) (see e.g. Definition 6.2 in [59]), that is

$$H(d(x_{n+m}, y)) \leqslant G(d(x_n, y)) + \sum_{i=n}^{n+m-1} \varepsilon_i \text{ for all } n \geqslant m \text{ and } y \in \mathrm{zer}F,$$

where (ε_i) is a nonnegative sequence of summable errors, to the continuous setting, with errors taking the form of integrals instead of series.

However, similar as with the perturbation functions, the generality of the generic error function e compared to a specific representation such as the above allows to encompass even more general situations. Namely, beyond such rather simple situations as detailed above, many analyses of continuous-time dynamical systems do not rely on such simple functions $\mathcal{E}_y(t) = \rho(d(x(t), y))$ as above but instead consider much more extensive (families of) so-called energy functionals $\mathcal{E}_y(t)$, suitably chosen to allow for the analysis of the underlying dynamical system $x(t)$ at hand, often in a rather ad-hoc way (see e.g. the survey [73]). Regardless, such energy functions often decompose in special ways.

One such decomposition takes the form of

$$\mathcal{E}_y(t) = \rho(d(x(t), y)) + \mathcal{E}'(t).$$

Indeed, such energy functions and associated differential inequalities are rather common, and we refer to e.g. [1, 2, 6, 8, 22, 21, 24, 26, 27], just to name a few. Integration of the associated differential inequality property yields

$$\mathcal{E}_y(t) - \mathcal{E}_y(s) \leqslant \int_s^t \overline{e}(\tau)\, d\tau \text{ for all } t \geqslant s$$

as before and, in the case of the above decomposition $\mathcal{E}_y(t) = \rho(d(x(t), y)) + \mathcal{E}'(t)$, this inequality immediately results in the monotonicity property

$$\rho(d(x(t), y)) \leqslant \rho(d(x(s), y)) + \mathcal{E}'(s) - \mathcal{E}'(t) + \int_s^t \overline{e}(\tau)\, d\tau$$

for all $t \geqslant s$ and $y \in \mathrm{zer}F$, so that in case the associated remainder $\mathcal{E}'$ satisfies $\mathcal{E}'(t) \to 0$ for $t \to \infty$, the dynamical system $x(t)$ is (ρ, ρ)-quasi-Fejér monotone in the above sense with a generalized error function

$$e(s, t) := \mathcal{E}'(s) - \mathcal{E}'(t) + \int_s^t \overline{e}(\tau)\, d\tau.$$

Another type of decomposition that sometimes occurs in linear spaces takes the form of

$$\mathcal{E}_y(t) = \rho(\|x(t) - y + v(t)\|)$$

for some additional function $v(t)$ with $\lim_{t\to\infty} v(t) = 0$. Such energy functions e.g. occur in [3, 11, 12]. At least in situations where $\|x(t) - y\|$, $\|x(t) - y + v(t)\|$ are bounded (as can generally be established without great difficulty in such situations) and ρ is Lipschitz on bounded sets (as is the case for e.g. $\rho = (\cdot)^2$), say with constant $L > 0$ for a respective set containing $\|x(t) - y\|$ and $\|x(t) - y + v(t)\|$, we can use the associated differential inequality as follows: Integration as before yields

$$\rho(\|x(t) - y + v(t)\|) \leqslant \rho(\|x(s) - y + v(s)\|) + \int_s^t \overline{e}(\tau)\, d\tau.$$

Now, using the Lipschitz continuity of ρ on the respective (bounded) set yields

$$\rho(\|x(t) - y\|) \leqslant \rho(\|x(t) - y + v(t)\|) + L\|v(t)\|$$

$$\leqslant \rho(\|x(s) - y + v(s)\|) + \int_s^t \overline{e}(\tau)\, d\tau + L\|v(t)\|$$

$$\leqslant \rho(\|x(s) - y\|) + \int_s^t \overline{e}(\tau)\, d\tau + L(\|v(t)\| + \|v(s)\|).$$

In that way, the dynamical system $x(t)$ then is (ρ, ρ)-quasi-Fejér monotone in the above sense with a generalized error function

$$e(s, t) := \int_s^t \overline{e}(\tau)\, d\tau + L(\|v(t)\| + \|v(s)\|).$$

Naturally, choosing such an energy function appropriately is quite difficult and lies at the heart of a convergence analysis of various continuous time methods. We in that way do not at all want to argue that the present results make such an approach obsolete but rather that once such a suitable function with a corresponding differential inequality has been isolated, it allows for a uniform quantitative analysis of the associated convergence results via the above notion of quasi-Fejér monotonicity for continuous-time dynamical systems and it is that quantitative perspective, as already discussed in the introduction, that justifies its use.

2.2. **Quantitative results for differential inequalities.** We have seen above how quasi-Fejér monotonicity for continuous-time dynamical systems often comes in the guise of differential inequalites. These inequalities have various key asymptotic results attached to them that are fundamental for most analyses in the literature. As such, we now discuss these results here, and in particular provide quantitative variants, on which we later rely in the context of our applications.

The main results with which we are concerned are the following fundamental lemmas (see e.g. Lemmas 5.1 and 5.2 in [2]):

Lemma 2.2. *Assume $\mathcal{E} : [0, \infty) \to \mathbb{R}$ is locally absolutely continuous and bounded below such that*

$$\frac{d}{dt}\mathcal{E}(t) \leqslant \overline{e}(t)$$

almost everywhere for some $\overline{e} \in L^1([0, \infty))$. Then $\lim_{t\to\infty} \mathcal{E}(t)$ exists.

If further $\mathcal{E}$ is nonnegative and $\mathcal{E} \in L^p([0, \infty))$ for $p \in [1, +\infty)$, then $\lim_{t\to\infty} \mathcal{E}(t) = 0$. In that case, we may also have $\overline{e} \in L^r([0, \infty))$ for $r \in [1, +\infty]$.

We begin with an analysis of the first part of the above lemma:

Lemma 2.3. *Assume* $\mathcal{E} : [0, \infty) \to \mathbb{R}$ *is locally absolutely continuous and bounded below with* $\mathcal{E}(t) \geqslant b$ *for all* $t \in [0, \infty)$ *and* $c \geqslant \mathcal{E}(0)$. *Further assume we have*

$$\frac{d}{dt}\mathcal{E}(t) \leqslant \overline{e}(t)$$

almost everywhere for some $\overline{e} \in L^1([0, \infty))$ *with* $\|\overline{e}\|_1 \leqslant B$. *Then for all* $\varepsilon > 0$ *and* $f : \mathbb{N} \to \mathbb{N}$, *it holds that*

$$\exists n \leqslant \tilde{f}^{(\omega_{b,c,B}(\varepsilon))}(0) \; \forall s, t \in [n, n + f(n)] \, (|\mathcal{E}(s) - \mathcal{E}(t)| \leqslant \varepsilon),$$

where $\tilde{f}(n) := n + f(n)$ *and*

$$\omega_{b,c,B}(\varepsilon) := \left\lceil \frac{2B}{\varepsilon} \right\rceil \cdot \left\lceil \frac{c + B - b}{\varepsilon} \right\rceil.$$

Proof. By the proof of [2, Lemma 5.1], the map $t \mapsto \mathcal{E}(t) - \int_0^t \overline{e}(\tau)\, d\tau$ is non-increasing with

$$D(t) := \mathcal{E}(t) - \int_0^t \overline{e}(\tau)\, d\tau \in [b - B, c] \text{ for all } t \in [0, \infty).$$

We define $H(m) = \tilde{f}^{(\lceil \frac{2B}{\varepsilon} \rceil \cdot m)}(0)$. It is clearly impossible for D to descend by more than $\varepsilon/2$ in all intervals $[H(m), H(m + 1)]$ with $m < \lceil 2(c + B - b)/\varepsilon \rceil$. So for one of the indicated m, we have

$$0 \leqslant D(s) - D(t) \leqslant \frac{\varepsilon}{2} \quad \text{whenever } H(m) \leqslant s \leqslant t \leqslant H(m + 1).$$

On the other hand, the integral over $|\overline{e}|$ cannot exceed $\frac{\varepsilon}{2}$ on all intervals $[\tilde{f}^{(i)}(H(m)), \tilde{f}^{(i+1)}(H(m))]$ with $i < \lceil \frac{2B}{\varepsilon} \rceil$. For an i where the integral is bounded by $\frac{\varepsilon}{2}$, we put

$$r = \left\lceil \frac{2B}{\varepsilon} \right\rceil \cdot m + i \quad \text{and} \quad n = \tilde{h}^{(r)}(0).$$

Observe that we have

$$[n, n + f(n)] = \left[\tilde{f}^{(r)}(0), \tilde{f}^{(r+1)}(0)\right] = \left[\tilde{f}^{(i)}(H(m)), \tilde{f}^{(i+1)}(H(m))\right] \subseteq [H(m), H(m + 1)].$$

For $n \leqslant s \leqslant t \leqslant n + f(n)$, we thus get

$$|\mathcal{E}(s) - \mathcal{E}(t)| \leqslant D(s) - D(t) + \int_s^t |G(\tau)|\, d\tau \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

as desired. □

We continue with an analysis of [2, Lemma 5.2]:

Lemma 2.4. *Assume* $\mathcal{E} \in L^p([0, \infty))$ *is locally absolutely continuous and non-negative, with* $1 \leqslant p < \infty$, *and where* $c \geqslant \mathcal{E}(0)$. *Further assume we have*

$$\frac{d}{dt}\mathcal{E}(t) \leqslant \overline{e}(t)$$

almost everywhere, for some $\overline{e} \in L^r([0, \infty))$ *with* $1 \leqslant r \leqslant \infty$. *Let* $\|\mathcal{E}\|_p \leqslant A$ *and* $\|\overline{e}\|_r \leqslant B$. *Then for all* $\varepsilon > 0$ *and* $f : \mathbb{N} \to \mathbb{N}$, *it holds that*

$$\exists n \leqslant \tilde{f}'^{(\varpi_{c,A,B}(\varepsilon))}(0) \; \forall t \in [n, n + f(n)] \, (\mathcal{E}(t) < \varepsilon),$$

where $\tilde{f}'(n) := n + f'(n)$ *as before, with* $f'(n) := \max\{f(n), \lceil (3A/\varepsilon)^p \rceil\}$, *and*

$$\varpi_{c,A,B}(\varepsilon) := \left\lceil \frac{2^{q+1} q A^{q-1} B}{\varepsilon^q (2^q - 1)} \right\rceil \cdot \left\lceil \frac{2^q (c^q + q A^{q-1} B)}{\varepsilon^q (2^q - 1)} \right\rceil \quad \text{with} \quad q = 1 + p \cdot \left(1 - \frac{1}{r}\right).$$

Proof. Let $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$ be given. We first show that

$$\exists n \leqslant \tilde{f}^{(\varpi_{c,A,B}(\varepsilon))}(0) \ \forall t \in [n, n + f(n)] \left(\mathcal{E}(t) < \varepsilon\right)$$

for $\tilde{f}(n) := n + f(n)$, if $f(n) > (2A/\varepsilon)^p$ for all $n \in \mathbb{N}$. Since $r' = \frac{p}{q-1}$ is conjugate to r and $\mathcal{E}^{q-1}$ is in $L^{r'}([0, \infty))$, Hölder's inequality yields

$$\|\mathcal{E}^{q-1} \cdot \overline{e}\|_1 \leqslant A^{q-1} \cdot B.$$

Except on a null set, we have

$$\frac{d}{dt}\mathcal{E}^q(t) = q \cdot \mathcal{E}^{q-1}(t) \cdot \frac{d}{dt}\mathcal{E}(t) \leqslant q \cdot \mathcal{E}^{q-1}(t) \cdot \overline{e}(t).$$

As $f(n) > (2A/\varepsilon)^p$ for all $n \in \mathbb{N}$, Lemma 2.3 shows that

$$|\mathcal{E}^q(s) - \mathcal{E}^q(t)| \leqslant \frac{\varepsilon^q(2^q - 1)}{2^q} \quad \text{for all } s, t \in [n, n + f(n)]$$

for some $n \leqslant \tilde{f}^{(\varpi_{c,A,B}(\varepsilon))}(0)$ as in the current lemma. Now there is some $s_0 \in [n, n + f(n)]$ with $\mathcal{E}(s_0) < \varepsilon/2$, since we would otherwise get

$$A^p \geqslant \int_n^{n+f(n)} \mathcal{E}^p(t)\, dt \geqslant \frac{\varepsilon^p}{2^p} \cdot f(n) > A^p.$$

We now show that $\mathcal{E}(t) < \varepsilon$ holds for any $t \in [n, n + f(n)]$. If this was not true, we would get

$$|\mathcal{E}^q(s_0) - \mathcal{E}^q(t)| > \varepsilon^q - \frac{\varepsilon^q}{2^q} = \frac{\varepsilon^q(2^q - 1)}{2^q},$$

against the above. If $f : \mathbb{N} \to \mathbb{N}$ is now a general function, define $f'(n) := \max\{f(n), \lceil (3A/\varepsilon)^p \rceil\}$. By definition $f'(n) > (2A/\varepsilon)^p$ and so the above implies

$$\exists n \leqslant \tilde{f'}^{(\varpi_{c,A,B}(\varepsilon))}(0) \ \forall t \in [n, n + f'(n)] \left(\mathcal{E}(t) < \varepsilon\right).$$

The result now follows as $f(n) \leqslant f'(n)$ and so $[n, n + f(n)] \subseteq [n, n + f'(n)]$. $\square$

Remark 2.5. In the above Lemmas 2.3 and 2.4, we have always quantitatively resolved integrability assumptions via upper bounds on the involved p-norms, e.g. assuming for a given $\overline{e} \in L^1([0, \infty))$ a bound $B \geqslant \|\overline{e}\|_1$. It is clear from the proofs above that it actually suffices to have rates of metastability for the integrals in question, e.g. in the case of $\overline{e} \in L^1([0, \infty))$ above taking the form of a function $\xi(\varepsilon, f)$ such that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, one has

$$\exists n \leqslant \xi(\varepsilon, f) \left(\int_n^{n+f(n)} |\overline{e}| < \varepsilon\right).$$

But then the results become combinatorially more involved. As one is commonly supplied with such a norm upper bound, we keep the results at that generality.

3. A quantitative convergence theory for quasi-Fejér monotone dynamical systems

We now turn to our main abstract results, presenting a quantitative theory of convergence for quasi-Fejér monotone dynamical systems in continuous time. As discussed before, we will be working over general metric spaces. Convergence results in these rather general contexts commonly fall in two categories:

(1) Strong (metric) convergence results under a (local) compactness assumption, for general problems $\text{zer}F$.

(2) Strong (metric) convergence results in the absence of compactness, but for that under an additional (local) regularity assumption on the problem $\mathrm{zer}F$.

Weak convergence results, which are quite characteristic for Fejér monotonicity, in general are not available for general metric spaces and require additional structure, commonly working over Hilbert spaces or Hadamard spaces (complete geodesic spaces of nonpositive curvature). In that way, these results shall not be our concern here, although the present work can be easily adapted also to these cases.

In either case, the only other property next to the quasi-Fejér monotonicity that is required to induce convergence is the rather weak approximation property that the sequence in question contains arbitrarily good approximate solutions along its run.

We here consider quantitative variants of both items (1) and (2) above, extending the previous seminal works in that regard [59] and [60] (see also [80]) from discrete time to continuous time. We begin with the former.

3.1. **Convergence under relative compactness assumptions.** We recall the basic setup. Let (X, d) be a given metric space and let $F : X \to [0, +\infty]$ be a given function, so that $\mathrm{zer}F := \{x \in X \mid F(x) = 0\}$ serves as an abstract representation of a solution set and so that each sublevel set

$$\mathrm{lev}_{\leqslant \varepsilon}F := \{x \in X \mid F(x) \leqslant \varepsilon\},$$

with $\varepsilon > 0$, serves as an abstract representation of a set of ε-approximate solutions.

We will be concerned with a quantitative convergence result for quasi-Fejér monotone dynamical systems in the context where X is totally bounded. To arrive at such a result, we require three central ingredients.

The first is, as mentioned before, a (quantitative) approximation property for the sequence in question that suffices to guarantee convergence.

Definition 3.1. Let $x : [0, \infty) \to X$ be a dynamical system. We say that x has the $\liminf$-property w.r.t. F if $\liminf_{t \to \infty} F(x(t)) = 0$. Quantitatively, we say that $\varphi : (0, \infty) \times \mathbb{N} \to \mathbb{N}$ is a $\liminf$-bound for x w.r.t. F if

$$\forall \varepsilon > 0 \ \forall n \in \mathbb{N} \exists t \in [n, \varphi(\varepsilon, n)] \left(F(x(t)) < \varepsilon \right).$$

Remark 3.2. We will later commonly assume that φ is monotone in the sense that $\varphi(\varepsilon, n) \geqslant \varphi(\delta, n)$ whenever $\varepsilon \leqslant \delta$. This can actually be assumed without loss of generality as if a general $\varphi(\varepsilon, n)$ is given, we can define

$$\widehat{\varphi}(\varepsilon, n) := \max\{\varphi(1/(k + 1), n) \mid k \leqslant \lceil 1/\varepsilon \rceil\}.$$

Indeed, then $\widehat{\varphi}$ is still a $\liminf$-bound as given $\varepsilon > 0$ and $n \in \mathbb{N}$, we have

$$\exists t \in [n, \varphi(1/(\lceil 1/\varepsilon \rceil + 1), n)] \left(F(x(t)) < \frac{1}{\lceil 1/\varepsilon \rceil + 1} \leqslant \varepsilon \right)$$

and $\varphi(1/(\lceil 1/\varepsilon \rceil + 1), n) \leqslant \widehat{\varphi}(\varepsilon, n)$, and if $\varepsilon \leqslant \delta$, then $\lceil 1/\varepsilon \rceil \geqslant \lceil 1/\delta \rceil$ and so

$$\widehat{\varphi}(\varepsilon, n) = \max\{\varphi(1/(k + 1), n) \mid k \leqslant \lceil 1/\varepsilon \rceil\} \geqslant \max\{\varphi(1/(k + 1), n) \mid k \leqslant \lceil 1/\delta \rceil\} = \widehat{\varphi}(\delta, n).$$

Hence $\widehat{\varphi}$ is also monotone in the above sense.

The second ingredient we will need is a uniform quantitative strengthening of the quasi-Fejér property. For that, consider the property of (G, H)-quasi-Fejér monotonicity unwound into its logical structure:

$$\forall y \in X \ (\forall \delta > 0 (y \in \mathrm{lev}_{\leqslant \delta}F) \to \forall \varepsilon > 0 \ \forall s \leqslant t (H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s, t) + \varepsilon)).$$

With the quantifiers suitably prenexed (and rewriting $\forall s \leqslant t$ into $\forall n, m \in \mathbb{N} \, \forall s \leqslant t \in [n, n+m]$) the above is equivalent to

$$\forall \varepsilon > 0 \, \forall n, m \in \mathbb{N} \, \forall y \in X \, \exists \delta > 0 (y \in \mathrm{lev}_{\leqslant \delta} F$$
$$\to \forall s \leqslant t \in [n, n+m] \, (H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s, t) + \varepsilon)).$$

We now call a sequence uniformly (G, H)-quasi-Fejér monotone if we can bound (and hence witness) the above δ in terms of ε, n, m, uniformly for all $y \in X$:

Definition 3.3. Let $x : [0, \infty) \to X$ be a dynamical system. We say that x is uniformly (G, H)-quasi-Fejér monotone w.r.t. F with modulus $\chi : (0, \infty) \times \mathbb{N}^2 \to (0, \infty)$ if

$$\forall \varepsilon > 0 \, \forall n, m \in \mathbb{N} \, \forall y \in \mathrm{lev}_{\leqslant \chi(\varepsilon, n, m)} F \, \forall s \leqslant t \in [n, n+m]$$
$$(H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s, t) + \varepsilon)$$

for some error function e with $\lim_{t \geqslant s \to \infty} e(s, t) = 0$. If $e \equiv 0$, then we call x uniformly (G, H)-Fejér monotone w.r.t. F.

We have two remarks on this notion:

Remark 3.4. A standard compactness argument shows that if X is compact and F as well as G, H are continuous, then any continuous dynamical system x which is (G, H)-quasi-Fejér monotone w.r.t. F and some continuous error function e is also already uniformly (G, H)-quasi-Fejér monotone w.r.t. F and that same error function.

Remark 3.5 (For logicians). As in the case of the associated notion of uniform (G, H)-quasi-Fejér monotonicity for discrete-time sequences given in [59] (see Remark 4.7 therein), the two notions above, plain and uniform quasi-Fejér monotonicity for continuous-time sequences, can be proven to be equivalent in theories for bounded metric spaces used in proof mining (see [55]) augmented with a nonstandard uniform boundedness principle (see [56]). In particular, associated logical metatheorems for these system due to Kohlenbach (recall the previous references and see further [57]) in that way guarantee that such a modulus χ can be extracted from large classes of proofs of the associated, weaker, quasi-Fejér monotonicity property, over bounded metric spaces, making it particularly natural for applications.

The last central ingredient will be a quantitative rendering of the compactness assumption on the space. Concretely, in the following we will work under a total boundedness assumption, which we quantitatively resolve following the work of [51]:

Definition 3.6 (essentially [51]). Let $A \subseteq X$ be nonempty. A function $\gamma : (0, \infty) \to \mathbb{N}$ is a modulus of total boundedness for A if for any $\varepsilon > 0$ and any sequence $(x_k) \subseteq A$:

$$\exists 0 \leqslant i < j \leqslant \gamma(\varepsilon) \, (d(x_i, x_j) \leqslant \varepsilon).$$

It follows rather immediately that a set is totally bounded if, and only if, it has a modulus of total boundedness. We refer to [51] and [59] for further discussions of this property.

Besides these three central ingredients, we are left with quantitatively resolving the minor assumptions around our sequence.

Let us first consider the perturbation functions $G, H : [0, \infty) \to [0, \infty)$. By assumption, we have $a \to 0$ whenever $H(a) \to 0$, and this is equivalent with assuming the existence of a quantitative modulus $h : (0, \infty) \to (0, \infty)$ such that

$$H(a) < h(\varepsilon) \to a < \varepsilon$$

for all $a \geqslant 0$ and $\varepsilon > 0$. Similarly, the property that $G(a) \to 0$ whenever $a \to 0$ is resolved by considering as modulus g with

$$a < g(\varepsilon) \to G(a) < \varepsilon$$

for all $a \geqslant 0$ and $\varepsilon > 0$.

The last minor quantitative detail we need to resolve is the error property $\lim_{t \geqslant s \to \infty} e(s,t) = 0$. Most conveniently, we would resolve this assumption with a rate of convergence (similar to how errors for quasi-Fejér monotone sequences are resolved in [59]), that is a function $\eta(\varepsilon)$ such that

$$\forall \varepsilon > 0 \; \forall t \geqslant s \geqslant \eta(\varepsilon) \, (|e(s,t)| \leqslant \varepsilon) .$$

This assumption is particularly mild in cases where the errors are effectively composed of parameters of a problem which can be freely chosen, so that such a rate can be readily supplied, as is often the case.

However, this will crucially not be the case for some applications later where we will only be able to derive a rate of metastability for the above limit, that is a function $\eta(\varepsilon, f)$ such that

$$\forall \varepsilon > 0 \; \forall f : \mathbb{N} \to \mathbb{N} \; \exists n \leqslant \eta(\varepsilon, f) \; \forall s \leqslant t \in [n, n + f(n)] \, (|e(s,t)| \leqslant \varepsilon) .$$

This generality however comes with the price of some added combinatorial complexity of the final rate.

Combining all of these ingredients, we then get the following central quantitative result on the convergence of quasi-Fejér monotone dynamical systems in totally bounded spaces.

Theorem 3.7. *Let X be totally bounded with a modulus γ. Assume that $G, H : [0, \infty) \to [0, \infty)$ are such that we have moduli g, h with*

$$H(a) < h(\varepsilon) \to a < \varepsilon \text{ and } a < g(\varepsilon) \to G(a) < \varepsilon$$

for all $a \geqslant 0$ and $\varepsilon > 0$. Also, let $e(s,t) \to 0$ as $s \leqslant t \to \infty$ with a rate of metastability η, i.e.

$$\forall \varepsilon > 0 \; \forall f : \mathbb{N} \to \mathbb{N} \; \exists n \leqslant \eta(\varepsilon, f) \; \forall s \leqslant t \in [n, n + f(n)] \, (|e(s,t)| < \varepsilon) .$$

Lastly, let $x : [0, \infty) \to X$ be a dynamical system such that

(1) x has the $\liminf$-property w.r.t. F with bound $\varphi : (0, \infty) \times \mathbb{N} \to \mathbb{N}$, i.e.

$$\forall \varepsilon > 0 \; \forall n \in \mathbb{N} \; \exists t \in [n, \varphi(\varepsilon, n)] \, (F(x(t)) < \varepsilon) .$$

Further, we assume (w.l.o.g.) that φ is monotone in the sense that $\varphi(\varepsilon, n) \geqslant \varphi(\delta, n)$ whenever $\varepsilon \leqslant \delta$.

(2) x is uniformly (G, H)-quasi-Fejér monotone w.r.t. F with error e and with modulus $\chi : (0, \infty) \times \mathbb{N}^2 \to (0, \infty)$, i.e.

$$\forall \varepsilon > 0 \; \forall n, m \in \mathbb{N} \; \forall y \in \text{lev}_{\leqslant \chi(\varepsilon, n, m)} F \; \forall s \leqslant t \in [n, n + m]$$
$$(H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s,t) + \varepsilon) .$$

Then x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon, f) \; \forall s, t \in [n, n + f(n)] \, (d(x(s), x(t)) \leqslant \varepsilon) ,$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described by

$$\Delta(\varepsilon, f) := \max\{\Delta(i, \varepsilon, f) \mid i \leqslant P\} + 1$$

where $P := \gamma(g(h(\varepsilon/2)/3)) + 1$ and $\Delta(0, \varepsilon, f) := 0$ as well as

$$\Delta(j, \varepsilon, f) := \max\{\varphi(\widehat{\varepsilon}_j, n) \mid n \leqslant \eta(h(\varepsilon/2)/3, f_{\varphi, \widehat{\varepsilon}_j})\}$$

with

$$\widehat{\varepsilon}_j = \min\{\chi_f^M(h(\varepsilon/2)/3, \Delta(i, \varepsilon, f)) \mid i < j\}$$

for $j \geqslant 1$, where

$$\chi_f(\varepsilon, n) := \chi(\varepsilon, n, f(n+1)+1),$$
$$\chi_f^M(\varepsilon, n) := \min\{\chi_f(\varepsilon, m) \mid m \leqslant n\},$$

as well as

$$f_{\varphi,\varepsilon}(n) := \max\{m+1+f(m+1) \mid m \leqslant \varphi(\varepsilon, n)\} \dotminus n.$$

Proof. Let φ_0 be a selection function for the $\liminf$-property which is bounded by φ, i.e. take $\varphi_0 : (0, \infty) \times \mathbb{N} \to [0, \infty)$ to be any function such that $x(\varphi_0(\varepsilon, n)) \in \mathrm{lev}_{\leqslant \varepsilon} F$ and $\varphi_0(\varepsilon, n) \in [n, \varphi(\varepsilon, n)]$ for all $\varepsilon > 0$ and $n \in \mathbb{N}$. Set $t_0 := 0$ and $n_0 := 0$. Then, given $f : \mathbb{N} \to \mathbb{N}$ and $\varepsilon > 0$, recursively define

$$\varepsilon_j := \min\{\chi(g(\varepsilon/2)/3, \lfloor t_i \rfloor, f(\lfloor t_i \rfloor + 1) + 1) \mid i < j\},$$

let $n_j \leqslant \eta(h(\varepsilon/2)/3, f_{\varphi,\widehat{\varepsilon}_j})$ be such that

$$\forall s \leqslant t \in [n_j, n_j + f_{\varphi,\widehat{\varepsilon}_j}(n_j)] \, (|e(s,t)| \leqslant h(\varepsilon/2)/3),$$

and set $t_j := \varphi_0(\varepsilon_j, n_j)$. We first note that $\varepsilon_j \geqslant \widehat{\varepsilon}_j$ for $j > 0$ as well as $\Delta(i, \varepsilon, f) \geqslant t_i$ for $i \geqslant 0$, which we show by induction. Clearly $\Delta(0, \varepsilon, f) = 0 = t_0$ and if $\Delta(i, \varepsilon, f) \geqslant t_i$ for all $i < j$, then

$$\widehat{\varepsilon}_j = \min\{\chi_f^M(h(\varepsilon/2)/3, \Delta(i, \varepsilon, f)) \mid i < j\}$$
$$\leqslant \min\{\chi_f(h(\varepsilon/2)/3, \lfloor t_i \rfloor) \mid i < j\} = \varepsilon_j$$

as $\lfloor t_i \rfloor \leqslant t_i \leqslant \Delta(i, \varepsilon, f)$. Then we get

$$t_j = \varphi_0(\varepsilon_j, n_j) \leqslant \varphi(\varepsilon_j, n_j) \leqslant \varphi(\widehat{\varepsilon}_j, n_j) \leqslant \max\{\varphi(\widehat{\varepsilon}_j, n) \mid n \leqslant \eta(h(\varepsilon/2)/3, f_{\varphi,\widehat{\varepsilon}_j})\} = \Delta(j, \varepsilon, f)$$

as $n_j \leqslant \eta(h(\varepsilon/2)/3, f_{\varphi,\widehat{\varepsilon}_j})$.

Now, using the modulus of total boundedness γ for the sequence $(x(t_j))$, we get indices

$$0 < I < J \leqslant \gamma(g(h(\varepsilon/2)/3)) + 1$$

such that $d(x(t_I), x(t_J)) \leqslant g(h(\varepsilon/2)/3)$. By definition, as $J > 0$ and $J > I$, we have

$$x(t_J) \in \mathrm{lev}_{\leqslant \varepsilon_J} F \subseteq \mathrm{lev}_{\leqslant \chi(h(\varepsilon/2)/3, \lfloor t_I \rfloor, f(\lfloor t_I \rfloor + 1) + 1)} F.$$

Also, as $I > 0$, we have $t_I \geqslant n_I$. The former now implies

$$H(d(x(t), x(t_J))) \leqslant G(d(x(s), x(t_J))) + e(s, t) + h(\varepsilon/2)/3$$

for all $s \leqslant t \in [\lfloor t_I \rfloor, \lfloor t_I \rfloor + 1 + f(\lfloor t_I \rfloor + 1)]$. At the same time, we have

$$\forall s \leqslant t \in [n_I, n_I + f_{\varphi,\widehat{\varepsilon}_I}(n_I)] \, (|e(s,t)| \leqslant h(\varepsilon/2)/3).$$

Now, we have $n_I \leqslant \lfloor t_I \rfloor$ as $n_I \leqslant t_I$ and n_I is a natural number, and so

$$n_I \leqslant \lfloor t_I \rfloor \leqslant t_I \leqslant \lfloor t_I \rfloor + 1 \leqslant \lfloor t_I \rfloor + 1 + f(\lfloor t_I \rfloor + 1)$$
$$\leqslant \max\{m + 1 + f(m+1) \mid m \leqslant \varphi(\widehat{\varepsilon}_I, n_I)\}$$
$$\leqslant n_I + f_{\varphi,\widehat{\varepsilon}_I}(n_I)$$

as $t_I = \varphi_0(\varepsilon_I, n_I) \leqslant \varphi(\varepsilon_I, n_I) \leqslant \varphi(\widehat{\varepsilon}_I, n_I)$ since $\widehat{\varepsilon}_I \leqslant \varepsilon_I$. Thus, in the above, we may pick $s = t_I$ to get

$$H(d(x(t), x(t_J))) \leqslant G(d(x(t_I), x(t_J))) + e(t_I, t) + h(\varepsilon/2)/3 \text{ and } |e(t_I, t)| \leqslant h(\varepsilon/2)/3$$

for all $t \in [\lfloor t_I \rfloor + 1, \lfloor t_I \rfloor + 1 + f(\lfloor t_I \rfloor + 1)]$ (as then also $t \geqslant t_I$). As $d(x(t_I), x(t_J)) \leqslant g(h(\varepsilon/2)/3)$, we have $G(d(x(t_I), x(t_J))) \leqslant h(\varepsilon/2)/3$ and so

$$H(d(x(t), x(t_J))) \leqslant |e(t_I, t)| + 2h(\varepsilon/2)/3 \leqslant h(\varepsilon/2)$$

for all such t. This implies $d(x(t), x(t_J)) \leqslant \varepsilon/2$ for all $t \in [n, n + f(n)]$ with $n = \lfloor t_I \rfloor + 1$, so that we obtain the result by the triangle inequality.

As seen above, we in particular have $\Delta(i, \varepsilon, f) \geqslant \lfloor t_i \rfloor$ so that $\Delta(\varepsilon, f) \geqslant \Delta(I, \varepsilon, f) + 1 \geqslant \lfloor t_i \rfloor + 1 = n$ as $I \leqslant P$. $\qquad\square$

The above result extends Theorem 6.4 in [59], the analogous result for discrete-time sequences, and in fact closely follows the proof strategy given there.

If F is uniformly continuous, then the above result can be further upgraded to also effectively guarantee that the dynamical system consists of approximate solutions.[1]

Theorem 3.8. *Under the assumptions of Theorem 3.7, assume that F is uniformly continuous with a modulus ω, that is $|F(x) - F(y)| \leqslant \varepsilon$ for all $x, y \in X$ with $d(x, y) \leqslant \omega(\varepsilon)$.*

Then for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\min\{\varepsilon, \omega(\varepsilon/2)\}, f) \; \forall s, t \in [n, n + f(n)] \, (d(x(s), x(t)) \leqslant \varepsilon \text{ and } F(x(t)) \leqslant \varepsilon),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described as in Theorem 3.7, except using $\widehat{\chi}_\varepsilon(\delta, n, m) := \min\{\varepsilon/2, \chi(\delta, n, m)\}$ in place of χ.

Proof. Note that $\widehat{\chi}_\varepsilon$ is also a modulus of uniform (G, H)-quasi-Fejér monotonicity. Now follow the proof of Theorem 3.7 with the respective bound. Writing $\varepsilon_0 := \min\{\varepsilon, \omega(\varepsilon/2)\}$, as therein we then derive $d(x(t), x(t_J)) \leqslant \varepsilon_0/2$ for all $t \in [n, n + f(n)]$ with $n = \lfloor t_I \rfloor + 1$, where

$$x(t_J) \in \text{lev}_{\leqslant \widehat{\chi}_\varepsilon(h(\varepsilon_0/2)/3, \lfloor t_I \rfloor, f(\lfloor t_I \rfloor + 1) + 1)} F \subseteq \text{lev}_{\leqslant \varepsilon/2} F.$$

The former gives $d(x(t), x(s)) \leqslant \varepsilon$ for all $s, t \in [n, n + f(n)]$ as before, and additionally that $d(x(t), x(t_J)) \leqslant \varepsilon_0/2 \leqslant \omega(\varepsilon/2)$. We hence have $|F(x(t)) - F(x(t_J))| \leqslant \varepsilon/2$. Now, the latter yields $F(x(t_J)) \leqslant \varepsilon/2$, so that we have

$$F(x(t)) \leqslant F(x(t_J)) + |F(x(t)) - F(x(t_J))| \leqslant \varepsilon$$

for all $t \in [n, n + f(n)]$. $\qquad\square$

If we are supplied with a rate of convergence for the error function e, or the errors disappear altogether, then the above construction becomes considerably simpler, which we record in the following theorem.

Theorem 3.9. *Under the assumptions of Theorem 3.7, assume that $e(s, t) \to 0$ as $s \leqslant t \to \infty$ with a rate of convergence η, i.e.*

$$\forall \varepsilon > 0 \; \forall t \geqslant s \geqslant \eta(\varepsilon) \, (|e(s, t)| < \varepsilon).$$

Then x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon, f) \; \forall s, t \in [n, n + f(n)] \, (d(x(s), x(t)) \leqslant \varepsilon),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described by

$$\Delta(\varepsilon, f) := \max\{\Delta(i, \varepsilon, f) \mid i \leqslant P\} + 1$$

with $P := \gamma(g(h(\varepsilon/2)/3)) + 1$ and $\Delta(0, \varepsilon, f) := 0$ as well as

$$\Delta(j, \varepsilon, f) := \max\{\varphi(\widehat{\varepsilon}_j, n) \mid n \leqslant \eta(h(\varepsilon/2)/3)\}$$

where $\widehat{\varepsilon}_j$ for $j \geqslant 1$ is as in Theorem 3.7.

---

[1]In fact, Theorem 3.8 can be established under the milder quantitative condition that $\text{zer} F$ is uniformly closed w.r.t. the level sets $\text{lev}_{\leqslant \varepsilon} F$ in the sense of [59], that is using moduli ω and δ such that for any $x, y \in X$, $d(x, y) \leqslant \omega(\varepsilon)$ and $F(x) \leqslant \delta(\varepsilon)$ implies $F(y) \leqslant \varepsilon$. For simplicity, and since it suffices for all our applications, we here restrict ourselves to the stronger assumption that F is uniformly continuous.

If $e \equiv 0$, then it suffices that x has the approximate F-point property (instead of the $\liminf$-property w.r.t. F), that is

$$\forall \varepsilon > 0 \; \exists t \leqslant \varphi(\varepsilon) \, (F(x(t)) < \varepsilon)$$

with a monotone bound $\varphi : (0, \infty) \to \mathbb{N}$, where Δ can be defined similar to above, with $\Delta(0, \varepsilon, f) := 0$ and $\Delta(j, \varepsilon, f) := \varphi(\widehat{\varepsilon}_j)$.

Lastly, if in any of the above cases F is uniformly continuous with a modulus ω, then for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\min\{\varepsilon, \omega(\varepsilon/2)\}, f) \; \forall s, t \in [n, n + f(n)] \, (d(x(s), x(t)) \leqslant \varepsilon \; \text{and} \; F(x(t)) \leqslant \varepsilon),$$

where the bound $\Delta(\varepsilon, f)$ is constructed as above, except using $\widehat{\chi}_\varepsilon(\delta, n, m) := \min\{\varepsilon/2, \chi(\delta, n, m)\}$ in place of χ.

Proof. As η is a rate of convergence, it follows immediately that $\widehat{\eta}(\varepsilon, f) := \eta(\varepsilon)$ is also a rate of metastability. If the errors are constantly 0, then $\eta \equiv 0$ is a corresponding rate of convergence. In any case, we can apply Theorem 3.7 to derive the above rates. That, in the case of vanishing errors, an approximate F-point bound suffices immediately follows from the proof of Theorem 3.7. The extensions for approximate solutions in the context of a uniformly continuous F can be accounted for as in the proof of Theorem 3.8. $\qquad\square$

Moreover, the above quantitative result trivially implies the following "usual" (that is, non-quantitative) convergence result for quasi-Fejér monotone dynamical systems in compact spaces.

Theorem 3.10. *Let X be compact and let $F : X \to [0, +\infty]$ be continuous. Assume that $G, H : [0, \infty) \to [0, \infty)$ are continuous such that $a \to 0$ whenever $H(a) \to 0$ and $G(a) \to 0$ whenever $a \to 0$. Let $x : [0, \infty) \to X$ be a continuous dynamical system such that x has the $\liminf$-property w.r.t. F and which is (G, H)-quasi-Fejér monotone w.r.t. F and a continuous error function.*

Then x converges to some point in $\mathrm{zer}F$.

Proof. All the present assumptions trivially (but nonconstructively) entail their quantitative variant. In particular, as X is compact, x is uniformly (G, H)-quasi-Fejér monotone w.r.t. F (recall Remark 3.4). Hence, Theorem 3.7 guarantees that x is metastable, i.e.

$$\forall \varepsilon > 0 \; \forall f : \mathbb{N} \to \mathbb{N} \; \exists n \in \mathbb{N} \; \forall s, t \in [n, n + f(n)] \, (d(x(s), x(t)) \leqslant \varepsilon).$$

Trivially, but noneffectively, this is equivalent to $x(t)$ being Cauchy for $t \to \infty$. Indeed, suppose for a contradiction that there is an $\varepsilon > 0$ such that for all $n \in \mathbb{N}$, there is an $m \in \mathbb{N}$ with

$$\exists s, t \in [n, n + m] \, (d(x(s), x(t)) > \varepsilon).$$

Define $f(n) := m$ for such an m. Then we have found an f such that

$$\forall n \in \mathbb{N} \; \exists s, t \in [n, n + f(n)] \, (d(x(s), x(t)) > \varepsilon).$$

This contradicts the above property. As x is Cauchy and X is compact, and hence complete, it has a limit z. We now show that $z \in \mathrm{zer}F$. For that, note that $F(z) = \lim_{t \to \infty} F(x(t))$ as F is continuous and since x has the $\liminf$-property w.r.t. F, we get $\liminf_{t \to \infty} F(x(t)) = 0$. Hence $F(z) = 0$. $\qquad\square$

Remark 3.11. Assume that G is inverse bounded, i.e.

$$G(a) \leqslant b \to a \leqslant g'(b)$$

for all $a \geqslant 0$ and $b > 0$ and additionally that

$$e^M := \sup_{t \geqslant 0} e(0, t) < \infty.$$

Then if x is (G, H)-quasi-Fejér monotone, we have

$$G(d(x(t), z)) \leqslant H(d(x(0), z)) + e(0, t)$$

for all $t \geqslant 0$ and so $G(d(x(t), z))$ and hence $x(t)$ is bounded with $d(x(t), z) \leqslant g'(b_0 + e^M)$ for $b_0 \geqslant H(d(x(0), z))$. A particular case where G is inverse bounded is for example when G is strictly increasing and continuous, where then g' is simply the inverse of G. The assumption that $\sup_{t \geqslant 0} e(0, t) < \infty$ is classically satisfied when e is of standard form, i.e. $e(s, t) = \int_s^t \overline{e}(\tau) \, d\tau$ for $\overline{e} \in L^1$, but further easily true in many other situations, such as those later discussed in this paper. In any case, in the context where the dynamical system is bounded, it of course thereby suffices that X is only proper in the above Theorem 3.10.

Remark 3.12. If one has asymptotic regularity w.r.t. F, that is $\lim_{t \to \infty} F(x(t)) = 0$, and for simplicity assuming that $e \equiv 0$, then the above qualitative theorem can be directly derived from the associated result for discrete-time dynamical systems (see e.g. Proposition 4.3 in [59]): Indeed, $x(t)$ converges to some point $z \in \mathrm{zer} F$ for $t \to \infty$ if, and only if, $x(t_n)$ converges to z for any (increasing) sequence (t_n) with $t_n \to \infty$ for $n \to \infty$. Using the asymptotic regularity w.r.t. F, we get that any such sequence has approximate F-points and (G, H)-Fejér monotonicity w.r.t. F for x transfers onto the respective sequence $x(t_n)$. The convergence of $x(t_n)$ now follows from the associated discrete-time result (recall Proposition 4.3 in [59]). This extends suitably to error sequences, using the $\liminf$-property w.r.t. F instead.

Exactly this usual reduction to the discrete-time setting under asymptotic regularity seems to be the reason why a qualitative theory for continuous-time Fejér monotone dynamical systems has received relatively little attention. However, this reduction proves to be problematic quantitatively, as one cannot immediately transfer rates of metastability for the resulting Cauchy property from the discrete-time subsequences to the full continuous-time sequence (as the respective point of metastability might move out further and further, and moreover, the resulting regions might not overlap).

Instead, the only quantitative approach known to us is to essentially redo the quantitative approach to discrete-time dynamical systems given in [59] for the continuous setting, and the proof for the above result essentially implements this argument, with various optimisations built in. However, while for metastability there seems to be no efficient way of bootstrapping the quantitative theory of continuous-time dynamical systems onto discrete-time dynamical systems, at least not in the generality of Theorem 3.7, we will later discuss how results on rates of convergence under regularity assumptions can indeed be transferred from the discrete to the continuous, at least in special cases.

3.2. **Convergence under metric regularity assumptions.** We will now be concerned with (quantitative) convergence results in the absence of compactness. Naturally, such results in turn require additional restrictions on the problem formulation given by F.

Such "regularity" conditions take various forms in the literature, and to cover these in a uniform manner that is moreover amenable to our quantitative pursuits, we follow the approach of the fundamental work of Kohlenbach, Lopéz-Acedo and Nicolae [60] and consider the rather broad notion of a modulus of regularity introduced therein.

Definition 3.13 (essentially [60]). Let $z \in \mathrm{zer} F$ and $r > 0$. A modulus of regularity for F w.r.t. $\overline{B}_r(z)$ is a function $\tau : (0, \infty) \to (0, \infty)$ such that

$$F(x) < \tau(\varepsilon) \to \mathrm{dist}(x, \mathrm{zer} F) < \varepsilon$$

for all $\varepsilon > 0$ and all $x \in \overline{B}_r(z)$.

Note that such a function τ always exists when X is proper and F is continuous (see Proposition 3.3 in [60]), although the description of τ is generally noneffective in this case.

Indeed, this notion is quite broad and, for various types of problems, covers a variety of different notions usually considered in the literature to induce (strong) convergence of Fejér monotone sequences, with simple moduli. We give a brief overview in the following examples as it pertains to our present work:

Example 3.14 ([60])**.** Let (X, d) be a metric space and $T : X \to X$ be a given mapping. We are interested in the problem of finding a fixed point $Tx = x$ of T. This problem can be equivalently expressed as a zero problem by setting $F(x) := d(x, Tx)$, where we then have $\mathrm{zer}F = \mathrm{Fix}T$, which we assume to be non-empty.

Situations under which we then can obtain a particularly simple and explicit moduli of regularity for the above F in particular include the following:

(1) Let T be a quasi-contraction, i.e. for some $z \in \mathrm{Fix}T$ and $c \in [0, 1)$ it holds that

$$d(Tx, z) \leqslant cd(x, z)$$

for all $x \in X$. Then $\tau(\varepsilon) := (1-c)\varepsilon$ is a modulus of regularity (and in fact of uniqueness) for F w.r.t. $\overline{B}_r(z)$, for any $r > 0$.

(2) Let T be a continuous orbital contraction, i.e. T is continuous and there is a $c \in [0, 1)$ such that

$$d(Tx, T^2 x) \leqslant cd(x, Tx)$$

for all $x \in X$. Then $\tau(\varepsilon) := (1 - c)\varepsilon$ is a modulus of regularity for F w.r.t. $\overline{B}_r(z)$, for any $r > 0$ and $z \in \mathrm{Fix}T$. In particular, there are examples of orbital contractions that do not have unique fixed points (see [60]).

(3) Let $T : X \to C \subseteq X$ be a retraction. Then $\tau(\varepsilon) = \varepsilon$ is a modulus of regularity for F w.r.t. $\overline{B}_r(z)$, for any $r > 0$ and $z \in \mathrm{Fix}T$.

Example 3.15 ([60])**.** Let $(X, \|\cdot\|)$ be a Banach space and $A : X \to 2^X$ be a given set-valued operator with

$$\mathrm{dist}(0, Ax) = 0 \to x \in \mathrm{zer}A$$

for all $x \in X$ (such as when A is closed, which is e.g. the case if A is m-accretive). We are interested in the problem of solving the inclusion $0 \in Ax$ for A. Due to the above assumption, this problem can be equivalently expressed as a zero problem by setting $F(x) := \mathrm{dist}(0, Ax)$, where we then have $\mathrm{zer}F = \mathrm{zer}A$, which we assume to be non-empty.

Situations under which we then can obtain a modulus of regularity for the above F in particular include the following:

(1) Let A be a τ-strongly accretive operator, i.e.

$$\langle x - y, u - v \rangle_+ \geqslant \tau(\|x - y\|)\|x - y\|$$

for all $(x, u), (y, v) \in A$, where τ is strictly increasing and $\tau(0) = 0$, and with

$$\langle x, y \rangle_+ := \max\{j(y) \mid j \in J(x)\}$$

where J is the normalized duality mapping of X (see e.g. [89]). Then τ is a modulus of regularity for F w.r.t. $\overline{B}_r(z)$, for any $r > 0$ and $z \in \mathrm{zer}A$.

This in particular includes the case of β-strongly accretive operators, i.e. operators A where

$$\langle x - y, u - v \rangle_+ \geqslant \beta\|x - y\|^2$$

for all $(x, u), (y, v) \in A$ for some $\beta > 0$, in which case the modulus takes the form of $\tau(\varepsilon) := \beta\varepsilon$.

(2) Let A be metrically subregular at $z \in \mathrm{zer}A$, i.e. there are $k, r > 0$ such that

$$\mathrm{dist}(x, \mathrm{zer}A) \leqslant k\mathrm{dist}(0, A(x))$$

for all $x \in \overline{B}_r(z)$. Then $\tau(\varepsilon) = \varepsilon/k$ is a modulus of regularity for F w.r.t. $\overline{B}_r(z)$.

Example 3.16 ([60])**.** Let (X, d) be a metric space and $\phi : X \to (-\infty, +\infty]$ be a given function. We are interested in the problem of finding a minimizer of ϕ. Assuming that $\mu := \inf_{x \in X} \phi(x)$ exists, this problem can be equivalently expressed as a zero problem by setting $F(x) := \phi(x) - \mu$ where we then have $\mathrm{zer}F = \arg\min \phi$, which we assume to be non-empty.

Situations under which we then can obtain a modulus of regularity for the above F in particular include functions ϕ that have τ-global weak sharp minima, i.e. where it holds that

$$\phi(x) \geqslant m + \tau(\mathrm{dist}(x, \arg\min\phi))$$

for all $x \in X$, where $\tau : [0, \infty) \to [0, \infty)$ is strictly increasing with $\tau(0) = 0$. Then τ is also a modulus of regularity for F w.r.t. $\overline{B}_r(z)$, for any $r > 0$ and $z \in \arg\min\phi$. This notion of weak sharp minima is intimately related to the notion of error bounds (see e.g. the discussion in [60]).

A particular situation where a function ϕ has weak sharp minima is if it is also uniformly quasiconvex. For this, let X be a uniquely geodesic metric space and, given $x, y \in X$ and $\lambda \in [0, 1]$, write $\lambda x \oplus (1 - \lambda)y$ for the unique point z on the geodesic connecting x and y that satisfies $d(x, z) = \lambda d(x, y)$ and $d(y, z) = (1 - \lambda)d(x, y)$. Assuming that ϕ satisfies

$$\phi(\lambda x \oplus (1 - \lambda)y) \leqslant \max\{\phi(x), \phi(y)\} - \lambda(1 - \lambda)\psi(d(x, y))$$

for all $x, y \in X$ and $\lambda \in [0, 1]$ for a τ as above, one derives

$$\mu \leqslant \phi\left(\frac{1}{2}x \oplus \frac{1}{2}z\right) \leqslant \max\{\phi(x), \phi(z)\} - \frac{1}{4}\psi(d(x, z)) \leqslant \phi(x) - \frac{1}{4}\tau(d(x, z))$$

for $z \in \arg\min\phi$ and $x \in X$, which implies that ϕ has $\frac{1}{4}\tau$-global weak sharp minimas. In the case of a strongly quasiconvex function, i.e. where

$$\phi(\lambda x \oplus (1 - \lambda)y) \leqslant \max\{\phi(x), \phi(y)\} - \lambda(1 - \lambda)\frac{\rho}{2}d(x, y)^2,$$

we simply have $\tau(\varepsilon) := \frac{\rho}{2}\varepsilon^2$.

Beyond these examples, the general formulation via a function F further accommodates various other problems (see [60]). In particular, as also illustrated by the above examples, the above moduli of regularity generalize quantitative renderings of uniqueness of solutions, so-called moduli of uniqueness, to situations with multiple solutions.

Next to the regularity of the problem and the quasi-Fejér monotonicity of the dynamical system, the only further property needed to induce convergence is the lim inf-property, as before. In that context, a modulus of regularity carries further benefits than just removing an otherwise necessary compactness assumption: it in principle allows for the construction of rates of convergence, not just rates of metastability, for the dynamical system.

However, we say "in principle" here as such a construction is only possible if all involved data also already come equipped with strong enough quantitative data. In the general case considered in Theorem 3.7 above, where the convergence of the error function e was resolved *only* with a rate of metastability, we will not be able to derive a rate of convergence, just a rate of metastability, even under the assumption of a modulus of regularity. We nevertheless begin with this case since, as mentioned before, for some applications later we will only be able to derive a rate of metastability for the errors.

In that way, we arrive at the following first quantitative convergence result.

Theorem 3.17. *Let $x : [0, \infty) \to X$ be a dynamical system and assume that F has a modulus of regularity τ w.r.t. $\overline{B}_b(z)$, where $z \in \mathrm{zer}F$ and $b \geqslant d(x(t), z)$ for all $t \geqslant 0$. Assume that $G, H : [0, \infty) \to [0, \infty)$ are such that we have moduli g, h with*

$$a < g(\varepsilon) \to G(a) < \varepsilon \text{ and } H(a) < h(\varepsilon) \to a < \varepsilon$$

for all $a \geqslant 0$ and $\varepsilon > 0$. Assume that x has the $\liminf$-property w.r.t. F with bound $\varphi : (0, \infty) \times \mathbb{N} \to \mathbb{N}$, i.e.

$$\forall \varepsilon > 0 \, \forall n \in \mathbb{N} \exists t \in [n, \varphi(\varepsilon, n)] \, (F(x(t)) < \varepsilon) \, ,$$

and further that x is (G, H)-quasi-Fejér monotone w.r.t. F, i.e.

$$\forall y \in \mathrm{zer}F \, \forall t \geqslant s \, (H(d(x(t), y)) \leqslant G(d(x(s), y)) + e(s, t)) \, .$$

Lastly, let $e(s, t) \to 0$ as $s \leqslant t \to \infty$ with a rate of metastability η, i.e.

$$\forall \varepsilon > 0 \, \forall f : \mathbb{N} \to \mathbb{N} \, \exists n \leqslant \eta(\varepsilon, f) \, \forall s \leqslant t \in [n, n + f(n)] \, (|e(s, t)| < \varepsilon) \, .$$

   Then x satisfies

$$\forall \varepsilon > 0 \, \forall f : \mathbb{N} \to \mathbb{N} \, \exists n \leqslant \rho(\varepsilon, f) \, \forall t \in [n, n + f(n)] \, (\mathrm{dist}(x(t), \mathrm{zer}F) < \varepsilon)$$

and further x is Cauchy with

$$\forall \varepsilon > 0 \, \forall f : \mathbb{N} \to \mathbb{N} \, \exists n \leqslant \rho(\varepsilon/2, f) \, \forall s, t \in [n, n + f(n)] \, (d(x(t), x(s)) < \varepsilon) \, .$$

   Here, ρ is defined by

$$\rho(\varepsilon, f) := \max\{\varphi(\tau(g(h(\varepsilon)/2)), n) \mid n \leqslant \eta(h(\varepsilon)/2, f_{\varphi, h(\varepsilon)})\} + 1$$

where

$$f_{\varphi, \delta}(n) = \max\{m + 1 + f(m + 1) \mid m \leqslant \varphi(\tau(g(h(\varepsilon)/2)), n)\} \dot- n.$$

Proof. Let $n_0 \leqslant \eta(h(\varepsilon)/2, f_{\varphi, h(\varepsilon)})$ be such that

$$\forall s \leqslant t \in [n_0, n_0 + f_{\varphi, h(\varepsilon)}(n_0)] \, (|e(s, t)| < h(\varepsilon)/2) \, .$$

Take now $t_0 \in [n_0, \varphi(\tau(f(h(\varepsilon)/2)), n_0)]$ such that $F(x(t_0)) < \tau(g(h(\varepsilon)/2))$. Define $n = \lfloor t_0 \rfloor$ and note that $n_0 \leqslant n$ as $n_0 \leqslant t_0$ and n_0 is a natural number. Further, we have $n \leqslant t_0 \leqslant \varphi(\tau(g(h(\varepsilon)/2)), n_0)$ and so

$$n + 1 + f(n + 1) \leqslant n_0 + f_{\varphi, h(\varepsilon)}(n_0).$$

Thus, we have $|e(s, t)| < h(\varepsilon)/2$ for all

$$s \leqslant t \in [n, n + 1 + f(n + 1)] \subseteq [n_0, n_0 + f_{\varphi, h(\varepsilon)}(n_0)].$$

Using that τ is a modulus of regularity over $\overline{B}_b(z)$, which contains x, we have $\mathrm{dist}(x(t_0), \mathrm{zer}F) < g(h(\varepsilon)/2)$. Thus, there exists a $y \in \mathrm{zer}F$ such that $d(x(t_0), y) < g(h(\varepsilon)/2)$ and so we have $G(d(x(t_0), y)) < h(\varepsilon)/2$. Using the quasi-Fejér monotonicity, we have

$$H(d(x(t), y)) \leqslant G(d(x(t_0), y)) + e(t_0, t) < h(\varepsilon)/2 + |e(t_0, t)|$$

for all $t \geqslant t_0 \geqslant \lfloor t_0 \rfloor = n$. Thus, we get $H(d(x(t), y)) < h(\varepsilon)$ and so

$$\mathrm{dist}(x(t), \mathrm{zer}F) \leqslant d(x(t), y) < \varepsilon$$

for all $t \in [n + 1, n + 1 + f(n + 1)]$. This yields the first claim. For the second, note simply that choosing $\varepsilon/2$ instead of ε in the above then yields $d(x(t), y) < \varepsilon/2$ and so

$$d(x(t), x(s)) \leqslant d(x(t), y) + d(x(s), y) < \varepsilon$$

for all $t, s \in [n + 1, n + 1 + f(n + 1)]$. $\qquad\square$

The above result extends (parts of) Theorem 4.1 in [60] as well as Theorem 4.1 in [80], the analogous results for discrete-time sequences without and with errors, respectively, and in fact closely follows the proof strategy given there.

The following result now improves the conclusion of the above to a full rate of convergence, provided e already comes equipped with one.

Theorem 3.18. *Under the assumptions of Theorem 3.17, assume that $e(s,t) \to 0$ as $s \leqslant t \to \infty$ with a rate of convergence η, i.e.*

$$\forall \varepsilon > 0 \; \forall t \geqslant s \geqslant \eta(\varepsilon) \, (|e(s,t)| < \varepsilon) \, .$$

Then x satisfies

$$\forall \varepsilon > 0 \; \forall t \geqslant \rho(\varepsilon) \, (\mathrm{dist}(x(t), \mathrm{zer}F) < \varepsilon)$$

and further x is Cauchy with

$$\forall \varepsilon > 0 \; \forall s, t \geqslant \rho(\varepsilon/2) \, (d(x(t), x(s)) < \varepsilon) \, ,$$

where ρ is now defined by $\rho(\varepsilon) := \varphi(\tau(g(h(\varepsilon)/2)), \eta(h(\varepsilon)/2)) + 1$.

If $e \equiv 0$, then it suffices that x has the approximate F-point property with bound $\varphi : (0, \infty) \to \mathbb{N}$, i.e.

$$\forall \varepsilon > 0 \; \exists t \leqslant \varphi(\varepsilon) \, (F(x(t)) < \varepsilon) \, ,$$

where then $\rho(\varepsilon) := \varphi(\tau(g(h(\varepsilon)))) + 1$ suffices for the above claims.

Lastly, if X is complete and $\mathrm{zer}F$ is closed, then in each of the above cases, $x(t)$ converges to some point in $\mathrm{zer}F$ with rate $\rho(\varepsilon/2)$, with the respective ρ above.

Proof. As η is a rate of convergence, it follows immediately that $\widehat{\eta}(\varepsilon, f) := \eta(\varepsilon)$ is also a rate of metastability. Theorem 3.17 then yields a rate of metastability of the form

$$\rho(\varepsilon, f) := \max\{\varphi(\tau(g(h(\varepsilon)/2)), n) \mid n \leqslant \eta(h(\varepsilon)/2)\} + 1.$$

As ρ is independent of f, it follows that the resulting function is also a rate of convergence in both cases (see e.g. Proposition 2.6 in [62]). That the above optimized rate $\rho(\varepsilon) := \varphi(\tau(g(h(\varepsilon)/2)), \eta(h(\varepsilon)/2)) + 1$, which dispenses of the maximum, is already sufficient follows by a simple inspection of the proof of Theorem 3.17.

If the errors are constantly 0, then $\eta \equiv 0$ is a corresponding rate of convergence. The result again follows from Theorem 3.17 as above. That, in the case of vanishing errors, an approximate F-point bound suffices immediately follows from the proof of Theorem 3.17, as does the fact that the above optimized rate $\rho(\varepsilon) := \varphi(\tau(g(h(\varepsilon)))) + 1$, which dispenses of the division by 2, suffices. The last claim is immediate. $\qquad\square$

We could have similarly slightly optimized the bounds in Theorem 3.9 as we have done for the above Theorem 3.18, but have chosen to only carry this out in the context of rates of convergence where optimality is much more critical.

We also here want to remind of Remark 3.11, where it is discussed how the boundedness assumption featuring in Theorem 3.17 can be derived from the (G, H)-quasi-Fejér monotonicity under suitable assumptions.

We now want to recall Remark 3.12, were the relationship between quantitative results for the discrete-time setting and the continuous-time setting have been discussed before. Contrary to the previous general results on rates of metastability, at least parts of the present results on rates of convergence under regularity assumptions can indeed be derived from corresponding results for rates of convergence in the discrete-time setting as given in [60] (see also [80]), at least under additional assumptions (such as suitable, e.g. vanishing, errors or in the context of rates of asymptotic regularity w.r.t. F).

As all these alternative proofs nevertheless retain a certain amount of pre- and postprocessing, which makes them practically not much more optimal than the tailored arguments presented above, we do not spell this out in detail for the previous theorems, and instead focus on one particular additional result for fast rates of convergence that we can derive in that way, which will be useful for our applications later.

Lemma 3.19 (essentially Theorem 4.5 in [60]). *Let (x_n) be a sequence in X and assume that F has a modulus of regularity $\tau(\varepsilon) = k\varepsilon$ for $k > 0$ w.r.t. $\overline{B}_b(z)$, where $z \in \mathrm{zer}F$ and $b \geqslant d(x_0, z)$. Assume further*

$$\forall y \in \mathrm{zer}F \ \forall n \in \mathbb{N} \left(d^p(x_{n+1}, y) \leqslant d^p(x_n, y) - \beta(F(x_{n+1}))^p \right).$$

Then

$$\mathrm{dist}(x_n, \mathrm{zer}F) \leqslant c^n \mathrm{dist}(x_0, \mathrm{zer}F) \ \text{and} \ d(x_n, x_{n+k}) \leqslant 2c^n \mathrm{dist}(x_0, \mathrm{zer}F)$$

for all $n, k \in \mathbb{N}$, where $c := (1 + \beta k^p)^{-1/p} \in (0, 1)$.

Note that the above result is slightly different to Theorem 4.5 presented in [60], which features $\beta(F(x_n))^p$ instead of $\beta(F(x_{n+1}))^p$. By inspecting the proof given in [60], it can however easily be seen that the above Lemma holds true as well.

In any way, a lift to the continuous setting yields the following:

Theorem 3.20. *Let $x : [0, +\infty) \to X$ be a dynamical system and assume that F has a modulus of regularity $\tau(\varepsilon) = k\varepsilon$ for $k > 0$ w.r.t. $\overline{B}_b(z)$, where $z \in \mathrm{zer}F$ and $b \geqslant d(x(0), z)$. Assume further*

$$\forall y \in \mathrm{zer}F \ \forall n \in \mathbb{N} \left(d^p(x(n + 1), y) \leqslant d^p(x(n), y) - \beta(F(x(n + 1)))^p \right)$$

and that x is Fejér monotone w.r.t. F. Then

$$\mathrm{dist}(x(t), \mathrm{zer}F) \leqslant c^{\lfloor t \rfloor} \mathrm{dist}(x_0, \mathrm{zer}F) \ \text{and} \ d(x(t), x(t + s)) \leqslant 2c^{\lfloor t \rfloor} \mathrm{dist}(x_0, \mathrm{zer}F)$$

for all $t, s \geqslant 0$, where $c := (1 + \beta k^p)^{-1/p} \in (0, 1)$.

Proof. The above Lemma 3.19 in combination with the first assumption in particular yields $\mathrm{dist}(x(n), \mathrm{zer}F) \leqslant c^n \mathrm{dist}(x(0), \mathrm{zer}F)$ for all $n \in \mathbb{N}$. Now, for arbitrary $y \in \mathrm{zer}F$, the Fejér monotonicity yields

$$\mathrm{dist}(x(t), \mathrm{zer}F) \leqslant d(x(t), y) \leqslant d(x(\lfloor t \rfloor), y),$$

so that

$$\mathrm{dist}(x(t), \mathrm{zer}F) \leqslant \mathrm{dist}(x(\lfloor t \rfloor), \mathrm{zer}F) \leqslant c^{\lfloor t \rfloor} \mathrm{dist}(x(0), \mathrm{zer}F).$$

Further, we now also have

$$d(x(t), x(t + s)) \leqslant d(x(t), y) + d(x(t + s), y) \leqslant 2d(x(t), y)$$

and so get

$$d(x(t), x(t + s)) \leqslant 2\mathrm{dist}(x(t), \mathrm{zer}F) \leqslant 2c^{\lfloor t \rfloor} \mathrm{dist}(x(0), \mathrm{zer}F). \qquad \square$$

The above result can also be extended to incorporate sufficiently fastly decreasing errors, in which case it can no longer be lifted from the discrete-time results but has to be re-derived. We do not spell this out here any further.

4. Applications to dynamical systems

To illustrate the breadth of the above rather general approach, we are now concerned with applying the previous results to concrete dynamical systems from the literature.

Two examples will focus on dynamical systems over Hilbert spaces. The first presents a quantitative analysis of the work [23] by Boţ and Csetnek for a first-order dynamical system over a nonexpansive operator. In particular, we derive rates of convergence under a metric regularity in that case, which in the special case of a continuous variant of a forward-backward scheme extend those derived in [25] using quite different methods to a broader setting. The second presents a quantitative analysis of the work [22] by Boţ and Csetnek for a second-order dynamical system over a cocoercive operator. In particular, this system presents an example where our approach, even in the context of metric regularity assumptions, cannot provide rates of convergence as we can only provide a rate of metastability for the respective error function. This, we think, in particular presents a novel structural reason for why no rates of convergence have appeared for that particular scheme, at that level of generality, which provides deeper insight into the nature of the scheme presented in [22]. In particular, the present rates of metastability seem to be the first structural quantitative results for the method presented in [22]. To illustrate the metric generality, our last case study is concerned with (generalized) gradient flows and associated semigroups in Hadamard spaces, that is geodesic metric spaces of nonpositive curvature. In particular, we study both the classical example of the gradient flow semigroup in these spaces, following the well-known work of Mayer [71] and Bačák [16], and then further provide a first quantitative study of a related nonlinear semigroup generated by a nonexpansive mapping due to Stojkovic [87] and later studied by Bačák and Reich [18]. In both cases, the quantitative results we provide are either completely novel, in particular relating to general rates of metastability, or extend previously known results on rates of convergence from specific ad-hoc assumption to broader metric regularity conditions.

4.1. **First-order systems over nonexpansive operators.** We begin with the essential setup from [23]. Let X be a Hilbert space with norm $\| \cdot \|$ and inner product $\langle \cdot, \cdot \rangle$ and let $T : X \to X$ be a nonexpansive map. For a Lebesgue measurable function $\lambda : [0, +\infty) \to [0, 1]$ and $x_0 \in X$, we consider the dynamical system

$(*)_1$
$$\begin{cases} \dot{x}(t) = \lambda(t) \left(T(x(t)) - x(t) \right), \\ x(0) = x_0. \end{cases}$$

This system is motivated by the first-order dynamical system

$$\dot{x} + x = \mathrm{prox}_{\mu f}(x - \mu B(x))$$

for a given proper, convex and lower-semicontinuous function f, where $\mathrm{prox}_{\mu f}$ is the corresponding proximal map, and a cocoercive operator B, as studied by Abbas and Attouch [1] (and in fact captures that system, if extended appropriately to averaged maps). This system in turn arose as an abstraction of the first-order dynamical system

$$\dot{x} + x = P_C(x - \mu \nabla \phi(x))$$

for a given closed convex nonempty set C and a convex C^1 function ϕ, first studied by Bolte [32] (see also [6]), and indeed captures an even more general continuous-time variant of the forward-backward method as discussed later. We refer to the survey [42] for further discussions along these lines.

Following [23] (see Definition 2 therein), we call a dynamical system $x : [0, +\infty) \to X$ a strong global solution of $(*)_1$ if x satisfies the respective equation almost everywhere on

$[0, +\infty)$ and x is absolutely continuous on each interval $[0, b]$ with $b \in (0, \infty)$. The asymptotic behavior of such strong global solutions is studied in [23] (see Theorem 6 therein), in particular establishing a continuous form of asymptotic regularity of the dynamical system together with weak convergence towards a fixed point of T. In this section, we want to study these results from a quantitative perspective.

As discussed in Remark 8 in [23], a discretization of the above equation w.r.t. the time variable in particular yields the well-known Krasnosel'skii-Mann method [65, 68] (see also [15]). Also, the above system immediately allows for extensions to averaged mappings (see Corollary 9 in [23] and the surrounding discussions).

At first, it follows from the above assumptions that the system $(*)_1$ has a unique strong global solution (see the discussion given in Section 2 in [23]). Further, that solution is immediately bounded (see item (i) of [23, Theorem 6]). We recall that result here, and in particular provide a more quantitative perspective on it in the following lemma:

Lemma 4.1. *Let $y \in \mathrm{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Let x be the unique strong global solution to $(*)_1$. Then x is bounded with $\|x(t) - y\| \leqslant b$.*

Proof. As in [23], Eq. (8), we have

$$\frac{d}{dt}\|x(t) - y\|^2 + \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 + \|\dot{x}(t)\|^2 \leqslant 0$$

so that, as $\lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2$ and $\|\dot{x}(t)\|^2$ are nonnegative, we have

$$\frac{d}{dt}\|x(t) - y\|^2 \leqslant 0$$

so that $t \mapsto \|x(t) - y\|$ is decreasing. In particular, we have

$$\|x(t) - y\| \leqslant \|x(0) - y\| \leqslant b$$

for all $t \geqslant 0$. $\qquad\square$

To capture the corresponding fixed point problem in the setup of the previous sections, we set $X_0 = \overline{B}_b(y)$ and

$$F(z) = \begin{cases} \|z - Tz\| & \text{if } z \in X_0, \\ +\infty & \text{otherwise,} \end{cases}$$

for a fixed solution $y \in \mathrm{Fix}T$ so that $\mathrm{zer}F = \mathrm{Fix}T \cap X_0$ and

$$\mathrm{lev}_{\leqslant\varepsilon}F = \{z \in X_0 \mid \|z - Tz\| \leqslant \varepsilon\}.$$

We now turn to the asymptotic regularity of the dynamical system, that is the property that

$$\lim_{t\to\infty}\|T(x(t)) - x(t)\| = 0.$$

A quantitative version of that property given in the following will in particular ensure that the dynamical system x has approximate F-points for the F defined above, so that in particular our previous general and abstract results apply to it. Generally of course, results of the above form and their quantitative renderings are also of independent interest, and rates of asymptotic regularity for the dynamical system x as above in particular provide continuous-time analogs of respective rates for the so-called T-asymptotic regularity of the Krasnosel'skii-Mann method (x_n), that is rates for the respective limit $\lim_{n\to\infty}\|Tx_n - x_n\| = 0$ (see e.g. [53, 54]). Further, note that by x solving the equations $(*)_1$, we in particular always get

$$\|\dot{x}(t)\| \leqslant \|T(x(t)) - x(t)\|,$$

so that any quantitative rendering of the latter also provides a similar such result for the former.

In [23] (see item (ii) of Theorem 6 therein), this result is now established under two different (and in fact independent [23, Remark 7]) conditions:

$$\int_0^\infty \lambda(t)(1 - \lambda(t))dt = +\infty \text{ and } \inf_{t \geqslant 0} \lambda(t) > 0.$$

We now provide quantitative versions of these results by giving rates of convergence in each case. We begin with the former assumption:

Lemma 4.2. *Let* $y \in \mathrm{Fix}\,T$ *and let* $b \in \mathbb{N}$ *with* $\|x_0 - y\| \leqslant b$. *Let further* $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))\,dt = +\infty$ *with a corresponding divergence modulus* $\eta : (0, \infty) \to \mathbb{N}$, *i.e. where*

$$\int_0^r \lambda(t)(1 - \lambda(t))\,dt \geqslant K$$

for any $r \geqslant \eta(K)$ *and* $K > 0$. *Then* x *is* T-*asymptotically regular, i.e.*

$$\forall \varepsilon > 0 \ \forall t \geqslant \varphi(\varepsilon) \left(\|T(x(t)) - x(t)\| \leqslant \varepsilon \right)$$

with rate $\varphi(\varepsilon) := \eta(b^2/\varepsilon^2)$. *In particular,* x *has approximate F-points with modulus* φ.

Proof. As in [23], Eq. (8), we have

$$\frac{d}{dt}\|x(t) - y\|^2 + \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 + \|\dot{x}(t)\|^2 \leqslant 0.$$

This yields, as $\|\dot{x}(t)\|^2$ is nonnegative, that

$$\lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 \leqslant -\frac{d}{dt}\|x(t) - y\|^2$$

and so that

$$
\begin{aligned}
\int_0^r \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2\,dt &\leqslant -\int_0^r \frac{d}{dt}\|x(t) - y\|^2\,dt \\
&= \|x(0) - y\|^2 - \|x(r) - y\|^2 \\
&\leqslant \|x(0) - y\|^2 \leqslant b^2
\end{aligned}
$$

for $r \geqslant 0$. As $t \mapsto \|T(x(t)) - x(t)\|^2$ is nonincreasing, we get

$$\|T(x(r)) - x(r)\|^2 \int_0^r \lambda(t)(1 - \lambda(t))\,dt \leqslant \int_0^r \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2\,dt \leqslant b^2.$$

For $r \geqslant \eta(b^2/\varepsilon^2) = \varphi(\varepsilon)$, we get

$$\|T(x(r)) - x(r)\|^2 \cdot \frac{b^2}{\varepsilon^2} \leqslant \|T(x(r)) - x(r)\|^2 \int_0^r \lambda(t)(1 - \lambda(t))\,dt \leqslant b^2$$

and so $\|T(x(r)) - x(r)\| \leqslant \varepsilon$. That x has approximate zeros w.r.t. F with the same modulus follows from Lemma 4.1. $\qquad\qquad\square$

A particular assumption under which not only $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))\,dt = +\infty$ can be guaranteed but where a corresponding divergence modulus η can be explicitly constructed is if

$$\underline{\tau} = \inf_{t \geqslant 0} \lambda(t)(1 - \lambda(t)) > 0.$$

Then $\eta(K) := \frac{K}{\underline{\tau}}$ is a corresponding divergence modulus. In particular, the above Lemma 4.2 yields

$$\|T(x(t)) - x(t)\| \leqslant \frac{b}{\sqrt{\underline{\tau}t}}$$

for any $t \geqslant 0$ and any $b \geqslant \|x_0 - y\|$. In particular, for $b = \|x_0 - y\|$, and taking the infimum over y, we obtain

$$\|T(x(t)) - x(t)\| \leqslant \frac{d(x_0, \mathrm{Fix}T)}{\sqrt{\tau t}},$$

which was already obtained as Theorem 10 in [23]. In fact, Boţ and Csetnek afterwards show that the convergence can be improved to $o(\frac{1}{\sqrt{t}})$ (see Theorem 11 in [23]), but we will not be concerned with this further here.

We now turn to the alternative assumption $\inf_{t \geqslant 0} \lambda(t) > 0$ for the asymptotic regularity of the dynamical system, which thereby essentially presents itself as a weakening of the above property.

Lemma 4.3. *Let* $y \in \mathrm{Fix}T$ *and let* $b \in \mathbb{N}$ *with* $\|x_0 - y\| \leqslant b$. *Let further* $\inf_{t \geqslant 0} \lambda(t) > 0$ *with a corresponding witness* $\underline{\lambda} > 0$, *i.e. where*

$$\lambda(t) \geqslant \underline{\lambda} \text{ for all } t \in [0, \infty).$$

Then x *is* T-*asymptotically regular, i.e.*

$$\forall \varepsilon > 0 \; \forall t \geqslant \varphi(\varepsilon) \left(\|T(x(t)) - x(t)\| \leqslant \varepsilon \right)$$

with rate $\varphi(\varepsilon) := 4b^4/\underline{\lambda}^2 \varepsilon^2$. *In particular,* x *has approximate* F-*points with modulus* φ.

Proof. Completely parallel to the first part of the proof of Lemma 4.2, we get

$$\int_0^\infty \|\dot{x}(t)\| \, dt \leqslant b^2.$$

Given that x is the unique global solution of $(*)_1$, this yields

$$\int_0^\infty \frac{1}{2} \|T(x(t)) - x(t)\|^2 \, dt = \int_0^\infty \frac{\|\dot{x}(t)\|^2}{2\lambda(t)^2} \, dt \leqslant \frac{b^2}{2\underline{\lambda}^2}.$$

As in the proof of Theorem 6 in [23], for almost all t we also have

$$\frac{d}{dt} \left(\frac{1}{2} \|T(x(t)) - x(t)\|^2 \right) \leqslant 0.$$

For any $\varepsilon > 0$ and with $f \equiv 0$, we can now apply Lemma 2.4 to get

$$\exists n \leqslant \tilde{f}'^{(\varpi(\varepsilon))}(0) \left(\frac{1}{2} \|T(x(n)) - x(n)\|^2 < \frac{\varepsilon^2}{2} \right),$$

where $\tilde{f}'(n) := n + \lceil 3b^2/2\underline{\lambda}^2 \varepsilon \rceil$ as $f \equiv 0$, and $\varpi(\varepsilon) := \lceil b^2/\varepsilon \rceil$. In particular, we have

$$\tilde{f}'^{(\varpi(\varepsilon))}(0) = \varpi(\varepsilon) \left\lceil \frac{3b^2}{2\underline{\lambda}^2 \varepsilon} \right\rceil \leqslant \frac{2b^2}{\varepsilon} \cdot \frac{2b^2}{\underline{\lambda}^2 \varepsilon} = \frac{4b^4}{\underline{\lambda}^2 \varepsilon^2}.$$

As $t \mapsto \|T(x(t)) - x(t)\|^2$ is nonincreasing, the latter even holds for all $t \geqslant n$, which yields the result. $\square$

Fejér monotonicity of the unique strong global solution x to the system $(*)_1$ w.r.t. F now comes to pass in exactly the way discussed in Section 2, through a differential inequality with an associated energy functional. Indeed, in [23] (see eq. (8) therein), it is established that

$$\frac{d}{dt} \|x(t) - y\|^2 \leqslant 0 \text{ for all } y \in \mathrm{Fix}T,$$

so that the inequality here is even of the most simple form discussed in Section 2. We now analyse the proof of this inequality as given in [23] to establish the associated uniform Fejér monotonicity and extract an associated modulus. Indeed, the following result now provides

a quantitative variant of the above differential inequality, relating approximate solutions to approximate inequalities.

Lemma 4.4. *Let $y \in \operatorname{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. For any $\varepsilon > 0$, if $z \in \operatorname{lev}_{\leqslant \varepsilon/4b}F$, then $\frac{d}{dt}\|x(t) - z\|^2 \leqslant \varepsilon$ for all $t \geqslant 0$.*

Proof. Take $z \in \operatorname{lev}_{\leqslant \varepsilon/4b}F$. Analogous to the proof of Theorem 6 in [23], we have

$$
\begin{aligned}
\frac{d}{dt}\|x(t) - z\|^2 &= 2\langle \dot{x}(t), x(t) - z \rangle \\
&= \|\dot{x}(t) + x(t) - z\|^2 - \|x(t) - z\|^2 - \|\dot{x}(t)\|^2 \\
&\leqslant \|\lambda(t)\,(Tx(t) - z) + (1 - \lambda(t))\,(x(t) - z)\,\|^2 - \|x(t) - z\|^2 \\
&\leqslant \|Tx(t) - z\|^2 - \|x(t) - z\|^2 \\
&= (\|Tx(t) - z\| - \|x(t) - z\|)(\|Tx(t) - z\| + \|x(t) - z\|).
\end{aligned}
$$

As $z \in X_0$, we have

$$
\|x(t) - z\| \leqslant \|x(t) - z\| + \|z - y\| \leqslant \|x_0 - z\| + \|z - y\| \leqslant 2b
$$

as well as

$$
\|Tx(t) - z\| \leqslant \|Tx(t) - y\| + \|z - y\| \leqslant \|x(t) - y\| + \|z - y\| \leqslant \|x_0 - z\| + \|z - y\| \leqslant 2b
$$

using the nonexpansivity of T and Lemma 4.1. Thus, we have

$$
\begin{aligned}
\frac{d}{dt}\|x(t) - z\|^2 &\leqslant (\|Tx(t) - z\| - \|x(t) - z\|)4b \\
&\leqslant (\|Tx(t) - Tz\| + \|Tz - z\| - \|x(t) - z\|)4b \\
&\leqslant \|Tz - z\|4b \leqslant \varepsilon
\end{aligned}
$$

as $\|Tz - z\| \leqslant \varepsilon/4b$. $\qquad\square$

An approximate differential inequality like the above then yields a modulus of uniform Fejér monotonicity (with $G = H = (\cdot)^2$) via the mean value theorem:

Lemma 4.5. *Let $y \in \operatorname{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Then x is uniformly (G, H)-Fejér monotone w.r.t. F and with $G = H = (\cdot)^2$, i.e.*

$$
\forall \varepsilon > 0 \ \forall n, m \in \mathbb{N} \ \forall z \in \operatorname{lev}_{\leqslant \chi(\varepsilon, n, m)}F \ \forall s \leqslant t \in [n, n + m] \ \left(\|x(t) - z\|^2 \leqslant \|x(s) - z\|^2 + \varepsilon\right),
$$

where $\chi(\varepsilon, n, m) := \varepsilon/(4bm)$.

Proof. Fix ε, n, m as well as $z \in \operatorname{lev}_{\leqslant \chi(\varepsilon, n, m)}F$ and $s, t \in [n, n + m]$. If $s = t$, the result is clear. If $s < t$, then the mean value theorem yields an $r \in (s, t)$ such that

$$
\frac{d}{dt}\|x(r) - z\|^2 = \frac{\|x(t) - z\|^2 - \|x(s) - z\|^2}{t - s}.
$$

Thus, as $z \in \operatorname{lev}_{\leqslant \chi(\varepsilon, n, m)}F$, we have

$$
\|x(t) - z\|^2 - \|x(s) - z\|^2 \leqslant \frac{\varepsilon}{m}(t - s) \leqslant \frac{\varepsilon}{m} \cdot m = \varepsilon
$$

by Lemma 4.4. $\qquad\square$

A direct application of Theorem 3.9 now yields the following quantitative result on the behavior of the solution x to $(*)_1$ under a finite-dimensionality (that is, local compactness) assumption on X, giving a rate of metastability for the convergence result established in Theorem 6 of [23] in that case:

Theorem 4.6. *Let X be a finite-dimensional Hilbert space with dimension d and let $T : X \to X$ be a nonexpansive map. Let $y \in \mathrm{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Further, assume that $\lambda : [0, +\infty) \to [0,1]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)_1$. Assume either:*

(1) $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))\, dt = +\infty$ with a divergence modulus $\eta : (0, \infty) \to \mathbb{N}$ with $\eta(K) \geqslant \eta(L)$ for $K \geqslant L$, in which case we set $\varphi(\varepsilon) := \eta(b^2/\varepsilon^2)$.

(2) $\inf_{t \geqslant 0} \lambda(t) > 0$ with a witness $\underline{\lambda} > 0$, in which case we set $\varphi(\varepsilon) := 4b^4/\underline{\lambda}^2\varepsilon^2$.

Then, in either case, x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon/4, f)\; \forall s, t \in [n, n + f(n)]\, (\|x(s) - x(t)\| \leqslant \varepsilon \text{ and } \|x(t) - T(x(t))\| \leqslant \varepsilon)\,,$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described by

$$\Delta(\varepsilon, f) := \max\{\Delta(i, \varepsilon, f) \mid i \leqslant P\} + 1$$

where $P := \left\lceil 2(\lceil \sqrt{12}\varepsilon^{-1}\rceil + 1)\sqrt{d}b \right\rceil^d + 1$ and $\Delta(0, \varepsilon, f) := 0$ and $\Delta(j, \varepsilon, f) := \varphi(\widehat{\varepsilon}_j)$ with

$$\widehat{\varepsilon}_j = \min\left\{\frac{\varepsilon}{2}, \frac{\varepsilon^2/12}{4b(f(m+1)+1)} \;\middle|\; m \leqslant \Delta(i, \varepsilon, f), i < j\right\}$$

for $j \geqslant 1$.

Proof. The result immediately follows by instantiating Theorem 3.9 over $X_0 = \overline{B}_b(y)$ and simplifying the resulting bound (recall that $x(t) \in X_0$ for all $t \geqslant 0$ by Lemma 4.1). In particular, we set $G := H := (\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon}$. That x is uniformly (G, H)-Fejér monotone w.r.t. F follows by Lemma 4.5, which in particular yields that $\chi(\varepsilon, n, m) := \varepsilon/(4bm)$ is an associated modulus. The definition of $\widehat{\varepsilon}_j$ is a resulting simplification. As we do not deal with errors, it suffices that x has the approximate F-point property. The above assumptions (1) and (2) guarantee that with respective bounds, using Lemmas 4.2 and 4.3. Note that the respective φ is immediately monotone under our assumptions. Also, note that X_0 is totally bounded with a modulus $\gamma(\varepsilon) := \left\lceil 2(\lceil \varepsilon^{-1}\rceil + 1)\sqrt{d}b \right\rceil^d$ by Example 2.8 of [59]. Lastly, note that F is uniformly continuous on X_0 with modulus $\omega(\varepsilon) := \varepsilon/2$, since we have $\|\|z - Tz\| - \|z' - Tz'\|\| \leqslant 2\|z - z'\|$ using the nonexpansivity of T. $\qquad\square$

Further, a direct application of Theorem 3.18 yields the following general construction of a rate of convergence under a regularity assumption on the mapping T and its associated fixed point set $\mathrm{Fix}T$:

Theorem 4.7. *Let X be a Hilbert space and let $T : X \to X$ be a nonexpansive map. Let $y \in \mathrm{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Assume that F has a modulus of regularity τ w.r.t. $\overline{B}_b(y)$. Further, assume that $\lambda : [0, +\infty) \to [0,1]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)_1$. Assume either:*

(1) $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))\, dt = +\infty$ with a divergence modulus $\eta : (0, \infty) \to \mathbb{N}$ with $\eta(K) \geqslant \eta(L)$ for $K \geqslant L$, in which case we set $\varphi(\varepsilon) := \eta(b^2/\varepsilon^2)$.

(2) $\inf_{t \geqslant 0} \lambda(t) > 0$ with a witness $\underline{\lambda} > 0$, in which case we set $\varphi(\varepsilon) := 4b^4/\underline{\lambda}^2\varepsilon^2$.

Then, in either case, x satisfies

$$\forall \varepsilon > 0\; \forall t \geqslant \rho(\varepsilon)\, (\mathrm{dist}(x(t), \mathrm{Fix}T) < \varepsilon)$$

and further x is Cauchy with

$$\forall \varepsilon > 0\; \forall s, t \geqslant \rho(\varepsilon/2)\, (\|x(t) - x(s)\| < \varepsilon)\,,$$

where ρ is defined by $\rho(\varepsilon) := \varphi(\tau(\varepsilon)/2) + 1$. In particular, x converges to some fixed point of T with that rate.

Proof. The result immediately follows by instantiating Theorem 3.18 and simplifying the resulting bound. In particular, we set $G := H := (\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon}$. That x is (G, H)-Fejér monotone w.r.t. F follows by eq. (8) in [23] (recall also Lemma 4.5). As we do not deal with errors, it suffices that x has the approximate F-point property. The above assumptions (1) and (2) guarantee that with respective bounds $\varphi(\varepsilon/2)$, using Lemmas 4.2 and 4.3 as before. $\square$

In the special case of operators T with linear moduli of regularity, we can further employ the previous Theorem 3.20 to derive exponentially fast rates of convergence for the solution of $(*)_1$:

Theorem 4.8. *Let X be a Hilbert space and let $T : X \to X$ be a nonexpansive map. Let $y \in \mathrm{Fix}T$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Further, assume that $\lambda : [0, +\infty) \to [0, 1]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)_1$. Assume that F has a modulus of regularity $\tau(\varepsilon) = k\varepsilon$ for $k > 0$ w.r.t. $\overline{B}_b(y)$ and further assume that $\underline{\tau} = \inf_{t \geqslant 0} \lambda(t)(1 - \lambda(t)) > 0$.*
Then

$$\mathrm{dist}(x(t), \mathrm{Fix}T) \leqslant c^{\lfloor t \rfloor}\mathrm{dist}(x_0, \mathrm{Fix}T) \text{ and } d(x(t), x(t + s)) \leqslant 2c^{\lfloor t \rfloor}\mathrm{dist}(x_0, \mathrm{Fix}T)$$

for all $t, s \geqslant 0$, where $c := (1 + \underline{\tau}k^p)^{-1/p} \in (0, 1)$.

Proof. As in [23] (see eq. (12) therein), for an arbitrary $y \in \mathrm{Fix}T$, we can derive

$$\frac{d}{dt}\|x(t) - y\|^2 + \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 \leqslant 0$$

for all $t \geqslant 0$. As before, this yields the Fejér monotonicity of x, but also integrating from n to $n + 1$ yields

$$\|x(n + 1) - y\|^2 \leqslant \|x(n) - y\|^2 - \underline{\tau}\int_n^{n+1} \|T(x(\tau)) - x(\tau)\|^2 \, d\tau.$$

As established in the proof of Theorem 6 in [23], $\|T(x(\tau)) - x(\tau)\|^2$ is decreasing. Thus, we have

$$\begin{aligned}
\|x(n + 1) - y\|^2 &\leqslant \|x(n) - y\|^2 - \underline{\tau}\int_n^{n+1} \|T(x(\tau)) - x(\tau)\|^2 \, d\tau \\
&\leqslant \|x(n) - y\|^2 - \underline{\tau}\|T(x(n + 1)) - x(n + 1)\|^2 \\
&\leqslant \|x(n) - y\|^2 - \underline{\tau}(F(x(n + 1)))^2.
\end{aligned}$$

Using Theorem 3.20, we get the result. $\square$

Combined with Example 3.14, the above result immediately allows for the construction of fast rates of convergence in the case where T is a quasi-contraction or a continuous orbital-contraction. However, the above is certainly not limited to that.

Let us now consider a particular instance of the system $(*)_1$ to which the above results apply: the continuous-time variant of the forward-backward method. We first recall the method from [23]: Over a Hilbert space X, consider a maximally monotone set-valued operator $A : X \to 2^X$ and a single-valued map $B : X \to X$ which is β-cocoercive, for some $\beta > 0$. For $\gamma \in (0, 2\beta)$ and $\delta := \min\{1, \beta/\gamma\} + 1/2$ as well as a Lebesgue measurable function $\lambda : [0, +\infty) \to [0, \delta]$ and a starting point $x_0 \in X$, we consider the dynamical system

$(*)_1'$
$$\begin{cases} \dot{x}(t) = \lambda(t)\left(J_{\gamma A}(x(t) - \gamma B(x(t))) - x(t)\right), \\ x(0) = x_0, \end{cases}$$

where $J_{\gamma A} := (\mathrm{Id} + \gamma A)^{-1}$ is the (single-valued and total) resolvent of γA. Indeed, the above system can be conceived of as an instantiation of the system $(*)_1$ using the map $T := J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$. As in the proof of Theorem 12 in [23], one can show that T is in fact $\frac{1}{\delta}$-averaged, that is by definition that there is a nonexpansive operator R such that

$$T = \left(1 - \frac{1}{\delta}\right)\mathrm{Id} + \frac{1}{\delta}R.$$

Thereby, the system $(*)_1'$ is further an instantiation of $(*)_1$ using the map R and the parameter sequence $\frac{1}{\delta}\lambda$ instead of λ (see also the proof of Corollary 9 in [23]). Note also that $\mathrm{Fix}\,T = \mathrm{zer}(A + B)$.

In that way, we immediately obtain the following result on the quantitative asymptotic behavior of the associated solution under a finite-dimensionality assumption on X, a quantitative variant of parts of Theorem 12 in [23]:

Theorem 4.9. *Let X be a finite-dimensional Hilbert space with dimension d and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Let $\gamma \in (0, 2\beta)$ and write $\delta := \min\{1, \beta/\gamma\} + 1/2$. Let $y \in \mathrm{zer}(A + B)$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Further, assume that $\lambda : [0, +\infty) \to [0, \delta]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)_1'$. Assume either:*

(1) $\int_0^{+\infty} \lambda(t)(\delta - \lambda(t))\,dt = +\infty$ with a divergence modulus $\eta : (0, \infty) \to \mathbb{N}$ with $\eta(K) \geqslant \eta(L)$ for $K \geqslant L$, in which case we set $\varphi(\varepsilon) := \eta(\delta b^2/\varepsilon^2)$.

(2) $\inf_{t \geqslant 0} \lambda(t) > 0$ with a witness $\underline{\lambda} > 0$, in which case we set $\varphi(\varepsilon) := 4b^4\delta^2/\underline{\lambda}^2\varepsilon^2$.

Then, in either case, x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon/4, f)\ \forall s, t \in [n, n + f(n)]\left(\|x(s) - x(t)\| \leqslant \varepsilon \ \text{and} \ \|x(t) - T(x(t))\| \leqslant \varepsilon\right),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described as in Theorem 4.6, using the above φ in either case.

Remark 4.10. Note that since $T(x) := J_{\gamma A}(x - \gamma B(x))$, we have $x - \gamma B(x) \in T(x) + \gamma A(T(x))$ by definition of the resolvent. In particular, we have $\gamma^{-1}(x - T(x)) - B(x) \in A(T(x))$. Combined, we hence obtain

$$\gamma^{-1}(x - T(x)) + B(T(x)) - B(x) \in (A + B)(T(x)).$$

Write $w := \gamma^{-1}(x - T(x)) + B(T(x)) - B(x)$ and $v := T(x)$, so that the above translates to $w \in (A + B)(v)$. Hence, if $\|x - T(x)\| \leqslant \min\{\varepsilon, \varepsilon/(\gamma^{-1} + \beta^{-1})\}$, we get $\|x - v\| = \|x - T(x)\| \leqslant \varepsilon$ as well as

$$\begin{aligned}
\|w\| &= \|\gamma^{-1}(x - T(x)) + B(T(x)) - B(x)\| \\
&\leqslant \gamma^{-1}\|x - T(x)\| + \|B(T(x)) - B(x)\| \\
&\leqslant \left(\gamma^{-1} + \beta^{-1}\right)\|x - T(x)\| \leqslant \varepsilon
\end{aligned}$$

using that B is $\frac{1}{\beta}$-cocoercive. As such, if the above Theorem 4.9 is applied with $\min\{\varepsilon, \varepsilon/(\gamma^{-1} + \beta^{-1})\}$ in place of ε, so that the property $\|x(t) - T(x(t))\| \leqslant \min\{\varepsilon, \varepsilon/(\gamma^{-1} + \beta^{-1})\}$ is maintained along the region of metastability, the dynamical system x in particular contains approximate zeros of $A + B$ in the above sense in that region, that is for any $t \in [n, n + f(n)]$, there are $w \in (A + B)(v)$ such that $\|x(t) - v\|, \|w\| \leqslant \varepsilon$. This argument already appears abstractly in [50] in the context of the discrete-time forward-backward method.

As is well-studied in the context of "ordinary" splitting methods, strong convergence can be guaranteed under a uniform monotonicity assumption for either of the operators A or B, in which case rates can also be extracted. For discrete-time dynamical systems, constructions of such rates were recently studied in very general situations in the work of Treusch and Kohlenbach [92]. For continuous-time dynamical systems, rates are presented under a strong monotonicity assumption in [25]. We here can obtain a result on rates of convergence under a uniform monotonicity assumption in a similar generality as the recent work [92], and in particular following their arguments.

First, we derive a quantitative result on the asymptotic behavior of $B(x(t))$. As such, we rely on the following key inequality also used again later:

Lemma 4.11. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Fix $\gamma > 0$ and define $T := J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$. For any $x \in \mathrm{zer}(A + B)$ and $z \in X$:*

$$\gamma\beta\|Bz - Bx\|^2 \leqslant \left(1 + \frac{\gamma}{\beta}\right)\|z - x\|\|Tz - z\|.$$

Proof. We proceed similar to the proof of Theorem 26.14 in [15], abstracting the proof of eq. (26.56) therein: For any $y \in X$, define $y^- := y - \gamma By$. Note that $Ty = J_{\gamma A}y^-$. By assumption, we also have $x = Tx = J_{\gamma A}x^-$. We then have

$$\langle Tz - x, z - Tz - \gamma(Bz - Bx)\rangle$$
$$= \langle Tz - x, z^- Tz + x - x^-\rangle$$
$$= \langle J_{\gamma A}z^- - J_{\gamma A}x^-, (\mathrm{Id} - J_{\gamma A})z^- - (\mathrm{Id} - J_{\gamma A})x^-\rangle \geqslant 0,$$

where the inequality follows from the monotonicity of A and the definition of the resolvent. Using the nonexpansivity of T and the β-cocoercivity of B, we in particular have that

$$\|z - x\|\|Tz - z\| \geqslant \|Tz - x\|\|Tz - z\|$$
$$\geqslant \langle Tz - x, z - Tz\rangle$$
$$\geqslant \gamma\langle Tz - x, Bz - Bx\rangle$$
$$= \gamma\left(\langle Tz - z, Bz - Bx\rangle + \langle z - x, Bz - Bx\rangle\right)$$
$$\geqslant -\gamma\|Tz - z\|\|Bz - Bx\| + \gamma\beta\|Bz - Bx\|^2$$
$$\geqslant -\frac{\gamma}{\beta}\|Tz - z\|\|z - x\| + \gamma\beta\|Bz - Bx\|^2,$$

where we have used the preceding inequality in the third line and in the last line in particular used that B is also $\frac{1}{\beta}$-Lipschitz. Rearranging the above inequality yields the result. $\square$

We then have the following result (cf. Theorem 3.3 in [92] for the discrete-time case, whose arguments we follows):

Proposition 4.12. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Let $\gamma \in (0, 2\beta)$ and write $\delta := \min\{1, \beta/\gamma\} + 1/2$. Let $y \in \mathrm{zer}(A + B)$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Further, assume that $\lambda : [0, +\infty) \to [0, \delta]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)'_1$. Assume either:*

(1) $\int_0^{+\infty} \lambda(t)(\delta - \lambda(t))\,dt = +\infty$ with a divergence modulus $\eta : (0, \infty) \to \mathbb{N}$ with $\eta(K) \geqslant \eta(L)$ for $K \geqslant L$, in which case we set $\varphi(\varepsilon) := \eta(\delta b^2/\varepsilon^2)$.

(2) $\inf_{t \geqslant 0} \lambda(t) > 0$ with a witness $\underline{\lambda} > 0$, in which case we set $\varphi(\varepsilon) := 4b^4\delta^2/\underline{\lambda}^2\varepsilon^2$.

Then, in either case, $\lim_{t\to\infty} B(x(t)) = B(y)$ *with*

$$\forall \varepsilon > 0 \ \forall s, t \geqslant \psi(\varepsilon) \left(\|B(x(t)) - B(y)\| \leqslant \varepsilon \right),$$

where $\psi(\varepsilon) := \varphi(\gamma \beta \varepsilon^2 / 3b)$ *in either case.*

Proof. By Lemma 4.11 and $\gamma \leqslant 2/\beta$, we obtain

$$\gamma \beta \|B(x(t)) - B(y)\|^2 \leqslant 3\|x(t) - y\| \|T(x(t)) - x(t)\|.$$

Lemma 4.1 then yields

$$\|B(x(t)) - B(y)\|^2 \leqslant \frac{3b}{\gamma\beta} \|T(x(t)) - x(t)\|$$

which implies the results using Lemmas 4.2 and 4.3. $\qquad\square$

The above result is obtained in [23] (see Theorem 12 therein) using quite a different argument, only under the assumption that $\inf_{t\geqslant 0} \lambda(t) > 0$ and in particular without any rates.

We now obtain a result on rates of convergence in the context of a uniform monotonicity assumption. Indeed, recall that a set-valued operator $M : X \to 2^X$ is called uniformly monotone if

$$\langle x - y, u - v \rangle \geqslant \phi(\|x - y\|)$$

for all $(x, u), (y, v) \in A$, where $\phi : [0, \infty) \to [0, \infty)$ is an increasing function which vanishes only at 0. Modelled after Theorem 3.4 in [92], the resulting construction of a rate of convergence relies on the following well-known result, which we only present here in the abstract as we reuse it later:[2]

Lemma 4.13. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Let $\gamma \in (0, 2\beta)$ and let $x \in X$ and $y \in \mathrm{zer}(A + B)$ with $b \in \mathbb{N}$ such that $\|x - y\| \leqslant b$. Let $\varepsilon > 0$ be arbitrary.*

(i) If A is uniformly monotone with function ϕ and

$$\|x - T(x)\| \leqslant \min\left\{ \frac{\gamma\phi(\varepsilon/2)}{2b}, \frac{\varepsilon}{2} \right\} \ \text{ and } \ \|B(x) - B(y)\| \leqslant \frac{\phi(\varepsilon/2)}{2b},$$

then $\|x - y\| \leqslant \varepsilon$.

(ii) If B is uniformly monotone with function ϕ and

$$\|B(x) - B(y)\| \leqslant \frac{\phi(\varepsilon)}{b},$$

then $\|x - y\| \leqslant \varepsilon$.

Proof. For the case where A is uniformly monotone with function ϕ, note first that using Cauchy-Schwarz, and the nonexpansivity of T, we have

$$\langle T(x) - y, \gamma^{-1}(x - T(x)) - B(x) + B(y) \rangle$$
$$\leqslant \|T(x) - y\|(\gamma^{-1}\|x - T(x)\| + \|B(x) - B(y)\|)$$
$$\leqslant \frac{b}{\gamma} \|x(t) - T(x)\| + b\|B(x) - B(y)\|.$$

As we have

$$\|x - T(x)\| \leqslant \min\left\{ \frac{\gamma\phi(\varepsilon/2)}{2b}, \frac{\varepsilon}{2} \right\} \ \text{ and } \ \|B(x) - B(y)\| \leqslant \frac{\phi(\varepsilon/2)}{2b}$$

---

[2]Note that Theorem 3.4 in [92] is formulated using a so-called modulus of uniform monotonicity, which we chose to avoid in the presentation of this paper.

by assumption, we get
$$\langle T(x) - y, \gamma^{-1}(x - T(x)) - B(x) + B(y)\rangle \leqslant \phi(\varepsilon/2).$$
Now note that $-B(y) \in A(y)$ as $y \in \mathrm{zer}(A + B)$ and that
$$\gamma^{-1}(x - T(x)) - B(x) \in A(T(x))$$
so that we get
$$\phi(\|T(x) - y\|) \leqslant \langle T(x) - y, \gamma^{-1}(x - T(x)) - B(x) + B(y)\rangle \leqslant \phi(\varepsilon/2).$$
As ϕ is increasing, we get $\|T(x) - y\| \leqslant \varepsilon/2$. Finally, we get
$$\|x - y\| \leqslant \|x - T(x)\| + \|T(x) - y\| \leqslant \varepsilon.$$

For the case where B is uniformly monotone with function ϕ, simply note that using Cauchy-Schwarz, we have
$$\langle x - y, B(x) - B(y)\rangle \leqslant \|x - y\|\|B(x) - B(y)\|$$
so that using the uniform monotonicity of B, we get
$$\phi(\|x - y\|) \leqslant \langle x - y, B(x) - B(y)\rangle \leqslant b\|B(x) - B(y)\| \leqslant \phi(\varepsilon)$$
since $\|B(x) - B(y)\| \leqslant \phi(\varepsilon)/b$. As ϕ is increasing, the result follows. $\square$

Combined with the above, this immediately yields the following rates of convergence:

Theorem 4.14. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Let $\gamma \in (0, 2\beta)$ and write $\delta := \min\{1, \beta/\gamma\} + 1/2$. Let $y \in \mathrm{zer}(A + B)$ and let $b \in \mathbb{N}$ with $\|x_0 - y\| \leqslant b$. Further, assume that $\lambda : [0, +\infty) \to [0, \delta]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)'_1$. Assume either:*

(1) $\int_0^{+\infty} \lambda(t)(\delta - \lambda(t))\,dt = +\infty$ with a divergence modulus $\eta : (0, \infty) \to \mathbb{N}$ with $\eta(K) \geqslant \eta(L)$ for $K \geqslant L$, in which case we set $\varphi(\varepsilon) := \eta(\delta b^2/\varepsilon^2)$ and ψ as in Proposition 4.12.

(2) $\inf_{t \geqslant 0} \lambda(t) > 0$ with a witness $\underline{\lambda} > 0$, in which case we set $\varphi(\varepsilon) := 4b^4\delta^2/\underline{\lambda}^2\varepsilon^2$ and ψ as in Proposition 4.12.

Then, if either A or B is uniformly monotone with function ϕ, and in either case (1) or (2) above, x converges to y. Further, we have the rates
$$\forall \varepsilon > 0 \; \forall t \geqslant \rho(\varepsilon)\,(\|x(t) - y\| \leqslant \varepsilon),$$
where ρ is defined as follows:

(i) If A is uniformly monotone with function ϕ, then
$$\rho(\varepsilon) := \min\left\{\varphi\left(\min\left\{\frac{\gamma\phi(\varepsilon/2)}{2b}, \frac{\varepsilon}{2}\right\}\right), \psi\left(\frac{\phi(\varepsilon/2)}{2b}\right)\right\}.$$

(ii) If B is uniformly monotone with function ϕ, then
$$\rho(\varepsilon) := \psi\left(\frac{\phi(\varepsilon)}{b}\right).$$

Proof. For item (i), given $t \geqslant \rho(\varepsilon)$, we have
$$\|x(t) - T(x(t))\| \leqslant \min\left\{\frac{\gamma\phi(\varepsilon/2)}{2b}, \frac{\varepsilon}{2}\right\} \text{ and } \|B(x(t)) - B(y)\| \leqslant \frac{\phi(\varepsilon/2)}{2b}$$
using Lemmas 4.2 or 4.3, respectively, and Proposition 4.12. Noting Lemma 4.1, we get $\|x(t) - y\| \leqslant \varepsilon$ for any $t \geqslant \rho(\varepsilon)$ by Lemma 4.13. Item (ii) is similarly a combination of Proposition 4.12, Lemma 4.1 and Lemma 4.13. $\square$

If A or B are strongly monotone, that is

$$\langle x - y, u - v \rangle \geqslant \rho \|x - y\|^2$$

for some $\rho > 0$ and all $(x, u), (y, v) \in A$, or similarly for B, then they are immediately uniformly monotone with function $\phi(\varepsilon) := \rho\varepsilon^2$. In particular, the above rates apply in that case, complementing the rates given in [25]. Moreover, using the previous result on fast rates for $(*)_1$ given by Theorem 4.8, we can also derive exponentially fast rates under special parameter selections in that case. In particular, our result is slightly different from that given in [25], working over a slightly stronger (yet still canonical) assumption on the operators but for that following the traditional argument known from the discrete-time setting, leveraging the previous results of this section instead of relying on arguments related to differential inequalities:

Theorem 4.15. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. Let $\gamma \in (0, 2\beta)$ and write $\delta := \min\{1, \beta/\gamma\} + 1/2$. Further, assume that $\lambda : [0, +\infty) \to [0, \delta]$ is Lebesgue measurable and fix $x_0 \in X$. Let x be the unique strong global solution to $(*)_1'$. Further assume that $\underline{\tau} = \inf_{t \geqslant 0} \lambda(t)(\delta - \lambda(t)) > 0$ and that A is strongly monotone with constant α.*

Then, for the unique point $y \in \mathrm{zer}(A + B)$, we have

$$d(x(t), y) \leqslant c^{\lfloor t \rfloor} d(x(0), y)$$

for all $t \geqslant 0$, where $c := (1 + \underline{\tau} k^p)^{-1/p} \in (0, 1)$ and $k := 1/(\alpha\gamma + 1)$.

Proof. It is well-known (see e.g. the proof of Proposition 26.16 in [15]) that under the present assumptions, the map $T := J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$ is a strict contraction with parameter $k \in [0, 1)$. The result now follows from Theorem 4.8. □

Following Remark 15 in [23], also we want to remark here that in the same way as the Douglas–Rachford splitting method arises as an instantiation of the Krasnosel'skii-Mann method in discrete-time contexts, the above system $(*)_1$ gives rise to a continuous-time variant. This dynamical system can hence be studied using the present results, and we think that it is moreover quite likely that, following a similar approach as in [92], one can also obtain particularly broad and uniform rates in that case.

4.2. **Second-order dynamical systems for cocoercive operators.** We again begin with the essential setup, now from [22]. Let X be a Hilbert space with norm $\|\cdot\|$ and inner product $\langle \cdot, \cdot \rangle$ and let $B : X \to X$ be β-cocoercive, i.e. $\beta\|Bx - By\|^2 \leqslant \langle x - y, Bx - By \rangle$ for all $x, y \in X$. We write $\mathrm{Zer}B := \{x \in X \mid B(x) = 0\}$. Further, let $\lambda, \gamma : [0, +\infty) \to [0, +\infty)$ be Lebesgue measurable and $u_0, v_0 \in X$ be given. For these data, we consider the dynamical system

$$(*)_2 \qquad \begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0, \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases}$$

This system arises naturally as an extension of the second-order dynamical system

$$\ddot{x} + \gamma\dot{x} + x - Tx = 0$$

formulated for a nonexpansive operator $T : X \to X$ as considered in [8], extending previous work on a gradient-projection second-order dynamical system

$$\ddot{x} + \gamma\dot{x} + x - P_C(x - \eta\nabla f(x)) = 0$$

studied by Antipin [6] in Euclidean spaces and by Attouch and Alvarez [8] in general Hilbert spaces, which in turn is motivated by a study of the heavy ball with friction dynamical system

$$\ddot{x} + \gamma\dot{x} + \nabla f(x) = 0$$

studied by various authors (see e.g. [4, 6, 10, 52], among others). We again refer to the survey [42] for further discussions along these lines.

Similar to before (see also Definition 3 in [22]), we call a dynamical system $x : [0, +\infty) \to X$ a strong global solution of $(*)_2$ if x satisfies the respective equation almost everywhere on $[0, +\infty)$ and x is absolutely continuous on each interval $[0, b]$ with $b \in (0, \infty)$. By Theorem 4 in [22], the system $(*)_2$ has a unique strong global solution under the assumptions above, provided that additionally $\lambda, \gamma \in L^1([0, b])$ for every $b > 0$.

As discussed in [22], we also here want to highlight that a discretization of the above equation w.r.t. the time variable leads to methods which include inertia terms, that is which make use of the distance between the previous two iterates, which commonly yields accelerated schemas (we refer e.g. to [4, 5, 74] for works explicitly discussing this in connection to dynamical systems, as well as to the further references provided in [22]).

Similar to before, we are now interested in providing a quantitative account of the results given in [22] (see Theorem 8 therein) on the asymptotic behavior of such strong global solutions. For that, following [22] (see assumption (A1) therein), we now throughout assume further that λ and γ are locally absolutely continuous and that there exists a $\theta > 0$ such that

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{1 + \theta}{\beta}$$

for almost every $t \in [0, +\infty)$. In particular, there are $0 < \underline{\lambda}, \overline{\gamma}, \overline{\lambda}, \underline{\gamma}$ such that $\underline{\lambda} \leqslant \lambda(t) \leqslant \overline{\lambda}$ and $\underline{\gamma} \leqslant \gamma(t) \leqslant \overline{\gamma}$. We refer to [22] for further motivation and discussion on these conditions, including natural examples of parameters satisfying them, and here just note that in such a case the solution to the above system $(*)_2$ is a C^2-function (see p. 1427 in [22]).

We now begin with the boundedness of both the solution dynamical system as well as the associated derivatives and the image under B (as established in item (i) of [22, Theorem 8]). We begin with the boundedness of the dynamical system itself:

Lemma 4.16. *Let $z \in \mathrm{Zer}B$ and let $b, c \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$ and $\|v_0\| \leqslant c$. Let x be the unique strong global solution to $(*)_2$. Then x is bounded with $\|x(t) - z\| \leqslant \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$, where we set $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$.*

Proof. Write $h(t) = \frac{1}{2}\|x(t) - z\|^2$. As in the proof of [22, Theorem 8], we have $\dot{h}(t) = \langle x(t) - z, \dot{x}(t) \rangle$ and it follows from Eq. (13) therein that

$$\dot{h}(t) + \gamma(t)h(t) + \beta\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2 \leqslant \dot{h}(0) + \gamma(0)h(0) + \beta\frac{\gamma(0)}{\lambda(0)}\|\dot{x}(0)\|^2$$

$$\leqslant \|u_0 - z\|\|v_0\| + \frac{\overline{\gamma}}{2}\|u_0 - z\|^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}\|v_0\|^2$$

$$\leqslant bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2 = M.$$

In particular, we have $\dot{h}(t) + \gamma(t)h(t) \leqslant K$. Continuing as in the proof of [22, Theorem 8] until Eq. (14), we get that

$$h(T) \leqslant h(0)\exp(-\underline{\gamma}T) + M\underline{\gamma}^{-1}(1 - \exp(-\underline{\gamma}T))$$

for all $T > 0$, and so

$$h(t) \leqslant \frac{1}{2}b^2 + \underline{\gamma}^{-1}M.$$

In particular, we have

$$\|x(t) - z\| \leqslant \sqrt{b^2 + 2\underline{\gamma}^{-1}M}. \qquad \square$$

Before we carry on with the boundedness of the derivative, we require the following simple lemma on real inequalities:

Lemma 4.17. *Let $a, b, c > 0$ be given. For any $x \geq 0$, if $ax^2 \leq bx + c$, then $x \leq \lceil \frac{b+c}{a} \rceil$.*

Proof. Suppose for a contradiction that $x > \lceil \frac{b+c}{a} \rceil \geq 1$. Then

$$ax^2 \leq bx + c \leq (b + c)x$$

and so $ax \leq b + c$. But this is a contradiction as then $b + c \leq a \lceil \frac{b+c}{a} \rceil < ax \leq b + c$. $\square$

The boundedness of the first derivative is now readily derived:

Lemma 4.18. *Let $z \in \mathrm{Zer}B$ and let $b, c \in \mathbb{N}$ with $\|u_0 - z\| \leq b$ and $\|v_0\| \leq c$. Let x be the unique strong global solution to $(*)_2$. Then $\dot{x}$ is bounded with $\|\dot{x}(t)\| \leq \left\lceil \frac{K+M}{\beta \underline{\gamma} \overline{\lambda}^{-1}} \right\rceil$, where we set $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta \overline{\gamma} \underline{\lambda}^{-1} c^2$ and $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$.*

Proof. As in the proof of [22, Theorem 8] (see the last equation before Eq. (16)), we derive

$$\langle x(t) - z, \dot{x}(t) \rangle + \beta \underline{\gamma} \overline{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M$$

and so

$$\beta \underline{\gamma} \overline{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M + \|x(t) - z\| \|\dot{x}(t)\| \leq M + K\|\dot{x}(t)\|$$

by Lemma 4.16. Lemma 4.17 yields

$$\|\dot{x}(t)\| \leq \left\lceil \frac{K + M}{\beta \underline{\gamma} \overline{\lambda}^{-1}} \right\rceil$$

as claimed. $\square$

We now turn to the L^2-boundedness of the derivatives and the image under the operator B:

Lemma 4.19. *Let $z \in \mathrm{Zer}B$ and let $b, c \in \mathbb{N}$ with $\|u_0 - z\| \leq b$ and $\|v_0\| \leq c$. Let x be the unique strong global solution to $(*)_2$. Then*

$$\|\dot{x}\|_2 \leq \sqrt{\frac{KL}{\theta}}, \quad \|\ddot{x}\|_2 \leq \sqrt{\frac{KL}{\beta \overline{\lambda}^{-1}}}, \quad \|Bx\|_2 \leq \underline{\lambda}^{-1}(\|\ddot{x}\|_2 + \overline{\gamma}\|\dot{x}\|_2)$$

where we set $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta \overline{\gamma} \underline{\lambda}^{-1} c^2$, $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$ and $L := \left\lceil (K + M)\beta \underline{\gamma} \overline{\lambda}^{-1} \right\rceil$.

Proof. Write $h(t) = \frac{1}{2}\|x(t) - z\|^2$ and recall $\dot{h}(t) = \langle x(t) - z, \dot{x}(t) \rangle$. As in Eq. (12) in the proof of [22, Theorem 8], we derive

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta \frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) + \theta\|\dot{x}(t)\|^2 + \beta \overline{\lambda}^{-1}\|\ddot{x}(t)\|^2 \leq 0$$

so that integration yields

$$\theta \int_0^t \|\dot{x}(s)\|^2 \, ds + \beta \overline{\lambda}^{-1} \int_0^t \|\ddot{x}(s)\|^2 \, ds \leq |\dot{h}(t)| \leq \|x(t) - z\| \|\dot{x}(t)\| \leq KL$$

for all $t \geq 0$. This yields the bounds on the L^2-norms for $\dot{x}$ and $\ddot{x}$. For Bx, note that as x satisfies the equation $(*)_2$, we have

$$\underline{\lambda}\|Bx\|_2 \leq \|\lambda Bx\|_2 \leq \|\ddot{x}\|_2 + \|\gamma \dot{x}\|_2 \leq \|\ddot{x}\|_2 + \overline{\gamma}\|\dot{x}\|_2$$

using Minkowski's inequality, which yields the claim. $\square$

To capture the corresponding problem in the setup of the previous sections, we set $X_0 = \overline{B}_K(z)$ and

$$F(x^*) = \begin{cases} \|B(x^*)\| & \text{if } x^* \in X_0, \\ +\infty & \text{otherwise,} \end{cases}$$

for a fixed solution $z \in \mathrm{Zer}\,B$, and where K is as in Lemma 4.18. In particular, we have $\mathrm{zer}F = \mathrm{Zer}\,B \cap X_0$ and

$$\mathrm{lev}_{\leqslant \varepsilon}F = \{x^* \in X_0 \mid \|B(x^*)\| \leqslant \varepsilon\}.$$

Our next step is a quantitative result on the asymptotic regularity of the dynamical system. Concretely, in item (ii) of Theorem 8 in [22], it is established that

$$\lim_{t \to \infty} \|\dot{x}(t)\| = \lim_{t \to \infty} \|\ddot{x}(t)\| = \lim_{t \to \infty} \|B(x(t))\| = 0.$$

We now provide quantitative versions of these results by giving rates of metastability for each case. Our quantitative rendering of the limit $\lim_{t \to \infty} B(x(t)) = 0$ then in particular allows us to derive the $\liminf$-property w.r.t. the above F for the dynamical system x, together with a respective bound for it.

Lemma 4.20. *Let $z \in \mathrm{Zer}\,B$ and let $b, c, d \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|B(u_0)\| \leqslant d$. Let x be the unique strong global solution to $(*)_2$. Then*

$$\lim_{t \to \infty} \dot{x}(t) = \lim_{t \to \infty} \ddot{x}(t) = \lim_{t \to \infty} B(x(t)) = 0$$

where further, for any $f : \mathbb{N} \to \mathbb{N}$ and $\varepsilon > 0$, we have

$$\exists n \leqslant \Lambda(\varepsilon, f) \; \forall t \in [n, n + f(n)] \left(\|\dot{x}(t)\|, \|B(x(t))\| < \varepsilon\right)$$

as well as

$$\exists n \leqslant \Lambda(\varepsilon/2\max\{\overline{\gamma}, \overline{\lambda}\}, f) \; \forall t \in [n, n + f(n)] \left(\|\ddot{x}(t)\| < \varepsilon\right).$$

Here, $\Lambda(\varepsilon, f) := \tilde{f}'^{(\varpi_{C,A,B}(\varepsilon))}(0)$ where $\varpi_{C,A,B}$ is defined via

$$\varpi_{C,A,B}(\varepsilon) := \left\lceil \frac{16AB}{3\varepsilon^2} \right\rceil \cdot \left\lceil \frac{4(C^2 + 2AB)}{3\varepsilon^2} \right\rceil,$$

where $\tilde{f}'(n) := n + f'(n)$ as before, with $f'(n) := \max\{f(n), \lceil (3A/\varepsilon)^2 \rceil\}$, and where we set $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$, $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$ and $L := \left\lceil (K + M)\beta\underline{\gamma}\overline{\lambda}^{-1} \right\rceil$ as well as $a_0 := \sqrt{KL/\theta}$, $a_1 := \sqrt{KL/\beta\overline{\lambda}^{-1}}$ and $a_2 := \underline{\lambda}^{-1}(a_1 + \overline{\gamma}a_0)$ together with $A = \frac{1}{2}\max\{a_0, a_2\}$, $B = \frac{1}{2}\max\{a_0 + a_1, a_2 + \frac{1}{\beta^2}a_0\}$ and $C = \frac{1}{2}\max\{c, d\}$.

In particular, x has the $\liminf$-property w.r.t. F with bound

$$\varphi(\varepsilon, n) := \varpi_{C,A,B}(\varepsilon)\max\{n, \lceil (3A/\varepsilon)^2 \rceil\}.$$

Proof. Note that as in the proof of item (ii) of Theorem 8 in [22], we have

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2\right) \leqslant \frac{1}{2}\|\dot{x}(t)\|^2 + \frac{1}{2}\|\ddot{x}(t)\|^2$$

as well as

$$\frac{d}{dt}\left(\frac{1}{2}\|B(x(t))\|^2\right) \leqslant \frac{1}{2}\|B(x(t))\|^2 + \frac{1}{2\beta^2}\|\dot{x}(t)\|^2.$$

Lemma 4.19 now in particular yields $\|\dot{x}\|_2 \leqslant a_0$, $\|\ddot{x}\|_2 \leqslant a_1$ and $\|Bx\|_2 \leqslant a_2$. The two respective rates for $\lim_{t\to\infty} \dot{x}(t) = \lim_{t\to\infty} B(x(t)) = 0$ now follow from Lemma 2.4. For the rate for $\ddot{x}$, note similar to before that x satisfies the equation $(*)_2$ and so we have

$$\|\ddot{x}(t)\| \leqslant \overline{\gamma}\|\dot{x}(t)\| + \overline{\lambda}\|B(x(t))\|.$$

We obtain the $\liminf$-property by setting $f :\equiv n$ for a given $n \in \mathbb{N}$ in the above rate of metastability for $\lim_{t\to\infty} B(x(t)) = 0$. $\qquad\square$

In similarity to the previous application, we now also obtain the quasi-Fejér monotonicity of the unique strong global solution x to the system $(*)_2$ w.r.t. F. Again, this follows the pattern discussed in Section 2, that is through a differential inequality for a suitable energy functional which decomposes in the right way. Indeed, in [22] (see eq. (12) therein), it is established that

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) \leqslant 0,$$

where again $h(t) = \frac{1}{2}\|x(t) - z\|^2$ for a solution $z \in \mathrm{Zer}B$. In that way, distinct from the previous application, this inequality is no longer of the most simple form discussed in Section 2. Nevertheless, it fits into another pattern discussed therein, whereby the energy functional additively decomposes into a distance to a solution as well as associated errors terms. We now analyse the proof of this inequality as given in [22] to establish the associated uniform quasi-Fejér monotonicity and extract an associated modulus. The following result first provides a quantitative variant of the above differential inequality:

Lemma 4.21. *Let $z \in \mathrm{Zer}B$ and let $b, c \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$ and $\|v_0\| \leqslant c$. Let x be the unique strong global solution to $(*)_2$. For any $\varepsilon > 0$, if $x^* \in \mathrm{lev}_{\leqslant \varepsilon/(\overline{\lambda}4K)}F$, then*

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) \leqslant \varepsilon$$

for all $t \geqslant 0$, where $h(t) := \frac{1}{2}\|x(t) - x^*\|^2$. Here $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$ and $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$.

Proof. Given $x^* \in \mathrm{lev}_{\leqslant r(\varepsilon)}F$, write $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. As in the proof of item (i) of Theorem 8 in [22], we get

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\langle x(t) - x^*, B(x(t))\rangle = \|\dot{x}(t)\|^2.$$

In particular, we have

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\langle x(t) - x^*, B(x(t)) - B(x^*)\rangle + \lambda(t)\langle x(t) - x^*, B(x^*)\rangle = \|\dot{x}(t)\|^2$$

and so, using the cocoercivity of B and Lemma 4.16, we get

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\beta\|B(x(t)) - B(x^*)\|^2 \leqslant \|\dot{x}(t)\|^2 + \overline{\lambda}2K\|B(x^*)\|.$$

Further, using the reverse triangle inequality, we have

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\beta\|B(x(t))\|^2 \leqslant \|\dot{x}(t)\|^2 + \overline{\lambda}2K\|B(x^*)\| + 2\lambda(t)\beta\|B(x(t))\|\|B(x^*)\|.$$

As B is $\frac{1}{\beta}$-Lipschitz and as z is a zero of B, we get

$$\|B(x(t))\| = \|B(x(t)) - B(z)\| \leqslant \frac{1}{\beta}\|x(t) - z\| \leqslant \frac{K}{\beta}$$

using Lemma 4.16. In particular, we have

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\beta\|B(x(t))\|^2 \leqslant \|\dot{x}(t)\|^2 + \overline{\lambda}4K\|B(x^*)\|.$$

Proceeding as in the proof of item (i) of Theorem 8 in [22], we get

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta\frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2\right) \leqslant \overline{\lambda}4K\|B(x^*)\|$$

which yields the claim.   □

Again, an application of the mean value theorem allows us to derive a modulus of uniform (G, H)-quasi-Fejér monotonicity (with $H = (\cdot)^2$ and $G = \overline{\gamma}\underline{\gamma}^{-1}(\cdot)^2$) from the above approximate differential inequality.

Lemma 4.22. *Let $z \in \mathrm{Zer}B$ and let $b, c \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$ and $\|v_0\| \leqslant c$. Let x be the unique strong global solution to $(*)_2$. Then x is uniformly (G, H)-quasi-Fejér monotone w.r.t. $\mathrm{zer}F$ and F and with $H = (\cdot)^2$ and $G = \overline{\gamma}\underline{\gamma}^{-1}(\cdot)^2$, i.e.*

$$\forall\varepsilon > 0 \ \forall n, m \in \mathbb{N} \ \forall x^* \in \mathrm{lev}_{\leqslant\chi(\varepsilon,n,m)}F \ \forall s, t \in [n, n + m]$$
$$\left(s \leqslant t \to \|x(t) - x^*\|^2 \leqslant \overline{\gamma}\underline{\gamma}^{-1}\|x(s) - x^*\|^2 + \varepsilon(s, t) + \varepsilon\right),$$

where $\chi(\varepsilon, n, m) := (\underline{\gamma}\varepsilon)/(m\overline{\lambda}8K)$ and

$$\varepsilon(s, t) = 2\beta\underline{\gamma}^{-1}\frac{\overline{\gamma}}{\underline{\lambda}}\|\dot{x}(s)\|^2 + 2\underline{\gamma}^{-1}K\|\dot{x}(s)\| + 2\underline{\gamma}^{-1}K\|\dot{x}(t)\|.$$

Here $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$ and $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$.

Proof. Fix ε, n, m as well as $x^* \in \mathrm{lev}_{\leqslant\chi(\varepsilon,n,m)}F$ and $s, t \in [n, n + m]$. If $s = t$, the result is clear. If $s < t$, then since $\dot{h}(t) + \gamma(t)h(t) + \beta\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2$ is locally absolutely continuous, the mean value theorem yields an $r \in (s, t)$ such that

$$\ddot{h}(r) + \frac{d}{dt}(\gamma h)(r) + \beta\frac{d}{dt}\left(\frac{\gamma(r)}{\lambda(r)}\|\dot{x}(r)\|^2\right)$$
$$= \frac{d}{dt}\left(\dot{h}(r) + \gamma(r)h(r) + \beta\frac{\gamma(r)}{\lambda(r)}\|\dot{x}(r)\|^2\right)$$
$$= \frac{\dot{h}(t) + \gamma(t)h(t) + \beta\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2 - (\dot{h}(s) + \gamma(s)h(s) + \beta\frac{\gamma(s)}{\lambda(s)}\|\dot{x}(s)\|^2)}{t - s}.$$

Thus, as $x^* \in \mathrm{lev}_{\leqslant\chi(\varepsilon,n,m)}F$, we have

$$\gamma(t)h(t) - \gamma(s)h(s) \leqslant \beta\frac{\gamma(s)}{\lambda(s)}\|\dot{x}(s)\|^2 - \beta\frac{\gamma(t)}{\lambda(t)}\|\dot{x}(t)\|^2 + \dot{h}(s) - \dot{h}(t) + \varepsilon$$
$$\leqslant \beta\frac{\overline{\gamma}}{\underline{\lambda}}\|\dot{x}(s)\|^2 + K\|\dot{x}(s)\| + K\|\dot{x}(t)\| + \underline{\gamma}\frac{\varepsilon}{2}$$

by Lemma 4.4. In particular, we get

$$h(t) \leqslant \overline{\gamma}\underline{\gamma}^{-1}h(s) + \beta\underline{\gamma}^{-1}\frac{\overline{\gamma}}{\underline{\lambda}}\|\dot{x}(s)\|^2 + \underline{\gamma}^{-1}K\|\dot{x}(s)\| + \underline{\gamma}^{-1}K\|\dot{x}(t)\| + \frac{\varepsilon}{2}$$

which yields the claim.   □

Another direct application of Theorem 3.9 now yields the following quantitative result on the behavior of the solution x to $(*)_2$, again under a finite-dimensionality assumption on X, giving a rate of metastability for the convergence result established in Theorem 8 of [22] in that case:

Theorem 4.23. *Let X be a finite-dimensional Hilbert space with dimension d and let $B : X \to X$ be β-cocoercive. Let $u_0, v_0 \in X$ be given and $z \in \mathrm{Zer}B$ and let $b, c, d \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|B(u_0)\| \leqslant d$. Further, let $\lambda, \gamma : [0, +\infty) \to [0, +\infty)$ be Lebesgue measurable and locally absolutely continuous and assume that there exists a $\theta > 0$ such that*

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \ and \ \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{1+\theta}{\beta}$$

for almost every $t \in [0, +\infty)$. Let $0 < \underline{\lambda}, \overline{\gamma}, \overline{\lambda}, \underline{\gamma}$ such that $\underline{\lambda} \leqslant \lambda(t) \leqslant \overline{\lambda}$ and $\underline{\gamma} \leqslant \gamma(t) \leqslant \overline{\gamma}$. Let x be the unique strong global solution to $(*)_2$.

Then x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\min\{\varepsilon, \beta\varepsilon/2\}, f) \ \forall s, t \in [n, n + f(n)] \ (\|x(s) - x(t)\| \leqslant \varepsilon \ and \ \|B(x(t))\| \leqslant \varepsilon),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described by

$$\Delta(\varepsilon, f) := \max\{\Delta(i, \varepsilon, f) \mid i \leqslant P\} + 1$$

where $\Delta(0, \varepsilon, f) := 0$ as well as

$$\Delta(j, \varepsilon, f) := \max\{\varphi(\widehat{\varepsilon}_j, n) \mid n \leqslant \eta(\varepsilon^2/12, f_{\varphi, \widehat{\varepsilon}_j})\}$$

with

$$\widehat{\varepsilon}_j = \min\{\varepsilon/2, (\underline{\gamma}\varepsilon^2)/12((f(m+1)+1)\overline{\lambda}8K) \mid m \leqslant \Delta(i, \varepsilon, f), i < j\}$$

for $j \geqslant 1$ as well as

$$f_{\varphi, \varepsilon}(n) := \max\{m + 1 + f(m+1) \mid m \leqslant \varphi(\varepsilon, n)\} \dot{-} n$$

and

$$P := \left\lceil 2 \left(\left\lceil \left(\varepsilon / \sqrt{12\overline{\gamma}\underline{\gamma}^{-1}} \right)^{-1} \right\rceil + 1 \right) \sqrt{d}b \right\rceil^d + 1$$

as well as

$$\eta(\varepsilon, f) := \Lambda \left(\min \left\{ \frac{\varepsilon}{6\underline{\gamma}^{-1}K}, \sqrt{\frac{\varepsilon}{6\beta\underline{\gamma}^{-1}\frac{\overline{\gamma}}{\underline{\lambda}}}} \right\}, f \right) \ and \ \varphi(\varepsilon, n) := \varpi_{C,A,B}(\varepsilon) \max\{n, \lceil (3A/\varepsilon)^2 \rceil\}.$$

Here, $\Lambda(\varepsilon, f) := \widetilde{f}'^{(\varpi_{C,A,B}(\varepsilon))}(0)$ and $\varpi_{C,A,B}$ as well as the other constants are defined as in Lemma 4.20.

Proof. The result immediately follows by instantiating Theorem 3.8 over $X_0 = \overline{B}_K(z)$ and simplifying the resulting bound (recall that $x(t) \in X_0$ for all $t \geqslant 0$ by Lemma 4.16). In particular, we set $H := (\cdot)^2$ and $G := \overline{\gamma}\underline{\gamma}^{-1}(\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon/\overline{\gamma}\underline{\gamma}^{-1}}$. That x is uniformly (G, H)-quasi-Fejér monotone w.r.t. F and errors

$$\varepsilon(s, t) = 2\beta\underline{\gamma}^{-1}\frac{\overline{\gamma}}{\underline{\lambda}}\|\dot{x}(s)\|^2 + 2\underline{\gamma}^{-1}K\|\dot{x}(s)\| + 2\underline{\gamma}^{-1}K\|\dot{x}(t)\|$$

follows by Lemma 4.22, which in particular yields that $\chi(\varepsilon, n, m) := (\underline{\gamma}\varepsilon)/(m\overline{\lambda}8K)$ is an associated modulus. The definition of $\widehat{\varepsilon}_j$ is a resulting simplification. The errors in particular satisfy $\varepsilon(s, t) \to 0$ as $s \leqslant t \to \infty$ with a rate of metastability η as defined above by Lemma 4.20. Also by Lemma 4.20, x has the $\liminf$-property w.r.t. F with a bound $\varphi(\varepsilon, n) := \varpi_{C,A,B}(\varepsilon) \max\{n, \lceil (3A/\varepsilon)^2 \rceil\}$, which can immediately be seen to be suitably monotone. As before, X_0 is totally bounded with a modulus $\gamma(\varepsilon) := \left[2(\lceil \varepsilon^{-1} \rceil + 1)\sqrt{d}b \right]^d$ by Example 2.8 of [59]. Lastly, note that F is uniformly continuous on X_0 with modulus $\omega(\varepsilon) := \beta\varepsilon$, since we have $|\|B(x)\| - \|B(y)\|| \leqslant \frac{1}{\beta}\|x - y\|$ using the fact that B is $\frac{1}{\beta}$-Lipschitz. $\square$

In the preceding application, Theorem 3.18 yielded rather general constructions for rates of convergence under a regularity assumption. While Theorem 3.18 can similarly be applied to the present case study to yield quantitative convergence guarantees for the strong convergence of the dynamical system under a regularity assumption, it does not yield rates in this general case.

This is due to the fact that we cannot supply the respective errors in the quasi-Fejér monotonicity property for the dynamical system at hand with a rate of convergence, but only with a rate of metastability. While this does not rule out constructions of rates of convergence for problem $(*)_2$ under regularity assumptions, it perhaps provides a structural reason for why such rates seem to be currently absent from the literature (to the best of our knowledge) at this level of generality. This issue seems to persist already for special cases of the above system $(*)_2$, such as the gradient-projection second-order dynamical system [8, 6] or its extensions to general nonexpansive mappings as discussed at the beginning of this subsection (indeed, the paper [22] uses essentially the same approach as [8], appropriately extended to its more general setting).

By comparison, Boţ and Csetnek [25] derive exponentially fast rates for the second-order forward-backward scheme, arising from the above by instantiating B appropriately and discussed in more detail in the following, under a strong monotonicity assumption on either of the involved operators and under suitable restrictions of the other parameters. Indeed, their argument essentially proceeds by utilizing the extremely strong premise of strong monotonicity to derive an alternative differential inequality that dispenses of the error induced by $\|\dot{x}(t)\|$ above. However, this inequality is then already so strong that it allows one to immediately derive exponential rates, so that no additional arguments beyond that are necessary. While it can be rather immediately seen that these arguments can be slightly generalized to the setting of $(*)_2$ for general B which are strict contractions, we do not record this here as no new arguments from this paper would be required for that. Instead, we focus on the general case of $(*)_2$ which however then restricts us to metastability.

Of course, by the modularity of our approach, if one would be supplied by such a rate by any other means, then Theorem 3.18 could be immediately used to derive corresponding rates for the strong convergence of the dynamical system under a regularity assumption. Outside of assumptions guaranteeing this however, the following theorem is the best we can currently provide.

Theorem 4.24. *Let X be a Hilbert space and let $B : X \to X$ be β-cocoercive. Let $u_0, v_0 \in X$ be given and $z \in \mathrm{Zer} B$ and let $b, c, d \in \mathbb{N}$ with $\|u_0 - z\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|B(u_0)\| \leqslant d$. Assume that F has a modulus of regularity τ w.r.t. $\overline{B}_K(z)$. Further, let $\lambda, \gamma : [0, +\infty) \to [0, +\infty)$ be Lebesgue measurable and locally absolutely continuous and assume that there exists a $\theta > 0$ such that*

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \ and \ \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{1 + \theta}{\beta}$$

for almost every $t \in [0, +\infty)$. Let $0 < \underline{\lambda}, \overline{\gamma}, \overline{\lambda}, \underline{\gamma}$ such that $\underline{\lambda} \leqslant \lambda(t) \leqslant \overline{\lambda}$ and $\underline{\gamma} \leqslant \gamma(t) \leqslant \overline{\gamma}$. Let x be the unique strong global solution to $(*)_2$.
Then x satisfies

$$\forall \varepsilon > 0 \ \forall f : \mathbb{N} \to \mathbb{N} \ \exists n \leqslant \rho(\varepsilon, f) \ \forall t \in [n, n + f(n)] \left(\mathrm{dist}(x(t), \mathrm{Zer} B) < \varepsilon \right)$$

and further x is Cauchy with

$$\forall \varepsilon > 0 \ \forall f : \mathbb{N} \to \mathbb{N} \ \exists n \leqslant \rho(\varepsilon/2, f) \ \forall t \in [n, n + f(n)] \left(\|x(t) - x(s)\| < \varepsilon \right).$$

Here, ρ is defined by

$$\rho(\varepsilon, f) := \max\left\{\varphi\left(\tau\left(\varepsilon/\sqrt{2\overline{\gamma}\underline{\gamma}^{-1}}\right), n\right) \mid n \leqslant \eta(\varepsilon^2, f_{\varphi,\varepsilon^2})\right\} + 1$$

where

$$f_{\varphi,\delta}(n) = \max\left\{m + 1 + f(m+1) \mid m \leqslant \varphi\left(\tau\left(\varepsilon/\sqrt{2\overline{\gamma}\underline{\gamma}^{-1}}\right), n\right)\right\} \dot{-} n$$

with

$$\eta(\varepsilon, f) := \Lambda\left(\min\left\{\frac{\varepsilon}{6\underline{\gamma}^{-1}K}, \sqrt{\frac{\varepsilon}{6\beta\underline{\gamma}^{-1}\frac{\overline{\gamma}}{\underline{\lambda}}}}\right\}, f\right) \quad and \quad \varphi(\varepsilon, n) := \varpi_{C,A,B}(\varepsilon)\max\{n, \lceil (3A/\varepsilon)^2 \rceil\}.$$

Here, $\Lambda(\varepsilon, f) := \tilde{f}'^{(\varpi_{C,A,B}(\varepsilon))}(0)$ and $\varpi_{C,A,B}$ as well as the other constants are defined as in Lemma 4.20.

Proof. The result immediately follows by instantiating Theorem 3.17 and simplifying the resulting bound. In particular, we set $H := (\cdot)^2$ and $G := \overline{\gamma}\underline{\gamma}^{-1}(\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon/\overline{\gamma}\underline{\gamma}^{-1}}$. That x is (G, H)-quasi-Fejér monotone w.r.t. F follows by Lemma 4.22. That x has the lim inf-property w.r.t. F with φ as the respective bound follows from Lemma 4.20. The rate of metastability η for the errors likewise follows from that lemma as before. $\square$

Similar to before, we here want to also study a particular instance of the system $(*)_2$ to which the above results apply. For that consider a continuous-time variant of the forward-backward method, this time defined via a second order equation as in $(*)_2$, as considered already in [22]. Let us recall the setup: Assume we are given a maximally monotone set-valued operator $A : X \to 2^X$ and a single-valued map $B : X \to X$ which is β-cocoercive, for some $\beta > 0$, such that $\mathrm{zer}(A + B) \neq \varnothing$. For $\eta \in (0, 2\beta)$ (following the notation of [22]) and $\delta := \frac{4\beta - \eta}{2\beta}$, let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be Lebesgue measurable functions which are locally absolutely continuous and satisfy

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \quad and \quad \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{2(1 + \theta)}{\delta}$$

almost everywhere, for some $\theta > 0$. For initial values $u_0, v_0 \in X$, we then consider the dynamical system

$$(*)_2' \qquad \begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\left[x(t) - J_{\eta A}\left(x(t) - \eta B(x(t))\right)\right] = 0, \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases}$$

In analogy to before, this system can be conceived of as an instantiation of the system $(*)_2$ using the map $T := J_{\eta A} \circ (\mathrm{Id} - \eta B)$, which is $\frac{1}{\delta}$-averaged and satisfies $\mathrm{Fix}T = \mathrm{zer}(A + B)$ as before (see the beginning of the proof of Theorem 12 in [22]). So writing

$$T = \left(1 - \frac{1}{\delta}\right)\mathrm{Id} + \frac{1}{\delta}R$$

with R nonexpansive, $(*)_2'$ is an instantiation of $(*)_2$ using the map R and the parameter sequence $\frac{1}{\delta}\lambda$ instead of λ (see also the proof of Corollary 11 in [22]).

We immediately obtain the following result on the quantitative asymptotic behavior of the associated solution under a finite-dimensionality assumption on X, a quantitative variant of parts of Theorem 12 in [22]:

Theorem 4.25. *Let X be a finite-dimensional Hilbert space with dimension d and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. For $\eta \in (0, 2\beta)$ and $\delta := \frac{4\beta - \eta}{2\beta}$, let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be Lebesgue measurable functions which are locally absolutely continuous and satisfy*

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \ \text{and} \ \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{2(1+\theta)}{\delta}$$

almost everywhere, for some $\theta > 0$. Fix $u_0, v_0 \in X$ and let $y \in \mathrm{zer}(A + B)$ with $b, c, d \in \mathbb{N}$ such that $\|u_0 - y\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|u_0 - T(u_0)\| \leqslant d$. Let x be the unique strong global solution to $(*)_2'$.

Then x is metastable in the sense that for all $\varepsilon > 0$ and $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\min\{\varepsilon, \beta\varepsilon/2\}, f) \ \forall s, t \in [n, n + f(n)] \left(\|x(s) - x(t)\| \leqslant \varepsilon \ \text{and} \ \|x(t) - T(x(t))\| \leqslant \varepsilon\right),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described as in Theorem 4.23.

Remark 4.26. Akin to Remark 4.10, if Theorem 4.25 is applied with $\min\{\varepsilon, \varepsilon/(\gamma^{-1} + \beta^{-1})\}$ in place of ε, then for any $t \in [n, n + f(n)]$ with the respective n, there are $w \in (A + B)(v)$ such that $\|x(t) - v\|, \|w\| \leqslant \varepsilon$.

We now move to the associated strong convergence result for $(*)_2'$ under a uniform monotonicity assumption for either of the operators A or B, again in similarity to before.

First, we have the following quantitative rendering of the associated limit $\lim_{t \to \infty} B(x(t)) = B(x^*)$ for $x^* \in \mathrm{zer}(A + B)$ as established in item (iv) of Theorem 12 in [22]:

Proposition 4.27. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. For $\eta \in (0, 2\beta)$ and $\delta := \frac{4\beta - \eta}{2\beta}$, let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be Lebesgue measurable functions which are locally absolutely continuous and satisfy*

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \ \text{and} \ \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{2(1+\theta)}{\delta}$$

almost everywhere, for some $\theta > 0$. Fix $u_0, v_0 \in X$ and let $y \in \mathrm{zer}(A + B)$ with $b, c, d \in \mathbb{N}$ such that $\|u_0 - y\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|u_0 - T(u_0)\| \leqslant d$. Let x be the unique strong global solution to $(*)_2'$. Then $\lim_{t \to \infty} B(x(t)) = B(y)$ and for any $f : \mathbb{N} \to \mathbb{N}$ and $\varepsilon > 0$, we have

$$\exists n \leqslant \Lambda(\varepsilon^2 \eta\beta/3K, f) \ \forall t \in [n, n + f(n)] \left(\|B(x(t)) - B(y)\| < \varepsilon\right)$$

where Λ is defined as in Lemma 4.20 and $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$ and $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$.

Proof. Using Lemmas 4.11 and 4.16, we again obtain

$$\eta\beta\|B(x(t)) - B(y)\|^2 \leqslant \left(1 + \frac{\eta}{\beta}\right)\|x(t) - y\|\|T(x(t)) - x(t)\| \leqslant \left(1 + \frac{\eta}{\beta}\right) K\|T(x(t)) - x(t)\|$$

and so, using $\eta \leqslant 2\beta$, we have

$$\|B(x(t)) - B(y)\| \leqslant \sqrt{\frac{3K}{\eta\beta}}\sqrt{\|T(x(t)) - x(t)\|}.$$

Lemma 4.20 now yields the result. $\qquad\square$

Remark 4.28. As in the case of the preceding case study on the first-order dynamical system, also here our approach towards the convergence of $B(x(t))$ to $B(y)$ for a solution $y \in \mathrm{zer}(A+B)$ deviates from the argument given in the respective paper that we analyze, that is [22] in this case. This is purely for reasons of simplicity, and we could have just as well analyzed the

concrete proof of this fact given in [22] (see the proof of item (iv) of Theorem 12 in [22]), which proceeds via a long chain of inequalities to establish the following key inequality (see the end of p. 1433 in [22]):

$$\frac{\beta\underline{\lambda}}{2}\int_0^T \|B(x(t)) - B(y)\|^2\, dt + \frac{1}{\eta}\left(\dot{h}(T) - \dot{h}(0) + \gamma(T)h(T) - \gamma(0)h(0)\right) \leqslant \frac{1}{\eta}\int_0^T \|\dot{x}(t)\|\, dt,$$

where $h(t) := \frac{1}{2}\|x(t) - y\|^2$. By this, it follows in particular that

$$\frac{\beta\underline{\lambda}}{2}\int_0^T \|B(x(t)) - B(y)\|^2\, dt \leqslant \frac{1}{\eta}\left(\int_0^T \|\dot{x}(t)\|\, dt + |\dot{h}(T)| + |\dot{h}(0)| + \gamma(0)h(0)\right)$$

which can be used to calculate an explicit bound on the L^2-norm of the dynamical system $\|B(x(t)) - B(y)\|^2$, using in particular Lemmas 4.16 and 4.18 to bound $|\dot{h}(T)|$ and Lemma 4.19 for a bound on the L^2-norm of $\|\dot{x}(t)\|$. Using the differential inequality

$$\frac{d}{dt}\left(\frac{1}{2}\|B(x(t)) - B(y)\|^2\right) \leqslant \frac{1}{2}\|B(x(t)) - B(y)\|^2 + \frac{1}{2\beta}\|\dot{x}(t)\|^2$$

established in [22] (see the first displayed equation on p. 1434 therein), the convergence of $B(x(t))$ to $B(y)$ follows. However, an analysis of this argument also "only" yields a rate of metastability as we apply Lemma 2.4, which generally only comes furnished with a rate of metastability, to exploit the above inequality with the respective bounds. In particular, it will be presumably of quite similar complexity as the rate presented in the preceding Lemma 4.12.

We now obtain a result on rates of metastability in the context of a uniform monotonicity assumption, being a quantitative rendering of item (v) from Theorem 12 in [22]:

Theorem 4.29. *Let X be a Hilbert space and let $A : X \to 2^X$ be maximally monotone and $B : X \to X$ be β-cocoercive for $\beta > 0$. For $\eta \in (0, 2\beta)$ and $\delta := \frac{4\beta - \eta}{2\beta}$, let $\lambda, \gamma : [0, +\infty) \to (0, +\infty)$ be Lebesgue measurable functions which are locally absolutely continuous and satisfy*

$$\dot{\gamma}(t) \leqslant 0 \leqslant \dot{\lambda}(t) \ \text{ and } \ \frac{\gamma^2(t)}{\lambda(t)} \geqslant \frac{2(1 + \theta)}{\delta}$$

almost everywhere, for some $\theta > 0$. Fix $u_0, v_0 \in X$ and let $y \in \mathrm{zer}(A + B)$ with $b, c, d \in \mathbb{N}$ such that $\|u_0 - y\| \leqslant b$, $\|v_0\| \leqslant c$ and $\|u_0 - T(u_0)\| \leqslant d$. Let x be the unique strong global solution to $(*)_2'$.

Then, if either A or B is uniformly monotone with function ϕ, x converges to y and for any $f : \mathbb{N} \to \mathbb{N}$ and $\varepsilon > 0$, we have

$$\exists n \leqslant \Theta(\varepsilon, f) \ \forall t \in [n, n + f(n)]\, (\|x(t) - y\| < \varepsilon)$$

where $\Theta(\varepsilon, f)$ is defined as follows:

(i) If A is uniformly monotone with function ϕ, then

$$\Theta(\varepsilon, f) := \Lambda\left(\min\left\{\left(\frac{\phi(\varepsilon/2)}{2K}\right)^2 \eta\beta/3K, \frac{\eta\phi(\varepsilon/2)}{2K}, \frac{\varepsilon}{2}\right\}, f\right).$$

(ii) If B is uniformly monotone with function ϕ, then

$$\Theta(\varepsilon, f) := \Lambda\left(\left(\frac{\phi(\varepsilon)}{K}\right)^2 \eta\beta/3K, f\right).$$

Here, Λ is defined as in Lemma 4.20 and $M := bc + \frac{\overline{\gamma}}{2}b^2 + \beta\overline{\gamma}\underline{\lambda}^{-1}c^2$ and $K := \sqrt{b^2 + 2\underline{\gamma}^{-1}M}$.

Proof. Items (i) and (ii) follow by a combination of Lemmas 4.13 as well as 4.16 and 4.20 together with Proposition 4.27. □

4.3. (Generalized) gradient flows in Hadamard spaces.

We now move to the topic of flows and associated semigroups in Hadamard spaces, and in particular to our applications to the classic gradient flow semigroup [16, 71] as well as nonlinear semigroup generated by a nonexpansive mapping [18, 87]. Throughout, if not stated otherwise, let X be a Hadamard space with metric d, that is (X, d) is a complete geodesic metric space of nonpositive curvature in the sense of Alexandrov. We refer to [17, 35] for further information on these spaces, and here just note that Hadamard spaces are uniquely geodesic. In particular, given $x, y \in X$ and $\lambda \in [0, 1]$, we write $(1 - \lambda)x \oplus \lambda y$ for the unique point z on the geodesic connecting x and y satisfying $d(x, z) = \lambda d(x, y)$ and $d(y, z) = (1 - \lambda)d(x, y)$.

To motivate our case studies, recall the inclusion

$$\begin{cases} -\dot{x}(t) \in A(x(t)), \\ x(0) = x_0, \end{cases}$$

where A is a maximally monotone operator on a Hilbert space X and $x_0 \in \text{dom}(A)$, as already discussed in Section 2 before. This inclusion problem not only generalizes the gradient flow equation $\dot{x} = \nabla\phi(x)$ but also captures the parabolic problem $-\dot{x} \in \partial\phi(x)$ associated with a convex lower-semicontinuous (lsc) function $\phi : X \to (-\infty, +\infty]$. By the well-known Crandall-Liggett theorem [40] (see also the preceding work [34, 41, 64]), the solutions of this problem induce a nonexpansive semigroup $S_t(x_0)$ and this semigroup can moreover be explicitly described by the exponential formula $S_t(x_0) := \lim_{n\to\infty}(J_{t/n})^{(n)}(x_0)$, where $J_\lambda(x) := (\text{Id} + \lambda A)^{-1}x$ is the resolvent of of A. This formula moreover generally produces a semigroup, even outside the solvability of the above problem. The study of the asymptotic behavior of solutions to the above problem can hence be addressed through general theory for these nonexpansive semigroups.

In particular, the above access to the inclusion problem via the exponential formula allows for an extension to the nonlinear context of Hadamard spaces. Given a Hadamard space X and a convex lsc function $\phi : X \to (-\infty, +\infty]$, a nonlinear version of the above gradient flow semigroup can be defined via

$$S_t(x) := \lim_{n\to\infty} \left(J_{t/n}\right)^{(n)}(x)$$

for $t \geqslant 0$ and $x \in \overline{\text{dom}\phi}$ for $\text{dom}\phi := \{x \in X \mid \phi(x) < +\infty\}$. Here, $J_t(x)$ is now the resolvent of ϕ defined by

$$J_t(x) := \arg\min_{y\in X} \left(\phi(y) + \frac{1}{2t}d^2(x, y)\right).$$

for $t > 0$ (and via $J_t(x) := x$ for $t = 0$). The study of this limit and the nonexpansive semigroup it generates in particular goes back to the work of Mayer [71] and was later extended by Bačák [16]. In particular, this semigroup essentially represents a continuous-time analog of the proximal point algorithm in Hadamard spaces (see [16]).

Another instance of the previous inclusion problem in Hilbert spaces is given by considering a monotone operator $A := \text{Id} - F$ defined via a nonexpansive map $F : X \to X$. A nonlinear variant of the resulting semigroup was subsequently considered by Stojkovic [87]: For a nonexpansive map $F : X \to X$ on a Hadamard space X and given $x \in X$ as well as $t > 0$, let $G_{x,t} : X \to X$ be defined by

$$G_{x,t}(y) := \frac{1}{1+t}x \oplus \frac{t}{1+t}F(y).$$

It can be easily seen that $G_{x,t}$ is a strict contraction, and so it has a unique fixed point $R_t(x)$. One calls $R_t : X \to X$ the resolvent of F. For $t \in [0, \infty)$ and $x \in X$, the resulting exponential formula

$$T_t(x) := \lim_{n \to \infty} \left(R_{t/n} \right)^{(n)} (x).$$

similarly induces a nonexpansive semigroup on X (see [87]). Similar definitions were already studied by Reich and Shafrir [85] for coaccretive operators on hyperbolic spaces, and the asymptotic study of the semigroup was in particular further developed by Bačák and Reich [18], extending previous results restricted to the Hilbert ball by Reich [84]. A practical use of the above semigroup is in particular illustrated in [18], where it is used to provide a novel access to the Hadamard-space-valued version of the L^2-Dirichlet problem on measure spaces with a symmetric Markov kernel as pioneered in the seminal work of Sturm [88] (see also [71] for related results).

The case studies here should be understood to be indicative, and we expect that other semigroups and flows on nonlinear spaces can be quantitatively approached in a similar manner. In particular we want to highlight the recent work of Chaipunya, Kohsaka and Kumam [36], where the notion of a monotone vector field known from Hadamard manifolds is extended to Hadamard spaces. Such fields carry a natural notion of a resolvent, and associated semigroups generated via an exponential formula are already discussed in [36]. Studying this semigroup could in particular provide a uniform (quantitative) study of both special cases studied in the present work, next to the associated monotone inclusions on Hilbert spaces and Hadamard manifolds, and beyond. We here however decided to only focus on the works [16, 71] and [18, 87] as they provide classic nontrivial examples that illustrate our metric generality without requiring too much new theory, in particular as the asymptotic study of the general semigroups from [36] is still a bit underdeveloped. We hence leave this for future work, as we do the study of flows in different geometric contexts, such as heat flows on Finsler manifolds as in [78] or gradient flows on Wasserstein spaces as in [77].

4.3.1. *The gradient flow of a convex function.* We first study the classical gradient flow over a Hadamard space as in [16, 71]. As before, we assume that X is a Hadamard space with metric d and we further fix a convex lsc $\phi : X \to (-\infty, +\infty]$, for which denote the associated gradient flow semigroup, defined as above, by $S_t(x)$. Throughout, we assume that $\arg \min \phi \neq \varnothing$ and write $\mu := \min_{x \in X} \phi(x)$.

The asymptotic behavior of these semigroups is studied in [71, 16]. In particular, strong convergence towards a minimizer of ϕ in proper Hadamard spaces is established in [71] (see Theorem 2.41 therein), together with strong convergence under a uniform convexity assumption on ϕ but for that over general Hadamard spaces (see Theorem 2.42 therein). Weak convergence towards a minimizer of ϕ was subsequently shown in [16] (see Theorem 1.5 therein). We now study these results from a quantitative perspective.

The key result that establishes both the continuous Fejér monotonicity of the semigroup as well as the associated approximation property is the following inequality:

Lemma 4.30 (Lemma 2.37 in [71])**.** *For any $x \in \mathrm{dom}\phi$, $z \in X$ and $t \geq s \geq 0$:*

$$d^2(S_t(x), z) \leq d^2(S_s(x), z) - 2(t-s)(\phi(S_t(x)) - \phi(z)).$$

Immediately, as $S_0(x) := x$, this yields the boundedness of the semigroup.

Lemma 4.31. *Let $y \in \arg \min \phi$, $x \in \mathrm{dom}\phi$ and let $b \in \mathbb{N}$ with $d(x, y) \leq b$. Then $S_t(x)$ is bounded with $d(S_t(x), y) \leq b$.*

To capture the corresponding minimization problem in the setup of the previous sections, we set $X_0 := \overline{B}_b(y) \cap \overline{\mathrm{dom}\phi}$ and

$$F(z) = \begin{cases} \phi(z) - \mu & \text{if } z \in X_0, \\ +\infty & \text{otherwise}, \end{cases}$$

for a fixed solution $y \in \arg\min\phi$ so that $\mathrm{zer}F = \arg\min\phi \cap X_0$ and

$$\mathrm{lev}_{\leqslant\varepsilon}F = \{z \in X_0 \mid \phi(z) - \mu \leqslant \varepsilon\}.$$

Mayer's inequality in particular also yields a modulus of uniform Fejér monotonicity (with $G = H = (\cdot)^2$):

Lemma 4.32. *Let $y \in \arg\min\phi$, $x \in \mathrm{dom}\phi$ and let $b \in \mathbb{N}$ with $d(x,y) \leqslant b$. Then $S_t(x)$ is uniformly (G,H)-Fejér monotone w.r.t. F and with $G = H = (\cdot)^2$, i.e.*

$$\forall\varepsilon > 0 \,\forall n,m \in \mathbb{N} \,\forall z \in \mathrm{lev}_{\leqslant\chi(\varepsilon,n,m)}F \,\forall s \leqslant t \in [n,n+m] \left(d^2(S_t(x),z) \leqslant d^2(S_s(x),z) + \varepsilon\right),$$

where $\chi(\varepsilon,n,m) := \varepsilon/2m$.

Proof. Fix ε,n,m and let $z \in \mathrm{lev}_{\leqslant\chi(\varepsilon,n,m)}F$, i.e. $z \in X_0$ with $\phi(z) - \mu \leqslant \chi(\varepsilon,n,m) \leqslant \varepsilon/2m$. Lemma 4.30 yields for $s \leqslant t \in [n,n+m]$ that

$$\begin{aligned} d^2(S_t(x),z) &\leqslant d^2(S_s(x),z) - 2(t-s)\left(\phi(S_t(x)) - \phi(z)\right) \\ &\leqslant d^2(S_s(x),z) - 2m\left(\mu - \phi(z)\right) \\ &\leqslant d^2(S_s(x),z) + \varepsilon. \qquad \square \end{aligned}$$

Also, we get the following approximation property from Mayer's inequality:

Lemma 4.33. *Let $y \in \arg\min\phi$, $x \in \mathrm{dom}\phi$ and let $b \in \mathbb{N}$ with $d(x,y) \leqslant b$. Then*

$$\forall\varepsilon > 0 \,\forall t \geqslant \varphi(\varepsilon) \left(\phi(S_t(x)) - \mu < \varepsilon\right)$$

with rate $\varphi(\varepsilon) := \lceil b^2/\varepsilon \rceil$. In particular, $S_t(x)$ has approximate F-points with modulus φ.

Proof. For $t \geqslant 0$, Lemma 4.30 yields

$$0 \leqslant d^2(S_t(x),y) \leqslant d^2(x,y) - 2t\left(\phi(S_t(x)) - \mu\right) \leqslant b^2 - 2t\left(\phi(S_t(x)) - \mu\right),$$

so that we get

$$\phi(S_t(x)) - \mu \leqslant \frac{b^2}{2t} \text{ for all } t > 0.$$

In particular, for $t \geqslant \lceil b^2/\varepsilon \rceil$, we have

$$\phi(S_t(x)) - \mu \leqslant \frac{b^2}{2\lceil b^2/\varepsilon \rceil} \leqslant \frac{\varepsilon}{2} < \varepsilon. \qquad \square$$

Combined, we can hence give the following result on the quantitative asymptotic behavior of the semigroup in the context of proper Hadamard spaces, a quantitative version of Theorem 2.41 in [71].

Theorem 4.34. *Consider an Hadamard space X and a convex lsc function $\phi : X \to (-\infty,+\infty]$. Let $y \in \arg\min\phi$, $x \in \mathrm{dom}\phi$ and let $b \in \mathbb{N}$ with $d(x,y) \leqslant b$. Define $X_0 := \overline{B}_b(y) \cap \overline{\mathrm{dom}\phi}$ and assume that γ is a modulus of total boundedness for X_0.*

Then $S_t(x)$ is metastable in the sense that for all $\varepsilon > 0$ and all nondecreasing $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon,f) \,\forall s,t \in [n,n+f(n)] \left(d(S_t(x),S_s(x)) \leqslant \varepsilon\right),$$

where moreover the bound $\Delta(\varepsilon, f)$ *can be explicitly described by* $\Delta(\varepsilon, f) := \Delta(\gamma(\varepsilon/\sqrt{12}) + 1, \varepsilon, f) + 1$ *with* $\Delta(0, \varepsilon, f) := 0$ *and*

$$\Delta(j + 1, \varepsilon, f) := \left\lceil 24b^2 \big(f(\Delta(j, \varepsilon, f) + 1) + 1 \big)/\varepsilon^2 \right\rceil.$$

for $j \geqslant 1$.

Proof. The result immediately follows by instantiating Theorem 3.9 over the given X_0 and simplifying the resulting bound (recall that $S_t(x) \in X_0$ for all $t \geqslant 0$ by Lemma 4.1 and since $S_t(x) \in \overline{\mathrm{dom}\phi}$). In particular, we set $G := H := (\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon}$. That $S_t(x)$ is uniformly (G, H)-Fejér monotone w.r.t. F follows by Lemma 4.32, which in particular yields that $\chi(\varepsilon, n, m) := \varepsilon/2m$ is an associated modulus. As we do not deal with errors, it suffices that $S_t(x)$ has the approximate F-point property, which follows by Lemma 4.33 with a respective bound $\varphi(\varepsilon) := \lceil b^2/\varepsilon \rceil$ that is immediately monotone under our assumptions. $\square$

If we assume that ϕ is uniformly continuous with a respective modulus, then this continuity property transfers to F, so that using Theorem 3.9 we can moreover provide a bound that guarantees that $\phi(S_t(x)) - \mu \leqslant \varepsilon$ holds along the region of metastability. We do not spell this out here any further.

Further, we can give the following result on rates of convergence under regularity assumptions in general Hadamard spaces.

Theorem 4.35. *Consider an Hadamard space X and a convex lsc function $\phi : X \to (-\infty, +\infty]$. Let $y \in \arg\min \phi$, $x \in \mathrm{dom}\phi$ and let $b \in \mathbb{N}$ with $d(x, y) \leqslant b$. Assume that F has a modulus of regularity τ w.r.t. $\overline{B}_b(y)$.*
Then $S_t(x)$ satisfies

$$\forall \varepsilon > 0 \ \forall t \geqslant \rho(\varepsilon) \ (\mathrm{dist}(S_t(x), \arg\min \phi) < \varepsilon)$$

and further $S_t(x)$ is Cauchy with

$$\forall \varepsilon > 0 \ \forall t, s \geqslant \rho(\varepsilon/2) \ (d(S_t(x), S_s(x)) < \varepsilon),$$

where ρ is defined by $\rho(\varepsilon) := \lceil b^2/\tau(\varepsilon) \rceil + 1$. In particular, $S_t(x)$ converges to a minimizer of ϕ with that rate.

Proof. The result immediately follows by instantiating Theorem 3.18 and simplifying the resulting bound. In particular, we set $G := H := (\cdot)^2$ as well as $h(\varepsilon) := \varepsilon^2$ and $g(\varepsilon) := \sqrt{\varepsilon}$. That $S_t(x)$ is (G, H)-Fejér monotone w.r.t. F follows from the fact that $S_t(x)$ is a nonexpansive semigroup (recall also Lemma 2.37 in [71] and Lemma 4.32). As we do not deal with errors, it suffices that $S_t(x)$ has the approximate F-point property which, with a respective bound $\varphi(\varepsilon) := \lceil b^2/\varepsilon \rceil$ follows using Lemma 4.33 as before. $\square$

The above in particular applies to functions ϕ with weak sharp minima (recall Example 3.16), and hence in particular yields a the strong convergence of the semigroup, together with an associated rate, when the function ϕ is uniformly convex. While derived here using our general methodology, such a rate of convergence for the uniformly convex case can in fact be easily read off from the convergence proof tailored to this special case in [16]. Concretely, a more direct inspection of the proof given in [16] gives a rate of convergence $2b^2/\tau(\varepsilon)$, where τ is the modulus of uniform convexity, which is virtually the same as the rate $\lceil 4b^2/\tau(\varepsilon/2) \rceil + 1$ derived from the above result.

4.3.2. *A class of nonlinear semigroups for a nonexpansive mapping.* We now move on to the semigroup generated by a nonexpansive map as in [87, 18]. As before, let X be a Hadamard space with metric d and now fix a nonexpansive map $F : X \to X$. Define the associated function $G_{x,t}(y)$ and is fixed point selection $R_t(x)$ as before, and denote the semigroup associated with $R_t(x)$ by $T_t(x)$. Throughout, we assume that $\text{Fix} F \neq \varnothing$.

The asymptotic behavior of these semigroups is studied in [18, 87]. In particular, it is established in [18] (see Theorem 1.6 therein) that $T_t(x)$ converges (in general weakly) to a fixed point of F, for any starting point $x \in X$. We now study this result from a quantitative perspective.

We begin with the boundedness of the semigroup.

Lemma 4.36. *Let* $y \in \text{Fix} F$, $x \in X$ *and let* $b \in \mathbb{N}$ *with* $d(x, y) \leqslant b$. *Then* $T_t(x)$ *is bounded with* $d(T_t(x), y) \leqslant b$.

Proof. Note that since $y \in \text{Fix} F$, we also have $T_t(y) := y$. Further, $T_t(x)$ is a nonexpansive semigroup (see Theorem 3.10 in [87] for both). Combined, we get

$$d(T_t(x), y) = d(T_t(x), T_t(y)) \leqslant d(x, y) \leqslant b. \qquad \square$$

To capture the corresponding fixed point problem in the setup of the previous sections, we set $X_0 := \overline{B}_b(y)$ and

$$\bar{F}(z) = \begin{cases} d(z, F(z)) & \text{if } z \in X_0, \\ +\infty & \text{otherwise,} \end{cases}$$

for a fixed solution $y \in \arg\min f$ so that $\text{zer}\bar{F} = \text{Fix} F \cap X_0$ and

$$\text{lev}_{\leqslant \varepsilon} \bar{F} = \{z \in X_0 \mid d(z, F(z)) \leqslant \varepsilon\}.$$

We now first derive the following approximation property:

Lemma 4.37. *Let* $y \in \text{Fix} F$, $x \in X$ *and let* $b \in \mathbb{N}$ *with* $d(x, y) \leqslant b$. *Then* $T_t(x)$ *has appproximate* $\bar{F}$-*points with modulus* φ, *i.e.*

$$\forall \varepsilon > 0 \, \exists t \leqslant \varphi(\varepsilon) \, (d(T_t(x), F(T_t(x))) < \varepsilon),$$

where we have $\varphi(\varepsilon) := \left\lceil (4b/\varepsilon) \cdot e^{4b/\varepsilon} \right\rceil$.

Proof. As $F(y) = (y)$ entails $T_t(y) = (y)$, the claim is immediate for $x = y$. In the following, we may thus assume $b > 0$ and hence $h := 4b/\varepsilon > 0$. Due to the semigroup property and since T_t is nonexpansive, we have

$$d(T_{t+h}(x), T_t(x)) = d(T_t(T_h(x)), T_t(x)) \leqslant d(T_h(x), x).$$

As in [18] (see the fourth equation on page 197 therein), we can hence derive

$$d(T_t(x), F(T_t(x))) \leqslant \frac{d(T_h(x), x)}{h} + \frac{(e^h - 1)\big(d(T_t x, FT_t x) - d(T_{t+h}(x), F(T_{t+h}(x)))\big)}{h}.$$

In view of $T_h(y) = y$ and since T_h is nonexpansive, we have

$$\frac{d(T_h(x), x)}{h} \leqslant \frac{d(T_h(x), T_h(y)) + d(y, x)}{h} \leqslant \frac{2b}{h} = \frac{\varepsilon}{2}.$$

So to complete the proof, it is enough to show that there is a $t \leqslant \varphi(\varepsilon)$ with

$$\frac{(e^h - 1)\big(d(T_t(x), F(T_t(x))) - d(T_{t+h}(x), F(T_{t+h}(x)))\big)}{h} < \frac{\varepsilon}{2}.$$

Towards a contradiction, we assume that all $t \leqslant \varphi(\varepsilon)$ validate

$$d(T_t(x), F(T_t(x))) - d(T_{t+h}(x), F(T_{t+h}(x))) \geqslant \frac{\varepsilon}{2} \cdot \frac{h}{e^h - 1} = \frac{2b}{e^h - 1}.$$

Observe that we have

$$(N-1)h \leqslant \varphi(\varepsilon) \text{ for } N := \lceil e^h \rceil.$$

Via a telescoping sum, we obtain

$$d(T_0(x), F(T_0(x))) - d(T_{N \cdot h}(x), F(T_{N \cdot h}(x))) \geqslant N \cdot \frac{2b}{e^h - 1} > 2b.$$

But we also have

$$d(T_0(x), F(T_0(x))) = d(x, F(x)) \leqslant d(x, y) + d(F(y), F(x)) \leqslant 2b,$$

which yields the desired contradiction. $\qquad\square$

Our next goal is (a quantitative version of) Fejér monotonicity for the semigroup $T_t(x)$. In a preparatory lemma, we first provide a quantitative version of the fact that fixed points of F are also fixed points of the semigroup.

Lemma 4.38. *For any $x \in X$ and $t, \delta > 0$, if $d(x, F(x)) \leqslant \delta$, then $d(x, T_t(x)) \leqslant \delta \cdot \frac{e^{2t}-1}{2}$.*

Proof. Write $x_k = (R_\lambda)^{(k)}(x)$. Assuming $d(x, F(x)) \leqslant \delta$, we shall establish

$$d(x_n, x_{n+1}) \leqslant \lambda(1 + 2\lambda)^n \cdot \delta$$

Lemma 3.4 in [87] (see also Lemma 2.2 in [18]) yields that

$$d(z, R_\lambda(z)) \leqslant \lambda d(z, F(z))$$

for all $z \in X$ and $\lambda > 0$, so that the above estimate reduces to

$$d(x_n, F(x_n)) \leqslant (1 + 2\lambda)^n \cdot \delta.$$

For $n = 0$, this holds by assumption. Using that F is nonexpansive, we inductively get

$$d(x_{n+1}, F(x_{n+1})) \leqslant d(x_{n+1}, x_n) + d(x_n, F(x_n)) + d(F(x_n), F(x_{n+1}))$$
$$\leqslant 2\lambda(1 + 2\lambda)^n \cdot \delta + (1 + 2\lambda)^n \cdot \delta = (1 + 2\lambda)^{n+1} \cdot \delta.$$

Let us now infer

$$d(x, x_n) \leqslant \sum_{k=0}^{n-1} d(x_k, x_{k+1}) \leqslant \lambda \cdot \delta \cdot \sum_{k=0}^{n-1} (1 + 2\lambda)^k = \delta \cdot \frac{(1 + 2\lambda)^n - 1}{2}.$$

As $T_t(x)$ is defined as the limit of $(R_{t/n})^{(n)}(x)$ for $n \to \infty$, we specialize to $\lambda = t/n$ and obtain

$$d(x, T_t(x)) = \lim_{n \to \infty} d\left(x, (R_{t/n})^{(n)}(x)\right) \leqslant \lim_{n \to \infty} \frac{\delta}{2} \cdot \left(\left(1 + \frac{2t}{n}\right)^n - 1\right) = \delta \cdot \frac{e^{2t} - 1}{2},$$

just like the lemma claims. $\qquad\square$

We can now derive the following quantitative result on the Fejér monotonicity of $T_t(x)$.

Lemma 4.39. *Let $y \in \mathrm{Fix} F$, $x \in X$ and let $b \in \mathbb{N}$ with $d(x, y) \leqslant b$. Then $T_t(x)$ is uniformly (G, H)-Fejér monotone w.r.t. $\bar{F}$ and with $G = H = \mathrm{Id}$, i.e.*

$$\forall \varepsilon > 0 \; \forall n, m \in \mathbb{N} \; \forall z \in \mathrm{lev}_{\leqslant \chi(\varepsilon, n, m)} \bar{F} \; \forall s \leqslant t \in [n, n + m] \left(d(T_t(x), z) \leqslant d(T_s(x), z) + \varepsilon\right),$$

where $\chi(\varepsilon, n, m) := 2\varepsilon/(e^{2m} - 1)$.

Proof. Let ε, n, m be given and let $z \in \mathrm{lev}_{\leqslant \chi(\varepsilon, n, m)}$. Using Lemma 4.38 with $\delta = 2\varepsilon/(e^{2m} - 1)$, we get $d(T_{t-s}(z), z) \leqslant \varepsilon$. Using that $T_t(x)$ is a nonexpansive semigroup, we get

$$d(T_t(x), z) \leqslant d(T_{t-s}(T_s(x)), T_{t-s}(z)) + d(T_{t-s}(z), z) \leqslant d(T_s(x), z) + \varepsilon. \qquad \square$$

We can now give the following result on the quantitative asymptotic behavior of the semigroup in the context of proper Hadamard spaces, a quantitative version of (parts of) Theorem 1.6 in [18].

Theorem 4.40. *Consider an Hadamard space X and a nonexpansive mapping $F : X \to X$. Let $y \in \mathrm{Fix}F$, $x \in X$ and let $b \in \mathbb{N}$ with $d(x, y) \leqslant b$. Define $X_0 := \overline{B}_b(y)$ and assume that γ is a modulus of total boundedness for X_0.*

Then $T_t(x)$ is metastable in the sense that for all $\varepsilon > 0$ and all nondecreasing $f : \mathbb{N} \to \mathbb{N}$, it holds that

$$\exists n \leqslant \Delta(\varepsilon, f) \; \forall s, t \in [n, n + f(n)] \, (d(T_t(x), T_s(x)) \leqslant \varepsilon),$$

where moreover the bound $\Delta(\varepsilon, f)$ can be explicitly described by $\Delta(\varepsilon, f) := \varphi(\widehat{\varepsilon}_P)$ for $P := \gamma(\varepsilon/6) + 1$ with $\widehat{\varepsilon}_1 := \chi_f(\varepsilon/6, 0)$ and

$$\widehat{\varepsilon}_j := \chi_f \left(\frac{\varepsilon}{6}, \varphi(\widehat{\varepsilon}_{j-1}) \right)$$

for $j \geqslant 1$, where

$$\varphi(\varepsilon) := \left\lceil \frac{4r}{\varepsilon} \cdot e^{4r/\varepsilon} \right\rceil \quad \text{and} \quad \chi_f(\varepsilon, n) := \frac{2\varepsilon}{e^{2(f(n+1)+1)} - 1}.$$

Proof. The result immediately follows by instantiating Theorem 3.9 over the given X_0 and simplifying the resulting bound (recall that $T_t(x) \in X_0$ for all $t \geqslant 0$ by Lemma 4.1). In particular, we set $G := H := \mathrm{Id}$ as well as $h := g := \mathrm{Id}$. That $T_t(x)$ is uniformly (G, H)-Fejér monotone w.r.t. $\bar{F}$ follows by Lemma 4.39, which in particular yields that $\chi(\varepsilon, n, m) := 2\varepsilon/(e^{2m} - 1)$ is an associated modulus. As we do not deal with errors, it suffices that $T_t(x)$ has the approximate $\bar{F}$-point property, which follows by Lemma 4.37 with a respective bound $\varphi(\varepsilon) := \left\lceil (4b/\varepsilon) \cdot e^{4b/\varepsilon} \right\rceil$ that is immediately monotone under our assumptions. $\qquad \square$

We also derive a quantitative result under regularity and without total boundedness.

Theorem 4.41. *Consider an Hadamard space X and a nonexpansive mapping $F : X \to X$. Let $y \in \mathrm{Fix}F$, $x \in X$ and let $b \in \mathbb{N}$ with $d(x, y) \leqslant b$. Assume that $\bar{F}$ has a modulus of regularity τ w.r.t. $\overline{B}_b(y)$.*

Then $T_t(x)$ satisfies

$$\forall \varepsilon > 0 \; \forall t \geqslant \rho(\varepsilon) \, (\mathrm{dist}(T_t(x), \mathrm{Fix}F) < \varepsilon)$$

and further $T_t(x)$ is Cauchy with

$$\forall \varepsilon > 0 \; \forall t, s \geqslant \rho(\varepsilon/2) \, (d(T_t(x), T_s(x)) < \varepsilon),$$

where ρ is defined by $\rho(\varepsilon) := \left\lceil (4b/\tau(\varepsilon)) \cdot e^{4b/\tau(\varepsilon)} \right\rceil + 1$. wIn particular, $T_t(x)$ converges to a fixed point of F with that rate.

Proof. The result immediately follows by instantiating Theorem 3.18 and simplifying the resulting bound. In particular, we set $G := H := \mathrm{Id}$ as well as $h := g := \mathrm{Id}$. That $T_t(x)$ is (G, H)-Fejér monotone w.r.t. $\bar{F}$ follows from the fact that $T_t(x)$ is a nonexpansive semigroup (recall also Lemma 4.39). As we do not deal with errors, it suffices that $T_t(x)$ has the approximate $\bar{F}$-point property which, with a respective bound $\varphi(\varepsilon) := \left\lceil (4b/\varepsilon) \cdot e^{4b/\varepsilon} \right\rceil$ follows using Lemma 4.37 as before. $\qquad \square$

References

[1] B. Abbas and H. Attouch. Dynamical systems and forward–backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator. *Optimization*, 64(10):2223–2252, 2015.

[2] B. Abbas, H. Attouch, and B.F. Svaiter. Newton-Like Dynamics and Forward-Backward Methods for Structured Monotone Inclusions in Hilbert Spaces. *Journal of Optimization Theory and Applications*, 161:331–360, 2014.

[3] C.D. Alecsa, S.C. László, and T. Pinţa. An extension of the second order dynamical system that models Nesterov's convex gradient method. *Applied Mathematics and Optimization*, 84(2):1687–1716, 2021.

[4] F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, 38:1102–1119, 2000.

[5] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis*, 9:3–11, 2001.

[6] A.S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Differential equations*, 30:1365–1375, 1994.

[7] A. Arakcheev and H.H. Bauschke. Fejér and Fejér* Monotonicity: New Results and Limiting Examples, 2025. Preprint, available at `https://arxiv.org/abs/2512.17039`.

[8] H. Attouch and F. Alvarez. The heavy ball with friction dynamical system for convex constrained minimization problems. In V.H. Nguyen, JJ. Strodiot, and P. Tossings, editors, *Optimization, Namur, 1998*, volume 481 of *Lecture Notes in Economics and Mathematical Systems*, pages 25–35. Springer, Berlin, 2000.

[9] H. Attouch and M.O. Czarnecki. Asymptotic behavior of coupled dynamical systems with multiscale aspects. *Journal of Differential Equations*, 248(6):1315–1344, 2010.

[10] H. Attouch, X. Goudou, and P. Redont. The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 2(1):1–34, 2000.

[11] H. Attouch and S.C. László. Continuous Newton-like inertial dynamics for monotone inclusions. *Set-valued and Variational Analysis*, 29(3):555–581, 2021.

[12] H. Attouch and B.F. Svaiter. A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM Journal on Control and Optimization*, 49(3):574–598, 2011.

[13] S. Banert and R.I. Boţ. A forward-backward-forward differential equation and its asymptotic properties. *Journal of Convex Analysis*, 25(2):371–388, 2018.

[14] H.H. Bauschke, J.M. Borwein, and P.L. Borwein. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[15] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, Cham, 2nd edition, 2017.

[16] M. Bačák. The Proximal Point Algorithm in Metric Spaces. *Israel Journal of Mathematics*, 194:689–701, 2013.

[17] M. Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter GmbH, Berlin/Boston, 2014.

[18] M. Bačák and S. Reich. The asymptotic behavior of a class of nonlinear semigroups in Hadamard spaces. *Journal of Fixed Point Theory and Applications*, 16:189–202, 2014.

[19] R. Behling, Y. Bello-Cruz, A. Iusem, D. Liu, and L.R. Santos. A finitely convergent circumcenter method for the convex feasibility problem. *SIAM Journal on Optimization*, 34:2535–2556, 2024.

[20] R. Behling, Y. Bello-Cruz, A.N. Iusem, A. Alves Ribeiro, and L.-R. Santos. Fejér* monotonicity in optimization algorithms, 2024. Preprint, available at `https://arxiv.org/abs/2410.08331`.

[21] R.I. Boţ and E.R. Csetnek. Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems. *Journal of Mathematical Analysis and Applications*, 435(2):1688–1700, 2016.

[22] R.I. Boţ and E.R. Csetnek. Second order forward-backward dynamical systems for monotone inclusion problems. *SIAM Journal on Control and Optimization*, 54(3):1423–1443, 2016.

[23] R.I. Boţ and E.R. Csetnek. A dynamical system associated with the fixed points set of a nonexpansive operator. *Journal of Dynamics and Differential Equations*, 29:155–168, 2017.

[24] R.I. Boţ and E.R. Csetnek. Second-order dynamical systems associated to variational inequalities. *Applicable Analysis*, 96(5):799–809, 2017.

[25] R.I. Boţ and E.R. Csetnek. Convergence rates for forward–backward dynamical systems associated with strongly monotone inclusions. *Journal of Mathematical Analysis and Applications*, 457(2):1135–1152, 2018.

[26] R.I. Boţ and E.R. Csetnek. A second-order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities. *Optimization*, 68(7):1265–1277, 2019.

[27] R.I. Boţ, E.R. Csetnek, and S.C. László. Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems. *Journal of Evolution Equations*, 18(3):1291–1318, 2018.

[28] R.I. Boţ, E.R. Csetnek, and S.C. László. A primal-dual dynamical approach to structured convex minimization problems. *Journal of Differential Equations*, 269(12):10717–10757, 2020.

[29] R.I. Boţ, E.R. Csetnek, and P.T. Vuong. The forward–backward–forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces. *European Journal of Operational Research*, 287(1):49–60, 2020.

[30] R.I. Boţ and C. Schindler. On a stochastic differential equation with correction term governed by a monotone and Lipschitz continuous operator, 2024. Preprint, available at `https://arxiv.org/abs/2404.17986`.

[31] R.I. Boţ and C. Schindler. Long-Time Analysis of Stochastic Heavy Ball Dynamics for Convex Optimization and Monotone Equations, 2025. Preprint, available at `https://arxiv.org/abs/2510.02951`.

[32] J. Bolte. Continuous gradient projection method in Hilbert spaces. *Journal of Optimization Theory and Applications*, 119(2):235–259, 2003.

[33] H. Brézis. *Opérateurs Maximaux Monotones Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert*, volume 5 of *North-Holland Mathematics Studies*. North-Holland, Amsterdam, 1973.

[34] H. Brezis and A. Pazy. Accretive sets and differential equations in Banach spaces. *Israel Journal of Mathematics*, 8:367–383, 1970.

[35] M.R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg, 1999.

[36] P. Chaipunya, F. Kohsaka, and P. Kumam. Monotone vector fields and generation of nonexpansive semi-groups in complete CAT(0) spaces. *Numerical Functional Analysis and Optimization*, 42(9):989–1018, 2021.

[37] P.L. Combettes. Quasi-Fejérian Analysis of Some Optimization Algorithms. In D. Butnariu, Y. Censor, and S. Reich, editors, *Studies in Computational Mathematics*, volume 8, pages 115–152. Elsevier, 2001.

[38] P.L. Combettes. Fejér monotonicity in convex optimization. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1016–1024. Springer, Boston, MA, 2009.

[39] P.L. Combettes and B.C. Vũ. Variable metric quasi-Fejér monotonicity. *Nonlinear Analysis: Theory, Methods & Applications*, 78:17–31, 2013.

[40] M.G. Crandall and T.M. Liggett. Generation of semi-groups of nonlinear transformations on general Banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.

[41] M.G. Crandall and A. Pazy. Semi-groups of nonlinear contractions and dissipative sets. *Journal of Functional Analysis*, 3(3):376–418, 1969.

[42] E.R. Csetnek. Continuous dynamics related to monotone inclusions and non-smooth optimization problems. *Set-Valued and Variational Analysis*, 28(4):611–642, 2020.

[43] E.R. Csetnek, Y. Malitsky, and M.K. Tam. Shadow Douglas–Rachford splitting for monotone inclusions. *Applied Mathematics and Optimization*, 80(3):665–678, 2019.

[44] I.I. Eremin. Methods of Fejér approximations in convex programming. *Mathematical Notes of the Academy of Sciences of the USSR*, 3:139–149, 1968.

[45] I.I. Eremin. On the speed of convergence in the method of Fejér approximations. *Mathematical Notes of the Academy of Sciences of the USSR*, 4:522–527, 1968.

[46] Y.M. Ermol'ev. On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics*, 5:208–220, 1969.

[47] Y.M. Ermol'ev. On convergence of random quasi-Fejér sequences. *Cybernetics*, 7:655–656, 1971.

[48] Yu.M. Ermol'ev and A.D. Tuniev. Random Fejér and quasi-Fejér sequences. In *Theory of Optimal Solutions*, volume 2, pages 76–83. Akademiya Nauk Ukrainskoi SSR Kiev, 1968. American Mathematical Society Selected Translations in Mathematical Statistics and Probability 13 (1973), 143–148.

[49] L. Fejér. Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen. *Mathematische Annalen*, 85:41–48, 1922.

[50] R. Findling. Logical analysis of a splitting algorithm for dual monotone inclusions. Master's thesis, Technische Universität Darmstadt, 2026.

[51] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49:431–446, 2008.

[52] A. Haraux and M.A. Jendoubi. Convergence of solutions of second order gradient like systems with analytic nonlinearities. *Journal of Differential Equations*, 144(2):313–320, 1998.

[53] U. Kohlenbach. On the computational content of the Krasnoselski and Ishikawa fixed point theorems. In J. Blanck, V. Brattka, and P. Hertling, editors, *Computability and complexity in analysis (CCA 2000)*, volume 2064 of *Lecture Notes in Computer Science*, pages 119–145, 2001.

[54] U. Kohlenbach. Uniform asymptotic regularity for Mann iterates. *Journal of Mathematical Analysis and Applications*, 279:531–544, 2003.

[55] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.

[56] U. Kohlenbach. A logical uniform boundedness principle for abstract metric and hyperbolic spaces. *Electronic Notes in Theoretical Computer Science*, 165:81–93, 2006.

[57] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer Berlin, Heidelberg, 2008.

[58] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proceedings of ICM 2018*, volume 2, pages 61–82. World Scientific, Singapure, 2019.

[59] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics*, 20(2), 2018. 1750015, 42pp.

[60] 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.

[61] 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.

[62] U. Kohlenbach and P. Pinto. Quantitative translations for viscosity approximation methods in hyperbolic spaces. *Journal of Mathematical Analysis and Applications*, 507, 2022. 125823, 33pp.

[63] U. Kohlenbach and P. Pinto. Fejér Monotone Sequences Revisited. *Journal of Convex Analysis*, 2025. Preprint available at `https://arxiv.org/abs/2310.06528`.

[64] Y. Komura. Nonlinear semi-groups in Hilbert space. *Journal of the Mathematical Society of Japan*, 19(4):493–507, 1967.

[65] M.A. Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1(63)):123–127, 1955.

[66] J. Liu, X.J. Long, X.S. Li, and N.J. Huang. Stochastic dual dynamical systems for linear equality constrained convex optimization problems. *Communications in Nonlinear Science and Numerical Simulation*, 154, 2026. 109538, 15pp.

[67] D.R. Luke, J.C. Schnebel, M. Staudigl, J. Peypouquet, and S. Qu. Asymptotic behaviour of coupled random dynamical systems with multiscale aspects, 2026. Preprint, available at `https://arxiv.org/abs/2601.15411`.

[68] W.R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.

[69] R. Maulen-Soto, J. Fadili, and H. Attouch. An stochastic differential equation perspective on stochastic convex optimization. *Mathematics of Operations Research*, 50(4):3190–3221, 2025.

[70] R. Maulen-Soto, J. Fadili, H. Attouch, and P. Ochs. Stochastic inertial dynamics via time scaling and averaging. *Stochastic Systems*, 2026. to appear.

[71] U. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps. *Communications in Analysis and Geometry*, 6:199–253, 1998.

[72] T. Motzkin and I. Schoenberg. The Relaxation Method for Linear Inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[73] C. Moucer, A. Taylor, and F. Bach. A systematic approach to Lyapunov analyses of continuous-time models in convex optimization. *SIAM Journal on Optimization*, 33(3):1558–1586, 2023.

[74] A. Moudafi and M. Oliny. Convergence of a splitting inertial proximal method for monotone operators. *Journal of Computational and Applied Mathematics*, 155(2):447–454, 2003.

[75] M. Neri, N. Pischke, and T. Powell. On the asymptotic behaviour of stochastic processes, with applications to supermartingale convergence, Dvoretzky's approximation theorem, and stochastic quasi-Fejér monotonicity, 2025. Preprint, available at `https://arxiv.org/abs/2504.12922`.

[76] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4:20), 2015. 44pp.

[77] S.I. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *American Journal of Mathematics*, 131(2):475–516, 2009.

[78] S.I. Ohta and K.-T. Sturm. Heat flow on Finsler manifolds. *Communications on Pure and Applied Mathematics*, 62(10):1386–1433, 2009.

[79] J. Peypouquet and S. Sorin. Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *Journal of Convex Analysis*, 17(3–4):1113–1163, 2010.

[80] 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.

[81] N. Pischke. Generalized Fejér monotone sequences and their finitary content. *Optimization*, 74(14):3771–3838, 2025.

[82] N. Pischke and T. Powell. Generalized fluctuation bounds for stochastic algorithms in the presence of compactness, 2026. Preprint, available at `https://arxiv.org/abs/2602.22741`.

[83] N. Pischke and T. Powell. On regularity for stochastic problems and rates for stochastic processes, 2026. Manuscript in preparation.

[84] S. Reich. The asymptotic behavior of a class of nonlinear semigroups in the Hilbert ball. *Journal of Mathematical Analysis and Applications*, 157(1):237–242, 1991.

[85] S. Reich and I. Shafrir. Nonexpansive iterations in hyperbolic spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 15(6):537–558, 1990.

[86] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *The Journal of Symbolic Logic*, 14:145–208, 1949.

[87] I. Stojkovic. Approximation for convex functionals on non-positively curved spaces and the Trotter-Kato product formula. *Advances in Calculus of Variations*, 5(1):77–126, 2012.

[88] K.-T. Sturm. Nonlinear Markov operators associated with symmetric Markov kernels and energy minimizing maps between singular spaces. *Calculus of Variations and Partial Differential Equations*, 12:317–357, 2001.

[89] W. Takahashi. *Nonlinear Functional Analysis*. Yokohama Publishers, Yokohama, Japan, 2000.

[90] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28(2):657–688, 2008.

[91] T. Tao. Soft Analysis, Hard Analysis, and the Finite Convergence Principle. In *Structure and Randomness: Pages from Year One of a Mathematical Blog*, pages 17–29. American Mathematical Society, Providence, RI, 2008.

[92] J. Treusch and U. Kohlenbach. Rates of convergence for splitting algorithms. *Israel Journal of Mathematics*, 2026. to appear.