ASYMPTOTIC REGULARITY OF A GENERALISED STOCHASTIC HALPERN SCHEME

NICHOLAS PISCHKE AND THOMAS POWELL

Department of Computer Science, University of Bath,
Claverton Down, Bath, BA2 7AY, United Kingdom,
E-mails: {nnp39,trjp20}@bath.ac.uk

Abstract. We provide abstract, general and highly uniform rates of asymptotic regularity for a generalized stochastic Halpern-style iteration, which incorporates a second mapping in the style of a Krasnoselskii-Mann iteration. This iteration is general in two ways: First, it incorporates stochasticity in a completely abstract way rather than fixing a sampling method; secondly, it includes as special cases stochastic versions of various schemes from the optimization literature, including Halpern's iteration as well as a Krasnoselskii-Mann iteration with Tikhonov regularization terms in the sense of Boţ, Csetnek and Meier. For these specific cases, we in particular obtain linear rates of asymptotic regularity, matching (or improving) the currently best known rates for these iterations in stochastic optimization, and quadratic rates of asymptotic regularity are obtained in the context of inner product spaces for the general iteration. At the end, we briefly sketch how the schemes presented here can be instantiated in the context of reinforcement learning to yield novel methods for Q-learning.

Keywords: Asymptotic regularity, Halpern iteration, Tikhonov regularization, proof mining
MSC2020 Classification: 47J25, 47H09, 62L20, 03F10

1. Introduction

Approximating fixed points of nonexpansive mappings is one of the most fundamental tasks in nonlinear analysis and optimization. The problem becomes particularly interesting when we only have noisy versions of those mappings, in which case the resulting approximation methods become stochastic processes. Concrete examples of this general situation include model-free reinforcement learning algorithms, where variants of Q-learning, for instance, can be viewed as stochastic methods for computing fixed points of nonexpansive operators.

To be more concrete, let $(X, \|\cdot\|)$ be a separable real-valued normed space and $T, U : X \to X$ be two nonexpansive mappings on X, i.e.

$$\|Tx - Ty\| \leqslant \|x - y\| \text{ and } \|Ux - Uy\| \leqslant \|x - y\|$$

for all $x, y \in X$. In order to approximate common fixed points of two such mappings under stochastic noise constraints, we introduce in this paper the so-called stochastic Halpern-Mann iteration, given by the schema

$$\text{(sHM)} \qquad \begin{cases} y_n := (1-\alpha_n)(Tx_n + \xi_n) + \alpha_n u, \\ x_{n+1} := (1-\beta_n)(Uy_n + \delta_n) + \beta_n y_n, \end{cases}$$

where, over some fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, x_0 and u are arbitrary X-valued random variables chosen as a fixed starting point and as an anchor of the iteration, respectively, $(\xi_n), (\delta_n)$

Date: July 7, 2025.

Funding: The first author was supported by the 'Deutsche Forschungsgemeinschaft' Project DFG KO 1737/6-2. The second author was supported by the EPSRC grant EP/W035847/1.
Conflicts of interests: The authors have no relevant financial or non-financial interests to disclose.
Data availability: Data sharing not applicable to this article as no data sets were generated or analyzed.

are sequences of X-valued random variables representing the stochastic noise, and $(\alpha_n), (\beta_n) \subseteq [0,1]$ are suitable *nonstochastic* parameter sequences.

Our main scheme (sHM) represents a stochastic analogue of the deterministic Halpern-Mann scheme for two operators, schematically appearing already in [37], and recently rediscovered in a much more generalized setting in [14]. This scheme integrates two of the most prominent methods for approximating fixed points for nonexpansive mappings, the Krasnoselskii-Mann method [24, 32] and Halpern's method [16] (see also [42] and [43]), with the intended gain of combining the beneficial features of both, in particular the strong convergence of Halpern's method even in infinite dimensional spaces. We conjecture that this combination gains further significance in the stochastic setting, where we anticipate that it can be used in particular to devise novel reinforcement learning algorithms or stochastic splitting methods.

In contrast to the deterministic scheme, (sHM) is designed to capture situations in which one does not have direct access to Tx_n (respectively Uy_n), but can only use noisy versions $\tilde{T}$ of T (and $\tilde{U}$ of U). Intuitively, ξ_n would then represent the difference between Tx_n and the corresponding approximation of Tx_n obtained from $\tilde{T}$ through a suitable sampling method (and similarly for δ_n and Uy_n), though our presentation is fully abstract and we will make no assumptions about ξ_n or δ_n, other than imposing controls on the means $\mathbb{E}[\|\xi_n\|]$ and $\mathbb{E}[\|\delta_n\|]$, which however are carefully chosen so that they can be achieved in concrete scenarios through sampling methods such as minibatching under suitable assumptions. In particular, we make no independence assumptions on the noise terms ξ_n and δ_n.

The main results of the paper establish general conditions under which we guarantee the asymptotic regularity of the scheme (sHM), both in the traditional sense of $\|x_n - x_{n+1}\|$ (sometimes called the discrete velocity [3]) and also relative to the mappings, i.e. considering the displacements $\|x_n - Tx_n\|$ and $\|x_n - Ux_n\|$. Furthermore, we establish these asymptotic regularity results both in expectation and almost surely, i.e. we show both

$$\mathbb{E}[\|x_n - x_{n+1}\|],\ \mathbb{E}[\|x_n - Tx_n\|],\ \mathbb{E}[\|x_n - Ux_n\|] \to 0$$

as well as

$$\|x_n - x_{n+1}\|\,,\ \|x_n - Tx_n\|\,,\ \|x_n - Ux_n\| \to 0 \text{ almost surely}$$

under suitable conditions on the auxiliary parameters as well as the noise. In the general case this requires a subtle geometric argument and involves a delicate interplay between the two modes of stochastic convergence, which manifest, respectively, in the additional assumptions that X is uniformly convex and that the sequence $(\|Uy_n - y_n\|)$ is uniformly integrable. Most importantly, in all cases we provide explicit convergence rates for these expressions, under very general conditions and dependent only on a few moduli witnessing quantitative aspects of our main assumptions. In particular, we identify natural circumstances under which these rates are very fast, reaching up to linear speed in special cases.

To date, only the very simple instance of (sHM) corresponding to a stochastic variant of Halpern's method [16] has been previously studied, where setting $U := \mathrm{Id}$ and $\delta_n := 0$ gives rise to the iteration

$$x_{n+1} = (1-\alpha_n)(Tx_n + \xi_n) + \alpha_n u, \tag{sH}$$

considered in [5] (for finite dimensional normed spaces), a scheme which in the Euclidean setting has recently received a great deal of attention in the context of stochastic monotone inclusion problems [6, 12, 31, 45]. In all cases, controlling the variance of the noise terms is crucial for convergence, and this is just one of several elements that makes the analysis of stochastic schemes such as (sH) markedly different from that of their nonstochastic counterparts, some of the others being a focus on *oracle* complexity, and the relevance of stochastic methods to machine learning.

Given this increasing interest in stochastic variants of classic methods, the purpose of the present paper is to broaden their current scope and provide a collection of generalised convergence results in which all of the aforementioned features (variance reduction, oracle complexity, applications in machine learning) are presented in the abstract.

While we focus on the special case of the stochastic Halpern iteration (sH) at several points in the paper, where it forms a useful example and where in particular our results on asymptotic regularity in that case generalize and improve those presented in [5], our method (sHM) is certainly not limited to this special case, and in line with [14] encompasses stochastic variants of other well-known deterministic methods.

An important example of this is represented by setting $T := \mathrm{Id}$ and $\xi_n := 0$, as well as $u := 0$ and $\gamma_n := 1 - \alpha_n$, whereby we obtain a version of the Krasnoselskii-Mann iteration with Tikhonov regularization terms considered in [2, 44] that now incorporates stochastic noise

$$x_{n+1} = (1 - \beta_n)(U(\gamma_n x_n) + \delta_n) + \beta_n(\gamma_n x_n) \tag{sKM-T}$$

or, in other words, the stochastic Krasnoselskii-Mann iteration as considered in [4] with Tikhonov regularization terms as considered in [2] (or, further, equivalently as a stochastic version of the modified Mann iteration [18]). This method is known to produce fast asymptotic behavior in the deterministic setting, and also benefits from strong convergence results similarly to Halpern's iteration [2]. To our knowledge, the stochastic variant (sKM-T) is introduced here for the first time, and our convergence results for this special case of (sHM), which for a particular choice of parameters and suitable conditions on the noise terms result in linear rates of asymptotic regularity, show that fast asymptotic behaviour is also inherited by the stochastic variant.

While the special cases of (sH) and (sKM-T) exhibit linear rates of asymptotic regularity under very mild (quantitative) assumptions on the space, parameters and noise, the general iteration (sHM) proves to be more complex as mentioned before, depending to a large degree on the geometry of the space: In the general case, we assume X to be uniformly convex, and the rate then depends on a corresponding modulus witnessing this uniform convexity. Our main geometric argument is carried out in a pointwise manner, yielding almost sure convergence (with rates). Nevertheless, the passage back to convergence in mean requires a uniform integrability assumption (which as we show can actually be imposed on the noise terms), and our rates then also depend on a corresponding modulus of uniform integrability, which we define and use for the first time in this paper. However, fast rates also apply to the general scheme (sHM), where in the special case of a uniformly convex space of power type $p \geqslant 2$, the complexity of those rates is of order p, and so in particular we get quadratic rates of asymptotic regularity for inner product spaces.

While all of our rates are novel, we emphasise that the "qualitative" asymptotic regularity results, namely convergence alone even without considering the quantitative aspects, are also to the best of our knowledge new, for any stochastic iteration contained in (sHM) that goes beyond the Halpern iteration.

All of the results obtained in this paper are motivated via the methodology of the proof mining program, a subfield of mathematical logic which combines an abstract approach to proofs in mainstream mathematics with the extraction of computational information, such as bounds or rates, from those proofs. We refer to the seminal monograph [20] for a comprehensive overview of both theoretical as well as applied aspects of this program, along with the survey [22] for an overview of more recent applications to nonlinear analysis. Proof mining has been widely applied in nonlinear analysis, and has found particular success in providing quantitative convergence results for Halpern's iteration and its many variants, with notable instances ranging from initial rates of asymptotic regularity for Halpern's iteration given by Leuştean [27] and the first analysis of Wittmann's proof of the strong convergence of Halpern's iteration given in the seminal paper of Kohlenbach [21], to the extensions of these results to nonlinear context such as CAT(0)-spaces as in [23] (by a logical analysis of a corresponding convergence proof by Saejung [39]). They also include extensions of the Halpern iteration [40] for the modified Mann iteration introduced in [18] (and extended to nonlinear spaces in [11]) as well as the Krasnoselskii-Mann iteration with Tikhonov regularization terms and its extensions as in [7, 8, 9, 13] (with [7] of particular note, as linear rates of asymptotic regularity are there obtained for the first time in the context of applications of proof mining). In particular, the definition of the deterministic Halpern-Mann method given in [14] and its corresponding convergence proof were motivated by these logical considerations, as were the recent rates of asymptotic regularity given for this iteration in [30].

The present work departs from the aforementioned case studies in nonlinear analysis in that it incorporates, for the first time, stochasticity. In this way it forms part of a recent advance of proof mining into probability theory, which comprises both new developments in the logical foundations of probability theory due to first author and Neri [33], together with applied results on the quantitative aspects of stochastic processes by the authors and Neri [34, 35, 36]. In particular, the present paper is one of the first applications of proof mining to stochastic optimization: It represents a particularly interesting case study in this respect, in that it does not readily follow from analogous quantitative results in the deterministic setting (such as in [14, 30]), but requires a substantial arsenal of new quantitative ideas for this stochastic setting, as will be introduced and discussed in detail throughout the paper. We stress that while this logical perspective was crucial in obtaining the present results, the paper does not rely on any notions from logic at all.

However, the main motivation of this paper is not so much its novelty within the proof mining program, but the real applicability of the main results in concrete areas. For this reason, our last section is dedicated to a brief outlook on further results arising from the present work, in particular concerning oracle complexity and reinforcement learning, and in the context of the latter how present schemes can yield novel methods for Q-learning, all of which will be discussed in further detail a forthcoming paper. While not discussed there, we also want to mention that another promising application is represented by stochastic versions of the splitting methods with Tikhonov regularization terms as studied in [2], but we leave this for future work, as we do the study of the convergence of the method (sHM), where we expect that strong convergence results hold for this method also in infinite dimensional spaces, analogously to the deterministic case.

2. Preliminaries and basic lemmas

We write $\mathbb{N}^*$ for $\mathbb{N}$ without 0. Throughout, if not stated otherwise, we fix an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and all probabilistic notions such as almost sureness refer to that space. Similarly, X will always denote, unless stated otherwise, a normed space with norm $\|\cdot\|$. We refer to measurable functions $\Omega \to \mathbb{R}$ as random variables, to measurable functions $\Omega \to X$ as X-valued random variables and we refer to sequences of random variables as stochastic processes. In order to ensure that basic properties enjoyed by real-valued random variables are also inherited by X-valued random variables, so that in particular our main scheme (sHM) is well-defined, one normally requires some further assumptions on the underlying space (as discussed in detail in [25]). The simplest option is to assume that X is separable, though if the reader prefers they can also just assume that X is finite dimensional. Equalities and inequalities involving random variables will always be understood to hold almost surely, even if not explicitly indicated.

Throughout the paper, we will be concerned with quantitative variants of various notions and we here now briefly the discuss the key definitions of the main quantitative notions used in the paper:

Given a non-negative sequence of reals (a_n), a rate of convergence for $a_n \to 0$ is a function $\varphi : (0,\infty) \to \mathbb{N}$ such that

$$\forall \varepsilon > 0\, \forall n \geqslant \varphi(\varepsilon)\, (a_n < \varepsilon)\,.$$

The immediate benefit of such a type of rate φ is that if it is invertible and decreasing, then we can even derive the non-asymptotic estimate $a_n < \varphi^{-1}(n)$ for all $n \in \mathbb{N}$, which of course further implies a complexity bound on the sequence in terms of the commonly used big O notation, namely $(a_n) = O(\varphi^{-1}(n))$.

Now, given a nonnegative stochastic process (X_n), a rate of convergence for $X_n \to 0$ almost surely is a function $\Phi : (0,\infty)^2 \to \mathbb{N}$ such that

$$\forall \lambda, \varepsilon > 0\, (\mathbb{P}\,(\exists n \geqslant \Phi(\lambda,\varepsilon)\,(X_n \geqslant \varepsilon)) < \lambda)\,.$$

We note that whenever Φ is a rate of convergence for $X_n \to 0$ almost surely, then for every $\varepsilon > 0$, $\Phi(\varepsilon,\cdot)$ is a rate of convergence for $\mathbb{P}\,(\sup_{n\geqslant N}(X_n \geqslant \varepsilon)) \to 0$ in the nonstochastic sense.

Further, given a non-negative sequence of reals (a_n), we later want to quantitatively witness the convergence or divergence of the series over that sequence. For that, if $\sum_{n=0}^{\infty} a_n < \infty$, we say that a function $\chi : (0,\infty) \to \mathbb{N}$ is a rate of convergence for that sum if

$$\forall \varepsilon > 0 \left(\sum_{n=\chi(\varepsilon)}^{\infty} a_n < \varepsilon \right).$$

If $\sum_{n=0}^{\infty} a_n = \infty$, we say that a function $\theta : \mathbb{N} \times (0,\infty) \to \mathbb{N}$ is a rate of divergence for that sum if

$$\forall b > 0\, \forall k \in \mathbb{N} \left(\sum_{n=k}^{\theta(k,b)} a_n \geqslant b \right).$$

Naturally, any such modulus satisfies $\theta(k,b) \geqslant k$ for any $k \in \mathbb{N}$ and $b > 0$.

We now collect some of the basic abstract convergence results that our paper relies on. The most crucial of these, on the asymptotic behavior of sequences of reals satisfying certain recursive inequalities is the following due to Xu [43], often called Xu's lemma:

Lemma 2.1 ([43]). *Suppose that $(s_n), (c_n) \subseteq [0,\infty)$ as well as $(a_n) \subseteq [0,1]$ and $(b_n) \subseteq \mathbb{R}$ satisfy*

$$s_{n+1} \leqslant (1-a_n)s_n + a_n b_n + c_n$$

for all $n \in \mathbb{N}$ where $\sum_{n=0}^{\infty} a_n = \infty$, $\limsup b_n \leqslant 0$ and $\sum_{n=0}^{\infty} c_n < \infty$. Then $s_n \to 0$.

We will in particular rely on a quantitative rendering of an instance of Xu's lemma which is represented by the following lemma. This result is contained in [23, 29] (up to the way the errors and the moduli are phrased) therefore for brevity we omit the proof.

Lemma 2.2 (essentially [23, 29]). *Suppose that $(s_n), (c_n) \subseteq [0,\infty)$ and $(a_n) \subseteq [0,1]$ satisfy*

$$s_{n+1} \leqslant (1-a_n)s_n + c_n$$

for all $n \in \mathbb{N}$, and furthermore, that $K > 0$ is an upper bound on (s_n), θ is a rate of divergence for $\sum_{n=0}^{\infty} a_n = \infty$ and χ a rate of convergence for $\sum_{n=0}^{\infty} c_n < \infty$. Then $s_n \to 0$ with rate

$$\varphi_{K,\theta,\chi}(\varepsilon) := \theta\left(\chi\left(\frac{\varepsilon}{2}\right), \ln\left(\frac{2K}{\varepsilon}\right)\right) + 1.$$

We now extend this lemma to a probabilistic variant. For that, we first consider the following result which allows us to transfer quantitative information from convergence in mean for "almost-monotone" sequences of random variables to rates of almost sure convergence.

Lemma 2.3. *Let (X_n), (C_n) be nonnegative stochastic processes satisfying*

$$X_{n+1} \leqslant X_n + C_n$$

almost surely for all $n \in \mathbb{N}$ and suppose furthermore that

(a) $\sum_{i=0}^{\infty} \mathbb{E}[C_i] < \infty$ with rate χ,
(b) $\mathbb{E}[X_n] \to 0$ with rate φ.

Then $X_n \to 0$ almost surely, and with rate

$$\psi(\lambda, \varepsilon) := \max\{\varphi(\lambda\varepsilon/2), \chi(\lambda\varepsilon/2)\}.$$

Proof. We first note that for any $n \in \mathbb{N}$, we have $\mathbb{E}\left[\sum_{i=n}^{\infty} C_i\right] = \sum_{i=n}^{\infty} \mathbb{E}[C_i]$ by the monotone convergence theorem. Now define a stochastic process (U_n) by $U_n := X_n + \sum_{i=n}^{\infty} C_i$. Then we have

$$U_{n+1} = X_{n+1} + \sum_{i=n+1}^{\infty} C_i \leqslant X_n + C_n + \sum_{i=n+1}^{\infty} C_i \leqslant X_n + \sum_{i=n}^{\infty} C_i = U_n$$

almost surely for any $n \in \mathbb{N}$, and therefore the events $(U_n \geqslant \varepsilon)$ are monotone decreasing in n. In particular, using Markov's inequality, we get for any $N \in \mathbb{N}$:

$$\mathbb{P}(\exists n \geqslant N(U_n \geqslant \varepsilon)) = \mathbb{P}(U_N \geqslant \varepsilon) \leqslant \frac{\mathbb{E}[U_N]}{\varepsilon} = \frac{\mathbb{E}[X_N] + \sum_{i=N}^{\infty} \mathbb{E}[C_i]}{\varepsilon}.$$

Therefore if $N = \psi(\lambda, \varepsilon)$, we have

$$\mathbb{P}(\exists n \geqslant N(U_n \geqslant \varepsilon)) \leqslant \frac{\mathbb{E}[X_N] + \sum_{i=N}^{\infty} \mathbb{E}[C_i]}{\varepsilon} < \frac{(\lambda\varepsilon/2 + \lambda\varepsilon/2)}{\varepsilon} = \lambda.$$

The result follows by observing that $X_n \leqslant U_n$ holds almost surely for all $n \in \mathbb{N}$ and thus

$$\mathbb{P}(\exists n \geqslant N(X_n \geqslant \varepsilon)) \leqslant \mathbb{P}(\exists n \geqslant N(U_n \geqslant \varepsilon)) < \lambda$$

for any N, and in particular for the N defined above. □

The above lemma now allows us to give a stochastic version of Lemma 2.2:

Lemma 2.4. *Suppose that (X_n), (C_n) are nonnegative stochastic processes satisfying*

$$X_{n+1} \leqslant (1 - a_n)X_n + C_n$$

almost surely for all $n \in \mathbb{N}$. Furthermore, suppose that

(a) $\mathbb{E}[X_n] \leqslant K$ for all $n \in \mathbb{N}$,
(b) $\sum_{i=0}^{\infty} a_i = \infty$ with rate of divergence θ,
(c) $\sum_{i=0}^{\infty} \mathbb{E}[C_i] < \infty$ with rate of convergence χ.

Then $\mathbb{E}[X_n] \to 0$ with rate $\varphi_{K,\theta,\chi}$ defined as in Lemma 2.2, i.e.

$$\varphi_{K,\theta,\chi}(\varepsilon) := \theta\left(\chi\left(\frac{\varepsilon}{2}\right), \ln\left(\frac{2K}{\varepsilon}\right)\right) + 1,$$

and further $X_n \to 0$ almost surely with rate

$$\psi_{K,\theta,\chi}(\lambda, \varepsilon) := \varphi_{K,\theta,\chi}\left(\frac{\lambda\varepsilon}{2}\right).$$

Proof. Taking expectations on both sides we have

$$\mathbb{E}[X_{n+1}] \leqslant (1-a_n)\mathbb{E}[X_n] + \mathbb{E}[C_n]$$

for all $n \in \mathbb{N}$ and so the rate for $\mathbb{E}[X_n] \to 0$ follows by Lemma 2.2. For the rate for the almost sure convergence, observe that $\chi(\lambda\varepsilon/4) \leqslant \varphi_{K,\theta,\chi}(\lambda\varepsilon/2)$ as $\theta(k,b) \geqslant k$. Hence, one can proceed as in the proof of Lemma 2.3 to show that

$$\mathbb{P}\left(\exists n \geqslant N(U_n \geqslant \varepsilon)\right) \leqslant \frac{\mathbb{E}[X_N] + \sum_{i=N}^{\infty} \mathbb{E}\left[C_i\right]}{\varepsilon}$$

for any N where, using $\chi(\lambda\varepsilon/4) \leqslant \varphi_{K,\theta,\chi}(\lambda\varepsilon/2)$, we then can conclude $\mathbb{P}\left(\exists n \geqslant N(U_n \geqslant \varepsilon)\right) < \lambda$ for $N = \varphi_{K,\theta,\chi}(\lambda\varepsilon/2)$. □

3. Quantitative asymptotic regularity for the generalized stochastic Halpern scheme

In this section we now outline our main theoretical results and derive rates of asymptotic regularity for the iterations generated by the generalized stochastic Halpern scheme.

3.1. **Basic results and rates of asymptotic regularity.** We begin with some fundamental recursive inequalities for the iterations generated by the iteration (sHM):

Lemma 3.1 (essentially [30]). *Let $(x_n), (y_n)$ be the sequences generated by* (sHM). *Then the following recurrence relations hold pointwise everywhere for all $n \in \mathbb{N}$:*

(1) $\|y_{n+1} - y_n\| \leqslant (1-\alpha_{n+1})\left(\|x_{n+1} - x_n\| + \|\xi_{n+1} - \xi_n\|\right) + |\alpha_{n+1} - \alpha_n| \cdot \|Tx_n + \xi_n - u\|,$

(2) $\|x_{n+2} - x_{n+1}\| \leqslant \|y_{n+1} - y_n\| + (1-\beta_{n+1})\|\delta_{n+1} - \delta_n\| + |\beta_{n+1} - \beta_n| \cdot \|Uy_n + \delta_n - y_n\|.$

Proof. For (1) we observe that

$$\begin{aligned}
\|y_{n+1} - y_n\| &= \|(1-\alpha_{n+1})(Tx_{n+1} + \xi_{n+1}) - (1-\alpha_n)(Tx_n + \xi_n) + (\alpha_{n+1} - \alpha_n)u\| \\
&\leqslant (1-\alpha_{n+1})\|(Tx_{n+1} + \xi_{n+1}) - (Tx_n + \xi_n)\| \\
&\quad + \|(\alpha_n - \alpha_{n+1})(Tx_n + \xi_n) - (\alpha_n - \alpha_{n+1})u\| \\
&\leqslant (1-\alpha_{n+1})\left(\|x_{n+1} - x_n\| + \|\xi_{n+1} - \xi_n\|\right) + |\alpha_{n+1} - \alpha_n| \cdot \|Tx_n + \xi_n - u\|
\end{aligned}$$

where for the last inequality we use that T is nonexpansive. Similarly for (2) we have

$$\begin{aligned}
\|x_{n+2} - x_{n+1}\| &= \|(1-\beta_{n+1})(Uy_{n+1} + \delta_{n+1}) + \beta_{n+1}y_{n+1} - (1-\beta_n)(Uy_n + \delta_n) - \beta_n y_n\| \\
&\leqslant \|(1-\beta_{n+1})(Uy_{n+1} + \delta_{n+1}) - (1-\beta_{n+1})(Uy_n + \delta_n) + \beta_{n+1}(y_{n+1} - y_n)\| \\
&\quad + \|(1-\beta_{n+1})(Uy_n + \delta_n) - (1-\beta_n)(Uy_n + \delta_n) - (\beta_n - \beta_{n+1})y_n\| \\
&\leqslant (1-\beta_{n+1})\left(\|Uy_{n+1} - Uy_n\| + \|\delta_{n+1} - \delta_n\|\right) + \beta_{n+1}\|y_{n+1} - y_n\| \\
&\quad + |\beta_{n+1} - \beta_n| \cdot \|Uy_n + \delta_n - y_n\| \\
&\leqslant \|y_{n+1} - y_n\| + (1-\beta_{n+1})\|\delta_{n+1} - \delta_n\| + |\beta_{n+1} - \beta_n| \cdot \|Uy_n + \delta_n - y_n\|
\end{aligned}$$

where again we use nonexpansivity of the operator in the last step. □

We now move to our first quantitative result which presents a rate of asymptotic regularity for the sequence (x_n), both in expectation and in probability. For that we introduce a first central assumption on the boundedness of the iteration (sHM) in expectation, as commonly made in the literature (see e.g. hypothesis (H_1) in [5] of which this assumption here is a natural extension to the generalised iteration (sHM)):

(Hyp) There exists a $K_0 \in \mathbb{N}^*$ such that for all $n \in \mathbb{N}$:

$$\mathbb{E}[\|Tx_n - u\|],\ \mathbb{E}[\|Uy_n - y_n\|],\ \mathbb{E}[\|Uu - u\|],\ \mathbb{E}[\|Uy_n - u\|] \leqslant K_0 < \infty.$$

Throughout, if not stated otherwise, we will assume the existence of such a K_0.

In the context of the asymptotic regularity results that hold almost surely, we will sometimes need to make a slightly stronger assumption that the random variables are actually L^1-bounded in the following sense:

(Hyp′) There exists a nonnegative random variable Y with $K_0 \geqslant \mathbb{E}[Y]$ for some $K_0 \in \mathbb{N}^*$ and for all $n \in \mathbb{N}$: $\|Tx_n - u\|, \|Uy_n - y_n\|, \|Uu - u\|, \|Uy_n - u\| \leqslant Y$ almost surely.

Contrary to the above (Hyp), which will essentially always be tacitly assumed, we will always be very explicit about when we actually need to assume the above hypothesis (Hyp′). It is to be noted that both hypotheses are guaranteed in the presence of a common fixed point of T and U, as will be later discussed in more detail (see Lemma 3.9).

In any case, under the assumption (Hyp), we can immediately derive a bound on the expectation of the discrete velocity and utilize that to derive our first rate of asymptotic regularity:

Theorem 3.2. *Let $(x_n), (y_n)$ be the sequences generated by* (sHM). *Suppose that $\sum_{n=0}^{\infty} \alpha_n = \infty$ with rate of divergence θ, that*

$$\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_{n+1} - \xi_n\|], \sum_{n=0}^{\infty} \mathbb{E}[\|\delta_{n+1} - \delta_n\|], \sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n|, \sum_{n=0}^{\infty} |\beta_{n+1} - \beta_n| < \infty$$

with moduli of convergence $\chi_1 - \chi_4$ and upper bounds $B_1 - B_4$, respectively, and that $\mathbb{E}[\|\xi_n\|] \leqslant E_0$ and $\mathbb{E}[\|\delta_n\|] \leqslant D_0$ for all $n \in \mathbb{N}$. Then $\mathbb{E}[\|x_{n+1} - x_n\|] \to 0$ with rate $\varphi_{K,\theta,\chi}$ as well as $\|x_{n+1} - x_n\| \to 0$ almost surely with rate $\psi_{K,\theta,\chi}$ with φ, ψ defined as in Lemma 2.4, i.e.

$$\varphi_{K,\theta,\chi}(\varepsilon) := \theta\left(\chi\left(\frac{\varepsilon}{2}\right), \ln\left(\frac{2K}{\varepsilon}\right)\right) + 1 \text{ and } \psi_{K,\theta,\chi}(\lambda, \varepsilon) := \varphi_{K,\theta,\chi}\left(\frac{\lambda\varepsilon}{2}\right),$$

where

$$\chi(\varepsilon) := \max\{\chi_1(\varepsilon/4), \chi_2(\varepsilon/4), \chi_3(\varepsilon/4(E_0 + K_0), \chi_4(\varepsilon/4(D_0 + K_0))\}$$

as well as $K := 2K_0 + E_0 + D_0 + B$ for $B := B_1 + B_2 + B_3(E_0 + K_0) + B_4(D_0 + K_0)$.

Proof. Using (1) and (2) of Lemma 3.1, we have that

$$\|x_{n+2} - x_{n+1}\| \leqslant (1 - \alpha_{n+1}) \|x_{n+1} - x_n\| + c_n$$

for all $n \in \mathbb{N}$ everywhere on Ω where

$$\begin{aligned} c_n := &\|\xi_{n+1} - \xi_n\| + |\alpha_{n+1} - \alpha_n|(\|Tx_n - u\| + \|\xi_n\|) \\ &+ \|\delta_{n+1} - \delta_n\| + |\beta_{n+1} - \beta_n|(\|Uy_n - y_n\| + \|\delta_n\|). \end{aligned}$$

It is immediate that

$$\mathbb{E}[c_n] \leqslant \mathbb{E}[\|\xi_{n+1} - \xi_n\|] + \mathbb{E}[\|\delta_{n+1} - \delta_n\|] + |\alpha_{n+1} - \alpha_n|(K_0 + E_0) + |\beta_{n+1} - \beta_n|(K_0 + D_0)$$

and so

$$\chi(\varepsilon) := \max\{\chi_1(\varepsilon/4), \chi_2(\varepsilon/4), \chi_3(\varepsilon/4(E_0 + K_0)), \chi_4(\varepsilon/4(D_0 + K_0))\}$$

is a rate of convergence for $\sum_{n=0}^{\infty} \mathbb{E}[c_n] < \infty$, while B as defined above is an upper bound for $\sum_{n=0}^{\infty} \mathbb{E}[c_n]$. Naturally, the above yields

$$\mathbb{E}[\|x_{n+1} - x_n\|] \leqslant \mathbb{E}[\|x_1 - x_0\|] + \sum_{i=0}^{n-1} \mathbb{E}[c_i] \leqslant \mathbb{E}[\|x_1 - x_0\|] + B$$

and we can then show that

$$\mathbb{E}[\|x_1 - x_0\|] \leqslant \mathbb{E}[\|y_0 - u\|] + \mathbb{E}[\|Uu - u\|] + \mathbb{E}[\|\xi_0\|] + \mathbb{E}[\|\delta_0\|] \leqslant 2K_0 + E_0 + D_0$$

so that $\mathbb{E}[\|x_{n+1} - x_n\|] \leqslant K := 2K_0 + E_0 + D_0 + B$. Using Lemma 2.4, we then get the desired rates. □

We can then immediately transfer that rate to the complementary sequence (y_n):

Theorem 3.3. *Under the assumptions of Theorem 3.2, we have that* $\mathbb{E}[\|y_{n+1} - y_n\|] \to 0$ *with rate*

$$\varphi'(\varepsilon) := \max\{\varphi(\varepsilon/3), \chi_1(\varepsilon/3), \chi_3(\varepsilon/3(K_0 + E_0))\}.$$

as well as $\|y_{n+1} - y_n\| \to 0$ *almost surely with rate*

$$\psi'(\lambda, \varepsilon) := \max\{\varphi'(\lambda\varepsilon/2), \chi(\lambda\varepsilon/2)\}$$

where φ *is a rate for* $\mathbb{E}[\|x_{n+1} - x_n\|] \to 0$ *and* χ *is as in Theorem 3.2.*

Proof. Using (1) of Lemma 3.1, we get that

$$\|y_{n+1} - y_n\| \leqslant \|x_{n+1} - x_n\| + \|\xi_{n+1} - \xi_n\| + |\alpha_{n+1} - \alpha_n|(\|Tx_n - u\| + \|\xi_n\|)$$

everywhere on Ω and for any $n \in \mathbb{N}$. Under expectation, we thus have

$$\mathbb{E}[\|y_{n+1} - y_n\|] \leqslant \mathbb{E}[\|x_{n+1} - x_n\|] + \mathbb{E}[\|\xi_{n+1} - \xi_n\|] + |\alpha_{n+1} - \alpha_n|(K_0 + E_0)$$

and from that the rate for $\mathbb{E}[\|y_{n+1} - y_n\|] \to 0$ immediately follows, noting that a rate of convergence χ for a series $\sum_{n=0}^{\infty} a_n < \infty$ yields that $\sum_{n=\chi(\varepsilon)}^{\infty} a_n < \varepsilon$ and so implies that $a_n < \varepsilon$ for any $n \geqslant \chi(\varepsilon)$. For the rate of $\|y_{n+1} - y_n\| \to 0$ almost surely, note that using both (1) and (2) of Lemma 3.1, we get that

$$\|y_{n+2} - y_{n+1}\| \leqslant \|y_{n+1} - y_n\| + d_n$$

for all $n \in \mathbb{N}$ everywhere on Ω where

$$\begin{aligned} d_n := &\|\xi_{n+2} - \xi_{n+1}\| + |\alpha_{n+2} - \alpha_{n+1}|(\|Tx_{n+1} - u\| + \|\xi_{n+1}\|) \\ &+ \|\delta_{n+1} - \delta_n\| + |\beta_{n+1} - \beta_n|(\|Uy_n - y_n\| + \|\delta_n\|). \end{aligned}$$

So, it is immediate that we have

$$\mathbb{E}[d_n] \leqslant \mathbb{E}[\|\xi_{n+2} - \xi_{n+1}\|] + \mathbb{E}[\|\delta_{n+1} - \delta_n\|] + |\alpha_{n+2} - \alpha_{n+1}|(K_0 + E_0) + |\beta_{n+1} - \beta_n|(K_0 + D_0).$$

and so that χ from Theorem 3.2 is a rate of convergence for $\sum_{n=0}^{\infty} \mathbb{E}[d_n] < \infty$ (noting that if χ is a rate of convergence for $\sum_{n=0}^{\infty} a_n < \infty$, then $\sum_{n=\chi(\varepsilon)} a_{n+1} = \sum_{n=\chi(\varepsilon)+1} a_n \leqslant \sum_{n=\chi(\varepsilon)} a_n < \varepsilon$ so that χ is also a rate of convergence for $\sum_{n=0}^{\infty} a_{n+1} < \infty$). Using Lemma 2.3, we get the desired rate for $\|y_{n+1} - y_n\| \to 0$ almost surely. □

3.2. **Asymptotic regularity relative to the mappings.** We now move on to establishing rates of asymptotic regularity for the iterations relative to the mappings. For that, we will actually see a crucial dichotomy, where results based on the use of just one of the mappings U or T are comparatively straightforward, whereas for the general case where neither U nor T trivialize, we rely on a geometric argument for establishing a rate of asymptotic regularity relative to U for the sequence (y_n), which requires both a uniform convexity of the space X along with a uniform integrability assumption on $(\|Uy_n - y_n\|)$. As such, before we move on to these results, we first give rates of asymptotic regularity relative to the mappings for the remaining cases, dependent on the relevant rates for $(\|Uy_n - y_n\|)$:

Theorem 3.4. *Assume that* $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ *with rate* φ*, and* $\mathbb{E}[\|\xi_n\|], \mathbb{E}[\|\delta_n\|], \alpha_n \to 0$ *with rates* ρ_1 – ρ_3*, respectively. Assume further that* $\mathbb{E}[\|x_{n+1} - x_n\|] \to 0$ *with a rate* φ_0*. Then*

(a) $\mathbb{E}[\|x_n - y_n\|] \to 0$ *with rate*

$$\varphi_1(\varepsilon) := \max\{\varphi_0(\varepsilon/3), \varphi(\varepsilon/3), \rho_2(\varepsilon/3)\},$$

(b) $\mathbb{E}[\|Ty_n - y_n\|] \to 0$ *with rate*

$$\varphi_2(\varepsilon) := \max\left\{\varphi_1(\varepsilon/3), \rho_3(\varepsilon/3K_0), \rho_1(\varepsilon/3)\right\},$$

(c) $\mathbb{E}[\|Ux_n - x_n\|] \to 0$ *with rate*

$$\varphi_3(\varepsilon) := \max\left\{\varphi_1(\varepsilon/3), \varphi(\varepsilon/3)\right\},$$

(d) $\mathbb{E}[\|Tx_n - x_n\|] \to 0$ *with rate*

$$\varphi_4(\varepsilon) := \max\left\{\varphi_1(\varepsilon/3), \varphi_2(\varepsilon/3)\right\}.$$

Proof. Related to (a) – (d), we can immediately establish the following inequalities:

$$\begin{aligned}
\|x_n - y_n\| &\leqslant \|x_{n+1} - x_n\| + \|x_{n+1} - y_n\| \\
&\leqslant \|x_{n+1} - x_n\| + \|Uy_n - y_n\| + \|\delta_n\|, \\
\|Ty_n - y_n\| &\leqslant \|Ty_n - Tx_n\| + \|Tx_n - y_n\| \\
&\leqslant \|y_n - x_n\| + \alpha_n \|Tx_n - u\| + \|\xi_n\|, \\
\|Ux_n - x_n\| &\leqslant \|Ux_n - Uy_n\| + \|Uy_n - y_n\| + \|y_n - x_n\| \\
&\leqslant 2\|x_n - y_n\| + \|Uy_n - y_n\|, \\
\|Tx_n - x_n\| &\leqslant \|Tx_n - Ty_n\| + \|Ty_n - y_n\| + \|y_n - x_n\| \\
&\leqslant 2\|x_n - y_n\| + \|Ty_n - y_n\|.
\end{aligned}$$

By taking the expectation, the rates immediately follow. $\square$

We then can similarly give rates of asymptotic regularity almost surely under a slight extension of the previous conditions on the errors:

Theorem 3.5. *Under the assumption* (Hyp′)*, assume that* $\|Uy_n - y_n\| \to 0$ *almost surely with rate* ψ*, and* $\alpha_n \to 0$ *with rate* ρ*. Further, assume that* $\|\xi_n\|, \|\delta_n\| \to 0$ *almost surely with rates* ϕ_1, ϕ_2*, respectively, and that* $\|x_{n+1} - x_n\| \to 0$ *almost surely with a rate* ψ_0*. Then*

(a) $\|x_n - y_n\| \to 0$ *almost surely with rate*

$$\psi_1(\lambda, \varepsilon) := \max\left\{\psi_0(\lambda/3, \varepsilon/3), \psi(\lambda/3, \varepsilon/3), \phi_2(\lambda/3, \varepsilon/3)\right\},$$

(b) $\|Ty_n - y_n\| \to 0$ *almost surely with rate*

$$\psi_2(\lambda, \varepsilon) := \max\left\{\psi_1(\lambda/3, \varepsilon/3), \rho(\varepsilon\lambda/9K_0), \phi_1(\lambda/3, \varepsilon/3)\right\},$$

(c) $\|Ux_n - x_n\| \to 0$ *almost surely with rate*

$$\psi_3(\lambda, \varepsilon) := \max\left\{\psi_1(\lambda/2, \varepsilon/3), \psi(\lambda/2, \varepsilon/3)\right\},$$

(d) $\|Tx_n - x_n\| \to 0$ *almost surely with rate*

$$\psi_4(\lambda, \varepsilon) := \max\left\{\psi_1(\lambda/2, \varepsilon/3), \psi_2(\lambda/2, \varepsilon/3)\right\}.$$

Proof. The results follow immediately by the same inequalities established in the proof of Theorem 3.4 where in the case (b) one just needs the following additional consideration, giving a rate for $\alpha_n \|Tx_n - u\| \to 0$ almost surely: Using Markov's inequality we have

$$\mathbb{P}(\exists n\, (\|Tx_n - u\| \geqslant K_0/\lambda)) \leqslant \mathbb{P}(Y \geqslant K_0/\lambda) \leqslant \lambda$$

for Y as in (Hyp′). Now noting that if ω is such that $\|Tx_n(\omega) - u(\omega)\| < K_0/\lambda$ for all $n \in \mathbb{N}$, then $\alpha_n \|Tx_n(\omega) - u(\omega)\| < \varepsilon$ for any $n \geqslant \rho(\varepsilon\lambda/K_0)$, and therefore we have

$$\mathbb{P}(\exists n \geqslant \rho(\varepsilon\lambda/K_0)(\alpha_n \|Tx_n - u\| \geqslant \varepsilon)) \leqslant \mathbb{P}(\exists n\, (\|Tx_n - u\| \geqslant K_0/\lambda)) \leqslant \lambda$$

for any $\varepsilon, \lambda > 0$ and this suffices to establish the claim in this case. $\square$

Remark 3.6. If we assume that $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$, $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$ with rates of convergence χ_1, χ_2, respectively, then we can immediately derive rates ϕ_1, ϕ_2 for $\|\xi_n\|, \|\delta_n\| \to 0$ almost surely: Since $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < \infty$ with rate χ_1, we get

$$\mathbb{P}(\exists n \geqslant \chi_1(\lambda\varepsilon)(\|\xi_n\| \geqslant \varepsilon)) \leqslant \sum_{n=\chi_1(\lambda\varepsilon)}^{\infty} \mathbb{P}(\|\xi_n\| \geqslant \varepsilon) \leqslant \sum_{n=\chi_1(\lambda\varepsilon)}^{\infty} \frac{\mathbb{E}[\|\xi_n\|]}{\varepsilon} < \lambda$$

using Markov's inequality so that $\phi_1(\lambda, \varepsilon) := \chi_1(\lambda\varepsilon)$ is a rate for $\|\xi_n\| \to 0$ almost surely. Similarly for $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$ and χ_2, ϕ_2.

3.2.1. *Special cases of the Halpern iteration and the Krasnoselskii-Mann iteration with Tikhonov regularization terms.* In the special case of $U := \mathrm{Id}$ and $\delta_n := 0$, the iteration (sHM) collapses to the stochastic Halpern iteration (sH). We then have trivial rates for $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ and $\|Uy_n - y_n\| \to 0$ almost surely and so, in that case, we get under the assumptions of Theorems 3.2 and 3.4 (and also under suitable monotonicity assumptions of the rates involved) that

$$\mathbb{E}[\|Tx_n - x_n\|] \to 0$$

with a rate

$$\varphi(\varepsilon) := \max\{\varphi_{K,\theta,\chi}(\varepsilon/27), \rho_3(\varepsilon/9K_0), \rho_1(\varepsilon/9)\}$$

with $\varphi_{K,\theta,\chi}$ defined as in Theorem 3.2 and ρ_1, ρ_3 as in Theorem 3.4. Note that this generalises known rates in this case [5, Theorem 3.3], which apply only to specific choices of the parameters. In a similar way, we get a new rate for $\|Tx_n - x_n\| \to 0$ almost surely, though we do not spell it out here.

In the special case of $T := \mathrm{Id}$ and $\xi_n := 0$, the iteration (sHM) collapses to a stochastic variant of the Krasnoselskii-Mann iteration with Tikhonov regularization terms (sKM-T) (and even a slight extension by allowing general anchors u). In that case, we do not need to rely on the geometric arguments discussed in the next part of this section and can instead directly derive rates of convergence for $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ and $\|Uy_n - y_n\| \to 0$ almost surely, essentially following the approach of [40] (see also [7]).

Lemma 3.7 (essentially [40])**.** *Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *where $T = \mathrm{Id}$ and $\xi_n = 0$ for all $n \in \mathbb{N}$. Then the following recurrence relation holds everywhere on Ω for all $n \in \mathbb{N}$:*

$$\|Uy_{n+1} - y_{n+1}\| \leqslant 2\|y_n - y_{n+1}\| + \alpha_{n+1}\|Uy_{n+1} - u\| + \|\delta_n\| + \beta_n\|Uy_{n+1} - y_{n+1}\|.$$

Proof. Given $n \in \mathbb{N}$, we have

$$\begin{aligned}
\|Uy_{n+1} - y_{n+1}\| &= \|Uy_{n+1} - (1-\alpha_{n+1})x_{n+1} - \alpha_{n+1}u\| \\
&\leqslant \|Uy_{n+1} - x_{n+1}\| + \alpha_{n+1}\|Uy_{n+1} - u\| \\
&\leqslant \|Uy_{n+1} - Uy_n\| + \beta_n\|Uy_{n+1} - y_n\| + \|\delta_n\| + \alpha_{n+1}\|Uy_{n+1} - u\| \\
&\leqslant 2\|y_{n+1} - y_n\| + \beta_n\|Uy_{n+1} - y_{n+1}\| + \|\delta_n\| + \alpha_{n+1}\|Uy_{n+1} - u\|
\end{aligned}$$

pointwise everywhere. □

From that inequality, the following rates follow in a straightforward way:

Theorem 3.8. *Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *where $T = \mathrm{Id}$ and $\xi_n = 0$ for all $n \in \mathbb{N}$. Also, let $\Lambda > 0$ be such that $\Lambda \leqslant \beta_n \leqslant 1 - \Lambda$ for all $n \in \mathbb{N}$. If $\mathbb{E}[\|y_n - y_{n+1}\|] \to 0$ with rate φ, $\alpha_n \to 0$ with rate ρ and $\mathbb{E}[\|\delta_n\|] \to 0$ with rate χ, then $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ with rate*

$$\kappa(\varepsilon) := \max\{\varphi(\Lambda\varepsilon/4), \rho(\Lambda\varepsilon/4K_0), \chi(\Lambda\varepsilon/4)\} + 1.$$

Under the alternative hypothesis (Hyp′) *and assuming* $\|y_n - y_{n+1}\| \to 0$ *almost surely with rate* ψ, $\alpha_n \to 0$ *with rate* ρ *and* $\|\delta_n\| \to 0$ *almost surely with rate* ϕ, *then* $\|Uy_n - y_n\| \to 0$ *almost surely with rate*

$$\zeta(\lambda, \varepsilon) := \max\{\psi(\lambda/3, \Lambda\varepsilon/4), \rho(\Lambda\lambda\varepsilon/4K_0), \phi(\lambda/3, \Lambda\varepsilon/4)\} + 1.$$

Proof. By the Lemma 3.7 above, we have

$$(1 - \beta_n) \|Uy_{n+1} - y_{n+1}\| \leqslant 2 \|y_n - y_{n+1}\| + \alpha_{n+1} \|Uy_{n+1} - u\| + \|\delta_n\|$$

pointwise everywhere for any $n \in \mathbb{N}$. After taking expectations, we have

$$\mathbb{E}[\|Uy_{n+1} - y_{n+1}\|] \leqslant \frac{1}{\Lambda} (2\mathbb{E}[\|y_n - y_{n+1}\|] + \alpha_{n+1}K_0 + \mathbb{E}[\|\delta_n\|])$$

and from that the first rate immediately follows. The second rate follows rather similarly from the above relation: Using Markov's inequality, it holds that

$$\mathbb{P}\left(\exists n \left(\|Uy_{n+1} - u\| \geqslant \frac{K_0}{\lambda}\right)\right) \leqslant \mathbb{P}\left(Y \geqslant \frac{K_0}{\lambda}\right) \leqslant \lambda$$

for all $\lambda > 0$. Let now $\lambda, \varepsilon > 0$ be given. Take then ω such that $\|y_n(\omega) - y_{n+1}(\omega)\| < \Lambda\varepsilon/4$ for all $n \geqslant \psi(\lambda/3, \Lambda\varepsilon/4)$ and $\|\delta_n(\omega)\| < \Lambda\varepsilon/4$ for all $n \geqslant \phi(\lambda/3, \Lambda\varepsilon/4)$ as well as $\|Uy_{n+1}(\omega) - u(\omega)\| \leqslant \frac{K_0}{\lambda}$ for all n. Then for $n \geqslant \zeta(\lambda, \varepsilon) - 1$, it follows from the above inequality that $\|Uy_{n+1}(\omega) - y_{n+1}(\omega)\| < \varepsilon$. This immediately yields $\mathbb{P}(\exists n \geqslant \zeta(\lambda, \varepsilon) (\|Uy_n - y_n\| \geqslant \varepsilon)) < \lambda$ which completes the proof. ☐

3.2.2. *The general case.* We now discuss an alternative scenario where, in particular, a random variable Y satisfying (Hyp′) can be explicitly constructed if our mappings possess a common fixed point. To be more precise, let us assume that $\mathrm{Fix}T \cap \mathrm{Fix}U \neq \varnothing$ and that p is a common fixed point of T and U. Further, instead of making the assumptions (Hyp) or (Hyp′), for the rest of this section we now fix

$$D \geqslant \sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] \text{ and } E \geqslant \sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$$

as well as a K_0 such that $K_0 \geqslant \mathbb{E}[\|x_0 - p\|], \mathbb{E}[\|u - p\|]$. Using these data, we immediately get the following extended result on bounds:

Lemma 3.9. *For all* $n \in \mathbb{N}$, $\|x_n - p\| \leqslant Y' \leqslant 2Y' =: Y$ *pointwise everywhere, where*

$$Y' := \|x_0 - p\| + \|u - p\| + \sum_{i=0}^{\infty} (\|\xi_n\| + \|\delta_n\|)$$

and furthermore $\mathbb{E}[Y] \leqslant K := 4K_0 + 2D + 2E$. *The sequences*

$$\|y_n - p\|, \|x_{n+1} - x_n\|, \|y_{n+1} - y_n\|, \|Tx_n - u\|, \|Uy_n - y_n\|, \|Uy_n - u\|$$

are "L_1-dominated" by Y *in a similar way.*

Proof. Pointwise everywhere it holds that

$$\begin{aligned}
\|x_{n+1} - p\| &= \|(1 - \beta_n)(Uy_n + \delta_n) + \beta_n y_n - p\| \\
&\leqslant (1 - \beta_n) \|Uy_n - p\| + \beta_n \|y_n - p\| + \|\delta_n\| \\
&\leqslant \|y_n - p\| + \|\delta_n\| \\
&= \|(1 - \alpha_n)(Tx_n + \xi_n) + \alpha_n u - p\| + \|\delta_n\| \\
&\leqslant (1 - \alpha_n) \|Tx_n - p\| + \alpha_n \|u - p\| + \|\xi_n\| + \|\delta_n\| \\
&\leqslant (1 - \alpha_n) \|x_n - p\| + \alpha_n \|u - p\| + \|\xi_n\| + \|\delta_n\|.
\end{aligned}$$

It follows immediately by induction that

$$\|x_{n+1} - p\| \leqslant Y'_n := \|x_0 - p\| + \|u - p\| + \sum_{i=0}^{n} (\|\xi_n\| + \|\delta_n\|)$$

holds pointwise everywhere. Since the Y'_n are pointwise monotone, defining $Y' := \sup_{n \in \mathbb{N}} Y'_n$ yields that $\|x_n - p\| \leqslant Y'_n \leqslant Y'$ pointwise everywhere for all $n \in \mathbb{N}$, and by the monotone convergence theorem we have

$$\mathbb{E}[Y'] = \mathbb{E}[\|x_0 - p\|] + \mathbb{E}[\|u - p\|] + \sum_{i=0}^{\infty} (\mathbb{E}[\|\xi_n\|] + \mathbb{E}[\|\delta_n\|]) \leqslant 2K_0 + D + E.$$

Therefore immediately $\mathbb{E}[Y] \leqslant K$. By the above inequalities, one also has $\|y_n - p\| \leqslant Y'_n \leqslant Y'$, and the rest of the bounds follow by the triangle inequality. □

For the rest of this section, we will always assume the existence of a fixed point as above and use Y and K to refer to the quantities in Lemma 3.9. Note that these in particular validate the assumptions (Hyp) and (Hyp′).

We now move on to the asymptotic regularity relative to U of the sequence (y_n). For that, we initially establish $\|Uy_n - y_n\| \to 0$ almost surely using geometric properties of the underlying space and then use quantitative uniform integrability assumption for this random variables to recover the asymptotic regularity relative to U of (y_n) in expectation. First, we will make the following geometric assumption on the underlying normed space $(X, \|\cdot\|)$:

Definition 3.10. We say that $(X, \|\cdot\|)$ is uniformly convex ([10]) if for any $\varepsilon \in (0, 2]$, there exists a $\delta \in (0, 1]$ such that for all $x, y \in \overline{B}_1(0)$:

$$\|x - y\| \geqslant \varepsilon \text{ implies } \left\| \frac{x+y}{2} \right\| \leqslant 1 - \delta.$$

We call a modulus $\eta : (0, 2] \to (0, 1]$ witnessing such a δ in terms of ε a modulus of uniform convexity for X.

The above modulus also applies to closed balls of any radius centered at any point in the space and for arbitrary convex combinations:

Lemma 3.11. *Let η be a modulus of uniform convexity. For any $r > 0$ and $\varepsilon \in (0, 2]$, if $x, y \in \overline{B}_r(a)$ for $a \in X$ with $\|x - y\| \geqslant \varepsilon \cdot r$, then for all $\lambda \in [0, 1]$:*

$$\|(1 - \lambda)x + \lambda y - a\| \leqslant (1 - 2\lambda(1 - \lambda)\eta(\varepsilon))r.$$

The proof is straightforward and we hence omit it (but refer e.g. [28] for a proof of such a property even in the context of nonlinear uniformly convex hyperbolic spaces).

The proof of the following theorem now follows the outline of the proof of an analogous result for the Halpern-Mann iteration in uniformly convex hyperbolic spaces as given in [30] (though without errors, even nonstochastic ones):

Theorem 3.12. *Let $(X, \|\cdot\|)$ be uniformly convex with modulus η. Let $\|x_{n+1} - x_n\| \to 0$ almost surely with rate Δ. Also, let ρ be a rate for $\alpha_n \to 0$ and assume that $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$, $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$ with rates of convergence χ_1, χ_2, respectively. Lastly, let $\Lambda > 0$ be such that $\Lambda \leqslant \beta_n \leqslant 1 - \Lambda$. Then $\|Uy_n - y_n\| \to 0$ almost surely with rate*

$$\Gamma(\lambda, \varepsilon) := \max\{\Delta(\lambda/9, \widehat{\varepsilon}/4), \rho(\widehat{\varepsilon}/4K'), \chi_1(\lambda\widehat{\varepsilon}/36), \chi_2(\lambda\widehat{\varepsilon}/36)\}$$

for $\widehat{\varepsilon} := \varepsilon \cdot \Lambda^2 \cdot \eta(\varepsilon/K')$ and $K' := 3K/\lambda$.

Proof. Suppose for contradiction that

$$\mathbb{P}\left(\exists n \geqslant \Gamma(\lambda,\varepsilon)\left(\|Uy_n - y_n\| \geqslant \varepsilon\right)\right) \geqslant \lambda$$

and call the set inside the probability $B_{\lambda,\varepsilon}$. By Lemma 3.9 and Markov's inequality, we have

$$\mathbb{P}\left(\exists n\left(\|y_n - p\| > \frac{K}{\lambda}\right)\right) \leqslant \mathbb{P}\left(Y \geqslant \frac{K}{\lambda}\right) \leqslant \frac{\mathbb{E}[Y]}{K/\lambda} \leqslant \lambda$$

for any $\lambda > 0$ and so $\mathbb{P}\left(\exists n\left(\|y_n - p\| > K'\right)\right) \leqslant \lambda/3$ for $K' := 3K/\lambda$. Similarly for $\|u - p\|$. Thus, using the Fréchet inequalities, we have

$$\begin{aligned}
&\mathbb{P}(\exists n \geqslant \Gamma(\lambda,\varepsilon)(\|Uy_n - y_n\| \geqslant \varepsilon) \wedge \forall n\left(\|y_n - p\|, \|u - p\| \leqslant K'\right)) \\
&\qquad \geqslant \mathbb{P}(\exists n \geqslant \Gamma(\lambda,\varepsilon)(\|Uy_n - y_n\| > \varepsilon)) + \mathbb{P}(\forall n(\|y_n - p\| \leqslant K')) + \mathbb{P}(\forall n(\|u - p\| \leqslant K')) - 2 \\
&\qquad \geqslant \lambda + (1 - \lambda/3) + (1 - \lambda/3) - 2 = \lambda/3.
\end{aligned}$$

We denote that set measured in the above by $A_{\lambda,\varepsilon}$, and let $\omega \in A_{\lambda,\varepsilon}$, i.e. there exists some $n(\omega) \geqslant \Gamma(\lambda,\varepsilon)$ such that

$$\left\|Uy_{n(\omega)}(\omega) - y_{n(\omega)}(\omega)\right\| \geqslant \varepsilon \text{ and } \left\|y_{n(\omega)}(\omega) - p\right\|, \|u(\omega) - p\| \leqslant K'.$$

Writing n_0 for $n(\omega)$, we then have $\|Uy_{n_0}(\omega) - y_{n_0}(\omega)\| \leqslant 2\|y_{n_0}(\omega) - p\| \leqslant 2K'$ so that $\varepsilon/2 \leqslant \|y_{n_0}(\omega) - p\| \leqslant K'$. Also, we have $\|Uy_{n_0}(\omega) - p\| \leqslant \|y_{n_0}(\omega) - p\| \leqslant K'$ as well as

$$\|Uy_{n_0}(\omega) - y_{n_0}(\omega)\| \geqslant \varepsilon = \varepsilon/K' \cdot K' \geqslant \varepsilon/K' \|y_{n_0}(\omega) - p\|$$

and $\varepsilon/K' \leqslant 2$. So, we can apply Lemma 3.11 to derive

$$\begin{aligned}
\|x_{n_0+1}(\omega) - p\| &= \|(1 - \beta_{n_0})(Uy_{n_0}(\omega) + \delta_{n_0}(\omega)) + \beta_{n_0} y_{n_0}(\omega) - p\| \\
&\leqslant \|(1 - \beta_{n_0})Uy_{n_0}(\omega) + \beta_{n_0} y_{n_0}(\omega) - p\| + \|\delta_{n_0}(\omega)\| \\
&\leqslant (1 - 2\beta_{n_0}(1 - \beta_{n_0})\eta(\varepsilon/K'))\|y_{n_0}(\omega) - p\| + \|\delta_{n_0}(\omega)\| \\
&\leqslant \|y_{n_0}(\omega) - p\| - 2\|y_{n_0}(\omega) - p\| \Lambda^2 \eta(\varepsilon/K') + \|\delta_{n_0}(\omega)\| \\
&\leqslant \|y_{n_0}(\omega) - p\| - \varepsilon \cdot \Lambda^2 \cdot \eta(\varepsilon/K') + \|\delta_{n_0}(\omega)\|.
\end{aligned}$$

Now, we further have

$$\begin{aligned}
\|y_{n_0}(\omega) - p\| &\leqslant (1 - \alpha_{n_0})\|Tx_{n_0}(\omega) - p\| + \alpha_{n_0}\|u(\omega) - p\| + \|\xi_{n_0}(\omega)\| \\
&\leqslant \|x_{n_0}(\omega) - p\| + \alpha_{n_0}\|u(\omega) - p\| + \|\xi_{n_0}(\omega)\|
\end{aligned}$$

so that we can in particular derive

$$\|x_{n_0+1}(\omega) - p\| \leqslant \|x_{n_0}(\omega) - p\| + \alpha_{n_0}K' + \|\xi_{n_0}(\omega)\| + \|\delta_{n_0}(\omega)\| - \varepsilon \cdot \Lambda^2 \cdot \eta(\varepsilon/K').$$

So, in the end we have

$$\begin{aligned}
\widehat{\varepsilon} &= \varepsilon \cdot \Lambda^2 \cdot \eta(\varepsilon/K') \\
&\leqslant \|x_{n_0}(\omega) - p\| - \|x_{n_0+1}(\omega) - p\| + \alpha_{n_0}K' + \|\xi_{n_0}(\omega)\| + \|\delta_{n_0}(\omega)\| \\
&\leqslant \|x_{n_0+1}(\omega) - x_{n_0}(\omega)\| + \alpha_{n_0}K' + \|\xi_{n_0}(\omega)\| + \|\delta_{n_0}(\omega)\|.
\end{aligned}$$

Letting $V_n := \|x_{n+1} - x_n\| + \alpha_n K' + \|\xi_n\| + \|\delta_n\|$, we have shown that

$$A_{\lambda,\varepsilon} \subseteq \{\exists n \geqslant \Gamma(\lambda,\varepsilon)\left(V_n \geqslant \widehat{\varepsilon}\right)\}.$$

Similarly to in the proof of Theorem 3.5, we now have that $\chi_1(\lambda\varepsilon), \chi_2(\lambda\varepsilon)$ are rates for $\|\xi_n\|, \|\delta_n\| \to 0$, respectively. So we have

$$\begin{aligned}
\lambda/3 &\leqslant \mathbb{P}(A_{\lambda,\varepsilon}) \\
&\leqslant \mathbb{P}(\exists n \geqslant \Gamma(\lambda,\varepsilon)\,(V_n \geqslant \widehat{\varepsilon})) \\
&\leqslant \mathbb{P}\left(\exists n \geqslant \Gamma(\lambda,\varepsilon)\;((\|x_{n+1}-x_n\| \geqslant \widehat{\varepsilon}/4) \cup (\alpha_n K' \geqslant \widehat{\varepsilon}/4) \cup (\|\xi_n\| \geqslant \widehat{\varepsilon}/4) \cup (\|\delta_n\| \geqslant \widehat{\varepsilon}/4))\right) \\
&\leqslant \mathbb{P}\left(\exists n \geqslant \Delta(\lambda/9,\widehat{\varepsilon}/4)(\|x_{n+1}-x_n\| \geqslant \widehat{\varepsilon}/4)\right) + \mathbb{P}\left(\exists n \geqslant \rho(\widehat{\varepsilon}/4K')(\alpha_n K' \geqslant \widehat{\varepsilon}/4)\right) \\
&\quad + \mathbb{P}\left(\exists n \geqslant \chi_1(\lambda\widehat{\varepsilon}/36)(\|\xi_n\| \geqslant \widehat{\varepsilon}/4)\right) + \mathbb{P}\left(\exists n \geqslant \chi_2(\lambda\widehat{\varepsilon}/36)(\|\delta_n\| \geqslant \widehat{\varepsilon}/4)\right) \\
&< \frac{\lambda}{9} + 0 + \frac{\lambda}{9} + \frac{\lambda}{9} = \frac{\lambda}{3},
\end{aligned}$$

a contradiction. □

Remark 3.13. Using a slightly different argument first devised in [19, Theorem 3.4] (see also [26, Remark 15] or [30, Remark 3.7] for similar remarks in the context of nonlinear spaces), we can slightly optimize the above rate in the context of moduli of uniform convexity of a special form: Let $\eta(\varepsilon) = \varepsilon \cdot \tilde{\eta}(\varepsilon)$ where $\tilde{\eta}$ is increasing. Then above rate Γ holds even with $\widehat{\varepsilon} := \varepsilon \cdot \Lambda^2 \cdot \tilde{\eta}(\varepsilon/K')$.

To see this, follow the proof of Theorem 3.12 but replace ε/K' with $\varepsilon/\|y_{n_0}(\omega) - p\|$. Then also $\varepsilon/\|y_{n_0}(\omega) - p\| \leqslant 2$ as well as

$$\|Uy_{n_0}(\omega) - y_{n_0}(\omega)\| \geqslant \varepsilon/\|y_{n_0}(\omega) - p\| \cdot \|y_{n_0}(\omega) - p\|$$

and this leads to

$$\begin{aligned}
\|x_{n_0+1}(\omega) - p\| &\leqslant \|y_{n_0}(\omega) - p\| - 2\cdot\varepsilon\cdot\Lambda^2\cdot\tilde{\eta}(\varepsilon/\|y_{n_0}(\omega) - p\|) + \|\delta_{n_0}(\omega)\| \\
&\leqslant \|y_{n_0}(\omega) - p\| - \varepsilon\cdot\Lambda^2\cdot\tilde{\eta}(\varepsilon/\|y_{n_0}(\omega) - p\|) + \|\delta_{n_0}(\omega)\| \\
&\leqslant \|y_{n_0}(\omega) - p\| - \varepsilon\cdot\Lambda^2\cdot\tilde{\eta}(\varepsilon/K') + \|\delta_{n_0}(\omega)\|.
\end{aligned}$$

using that $\tilde{\eta}(\varepsilon/K') \leqslant \tilde{\eta}(\varepsilon/\|y_{n_0}(\omega) - p\|)$ as $\tilde{\eta}(\varepsilon/\|y_{n_0}(\omega) - p\|) \leqslant K'$ and since $\tilde{\eta}$ is increasing. Then the proof continuous as before.

We now discuss the assumptions from the quantitative theory of expected values that we require to establish an analogous result on the asymptotic regularity relative to U of (y_n) in mean.

Definition 3.14. Let X be an integrable random variable. We call a function $\mu : (0,\infty) \to (0,\infty)$ such that

$$\forall \varepsilon > 0\, \forall A \in \mathcal{F}\, (\mathbb{P}(A) \leqslant \mu(\varepsilon) \to \mathbb{E}[|X|1_A] \leqslant \varepsilon)$$

a modulus of absolute continuity for X.

Lemma 3.15. *Let X be an integrable random variable and μ a modulus of absolute continuity for X. For any $a, \varepsilon \in (0,\infty)$, we have that*

$$\mathbb{E}[|X|] \geqslant a + \varepsilon \text{ implies } \mathbb{P}(|X| > a) > \mu(\varepsilon/2).$$

In particular, we have that $\mathbb{E}[|X|] \geqslant \varepsilon$ implies $\mathbb{P}(|X| > \varepsilon/2) > \mu(\varepsilon/4)$.

Proof. Suppose $\mathbb{P}(|X| > a) \leqslant \mu(\varepsilon/2)$. Then we have

$$\mathbb{E}[|X|] = \mathbb{E}[|X|1_{|X|\leqslant a}] + \mathbb{E}[|X|1_{|X|>a}] \leqslant a + \varepsilon/2 < a + \varepsilon$$

which is the claim. □

Definition 3.16. A sequence of random variables (X_n) is called uniformly integrable if both $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < \infty$ and for any $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\forall n \in \mathbb{N}\, \forall A \in \mathcal{F}\, (\mathbb{P}(A) \leqslant \delta \to \mathbb{E}[|X_n| 1_A] \leqslant \varepsilon)\,.$$

We call a function μ that witnesses such a δ in terms of ε a modulus of uniform integrability for (X_n).

Note μ is a modulus of uniform integrability for (X_n) exactly when μ is a modulus of absolute continuity for any X_n. The main use that a modulus of uniform integrability has for a stochastic process is that with it, we can transfer a rate of almost-sure convergence to a rate of convergence in mean:

Lemma 3.17. *Let (X_n) be a sequence of nonnegative random variables such that $X_n \to 0$ almost surely with rate φ and such that μ is modulus of uniform integrability for (X_n). Then $\mathbb{E}[X_n] \to 0$ with rate*

$$\Gamma(\varepsilon) := \varphi\left(\mu\left(\frac{\varepsilon}{4}\right), \frac{\varepsilon}{2}\right).$$

Proof. Suppose for contradiction that there exists some $n_0 \geqslant \Gamma(\varepsilon)$ with $\mathbb{E}[X_{n_0}] \geqslant \varepsilon$. Then by Lemma 3.15 we have $\mathbb{P}(X_{n_0} > \varepsilon/2) > \mu(\varepsilon/4)$ and hence $\mathbb{P}\left(\exists n \geqslant \Gamma(\varepsilon)(X_n \geqslant \varepsilon/2)\right) \geqslant \mathbb{P}(X_{n_0} > \varepsilon/2) > \mu(\varepsilon/4)$, a contradiction. □

Theorem 3.18. *Let $(X, \|\cdot\|)$ be uniformly convex with modulus η. Under the assumptions of Theorem 3.2, let $\mathbb{E}[\|x_{n+1} - x_n\|] \to 0$ with rate Δ from Theorem 3.2. Also, let ρ be a rate for $\alpha_n \to 0$ and assume that $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$, $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$ with rates of convergence χ_1, χ_2, respectively. Also, let $\Lambda > 0$ be such that $\Lambda \leqslant \beta_n \leqslant 1 - \Lambda$. Lastly, let μ be a modulus of uniform integrability for $(\|Uy_n - y_n\|)$. Then $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ with rate*

$$\Gamma(\varepsilon) := \max\{\Delta(\overline{\varepsilon}), \rho(\widehat{\varepsilon}/4K'), \chi_1(\overline{\varepsilon}), \chi_2(\overline{\varepsilon})\}$$

where $\overline{\varepsilon} := \widehat{\varepsilon}\mu(\varepsilon/4)/36$ for $\widehat{\varepsilon} := \varepsilon/2 \cdot \Lambda^2 \cdot \eta(\varepsilon/2K')$ and $K' := 3K/\mu(\varepsilon/4)$.

Proof. By Theorem 3.2, $\Delta(\lambda\varepsilon/2)$ is a rate of almost sure convergence for $\|x_{n+1} - x_n\| \to 0$. Therefore by Theorem 3.12, a rate of almost sure convergence for $\|Uy_n - y_n\| \to 0$ is given by

$$\tilde{\Gamma}(\lambda, \varepsilon) := \max\{\Delta(\lambda\tilde{\varepsilon}/36), \rho(\tilde{\varepsilon}/4K'), \chi_1(\lambda\tilde{\varepsilon}/36), \chi_2(\lambda\tilde{\varepsilon}/36)\}$$

for $\tilde{\varepsilon} := \varepsilon \cdot \Lambda^2 \cdot \eta(\varepsilon/K')$ and $K' := 3K/\lambda$. By Lemma 3.17 we have $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ with rate

$$\Gamma(\varepsilon) := \tilde{\Gamma}\left(\mu\left(\frac{\varepsilon}{4}\right), \frac{\varepsilon}{2}\right) = \max\{\Delta(\overline{\varepsilon}), \rho(\widehat{\varepsilon}/4K'), \chi_1(\overline{\varepsilon}), \chi_2(\overline{\varepsilon})\}$$

where $\overline{\varepsilon} := \widehat{\varepsilon}\mu(\varepsilon/4)/36$ now for $\widehat{\varepsilon} := \varepsilon/2 \cdot \Lambda^2 \cdot \eta(\varepsilon/2K')$ and $K' := 3K/\mu(\varepsilon/4)$. □

Remark 3.19. Using Remark 3.13, it follows that also here, if $\eta(\varepsilon) = \varepsilon \cdot \tilde{\eta}(\varepsilon)$ where $\tilde{\eta}$ is increasing, then above rate Γ holds even with $\widehat{\varepsilon}$ defined as $\widehat{\varepsilon} := \varepsilon/2 \cdot \Lambda^2 \cdot \tilde{\eta}(\varepsilon/2K')$.

Before moving on to the case of fast rates for special choices of scalars and errors, we briefly discuss some natural conditions under which the previous additional assumption of a modulus of uniform integrability for the sequence $(\|Uy_n - y_n\|)$ can be obtained (cf. [15, Chapter 5.4]), and indicate how a modulus arises from these.

A first common assumption to establish uniform integrability is to assume a uniform finite p-th moment for $p > 1$. Now, in terms of an upper bound K witnessing that assumption, i.e. $\sup_{n\in\mathbb{N}} \mathbb{E}[\|Uy_n - y_n\|^p] < K$ for $p > 1$, we can actually compute a corresponding modulus μ. To see this, note that for any $\varepsilon > 0$, setting $X_n := \|Uy_n - y_n\|$, we have

$$\mathbb{E}[|X_n| 1_A] \leqslant \mathbb{E}[|X_n| 1_{A\cap(|X_n|\leqslant a)}] + \mathbb{E}[|X_n| 1_{A\cap(|X_n|>a)}] \leqslant a\mathbb{P}(A) + a^{1-p}\mathbb{E}[|X_n|^p] \leqslant a\mathbb{P}(A) + \varepsilon/2$$

for $a := (2K/\varepsilon)^{1/(p-1)}$ and $A \in \mathcal{F}$ and thus a modulus of uniform integrability is given by

$$\mu(\varepsilon) := \frac{\varepsilon}{2}\left(\frac{\varepsilon}{2K}\right)^{1/(p-1)}.$$

Another growth condition commonly imposed is to assume uniform finite expectation of the random variables under a supercoercive function. Then, in terms of an upper bound K witnessing that property, i.e. $\sup_{n\in\mathbb{N}} \mathbb{E}[g(\|Uy_n - y_n\|)] < K$ where $g : [0,\infty) \to [0,\infty)$ is supercoercive, i.e. $g(x)/x \to \infty$ as $x \to \infty$, with a rate of divergence $\kappa : (0,\infty) \to (0,\infty)$, i.e.

$$\forall a > 0\, \forall x \geqslant \kappa(a) \left(\frac{g(x)}{x} \geqslant a\right),$$

a corresponding modulus μ can again be easily computed: Similarly to above we have

$$\mathbb{E}[|X_n|1_A] \leqslant \kappa(a)\mathbb{P}(A) + \mathbb{E}[g\left(|X_n|\right)]/a \leqslant \kappa(a)\mathbb{P}(A) + \varepsilon/2$$

for $a := 2K/\varepsilon$ and thus a modulus of uniform integrability is now given by $\mu(\varepsilon) := \varepsilon/2\kappa\left(2K/\varepsilon\right)$.

Most importantly, we want to show that the required modulus of uniform integrability can be obtained through quantitative integrability assumptions on the errors. Concretely, note that we have $\|Uy_n - y_n\| \leqslant Y$ pointwise everywhere by Lemma 3.9, so a modulus of integrability for Y immediately yields a modulus of uniform integrability for $(\|Uy_n - y_n\|)$. In particular, if u and x_0 are chosen to be constant and $K > 0$ is such that $\|x_0 - p\|, \|u - p\| < K$, and in addition both $\sum_{i=0}^{\infty} \|\xi_i\|$ and $\sum_{i=0}^{\infty} \|\delta_i\|$ are integrable with moduli of absolute continuity μ_1 and μ_2 respectively, by Lemma 3.9 we have

$$\mathbb{E}[|X_n|1_A] \leqslant 2K\mathbb{P}(A) + 2\left(\sum_{i=0}^{\infty}\left(\mathbb{E}[\|\xi_i\|\, 1_A] + \mathbb{E}[\|\delta_i\|\, 1_A]\right)\right)$$

and thus a modulus of uniform integrability is now given by

$$\mu(\varepsilon) := \min\left\{\frac{\varepsilon}{4K}, \mu_1\left(\frac{\varepsilon}{8}\right), \mu_2\left(\frac{\varepsilon}{8}\right)\right\}.$$

In this way, we demonstrate that it is possible to shift the integrability condition on $(\|Uy_n - y_n\|)$ to appropriate integrability conditions on the sums of the error terms.

4. Fast rates of asymptotic regularity

In this section, we focus on particular instantiations of the parameters together with suitable growth conditions on the errors that allow for fast rates of asymptotic regularity for the above iteration(s). For that, we begin with some general results on deriving linear rates of convergence for sequences of real numbers satisfying a general recursive inequality and we subsequently extend this to sequences of random variables and utilize these general results then to in turn derive the fast rates. Throughout the section, we will be very explicit about the exact kind of assumptions (i.e. (Hyp), or (Hyp′), or the existence of common fixed points in the last part of the previous section) that are placed on the iterations in question.

4.1. **General results on linear rates.** We begin with the crucial result on deriving fast rates of convergence for Halpern-style iterations in nonlinear optimization. This result is closely modelled after a seminal lemma by Sabach and Shtern [38], first utilised in the context of proof mining in [7]. Here we formulate the idea behind the lemma in a slightly different style to fit the iterations considered in this paper, and in this way our presentation is closer to the explicit closed-form bounds in [5].

Lemma 4.1 (essentially [38])**.** *Suppose that $(s_n), (c_n)$ are sequences of nonnegative real numbers satisfying*

$$s_{n+1} \leqslant (1-a_n)s_n + c_n$$

for all $n \in \mathbb{N}$ where $(a_n) \subseteq [0,1]$. Then for all $m, K \in \mathbb{N}$ we have

$$s_{K+m+1} \leqslant A_K^{K+m} s_K + \sum_{i=K}^{K+m} A_{i+1}^{K+m} c_i$$

for $A_j^k := \prod_{i=j}^k (1-a_i)$, with $A_j^k := 1$ for $j > k$. In the special case that $a_n := \alpha_{n+1}$ and $c_n \leqslant (\alpha_n - \alpha_{n+1})L$ for some $(\alpha_n) \subseteq [0,1]$ and $L > 0$ we have

$$s_n \leqslant \tilde{A}_1^n s_0 + L\sum_{i=1}^{n}(\alpha_{i-1} - \alpha_i)\tilde{A}_{i+1}^n$$

for all $n \in \mathbb{N}$ where $\tilde{A}_j^k := \prod_{i=j}^k (1-\alpha_i)$, with $\tilde{A}_j^k := 1$ for $j > k$. If we furthermore define $\alpha_n := 2/(n+2)$ and assume that $s_0 \leqslant L$, then $s_n \leqslant 2L/(n+2)$ for all $n \in \mathbb{N}$.

Proof. The first inequality follows for all $m, K \in \mathbb{N}$ as in Lemma 2.2 immediately by induction. For the second part of the above lemma, we note that $A_j^k = \tilde{A}_{j+1}^{k+1}$ and thus from this first inequality (after setting $K = 0$), we have

$$s_{n+1} \leqslant A_0^n s_0 + \sum_{i=0}^{n} A_{i+1}^n c_i = \tilde{A}_1^{n+1} s_0 + \sum_{i=0}^{n} \tilde{A}_{i+2}^{n+1}(\alpha_i - \alpha_{i+1})L \leqslant \tilde{A}_1^{n+1} s_0 + L\sum_{i=1}^{n+1} \tilde{A}_{i+1}^{n+1}(\alpha_{i-1} - \alpha_i)$$

and also $s_0 = \tilde{A}_1^0 s_0$ by definition. For the final part, we observe that

$$\tilde{A}_i^n = \prod_{j=i}^{n}(1-\alpha_j) = \prod_{j=i}^{n} \frac{j}{j+2} = \frac{i}{i+2} \cdot \frac{i+1}{i+3} \cdot \ldots \cdot \frac{n-1}{n+1} \cdot \frac{n}{n+2} = \frac{i(i+1)}{(n+1)(n+2)}$$

for $i \leqslant n$ (noting that this also holds for $i = n$), and therefore

$$\begin{aligned}
s_n &\leqslant \frac{2s_0}{(n+1)(n+2)} + L\sum_{i=1}^{n}\left(\frac{2}{i+1} - \frac{2}{i+2}\right)\frac{(i+1)(i+2)}{(n+1)(n+2)} \\
&= \frac{2}{(n+1)(n+2)}\left(s_0 + L\sum_{i=1}^{n}\left(\frac{1}{i+1} - \frac{1}{i+2}\right)(i+1)(i+2)\right) \\
&\leqslant \frac{2L}{(n+1)(n+2)}\left(1 + \sum_{i=1}^{n} 1\right) = \frac{2L}{n+2}
\end{aligned}$$

which completes the proof. □

The following is an adaptation of the special case of the previous lemma concerning fast rates to sequences of random variables and as we will see in the following, it assumes a similarly important role for deriving linear rates of almost sure convergence.

Lemma 4.2. *Suppose that (X_n), (C_n) are nonnegative stochastic processes satisfying*

$$X_{n+1} \leqslant (1-\alpha_{n+1})X_n + C_n$$

almost surely for any $n \in \mathbb{N}$ where $\alpha_n := 2/(n+2)$ and where $\mathbb{E}[C_n] \leqslant (\alpha_n - \alpha_{n+1})L$ almost surely for all $n \in \mathbb{N}$ where $L \geqslant \mathbb{E}[X_0]$. Then

$$\mathbb{E}[X_n] \leqslant \frac{2L}{n+2} \text{ and } \mathbb{P}\left(\exists i \geqslant n\,(X_i \geqslant \varepsilon)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{n+2}$$

for all $n \in \mathbb{N}$.

Proof. From the fact that $X_{n+1} \leqslant (1-\alpha_{n+1})X_n + C_n$ holds almost surely, we immediately derive $\mathbb{E}[X_{n+1}] \leqslant (1-\alpha_{n+1})\mathbb{E}[X_n] + \mathbb{E}[C_n]$ and Lemma 4.1 yields $\mathbb{E}[X_n] \leqslant 2L/(n+2)$. Proceeding as in the proof of Lemma 2.3, noting that we in particular have $X_{n+1} \leqslant X_n + C_n$ almost surely, we similarly derive

$$\mathbb{P}(\exists n \geqslant N(U_n \geqslant \varepsilon)) \leqslant \frac{1}{\varepsilon}\left(\mathbb{E}[X_N] + \sum_{i=N}^{\infty} \mathbb{E}[C_i]\right)$$

for $U_n := X_n + \sum_{i=n}^{\infty} C_i$. In particular, we have $\mathbb{E}[X_N] \leqslant 2L/(N+2)$ and

$$\sum_{i=N}^{\infty} \mathbb{E}[C_i] \leqslant L \sum_{i=N}^{\infty} (\alpha_i - \alpha_{i+1}) = L\alpha_N = \frac{2L}{N+2}$$

so that $\mathbb{P}(\exists n \geqslant N(U_n \geqslant \varepsilon)) \leqslant \frac{1}{\varepsilon}\frac{4L}{N+2}$. This gives

$$\mathbb{P}\left(\exists n \geqslant N\left(X_n \geqslant \varepsilon\right)\right) \leqslant \mathbb{P}\left(\exists n \geqslant N\left(U_n \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{N+2}$$

again as in Lemma 2.3. □

Remark 4.3. Note that from the conclusions of Lemma 4.2, it is rather immediate to give corresponding rates for $\mathbb{E}[X_n] \to 0$ and $X_n \to 0$ almost surely, e.g. by setting $\Phi(\lambda, \varepsilon) := \lceil 4L/\varepsilon\lambda \rceil$ for the latter, but we prefer the above formulations in this section to make the constants very explicit.

4.2. **Linear rates of asymptotic regularity.** We now begin by establishing linear rates of asymptotic regularity for the iterations (x_n) and (y_n) in the special case of parameters

$$\text{(Par)} \qquad \alpha_n = \frac{2}{n+2} \text{ and } \beta_n = \beta \in (0,1).$$

Theorem 4.4. *Let* $(x_n), (y_n)$ *be the sequences generated by* (sHM) *for the parameters as in* (Par). *Assume* (Hyp) *with constant* K_0. *Also, assume that* $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$ *and* $\mathbb{E}[\|\delta_n\|] \leqslant K_2/(n+2)^2$. *Then*

$$\mathbb{E}[\|x_n - x_{n+1}\|] \leqslant \frac{2L}{n+2} \text{ and } \mathbb{P}\left(\exists i \geqslant n\left(\|x_i - x_{i+1}\| \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{n+2}$$

for all $n \in \mathbb{N}$ *where* $L = 2K_0 + 2K_1 + 2K_2$ *in both cases.*

Proof. As in the proof of Theorem 3.2, we have $X_{n+1} \leqslant (1-\alpha_{n+1})X_n + C_n$ for $X_n := \|x_n - x_{n+1}\|$ and

$$C_n := \|\xi_{n+1} - \xi_n\| + \|\delta_{n+1} - \delta_n\| + (\alpha_n - \alpha_{n+1})(\|Tx_n - u\| + \|\xi_n\|).$$

Also following the proof of Theorem 3.2 we have

$$\mathbb{E}[X_0] = \mathbb{E}[\|x_0 - x_1\|] \leqslant 2K_0 + K_1 + K_2 \leqslant L.$$

So it remains to show that $\mathbb{E}[C_n] \leqslant (\alpha_n - \alpha_{n+1})L$, and for this it suffices to show that

$$\mathbb{E}[\|\xi_{n+1} - \xi_n\|] \leqslant (\alpha_n - \alpha_{n+1}) \cdot 2K_1 \text{ and } \mathbb{E}[\|\delta_{n+1} - \delta_n\|] \leqslant (\alpha_n - \alpha_{n+1}) \cdot 2K_2.$$

We conclude by observing that

$$\begin{aligned}\mathbb{E}[\|\xi_{n+1} - \xi_n\|] \leqslant \mathbb{E}[\|\xi_{n+1}\|] + \mathbb{E}[\|\xi_n\|] &= \frac{K_1}{(n+3)^2} + \frac{K_1}{(n+2)^2}\\ &\leqslant \frac{2K_1}{(n+2)^2} \leqslant \frac{4K_1}{(n+2)(n+3)} = (\alpha_n - \alpha_{n+1}) \cdot 2K_1\end{aligned}$$

and similarly for (δ_n) and K_2. The rates then follow from Lemma 4.2. □

Remark 4.5. Before moving to the other asymptotic regularity results, we just briefly note that the asymptotic condition $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$ naturally implies that $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < \infty$ with a rather simple rate of convergence that can be easily calculated from the fact that

$$\begin{aligned}\sum_{n=N}^{\infty} \mathbb{E}[\|\xi_n\|] &\leqslant K_1 \sum_{n=N}^{\infty} \frac{1}{(n+2)^2} \leqslant K_1 \sum_{n=N}^{\infty} \frac{1}{(n+1)(n+2)}\\ &= K_1 \sum_{n=N}^{\infty} \left(\frac{1}{n+1} - \frac{1}{n+2}\right) = \frac{K_1}{N+1}\end{aligned}$$

for $N \geqslant 1$. Similarly, this applies to δ_n and K_2. In particular, as highlighted before in Remark 3.6, we have

$$\mathbb{P}(\exists n \geqslant N(\|\xi_n\| \geqslant \varepsilon)) \leqslant \sum_{n=N}^{\infty} \mathbb{P}(\|\xi_n\| \geqslant \varepsilon) \leqslant \sum_{n=N}^{\infty} \frac{\mathbb{E}[\|\xi_n\|]}{\varepsilon} \leqslant \frac{1}{\varepsilon}\frac{K_1}{N+1} \leqslant \frac{1}{\varepsilon}\frac{2K_1}{N+2}.$$

Now, in the case of sequence (y_n), the above then immediately implies the following:

Theorem 4.6. *Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *for the parameters as in* (Par). *Assume* (Hyp) *with constant K_0. Also, assume that $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$ and $\mathbb{E}[\|\delta_n\|] \leqslant K_2/(n+2)^2$. Then*

$$\mathbb{E}[\|y_n - y_{n+1}\|] \leqslant \frac{2L}{n+2}$$

for all $n \in \mathbb{N}$ where L can be given as an integer linear combination of K_0, K_1 and K_2. If we assume (Hyp′) *with K_0 and Y, then*

$$\mathbb{P}\left(\exists i \geqslant n\left(\|y_i - y_{i+1}\| \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{n+2}$$

for all $n \in \mathbb{N}$ with a suitable L constructed similarly.

Proof. Using Lemma 3.1, (1), we have

$$\|y_{n+1} - y_n\| \leqslant \|x_{n+1} - x_n\| + \|\xi_{n+1} - \xi_n\| + \alpha_n\left(\|Tx_n - u\| + \|\xi_n\|\right)$$

for all $n \in \mathbb{N}$ pointwise everywhere. By taking expectations, we get

$$\mathbb{E}[\|y_{n+1} - y_n\|] \leqslant \mathbb{E}[\|x_{n+1} - x_n\|] + \mathbb{E}[\|\xi_{n+1} - \xi_n\|] + \alpha_n(K_0 + K_1).$$

From Theorem 4.4, we get $\mathbb{E}[\|x_n - x_{n+1}\|] \leqslant 2L/(n+2)$ and similar as in the proof thereof, we have $\mathbb{E}[\|\xi_{n+1} - \xi_n\|] \leqslant 2K_1/(n+2)^2 \leqslant 2K_1/(n+2)$. Combined with the definition of α_n, we get the first claim for a suitable L arising as an integer linear combination of K_0, K_1 and K_2. The second claim follows similarly, noting the above Remark 4.5 and the fact that, using Markov's inequality, we have

$$\begin{aligned}&\mathbb{P}\left(\exists n \geqslant N\left(\alpha_n(\|Tx_n - u\| + \|\xi_n\|) \geqslant \varepsilon\right)\right)\\ &\quad\leqslant \mathbb{P}\left(\exists n \geqslant N\left((\|Tx_n - u\| + \|\xi_n\|) \geqslant \varepsilon/\alpha_N\right)\right)\\ &\quad\leqslant \mathbb{P}\left(\exists n \geqslant N\left(Y \geqslant \varepsilon/2\alpha_N\right)\right) + \mathbb{P}\left(\exists n \geqslant N\left(\|\xi_n\| \geqslant \varepsilon/2\alpha_N\right)\right)\\ &\quad\leqslant \frac{1}{\varepsilon}\frac{4K_0}{N+2} + \frac{1}{\varepsilon}\frac{8K_1}{N+2},\end{aligned}$$

which, combined with the previous, rather immediately yields the result (which we therefore do not spell out any further). □

Remark 4.7. Reformulated, the above results in particular state that if $(x_n), (y_n)$ are the sequences generated by (sHM) for parameters as in (Par) under the assumption (Hyp) and $\mathbb{E}(\|\xi_n\|) = O(1/n^2)$ as well as $\mathbb{E}(\|\delta_n\|) = O(1/n^2)$, then $\mathbb{E}[\|x_n - x_{n+1}\|] = O(1/n)$ as well as $\mathbb{E}[\|y_n - y_{n+1}\|] = O(1/n)$.

4.3. **Linear rates of asymptotic regularity relative to the mappings in special cases.**

In the special case of the stochastic Halpern iteration, which we reobtain (as discussed before) by setting $U := \mathrm{Id}$ as well as $\delta_n := 0$, we get the following fast rates:

Theorem 4.8. *Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *for parameters as in* (Par) *and where $U := \mathrm{Id}$ and $\delta_n := 0$. Assume* (Hyp) *with constant K_0. Also, assume that $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$. Then*

$$\mathbb{E}[\|Tx_n - x_n\|] \leqslant \frac{2L}{n+2}$$

for all $n \in \mathbb{N}$ where L can be given as an integer linear combination of K_0 and K_1. If we assume (Hyp′) *with K_0 and Y, then*

$$\mathbb{P}\left(\exists i \geqslant n \left(\|Tx_i - x_i\| \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{n+2}$$

for all $n \in \mathbb{N}$ with a suitable L constructed similarly.

Proof. Using the inequalities listed in the proof of Theorem 3.4, we obtain

$$\|Tx_n - x_n\| \leqslant 3\,\|x_{n+1} - x_n\| + \alpha_n\,\|Tx_n - u\| + \|\xi_n\|$$

for all $n \in \mathbb{N}$ pointwise everywhere, in this special case where $U = \mathrm{Id}$ and $\delta_n = 0$. This immediately yields the above rates (using similar arguments as in Lemma 4.6 in the case of the almost sure convergence) using the previous Theorem 4.4 (noting that in this case $K_2 = 0$). □

Theorem 4.8 is closely related to [5, Theorem 3.3], but with adjusted step-sizes that now provide *exact* linear rates (without logarithmic factors).

In the special case of the stochastic Krasnoselskii-Mann iteration with Tikhonov regularization terms, which we re-obtain by setting $T := \mathrm{Id}$ as well as $\xi_n := 0$, we get the following fast rates in the above special case:

Theorem 4.9. *Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *for parameters as in* (Par) *and where $T := \mathrm{Id}$ and $\xi_n := 0$. Assume* (Hyp) *with constant K_0. Also, assume that $\mathbb{E}[\|\delta_n\|] \leqslant K_2/(n+2)^2$. Lastly, let $B \geqslant 1/(1-\beta)$. Then*

$$\mathbb{E}[\|Ux_n - x_n\|] \leqslant \frac{2L}{n+2}$$

for all $n \in \mathbb{N}^*$ where L can be constructed in terms of K_0, K_2 and B. If we assume (Hyp′) *with K_0 and Y, then*

$$\mathbb{P}\left(\exists i \geqslant n \left(\|Ux_i - x_i\| \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L}{n+2}$$

for all $n \in \mathbb{N}^*$ with a suitable L constructed similarly.

Proof. Using Lemma 3.7, we have

$$\|Uy_n - y_n\| \leqslant B\left(2\,\|y_{n-1} - y_n\| + \alpha_n\,\|Uy_n - u\| + \|\delta_{n-1}\|\right)$$

for $n \geqslant 1$ pointwise everywhere. Using the preceding Theorem 4.6, we immediately get that

$$\mathbb{E}[\|Uy_n - y_n\|] \leqslant \frac{2L_0}{n+2} \text{ and } \mathbb{P}\left(\exists i \geqslant n \left(\|Uy_i - y_i\| \geqslant \varepsilon\right)\right) \leqslant \frac{1}{\varepsilon}\frac{4L_0}{n+2}$$

for $n \geqslant 1$ and a suitable constant L_0 arising as an integer linear combination of K_0 and K_2. Using the inequalities from Theorem 3.4, we then further have

$$\|Ux_n - x_n\| \leqslant 2\|x_{n+1} - x_n\| + 3\|Uy_n - y_n\| + 2\|\delta_n\|$$

for all $n \in \mathbb{N}$ pointwise everywhere and so, using the previous results as well as Theorem 4.4, we get the desired rates. □

4.4. **Fast rates of asymptotic regularity relative to the mappings in the general case.** In the context of the above assumptions on the scalar sequences and the errors, we can still get rather sensible complexity estimates in the general case where neither mapping necessarily trivializes. While we could express this again using a general modulus of uniform convexity η for the underlying space, we here focus on the case where η is of *power type* p for $p \geqslant 2$, i.e. there exists a constant C such that $\eta(\varepsilon) = C\varepsilon^p$. Crucially, this is the case for inner product spaces for $p = 2$:

Lemma 4.10 (essentially [26])**.** *If X is a inner product space, then X is uniformly convex with a corresponding modulus $\eta(\varepsilon) = \varepsilon^2/8$.*

This allows for the following results on the asymptotic regularity in the general case. We begin by instantiating Theorem 3.12 on the asymptotic regularity of the sequence (y_n) relative to U almost surely and Theorem 3.18 for deriving the respective regularity result in expectation.

Lemma 4.11. *Let X be uniformly convex with a modulus η of power type p with constant C. Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *for parameters as in* (Par)*. Let K and Y be as in Lemma 3.9. Also, assume that $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$ and $\mathbb{E}[\|\delta_n\|] \leqslant K_2/(n+2)^2$. Lastly, let $\Lambda > 0$ be such that $\Lambda \leqslant \beta \leqslant 1 - \Lambda$. Then $\|Uy_n - y_n\| \to 0$ almost surely with rate*

$$\Gamma(\lambda, \varepsilon) := \left\lceil \frac{(3K)^{p-1}L}{C\Lambda^2\varepsilon^p\lambda^p} \right\rceil$$

for a suitable L arising as an integer linear combination of K, K_1 and K_2. Assuming the existence of a modulus μ of uniform integrability for $(\|Uy_n - y_n\|)$, we further get $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ with rate

$$\Gamma'(\varepsilon) := \left\lceil \frac{2^p(3K)^{p-1}L}{C\Lambda^2\varepsilon^p\mu(\varepsilon/4)^p} \right\rceil$$

where L is as above.

Proof. First note that in the context of moduli η of power type $p \geqslant 2$, we are actually in the setting of the previous Remarks 3.13 and 3.19 where $\tilde{\eta}(\varepsilon) = C\varepsilon^{p-1}$. Then the rate for $\|Uy_n - y_n\| \to 0$ almost surely follows by instantiating the rate given in Theorem 3.12 with the following moduli: With the above $\tilde{\eta}$, we have $\widehat{\varepsilon} := C\Lambda^2\varepsilon^p\lambda^{p-1}/(3K)^{p-1}$ and $K' := 3K/\lambda$. Using Theorem 4.4, we have $\mathbb{P}\left(\exists i \geqslant n\,(\|x_i - x_{i+1}\| \geqslant \varepsilon)\right) \leqslant \frac{1}{\varepsilon}\frac{4L_0}{n+2}$ for all $n \in \mathbb{N}$ and a suitable constant L_0 arising as an integer linear combination of K, K_1 and K_2. So we in particular have that $\Delta(\lambda, \varepsilon) = \lceil 4L_0/\varepsilon\lambda \rceil$ is a corresponding rate for $\|x_n - x_{n+1}\| \to 0$ almost surely. As $\alpha = 2/(n+2)$, we further have rather immediately that $\rho(\varepsilon) = \lceil 2/\varepsilon \rceil$ is a corresponding rate for $\alpha_n \to 0$. Lastly, using the assumptions on $\|\xi_n\|$ and $\|\delta_n\|$, note that as in Remark 4.5 we have $\sum_{n=N}^{\infty} \mathbb{E}[\|\xi_n\|] \leqslant K_1/(N+1)$ so that $\chi_1(\varepsilon) = \lceil K_1/\varepsilon \rceil$ is a corresponding rate of convergence for $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < \infty$. The rate of convergence $\chi_2(\varepsilon) = \lceil K_2/\varepsilon \rceil$ for $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$ follows similarly. Then instantiating Theorem 3.12 under Remark 3.13 gives us the rate

$$\max\left\{ \left\lceil \frac{144L_0(3K)^{p-1}}{C\Lambda^2\varepsilon^p\lambda^p} \right\rceil, \left\lceil \frac{24K(3K)^{p-1}}{C\Lambda^2\varepsilon^p\lambda^p} \right\rceil, \left\lceil \frac{36K_1(3K)^{p-1}}{C\Lambda^2\varepsilon^p\lambda^p} \right\rceil, \left\lceil \frac{36K_2(3K)^{p-1}}{C\Lambda^2\varepsilon^p\lambda^p} \right\rceil \right\} \leqslant \frac{(3K)^{p-1}L}{C\Lambda^2\varepsilon^p\lambda^p}$$

for $L := 144L_0$, noting that $K, K_1, K_2 \leqslant L_0$ and so the first part follows. For the second part, we just apply Lemma 3.17 directly. □

Using that, we can then employ the previous Theorems 3.4 and 3.5 to derive rates of asymptotic regularity also for the sequence (x_n) relative to the mappings U and T. For simplicity, we now focus on inner product spaces, i.e. where $p = 2$ and $C = 1/8$ by Lemma 4.10.

Theorem 4.12. *Let X be an inner product space. Let $(x_n), (y_n)$ be the sequences generated by* (sHM) *for parameters as in* (Par). *Let K and Y be as in Lemma 3.9. Also, assume that $\mathbb{E}[\|\xi_n\|] \leqslant K_1/(n+2)^2$ and $\mathbb{E}[\|\delta_n\|] \leqslant K_2/(n+2)^2$. Lastly, let $\Lambda > 0$ be such that $\Lambda \leqslant \beta \leqslant 1-\Lambda$. Then $\|Ux_n - x_n\| \to 0$ and $\|Tx_n - x_n\| \to 0$ almost surely with rates*

$$\Phi_1(\lambda, \varepsilon) := \left\lceil \frac{24KL}{\Lambda^2\varepsilon^2\lambda^2} \right\rceil \text{ and } \Phi_2(\lambda, \varepsilon) := \left\lceil \frac{72KL}{\Lambda^2\varepsilon^2\lambda^2} \right\rceil,$$

respectively, where L is as in Lemma 4.11. Assuming the existence of a modulus μ of uniform integrability for $(\|Uy_n - y_n\|)$, we further get $\mathbb{E}[\|Ux_n - x_n\|] \to 0$ and $\mathbb{E}[\|Tx_n - x_n\|] \to 0$ with respective rates

$$\varphi_1(\varepsilon) := \left\lceil \frac{96KL}{\Lambda^2\varepsilon^2\mu(\varepsilon/4)^2} \right\rceil \text{ and } \varphi_2(\varepsilon) := \left\lceil \frac{288KL}{\Lambda^2\varepsilon^2\mu(\varepsilon/4)^2} \right\rceil.$$

Proof. The rates follow immediately by instantiating Theorems 3.4 and 3.5 with the rates obtained from Theorem 4.4 and Lemma 4.11, noting in particular that the quadratic rates for $\|Uy_n - y_n\| \to 0$ and $\mathbb{E}[\|Uy_n - y_n\|] \to 0$ dominate. □

5. An outlook onto future work

We conclude the paper with a brief discussion on the potential of the schemas and results given here and, with that, an outlook onto forthcoming work. Concretely, in a forthcoming paper, we will outline applications of the present paper to problems in reinforcement learning, where a novel variant of the prominent model-free Q-learning algorithm [41] with Tikhonov regularization terms is presented, and the asymptotic regularity results from the main theoretical part of this paper are utilized to give very good oracle complexity estimates for a specific mini-batching procedure that matches those recently obtained for Halpern's iteration in [5]. In more detail, only very recently in [5] has the special case of the scheme (sHM) corresponding to Halpern's iteration (obtained by setting $U :=$ Id and $\delta_n := 0$ in (sHM)) been instantiated to yield a Halpern-type version of Q-learning, and in a similar way our general method represents an expanded class of learning algorithms. A key aspect of those variants, already motivating [5], is that they stay computationally effective over the averaged reward setting as the iterations allow for fast asymptotic behavior even in the presence of general nonexpansive maps.

In the case of our stochastic Krasnoselskii-Mann iteration with Tikhonov regularization terms (obtained by setting $T :=$ Id and $\xi_n := 0$ as well as $\gamma_n := 1 - \alpha_n$ and $u := 0$ in (sHM)), this in particular leads to the following novel Q-learning method *with Tikhonov regularization terms*: Let (S, A, r, p) be a given Markov decision process with a finite set of states S and actions A, where if we choose action a in state s, $r(s, a)$ represents an immediate reward and $p(s, a, t)$ the probability that we transition to state t. Under standard conditions on the Markov decision process and by instantiating the noise terms according to a regularized minibatch strategy as e.g. discussed in [5] for a Halpern-type version of Q-learning, we can then utilize variants of RVI-Q-learning [1] in the style of the Krasnoselskii-Mann iteration with Tikhonov regularization terms to produce the new scheme

$$Q_{n+1}(s,a) := (1-\beta_n)\left(r(s,a) + \frac{\gamma_n}{s_n}\sum_{j=1}^{s_n}\max_{b\in S} Q_n(\zeta_{n,j},b) - f(\gamma_n Q_n(s,a))\right) + \beta_n\gamma_n Q_n(s,a),$$

to quickly approximate the optimal policy in the average reward setting, where $\zeta_{n,1}\ldots,\zeta_{n,s_n}$ are i.i.d. random variables sampled from $p(s,a,\cdot)$, and f is a suitable function (see e.g. [1]).

A detailed exposition of these new methods, and of the use of the additional degree of freedom gained by the Tikhonov regularization terms, together with a careful and rather general examination of associated oracle complexities, will be given in a forthcoming paper. In that paper, we will also examine a novel use of the second mapping in the schema (sHM) in the style of double Q-learning [17], optimizing over two concurrent Markov decision processes in parallel environments.

References

[1] J. Abounadi, D.P. Bertsekas, and V.S. Borkar. Learning algorithms for Markov decision processes with average cost. *SIAM Journal on Control and Optimization*, 40(3):681–698, 2001.

[2] R.I. Boţ, E.R. Csetnek, and D. Meier. Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces. *Optimization Methods and Software*, 34:489–514, 2019.

[3] R.I. Boţ, M. Sedlmayer, and P. Tu Vuong. A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs. *Journal of Machine Learning Research*, 24(8):1–37, 2023.

[4] M. Bravo and R. Cominetti. Stochastic fixed-point iterations for nonexpansive maps: Convergence and error bounds. *SIAM Journal on Control and Optimization*, 62:191–219, 2024.

[5] M. Bravo and J.P. Contreras. Stochastic Halpern iteration in normed spaces and applications to reinforcement learning, 2024. arXiv e-print, math.OC, 2403.12338.

[6] X. Cai, C. Song, C. Guzmán, and J. Diakonikolas. Stochastic Halpern iteration with variance reduction for stochastic monotone inclusions. In *Advances in Neural Information Processing Systems*, volume 35, pages 24766–24779, 2022.

[7] H. Cheval, U. Kohlenbach, and L. Leuştean. On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications*, 197:233–251, 2023.

[8] H. Cheval and L. Leuştean. Quadratic rates of asymptotic regularity for the Tikhonov-Mann iteration. *Optimization Methods & Software*, 37(6):2225–2240, 2022.

[9] H. Cheval and L. Leuştean. Linear rates of asymptotic regularity for Halpern-type iterations. *Mathematics of Computation*, 2024. To appear, 11pp., doi:10.1090/mcom/3991.

[10] J.A. Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.

[11] A. Cuntavepanit and B. Panyanak. Strong convergence of modified Halpern iterations in CAT(0) spaces. *Fixed Point Theory and Applications*, 2011. Art. ID 869458, 11pp.

[12] J. Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Proceedings of 33rd Annual Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 1428–1451, 2020.

[13] B. Dinis and P. Pinto. On the convergence of algorithms with Tikhonov regularization terms. *Optimization Letters*, 15(4):1263–1276, 2021.

[14] B. Dinis and P. Pinto. Strong convergence for the alternating Halpern-Mann iteration in CAT(0) spaces. *SIAM Journal on Optimization*, 33(2):785–815, 2023.

[15] A. Gut. *Probability: A Graduate Course.* Springer Texts in Statistics. Springer, second edition, 2013.

[16] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.

[17] H.v. Hasselt. Double Q-learning. In *Advances in Neural Information Processing Systems*, volume 23, pages 2613–2621, 2010.

[18] T.-H. Kim and H.-K. Xu. Strong convergence of modified Mann iterations. *Nonlinear Analysis. Theory, Methods & Applications*, 61(1–2):51–60, 2005.

[19] U. Kohlenbach. Uniform asymptotic regularity for Mann iterates. *Journal of Mathematical Analysis and Applications*, 279(2):531–544, 2003.

[20] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.

[21] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226(3):2764–2795, 2011.

[22] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proc. ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.

[23] U. Kohlenbach and L. Leuştean. Effective metastability of Halpern iterates in CAT(0) spaces. *Advances in Mathematics*, 231:2526–2556, 2012.

[24] M.A. Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1(63)):123–127, 1955.

[25] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Classics in Mathematics. Springer-Verlag Berlin Heidelberg, 1991.

[26] L. Leuştean. A quadratic rate of asymptotic regularity for CAT(0)-spaces. *Journal of Mathematical Analysis and Applications*, 325:386–399, 2007.

[27] L. Leuştean. Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science*, 13(11):1680–1691, 2007.

[28] L. Leuştean. Nonexpansive iterations in uniformly convex W-hyperbolic spaces. In A. Leizarowitz, B.S. Mordukhovich, I. Shafrir, and A. Zaslavski, editors, *Nonlinear Analysis and Optimization I: Nonlinear Analysis*, pages 193–209. American Mathematical Society, 2010.

[29] L. Leuştean and P. Pinto. Quantitative results on a Halpern-type proximal point algorithm. *Computational Optimization and Applications*, 79(1):101–125, 2021.

[30] L. Leuştean and P. Pinto. Rates of asymptotic regularity for the alternating Halpern-Mann iteration. *Optimization Letters*, 18:529–543, 2024.

[31] Kim-Chuan Toh Ling Liang and Jia-Jie Zhu. An Inexact Halpern Iteration with Application to Distributionally Robust Optimization, 2024. arXiv e-print, math.OC, 2402.06033.

[32] W.R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.

[33] M. Neri and N. Pischke. Proof mining and probability theory, 2024. arXiv e-print, math.LO, 2403.00659.

[34] 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. arXiv e-print, math.OC, 2504.12922.

[35] M. Neri and T. Powell. A quantitative Robbins-Siegmund theorem, 2024. arXiv e-print, math.OC, 2410.15986.

[36] M. Neri and T. Powell. On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales. *Transactions of the American Mathematical Society*, 2025. to appear, arXiv e-print, math.PR, 2406.19979.

[37] B. Nguyen and L. Nguyen Duc. Hybrid descent-like Halpern iteration methods for two nonexpansive mappings and semigroups on two sets. *Theoretical Mathematics & Applications*, 2(3):23–38, 2012.

[38] S. Sabach and S. Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27:640–660, 2017.

[39] S. Saejung. Halpern iterations in CAT(0) spaces. *Fixed Point Theory and Applications*, 2010. Article ID 471781, 13pp.

[40] K. Schade and U. Kohlenbach. Effective metastability for modified Halpern iterations in CAT(0) spaces. *Fixed Point Theory and Applications*, 2012. Article no. 191, 19pp.

[41] C. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge, 1989.

[42] R. Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58(5):486–491, 1992.

[43] H.-K. Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66:240–256, 2002.

[44] Y. Yao, H. Zhou, and Y.-C. Liou. Strong convergence of a modified Krasnoselski-Mann iterative algorithm for non-expansive mappings. *Journal of Applied Mathematics and Computing*, 29(1–2):383–389, 2009.

[45] T. Yoon and E.K. Ryu. Accelerated algorithms for smooth convex-concave minimax problems with $\mathcal{O}(1/k^2)$ rate on squared gradient norm. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 12098–12109, 2021.