

# 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 completely abstractly, rather than fixing a sampling method; second, 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 (where this stochastic variant of the latter is considered for the first time in this paper). For these specific cases, we 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. We conclude by discussing how variance can be managed in practice through sampling methods in the style of minibatching, how our convergence rates can be adapted to provide oracle complexity bounds, and by sketching 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

**1.1. Background and motivation.** 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. Model-free reinforcement learning algorithms represent just one well-known example of this general situation, where variants of  $Q$ -learning, for instance, can be viewed as stochastic methods for computing fixed points of nonexpansive operators.

Let  $(X, \|\cdot\|)$  be a separable real-valued normed space and  $T, U : X \rightarrow X$  be two nonexpansive mappings on  $X$ , i.e.

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

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

---

*Date:* December 17, 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.

**Acknowledgments:** We want to thank the two anonymous referees for their helpful suggestions which improved the paper in various places.

**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.

iteration, given by the scheme

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

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,  $(\xi_n), (\delta_n)$  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 (HM), schematically appearing already in [40], and recently (re-)discovered in a much more generalized setting in [13]. This scheme integrates two of the most prominent methods for approximating fixed points for nonexpansive mappings, the Krasnoselskii-Mann method [24, 35] and Halpern's method [14] (see also [48] and [49]), 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 propose 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,  $\xi_n$  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  $\delta_n$  and  $Uy_n$ ), though our presentation is fully abstract and we will make no assumptions about  $\xi_n$  or  $\delta_n$ , other than imposing controls on  $\mathbb{E}[\|\xi_n\|]$  and  $\mathbb{E}[\|\delta_n\|]$ , in a way which can be easily achieved in concrete scenarios through sampling methods such as minibatching, as will also be discussed later.

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

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

considered on an abstract level in [4] (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 [5, 11, 31, 51]. 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 statistics and 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, our method (sHM) is certainly not limited to this special case, and in line with [13] encompasses stochastic variants of other well-known deterministic methods. An important example of this is represented by setting  $T := \text{Id}$  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 (KM-T) considered in [1, 50] that now

incorporates stochastic noise, taking the form of

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

Alternatively, the method can be seen as the stochastic Krasnoselskii-Mann iteration as considered in [3] with Tikhonov regularization terms as considered in [1]. This method is known to produce fast asymptotic behavior in the deterministic setting, and in contrast to the standard Krasnoselskii-Mann scheme, also benefits from strong convergence results similarly to Halpern's iteration [1]. As we will show in this paper, in particular the first of these features extends to the stochastic setting (while we also lay the foundation for the second, as discussed later). For a simple overview, we present the relationships between the different stochastic schemes considered here, as well as their deterministic counterparts, in Figure 1.

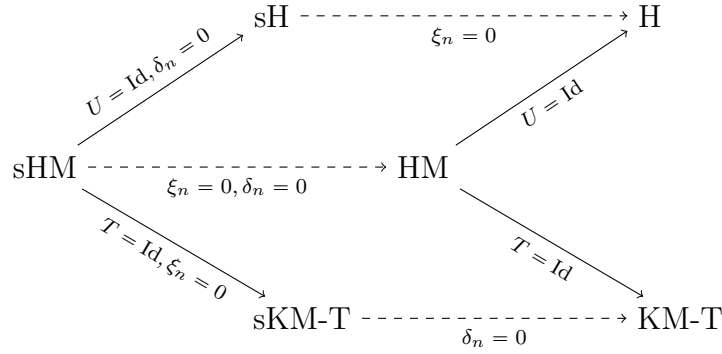


FIGURE 1. Relationships of the stochastic Halpern-type schemes and their deterministic variants

To our knowledge, each of the stochastic schemes (sHM) and (sKM-T) are introduced here for the first time.

**1.2. Overview of the main results.** Our main results comprise both abstract convergence theorems for the stochastic schemes in Figure 1, valid under very broad assumptions (and with convergence rates given at a corresponding level of generality), along with a series of special cases where fast, linear rates are possible, the latter demonstrating that our framework includes stochastic algorithms that come with state-of-the-art convergence guarantees of a kind only recently established for deterministic Halpern-type schemes [32, 43]. We now outline each of these in turn.

**1.2.1. General (quantitative) asymptotic regularity results.** The main results of the paper establish general conditions under which one can guarantee the asymptotic regularity of the scheme (sHM), both in the traditional sense of  $\|x_n - x_{n+1}\|$  (sometimes called the discrete velocity [2]) 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, that is we show both

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

as well as

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

Most importantly, in all cases we provide explicit convergence rates for these expressions. In the case of convergence in mean, these rates take the form of functions  $\varphi$  which guarantee

$$\mathbb{E}[\|x_n - x_{n+1}\|] < \varepsilon \text{ for all } \varepsilon > 0 \text{ and any } n \geq \varphi(\varepsilon),$$

and similarly for  $\mathbb{E}[\|x_n - Tx_n\|]$  and  $\mathbb{E}[\|x_n - Ux_n\|]$ . In the case of our almost-sure convergence results, our rates instead take the form of functions  $\Phi$  which guarantee

$$\mathbb{P}(\exists n \geq \Phi(\lambda, \varepsilon)(\|x_n - x_{n+1}\| \geq \varepsilon)) < \lambda$$

for all  $\varepsilon, \lambda > 0$  (and similarly for  $\|x_n - Tx_n\|$  and  $\|x_n - Ux_n\|$ ). In all cases, these rates are explicitly constructed under very general conditions and depend only on a few moduli witnessing quantitative aspects of our main assumptions.

These general asymptotic regularity results and their corresponding rates for the scheme (sHM), both in the sense of the discrete velocity and relative to the mappings, can be found in Theorems 3.2 – 3.5. While the present paper features various fine-grained discussions on the precise assumptions necessary for each of these results, they essentially amount to

- the existence of a common fixed point of  $T$  and  $U$ ,
- standard conditions on the parameters  $(\alpha_n)$  and  $(\beta_n)$ , and
- variance control in the form of  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < +\infty$  and  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < +\infty$ .

In their quantitative variants, the respective rates correspondingly rely on bounds and rates witnessing these assumptions in various ways. While the results on the discrete velocity as given in Theorems 3.2 and 3.3 are proved unconditionally, asymptotic regularity relative to the mappings turns out to be more involved, and the respective Theorems 3.4 and 3.5 depend a priori on the assumption that  $\|Uy_n - y_n\| \rightarrow 0$ , either in mean or almost surely. It is precisely at this point that our treatment divides according to the main schemes we consider in Figure 1. For (sH) and (sKM-T), this premise either trivializes or can be directly derived without additional conditions (see Theorem 3.8), but for the general case of (sHM), establishing  $\|Uy_n - y_n\| \rightarrow 0$  almost surely requires a subtle pointwise geometric argument based on the additional assumption that the underlying space  $X$  is uniformly convex (see Theorem 3.12). Our corresponding result in mean (see Theorem 3.18) is effectively obtained as a direct lift of the pointwise argument, which we achieve in an abstract way by showing that the sequence  $(\|Uy_n - y_n\|)$  is uniformly integrable given the existence of a common fixed point and the variance assumptions  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < +\infty$  and  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < +\infty$ .

Naturally, these geometric and probabilistic considerations also influence the corresponding rates, where the crucial assumption of uniform convexity and passage through uniform integrability result in a dependence of the rate on a modulus measuring the degree of uniform convexity of  $X$ , along with a similar modulus measuring the degree of uniform integrability of  $(\|Uy_n - y_n\|)$ . In particular, we show how such moduli can be explicitly constructed for particular spaces, and from natural quantitative integrability assumptions on the error terms, respectively.

To the best of our knowledge, all the respective rates for (sKM-T) and (sHM), as well as even the “qualitative” asymptotic regularity results, that is convergence alone even without considering the quantitative aspects, are completely novel. In terms of complexity and structure, they seem to match previously constructed general rates for their deterministic analogs (KM-T) and (HM) derived in [6, 7, 8, 12] and [13, 30], respectively, with the additional component of a modulus of uniform integrability featuring in the case of (sHM) as mentioned before. In the case of the stochastic Halpern iteration (sH), the only scheme already to have been studied, our convergence theorems are the best yet in that they both generalize and improve the asymptotic regularity results recently presented in [4].

**1.2.2. Fast rates of asymptotic regularity.** Beyond just presenting general and rather abstract convergence results and rates as outlined above, we also identify natural circumstances under which these rates are very fast, reaching up to linear speed in special cases. For the parameter

choices  $\alpha_n = 2/(n+2)$  and  $\beta_n = \beta \in (0, 1)$ , together with sufficiently fast decaying errors with  $\mathbb{E}[\|\xi_n\|], \mathbb{E}[\|\delta_n\|] \leq K/(n+2)^2$  for some constant  $K$ , we derive (see Theorem 4.4) linear rates for the discrete velocity of the general scheme (sHM) in the form of

$$\mathbb{E}[\|x_n - x_{n+1}\|] \leq \frac{K'}{n+2} \text{ and } \mathbb{P}(\exists i \geq n (\|x_i - x_{i+1}\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{K'}{n+2}$$

for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$ , and some suitable (explicitly definable) constant  $K'$ . Similar results also hold for the discrete velocity of  $(y_n)$  in mean and almost surely (see Theorem 4.6). In particular, we want to note that the above results, as well as all other linear rates, are actually given in the form of nonasymptotic guarantees, also in the case of the almost sure rates.

In the case of asymptotic regularity relative to the mappings, our results again become sensitive to the choice of the scheme: The special cases of (sH) and (sKM-T) exhibit linear nonasymptotic guarantees under the same assumptions on the parameters as detailed above. In the case of (sH) (see Theorem 4.8), we in particular have

$$\mathbb{E}[\|Tx_n - x_n\|] \leq \frac{K'}{n+2} \text{ and } \mathbb{P}(\exists i \geq n (\|Tx_i - x_i\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{K'}{n+2}$$

for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$  and some suitable (explicitly definable) constant  $K'$  similar to before. Effectively the same results hold for  $U$  in place of  $T$  in the case of (sKM-T) (see Theorem 4.9).

The results for the general scheme (sHM) again rely both on assumptions on the geometry of the space  $X$  and the uniform integrability of the sequence  $(\|Uy_n - y_n\|)$ . In the special case of a uniformly convex space of power type  $p \geq 2$ , our results in particular yield rates of order  $p$  so that we obtain quadratic rates in the case of inner product spaces, with order

$$\mathcal{O}(\varepsilon^{-2}\lambda^{-2}) \text{ or } \mathcal{O}(\varepsilon^{-2}\mu(\varepsilon/4)^{-2})$$

for  $\|Ux_n - x_n\|, \|Tx_n - x_n\| \rightarrow 0$  almost surely or in mean, respectively, where  $\mu$  is our modulus of uniform integrability for  $(\|Uy_n - y_n\|)$  (see Theorem 4.12 and Remark 4.13).

As with our generally constructed rates, our fast rates for the schemes (sHM) and (sKM-T) are to the best of our knowledge also novel. In regards to complexity, and seem to match the complexities observed in the deterministic case (we again refer to [13, 30] and [6, 7, 8, 12], respectively). For the scheme (sH), our results improve the fast rates presented in [4] by removing the respective logarithmic factors contained therein. In particular, by restricting our attention to the deterministic Halpern iteration (H), we reobtain the linear rates of asymptotic regularity previously derived in [32, 43] (albeit with different constants), which are known to be tight (see [32]). Indeed, our linear rates are derived by adapting the approach of [43]). Based on their inherent similarity, it can moreover be believed that the linear rates of asymptotic regularity observed for the deterministic Krasnoselskii-Mann iteration with Tikhonov regularization terms (KM-T) are also tight, in which case our corresponding results for the stochastic scheme (sKM-T) would be tight as well, although we are not aware of any results that prove this explicitly. The question for tight rates for the general Halpern-Mann scheme, both in its deterministic variant (HM) as well as in the stochastic version (sHM) presented here, remains an open problem.

**1.3. Future applications and developments.** Beyond our series of general theoretical results, one of the core motivations for this paper is the real applicability of those results in concrete areas, and we conclude with a section on how this might be achieved. First, we discuss how the requisite variance control can be managed in a practical way through the use of sampling techniques such as minibatching, and connected with this we show how our complexity results can be lifted to corresponding results on oracle complexity. We then outline several

concrete applications of our methods, sketching the particularly interesting case of reinforcement learning, where the novel schemes (sHM) and (sKM-T) can be instantiated in the style of Q-learning, as done recently for the stochastic Halpern iteration in [4]. A proper account of these applications will be provided in a forthcoming paper by the authors. Another crucial property of Halpern-type iterations not addressed in the present paper is the fact that they remain strongly convergent also in infinite dimensional spaces. These strong convergence results extend to the stochastic setting, and will be similarly addressed in forthcoming work which in particular relies on the asymptotic regularity results established here.

**1.4. This paper in connection with the proof mining program.** 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 [19] for a comprehensive overview of both theoretical as well as applied aspects of this program, along with the survey [21] 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 by Kohlenbach [20], to the extensions of these results to nonlinear context such as CAT(0)-spaces as in [22] (by a logical analysis of a corresponding convergence proof by Saejung [44]). They also include extensions of the Halpern iteration [45] for the modified Mann iteration introduced in [16] (and extended to nonlinear spaces in [10]) as well as the Krasnoselskii-Mann iteration with Tikhonov regularization terms and its extensions as in [6, 7, 8, 12] (with [6] 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 [13] 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 [36], together with applied results on the quantitative aspects of stochastic processes by the authors and Neri [37, 38, 39]. In particular, the present paper is one of the first applications of proof mining to stochastic optimization, and the very first to consider a concrete stochastic algorithm. 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 [13, 30]), but requires a substantial arsenal of new quantitative ideas for this stochastic setting. These include quantitative, stochastic variants of a crucial abstract lemma on recurrence inequalities due to Xu [49] (Lemma 2.4), including an adaptation of the “fast” variant due to Sabach and Shtern [43] (Lemma 4.2). Furthermore, proof mining also motivated the definition of a modulus of uniform integrability and its use as a suitable assumption to derive rates of asymptotic regularity in expectation in the general case of the iteration (sHM). We envisage that all of these tools will be relevant in subsequent applications of proof mining in stochastic optimization, just as their deterministic counterparts have been used repeatedly for proof mining in nonstochastic optimization.

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.

## 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  $\Omega \rightarrow \mathbb{R}$  as random variables, to measurable functions  $\Omega \rightarrow X$  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 a separable Banach space, 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 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  $a_n \rightarrow 0$  is a function  $\varphi : (0, \infty) \rightarrow \mathbb{N}$  such that

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

The immediate benefit of such a type of rate  $\varphi$  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  $X_n \rightarrow 0$  almost surely is a function  $\Phi : (0, \infty)^2 \rightarrow \mathbb{N}$  such that

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

We note that whenever  $\Phi$  is a rate of convergence for  $X_n \rightarrow 0$  almost surely, then for every  $\varepsilon > 0$ ,  $\Phi(\varepsilon, \cdot)$  is a rate of convergence for  $\mathbb{P}(\sup_{n \geq N} (X_n \geq \varepsilon)) \rightarrow 0$  as  $N \rightarrow \infty$  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  $\chi : (0, \infty) \rightarrow \mathbb{N}$  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  $\theta : \mathbb{N} \times (0, \infty) \rightarrow \mathbb{N}$  is a rate of divergence for that sum if

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

Naturally, any such modulus satisfies  $\theta(k, b) \geq k$  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 [49], often called Xu's lemma:

**Lemma 2.1** ([49]). *Suppose that  $(s_n), (c_n) \subseteq [0, \infty)$  as well as  $(a_n) \subseteq [0, 1]$  and  $(b_n) \subseteq \mathbb{R}$  satisfy*

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

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

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 [22, 29] (up to the way the errors and the moduli are phrased) therefore for brevity we omit the proof.

**Lemma 2.2** (essentially [22, 29]). *Suppose that  $(s_n), (c_n) \subseteq [0, \infty)$  and  $(a_n) \subseteq [0, 1]$  satisfy*

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

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

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

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

- (a)  $\sum_{i=0}^{\infty} \mathbb{E}[C_i] < \infty$  with rate  $\chi$ ,
- (b)  $\mathbb{E}[X_n] \rightarrow 0$  with rate  $\varphi$ .

*Then  $X_n \rightarrow 0$  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  $\mathbb{E}[\sum_{i=n}^{\infty} C_i] = \sum_{i=n}^{\infty} \mathbb{E}[C_i]$  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

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

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

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

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

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

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

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

for any  $N$ , and in particular for the  $N$  defined above.  $\square$



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*

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

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

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

*Then  $\mathbb{E}[X_n] \rightarrow 0$  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  $X_n \rightarrow 0$  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

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

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

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

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

### 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)  $\|y_{n+1} - y_n\| \leq (1 - \alpha_{n+1})(\|x_{n+1} - x_n\| + \|\xi_{n+1} - \xi_n\|) + |\alpha_{n+1} - \alpha_n| \cdot \|Tx_n + \xi_n - u\|,$
- (2)  $\|x_{n+2} - x_{n+1}\| \leq \|y_{n+1} - y_n\| + (1 - \beta_{n+1})\|\delta_{n+1} - \delta_n\| + |\beta_{n+1} - \beta_n| \cdot \|Uy_n + \delta_n - y_n\|.$

*Proof.* For (1) we observe that

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

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

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

where again we use nonexpansivity of the operator in the last step.  $\square$

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<sub>1</sub>) in [4] 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  $n \in \mathbb{N}$ :

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

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  $K_0 \geq \mathbb{E}[Y]$  for some  $K_0 \in \mathbb{N}^*$

and for all  $n \in \mathbb{N}$ :  $\|Tx_n - u\|, \|Uy_n - y_n\|, \|Uu - u\|, \|Uy_n - u\| \leq Y$  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  $\theta$ , 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 rates of convergence  $\chi_1 - \chi_4$  and upper bounds  $B_1 - B_4$ , respectively, and that  $\mathbb{E}[\|\xi_n\|] \leq E_0$  and  $\mathbb{E}[\|\delta_n\|] \leq D_0$  for all  $n \in \mathbb{N}$ . Then  $\mathbb{E}[\|x_{n+1} - x_n\|] \rightarrow 0$  with rate  $\varphi_{K,\theta,\chi}$  as well as  $\|x_{n+1} - x_n\| \rightarrow 0$  almost surely with rate  $\psi_{K,\theta,\chi}$  with  $\varphi, \psi$  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

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

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

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

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

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

and we can then show that

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

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

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  $\mathbb{E}[\|y_{n+1} - y_n\|] \rightarrow 0$  with rate*

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

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

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

where  $\varphi$  is a rate for  $\mathbb{E}[\|x_{n+1} - x_n\|] \rightarrow 0$  and  $\chi$  is as in Theorem 3.2.

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

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

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

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

and from that the rate for  $\mathbb{E}[\|y_{n+1} - y_n\|] \rightarrow 0$  immediately follows, noting that a rate of convergence  $\chi$  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  $n \geq \chi(\varepsilon)$ . For the rate of  $\|y_{n+1} - y_n\| \rightarrow 0$  almost surely, note that using both (1) and (2) of Lemma 3.1, we get that

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

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

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

and so that  $\chi$  from Theorem 3.2 is a rate of convergence for  $\sum_{n=0}^{\infty} \mathbb{E}[d_n] < \infty$  (noting that if  $\chi$  is a rate of convergence for  $\sum_{n=0}^{\infty} a_n < \infty$ , then  $\sum_{n=\chi(\varepsilon)}^{\infty} a_{n+1} = \sum_{n=\chi(\varepsilon)+1}^{\infty} a_n \leq \sum_{n=\chi(\varepsilon)}^{\infty} a_n < \varepsilon$  so that  $\chi$  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  $\|y_{n+1} - y_n\| \rightarrow 0$  almost surely.  $\square$

**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 a uniform convexity assumption on the space  $X$ , and in the case of asymptotic regularity in mean further relies on an essential use of the uniform integrability of  $(\|Uy_n - y_n\|)$  (which for example immediately derives from the assumption (Hyp'), as also discussed in detail later). 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  $(\|Uy_n - y_n\|)$ . As all proofs outside this geometric construction are rather routine, they are deferred to the appendix.

**Theorem 3.4.** *Assume that  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  with rate  $\varphi$ , and  $\mathbb{E}[\|\xi_n\|], \mathbb{E}[\|\delta_n\|], \alpha_n \rightarrow 0$  with rates  $\rho_1 - \rho_3$ , respectively. Assume further that  $\mathbb{E}[\|x_{n+1} - x_n\|] \rightarrow 0$  with a rate  $\varphi_0$ . Then*

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

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

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

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

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

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

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

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

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  $\|Uy_n - y_n\| \rightarrow 0$  almost surely with rate  $\psi$ , and  $\alpha_n \rightarrow 0$  with rate  $\rho$ . Further, assume that  $\|\xi_n\|, \|\delta_n\| \rightarrow 0$  almost surely with rates  $\phi_1, \phi_2$ , respectively, and that  $\|x_{n+1} - x_n\| \rightarrow 0$  almost surely with a rate  $\psi_0$ . Then*

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

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

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

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

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

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

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

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

*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  $\chi_1, \chi_2$ , respectively, then we can immediately derive rates  $\phi_1, \phi_2$  for  $\|\xi_n\|, \|\delta_n\| \rightarrow 0$  almost surely: Since  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] < \infty$  with rate  $\chi_1$ , we get

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

using Markov's inequality so that  $\phi_1(\lambda, \varepsilon) := \chi_1(\lambda\varepsilon)$  is a rate for  $\|\xi_n\| \rightarrow 0$  almost surely. Similarly for  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$  and  $\chi_2, \phi_2$ .

**3.2.1. Special cases of the Halpern iteration and the Krasnoselskii-Mann iteration with Tikhonov regularization terms.** In the special case of  $U := \text{Id}$  and  $\delta_n := 0$ , the iteration (sHM) collapses to the stochastic Halpern iteration (sH). We then have trivial rates for  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  and  $\|Uy_n - y_n\| \rightarrow 0$  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

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

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  $\rho_1, \rho_3$  as in Theorem 3.4. Note that this generalises known rates in this case [4, Theorem 3.3], which apply only to specific choices of the parameters. In a similar way, we get a new rate for  $\|Tx_n - x_n\| \rightarrow 0$  almost surely, though we do not spell it out here.

In the special case of  $T := \text{Id}$  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  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  and  $\|Uy_n - y_n\| \rightarrow 0$  almost surely, essentially following the approach of [45] (see also [6]).

**Lemma 3.7** (essentially [45]). *Let  $(x_n), (y_n)$  be the sequences generated by (sHM) where  $T = \text{Id}$  and  $\xi_n = 0$  for all  $n \in \mathbb{N}$ . Then the following recurrence relation holds everywhere on  $\Omega$  for all  $n \in \mathbb{N}$ :*

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

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  $T = \text{Id}$  and  $\xi_n = 0$  for all  $n \in \mathbb{N}$ . Also, let  $\Lambda > 0$  be such that  $\Lambda \leq \beta_n \leq 1 - \Lambda$  for all  $n \in \mathbb{N}$ . If  $\mathbb{E}[\|y_n - y_{n+1}\|] \rightarrow 0$  with rate  $\varphi$ ,  $\alpha_n \rightarrow 0$  with rate  $\rho$  and  $\mathbb{E}[\|\delta_n\|] \rightarrow 0$  with rate  $\chi$ , then  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  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  $\|y_n - y_{n+1}\| \rightarrow 0$  almost surely with rate  $\psi$ ,  $\alpha_n \rightarrow 0$  with rate  $\rho$  and  $\|\delta_n\| \rightarrow 0$  almost surely with rate  $\phi$ , then  $\|Uy_n - y_n\| \rightarrow 0$  almost surely with rate*

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

**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  $\text{Fix}T \cap \text{Fix}U \neq \emptyset$  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

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

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

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

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

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

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

It follows immediately by induction that

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

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

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

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

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  $(y_n)$ , which then in conjunction with Theorems 3.4 and 3.5 will allow us to derive all of the other regularity properties relative to mappings. For that, we initially establish  $\|Uy_n - y_n\| \rightarrow 0$  almost surely using geometric properties of the underlying space. Using that  $(\|Uy_n - y_n\|)$  is uniformly

integrable, as follows from (Hyp') and hence from Lemma 3.9 in the context of the present assumptions, we can then establish the asymptotic regularity relative to  $U$  of  $(y_n)$  in expectation. The quantitative result will then in particular depend on quantitative renderings of both the uniform convexity of the space and the uniform integrability of the sequence. We begin with the geometric assumption on the underlying normed space  $(X, \|\cdot\|)$ :

**Definition 3.10.** We say that  $(X, \|\cdot\|)$  is uniformly convex ([9]) if for any  $\varepsilon \in (0, 2]$ , there exists a  $\delta \in (0, 1]$  such that for all  $x, y \in \overline{B}_1(0)$ :

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

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

We will later discuss examples of such moduli of uniform convexity for particular (classes of) spaces (see in particular Lemma 4.10). 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  $\eta$  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  $\|x - y\| \geq \varepsilon \cdot r$ , then for all  $\lambda \in [0, 1]$ :*

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

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  $\eta$ . Let  $\|x_{n+1} - x_n\| \rightarrow 0$  almost surely with rate  $\Delta$ . Also, let  $\rho$  be a rate for  $\alpha_n \rightarrow 0$  and assume that  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$ ,  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$  with rates of convergence  $\chi_1, \chi_2$ , respectively. Lastly, let  $\Lambda > 0$  be such that  $\Lambda \leq \beta_n \leq 1 - \Lambda$ . Then  $\|Uy_n - y_n\| \rightarrow 0$  almost surely with rate*

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

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

*Proof.* Suppose for contradiction that

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

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

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

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

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

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

$$\|Uy_{n(\omega)}(\omega) - y_{n(\omega)}(\omega)\| \geq \varepsilon \text{ and } \|y_{n(\omega)}(\omega) - p\|, \|u(\omega) - p\| \leq K'.$$

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

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

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

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

Now, we further have

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

so that we can in particular derive

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

So, in the end we have

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

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

$$A_{\lambda, \varepsilon} \subseteq \{\exists n \geq \Gamma(\lambda, \varepsilon) (V_n \geq \hat{\varepsilon})\}.$$

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

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

a contradiction. □

*Remark 3.13.* Using a slightly different argument first devised in [18, 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 the above rate  $\Gamma$  holds even with  $\hat{\varepsilon} := \varepsilon \cdot \Lambda^2 \cdot \tilde{\eta}(\varepsilon/K')$ .

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

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



and this leads to

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

using that  $\tilde{\eta}(\varepsilon / K') \leq \tilde{\eta}(\varepsilon / \|y_{n_0}(\omega) - p\|)$  as  $\tilde{\eta}(\varepsilon / \|y_{n_0}(\omega) - p\|) \leq K'$  and since  $\tilde{\eta}$  is increasing. Then the proof continues 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  $\mu : (0, \infty) \rightarrow (0, \infty)$  such that

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

a modulus of absolute continuity for  $X$ .

By e.g. Lemma 13.1 in [47], such a modulus always exists for integrable  $X$ . However, it provides a quantitatively different rendering of that property than a simple upper bound on the mean and can in general not be derived from such a bound. We begin with a simple property of such moduli.

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

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

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

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

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

which is the claim.  $\square$

A modulus of absolute continuity now motivates the following new notion of a modulus of uniform integrability:

**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

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

We call a function  $\mu$  that witnesses such a  $\delta$  in terms of  $\varepsilon$  a modulus of uniform integrability for  $(X_n)$ .

Note  $\mu$  is a modulus of uniform integrability for  $(X_n)$  exactly when  $\mu$  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  $X_n \rightarrow 0$  almost surely with rate  $\varphi$  and such that  $\mu$  is modulus of uniform integrability for  $(X_n)$ . Then  $\mathbb{E}[X_n] \rightarrow 0$  with rate

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

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

We now obtain the following result on rates of asymptotic regularity in mean, derived from the previous Theorem 3.12 using a modulus of uniform integrability for  $(\|Uy_n - y_n\|)$ .

**Theorem 3.18.** *Let  $(X, \|\cdot\|)$  be uniformly convex with modulus  $\eta$ . Under the assumptions of Theorem 3.2, let  $\mathbb{E}[\|x_{n+1} - x_n\|] \rightarrow 0$  with rate  $\Delta$  from Theorem 3.2. Also, let  $\rho$  be a rate for  $\alpha_n \rightarrow 0$  and assume that  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$ ,  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$  with rates of convergence  $\chi_1$ ,  $\chi_2$ , respectively. Also, let  $\Lambda > 0$  be such that  $\Lambda \leq \beta_n \leq 1 - \Lambda$ . Lastly, let  $\mu$  be a modulus of uniform integrability for  $(\|Uy_n - y_n\|)$ . Then  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  with rate*

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

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

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

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

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

*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  $\Gamma$  holds even with  $\hat{\varepsilon}$  defined as  $\hat{\varepsilon} := \varepsilon/2 \cdot \Lambda^2 \cdot \tilde{\eta}(\varepsilon/2K')$ .

The following result collects the qualitative core of the above asymptotic regularity results for (sHM), void of any quantitative considerations, to illustrate the main assumptions more clearly:

*Corollary 3.20.* Let  $(X, \|\cdot\|)$  be uniformly convex and let  $(x_n), (y_n)$  be the sequences generated by (sHM) where  $\text{Fix}T \cap \text{Fix}U \neq \emptyset$ . Assume  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$ ,  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|] < \infty$  and  $\mathbb{E}[\|x_0 - p\|]$ ,  $\mathbb{E}[\|u - p\|] < \infty$  together with  $\sum_{n=0}^{\infty} \alpha_n = \infty$  and  $\alpha_n \rightarrow 0$  as well as

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

and  $\Lambda \leq \beta_n \leq 1 - \Lambda$  for some  $\Lambda > 0$ . Then  $\|Uy_n - y_n\| \rightarrow 0$  almost surely and in mean.

To be more precise, Corollary 3.20 follows from Theorems 3.12 and 3.18 in conjunction with Theorem 3.2. As emphasised already, the requirement in Theorem 3.18 that  $(\|Uy_n - y_n\|)$  comes equipped with a modulus of uniform integrability disappears in the qualitative result: Under the assumptions of this section, (Hyp') is automatically satisfied, and in particular  $(\|Uy_n - y_n\|)$  is dominated by some integrable random variable  $Y$  (explicitly definable by Lemma 3.9) and is thus automatically uniformly integrable. The modulus is only required in order to construct our rate of convergence.

Indeed, both Theorem 3.12 and 3.18 provide rather complex constructions for the corresponding rates of asymptotic regularity, featuring an interplay of many different moduli, in particular of the moduli of uniform convexity  $\eta$  and uniform integrability  $\mu$ . While concrete instantiations for the modulus  $\eta$  for special (classes of) spaces will be discussed later on (see Lemma 4.10)

in the context of standard assumptions on the errors and the parameters which result in particularly fast rates, here we illustrate how a corresponding modulus  $\mu$  can be derived in more concrete situations.

We note that under our general assumption (Hyp') whereby  $(\|Uy_n - y_n\|)$  is dominated by some  $Y$  with finite mean, a modulus of absolute continuity for  $Y$  (in the sense of Definition 3.14) is clearly also a modulus of uniform integrability for the sequence  $(\|Uy_n - y_n\|)$ , and our quantitative uniform integrability requirement is thus reduced to a quantitative continuity property of the bound  $Y$ . Here the bounds  $D, E$  for the summability assumption on the error terms do not suffice to derive such a modulus explicitly, as they do not allow for a construction of corresponding moduli of absolute continuity for the series, but provided these assumptions are extended to include such moduli of uniform integrability for  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|]$  and  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$ , then we can effectively construct our modulus  $\mu$  as follows:

**Lemma 3.21.** *Suppose that  $u$  and  $x_0$  are chosen to be constant, that  $K > 0$  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  $\mu_1$  and  $\mu_2$  respectively. Then a modulus of uniform integrability for  $(\|Uy_n - y_n\|)$  is given by*

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

*Proof.* By Lemma 3.9 we have  $\|Uy_n - y_n\| \leq Y$  pointwise everywhere for all  $n \in \mathbb{N}$ , where  $Y$  is defined as in Lemma 3.9, and so in particular we then have

$$\|Uy_n - y_n\| \leq 2 \left( 2K + \sum_{i=0}^{\infty} (\|\xi_n\| + \|\delta_n\|) \right).$$

for any  $n \in \mathbb{N}$ . Thus for any  $A \in \mathcal{F}$  and  $n \in \mathbb{N}$  it follow that

$$\mathbb{E}[\|Uy_n - y_n\| 1_A] \leq 4K\mathbb{P}(A) + 2 \left( \sum_{i=0}^{\infty} (\mathbb{E}[\|\xi_i\| 1_A] + \mathbb{E}[\|\delta_i\| 1_A]) \right)$$

and so the result follows by definition of  $\mu_1$  and  $\mu_2$ .  $\square$

While such moduli  $\mu_1, \mu_2$  always exist, the question of how these and related moduli of uniform integrability can actually be constructed in concrete situations still remains. First, we observe that in scenarios where  $(\|Uy_n - y_n\|)$ , or its  $L_1$ -bound  $Y$  or the series over the errors enjoy stronger properties, such as higher moment conditions, corresponding moduli can be given in a straightforward manner and moreover assume a very simple form.

For that, we first consider the following general result:

**Lemma 3.22.** *Let  $(X_n)$  be a sequence of random variables such that  $\sup_{n \in \mathbb{N}} \mathbb{E}[g(|X_n|)] < \infty$  for some measurable supercoercive  $g : [0, \infty) \rightarrow [0, \infty)$ , i.e.  $g(x)/x \rightarrow \infty$  as  $x \rightarrow \infty$ . Then  $(X_n)$  is uniformly integrable and  $\mu$  defined by*

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

*is a corresponding modulus in the sense of Definition 3.16, where  $\sup_{n \in \mathbb{N}} \mathbb{E}[g(|X_n|)] < K$  and  $\kappa : (0, \infty) \rightarrow (0, \infty)$  is a rate of divergence for  $g(x)/x \rightarrow \infty$ , that is*

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

*Proof.* Uniform integrability of  $(X_n)$  is a standard fact [41] (see also [17, Theorem 6.19]), and we simply need to verify the moduli. Fix  $\varepsilon > 0$ ,  $n \in \mathbb{N}$  and  $A \in \mathcal{F}$ . Note that for any  $a > 0$ , we have

$$\begin{aligned} \mathbb{E}[|X_n|1_A] &\leq \mathbb{E}[|X_n|1_{A \cap (|X_n| \leq \kappa(a))}] + \mathbb{E}[|X_n|1_{A \cap (|X_n| > \kappa(a))}] \\ &\leq \kappa(a)\mathbb{P}(A) + \mathbb{E}[|X_n|1_{A \cap (|X_n| > \kappa(a))}]. \end{aligned}$$

Using now that  $x > \kappa(a)$  implies  $g(x)/a \geq x$ , we have

$$\mathbb{E}[|X_n|1_A] \leq \kappa(a)\mathbb{P}(A) + a^{-1}\mathbb{E}[g(|X_n|)] < \kappa(a)\mathbb{P}(A) + a^{-1}K.$$

Setting  $a := 2K/\varepsilon$  yields  $\mathbb{E}[|X_n|1_A] \leq \kappa(2K/\varepsilon)\mathbb{P}(A) + \varepsilon/2$ , so that if  $\mathbb{P}(A) \leq \mu(\varepsilon)$  one has  $\mathbb{E}[|X_n|1_A] \leq \varepsilon$ .  $\square$

The above lemma is a quantitative variant of the fundamental de la Vallée-Poussin theorem [41] (see again also [17, Theorem 6.19]), and the existence of such a supercoercive function in fact characterizes uniformly integrable sequences of random variables. In particular, an immediate consequence is the following result for higher moment conditions:

**Lemma 3.23.** *Let  $(X_n)$  be a sequence of random variables such that  $\sup_{n \in \mathbb{N}} \mathbb{E}[|X_n|^p] < K$  for some  $K > 0$  and  $p > 1$ . Then  $(X_n)$  is uniformly integrable and  $\mu$  defined by*

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

*is a corresponding modulus in the sense of Definition 3.16.*

Lemmas 3.22 and 3.23 can be applied directly to  $(\|Uy_n - y_n\|)$ , and of course in the special case that  $X_n := Y$  become simpler results on absolute continuity that can be instantiated to produce moduli of continuity on  $Y$ , and thus a modulus of integrability for  $(\|Uy_n - y_n\|)$ . However, when reducing the problem to continuity properties of  $\sum_{i=0}^{\infty} \|\xi_i\|$  and  $\sum_{i=0}^{\infty} \|\delta_i\|$ , as possible under the assumptions of this section as shown in Lemma 3.21, we conjecture that stronger assumptions such as those of Lemma 3.23 are not even required for the quantitative result, and that for concrete instantiations of the error terms (via e.g. minibatching as discussed in Section 5.1), moduli of integrability for the sums  $\sum_{i=0}^{\infty} \|\xi_i\|$  and  $\sum_{i=0}^{\infty} \|\delta_i\|$  can be calculated explicitly, exploiting the fact that we have concrete knowledge of the distribution of the errors. However, we do not give further details here.

#### 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 as 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 [43], first utilised in the context of proof mining in [6]. 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 [4]. However, as the proof is still nothing more than a careful implementation of the arguments given in [43], we defer it to the appendix.

**Lemma 4.1** (essentially [43]). *Suppose that  $(s_n), (c_n)$  are sequences of nonnegative real numbers satisfying*

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

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

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

*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  $c_n \leq (\alpha_n - \alpha_{n+1})L$  for some  $(\alpha_n) \subseteq [0, 1]$  and  $L > 0$  we have*

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

*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  $s_0 \leq L$ , then  $s_n \leq 2L/(n+2)$  for all  $n \in \mathbb{N}$ .*

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*

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

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

$$\mathbb{E}[X_n] \leq \frac{2L}{n+2} \text{ and } \mathbb{P}(\exists i \geq n (X_i \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L}{n+2}$$

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

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

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

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

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

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

$$\mathbb{P}(\exists n \geq N (X_n \geq \varepsilon)) \leq \mathbb{P}(\exists n \geq N (U_n \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L}{N+2}$$

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  $\mathbb{E}[X_n] \rightarrow 0$  and  $X_n \rightarrow 0$  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

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

**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  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  and  $\mathbb{E}[\|\delta_n\|] \leq K_2/(n+2)^2$ . Then*

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

for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$ , where  $L = 2K_0 + 2K_1 + 2K_2$  in both cases.

*Proof.* As in the proof of Theorem 3.2, we have  $X_{n+1} \leq (1 - \alpha_{n+1})X_n + C_n$  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

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

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

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

We conclude by observing that

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

and similarly for  $(\delta_n)$  and  $K_2$ . The rates then follow from Lemma 4.2.  $\square$

*Remark 4.5.* Before moving to the other asymptotic regularity results, we just briefly note that the asymptotic condition  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  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

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

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

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

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  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  and  $\mathbb{E}[\|\delta_n\|] \leq K_2/(n+2)^2$ . Then*

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

*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*

$$\mathbb{P}(\exists i \geq n (\|y_i - y_{i+1}\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L}{n+2}$$

*for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$ , with a suitable  $L$  constructed similarly.*

The proof is rather routine and hence deferred to the appendix.

*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.

We now discuss fast rates for the special cases (sH) and (sKM-T). As they are also routine, all proofs in the present section are deferred to the appendix.

In the special case of the stochastic Halpern iteration, which we reobtain (as discussed before) by setting  $U := \text{Id}$  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  $U := \text{Id}$  and  $\delta_n := 0$ . Assume (Hyp) with constant  $K_0$ . Also, assume that  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$ . Then*

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

*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*

$$\mathbb{P}(\exists i \geq n (\|Tx_i - x_i\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L}{n+2}$$

*for all  $n \in \mathbb{N}$  and  $\varepsilon > 0$ , with a suitable  $L$  constructed similarly.*

Theorem 4.8 is closely related to [4, 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  $T := \text{Id}$  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  $T := \text{Id}$  and  $\xi_n := 0$ . Assume (Hyp) with constant  $K_0$ . Also, assume that  $\mathbb{E}[\|\delta_n\|] \leq K_2/(n+2)^2$ . Lastly, let  $B \geq 1/(1-\beta)$ . Then*

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

*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*

$$\mathbb{P}(\exists i \geq n (\|Ux_i - x_i\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L}{n+2}$$

*for all  $n \in \mathbb{N}^*$  and  $\varepsilon > 0$ , with a suitable  $L$  constructed similarly.*

#### 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  $\eta$  for the underlying space, we here focus on the case where  $\eta$  is of *power type*  $p$  for  $p \geq 2$ , i.e. there exists a constant  $C$  such that  $\eta(\varepsilon) = C\varepsilon^p$ . Crucially, this is the case for  $L^p$  spaces for general  $p \in (1, \infty)$ , and for arbitrary inner product spaces:

**Lemma 4.10** (essentially [9], see also [23, 33]). *If  $X$  is an inner product space, then  $X$  is uniformly convex with a corresponding modulus  $\eta(\varepsilon) = \varepsilon^2/8$ . Further, if  $X = L^p$  for  $p > 1$ , then  $X$  is uniformly convex with a corresponding modulus*

$$\eta(\varepsilon) = \begin{cases} \frac{p-1}{8}\varepsilon^2 & \text{if } 1 < p < 2, \\ \frac{1}{p^{2p}}\varepsilon^p & \text{if } 2 \leq p < \infty. \end{cases}$$

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  $\eta$  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  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  and  $\mathbb{E}[\|\delta_n\|] \leq K_2/(n+2)^2$ . Lastly, let  $\Lambda > 0$  be such that  $\Lambda \leq \beta \leq 1 - \Lambda$ . Then  $\|Uy_n - y_n\| \rightarrow 0$  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  $K_2$ . Given a modulus  $\mu$  of uniform integrability for  $(\|Uy_n - y_n\|)$ , we further get  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  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  $\eta$  of power type  $p \geq 2$ , 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  $\|Uy_n - y_n\| \rightarrow 0$  almost surely follows by instantiating the rate given in Theorem 3.12 with the following moduli: With the above  $\tilde{\eta}$ , we have  $\hat{\varepsilon} := C\Lambda^2\varepsilon^p\lambda^{p-1}/(3K)^{p-1}$  and  $K' := 3K/\lambda$ . Using Theorem 4.4, we have  $\mathbb{P}(\exists i \geq n (\|x_i - x_{i+1}\| \geq \varepsilon)) \leq \frac{1}{\varepsilon} \frac{4L_0}{n+2}$  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  $\|x_n - x_{n+1}\| \rightarrow 0$  almost surely. As  $\alpha = 2/(n+2)$ , we further have rather immediately that  $\rho(\varepsilon) = \lceil 2/\varepsilon \rceil$  is a corresponding rate for  $\alpha_n \rightarrow 0$ . Lastly, using the assumptions on  $\|\xi_n\|$  and  $\|\delta_n\|$ , note that as in Remark 4.5 we have  $\sum_{n=N}^{\infty} \mathbb{E}[\|\xi_n\|] \leq K_1/(N+1)$  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

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

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



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  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  and  $\mathbb{E}[\|\delta_n\|] \leq K_2/(n+2)^2$ . Lastly, let  $\Lambda > 0$  be such that  $\Lambda \leq \beta \leq 1 - \Lambda$ . Then  $\|Ux_n - x_n\| \rightarrow 0$  and  $\|Tx_n - x_n\| \rightarrow 0$  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. Given a modulus  $\mu$  of uniform integrability for  $(\|Uy_n - y_n\|)$ , we further get  $\mathbb{E}[\|Ux_n - x_n\|] \rightarrow 0$  and  $\mathbb{E}[\|Tx_n - x_n\|] \rightarrow 0$  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  $\|Uy_n - y_n\| \rightarrow 0$  and  $\mathbb{E}[\|Uy_n - y_n\|] \rightarrow 0$  dominate.  $\square$

*Remark 4.13.* Compared to the previous nonasymptotic guarantees for the special cases of (sH) and (sKM-T), the fast rates given for the general schema (sHM) given above might look slightly complex at first sight. However, note that the rates from Theorem 4.12 can be simply bounded as

$$\Phi_1(\lambda, \varepsilon), \Phi_2(\lambda, \varepsilon) \leq \frac{R}{\varepsilon^2 \lambda^2} + 1 \in \mathcal{O}(\varepsilon^{-2} \lambda^{-2})$$

and

$$\varphi_1(\varepsilon), \varphi_2(\varepsilon) \leq \frac{R}{\varepsilon^2 \mu(\varepsilon/4)^2} + 1 \in \mathcal{O}(\varepsilon^{-2} \mu(\varepsilon/4)^{-2}),$$

where, in each case,  $R$  is a suitably large constant and  $\mu$  is still a modulus of uniform integrability for  $(\|Uy_n - y_n\|)$ .

## 5. AN OUTLOOK ONTO APPLICATIONS

Our stochastic Halpern-Mann scheme (sHM) and the subsequent convergence analysis have been presented in a completely abstract way. While will consider this level of abstraction to be a virtue, in that it potentially encompasses many different scenarios and shows the interrelations of the different quantitative ingredients in the final rates, we nevertheless conclude with an informal discussion on how various aspects of our analysis can be both extended to encompass alternative notions of complexity and interpreted in a concrete way. We however emphasise that a more detailed study of the various applications of both (sKM-T) and (sHM) will be provided in a forthcoming paper, so that our comments remain at the level of an extended sketch.

**5.1. Oracle complexity and managing variance.** Recall that the intuition behind our abstract stochastic scheme is that  $\xi_n$  and  $\delta_n$  arise by evaluating stochastic oracles  $\tilde{T}$  and  $\tilde{U}$  for nonexpansive mappings  $T$  and  $U$  respectively. Though our primary focus has been on establishing direct convergence rates for our schemes, in concrete applications it is the resulting *oracle complexity* that may act as a more reliable estimate of the actual cost of running the algorithm. To address this at a level of generality in line with our overall approach, we propose an abstract characterisation of oracle complexity: We first introduce two sequences of natural

numbers  $(\#\xi_n)$  and  $(\#\delta_n)$ , with the intuition that  $\#\xi_n$  denotes the number of calls made to the stochastic oracle  $\tilde{T}$  in computing the approximation to  $Tx_n$  (with output  $Tx_n + \xi_n$ ), and  $\#\delta_n$  the number of calls to  $\tilde{U}$  in computing the approximation to  $Uy_n$ . With this intuition in mind, we say that  $\psi : \mathbb{N} \rightarrow \mathbb{N}$  is a bound on the oracle growth of the general scheme (sHM) if

$$\forall N \in \mathbb{N} \left( \#\xi_N + \sum_{n=0}^{N-1} (\#\xi_n + \#\delta_n) \leq \psi(N) \right),$$

so that the total number of oracle calls needed to compute  $x_N$  and  $y_N$  is bounded by  $\psi(N)$ . Note that the additional term  $\#\xi_N$  occurs as we also calculate  $y_N$ , which requires us to evaluate  $Tx_N + \xi_N$  as per the schema (sHM).

Then, for example, if

$$\mathbb{E}[\|Tx_n - x_n\|] \rightarrow 0 \text{ with rate } \varphi,$$

it follows that for any  $\varepsilon > 0$ , we can compute an approximant  $x_N$  such that

$$\mathbb{E}[\|Tx_N - x_N\|] < \varepsilon \text{ using at most } (\psi \circ \varphi)(\varepsilon) \text{ queries to the stochastic oracles.}$$

Naturally,  $N = \varphi(\varepsilon)$  suffices here, and this  $N$  further has the property that  $\mathbb{E}[\|Tx_n - x_n\|] < \varepsilon$  for all  $n \geq N$ . Indeed, the bounding function  $\psi$  can be used in a similar way to convert any of our convergence results to a corresponding characterisation of the overall oracle complexity in a completely general manner, which can then be appropriately instantiated in specific applications. This in particular applies to our almost sure convergence results, but here we focus solely on convergence in mean.

In practice, there are a number of techniques that would allow us to approximate the mappings  $T, U$  while at the same time achieving the necessary variance bounds needed to guarantee convergence within our framework. Minibatching is an obvious example: Here we assume that we have access to  $T$  and  $U$  via a pair of stochastic oracles  $\tilde{T}, \tilde{U} : X \times \Theta \rightarrow X$ , defined over some suitable measure space  $\Theta$ . These oracles then give rise to a concrete minibatched Halpern-Mann iteration via

$$\text{(sHM-mini)} \quad \begin{cases} y_n := (1 - \alpha_n) \frac{1}{k_n} \sum_{j=1}^{k_n} \tilde{T}(x_n, \zeta_{n,j}) + \alpha_n u, \\ x_{n+1} := (1 - \beta_n) \frac{1}{l_n} \sum_{j=1}^{l_n} \tilde{U}(y_n, \zeta'_{n,j}) + \beta_n y_n, \end{cases}$$

where  $(k_n)$  and  $(l_n)$  are sequences of batchsizes, and for each  $n \in \mathbb{N}$ ,  $\zeta_{n,1}, \dots, \zeta_{n,k_n}$  are independent samples drawn from some distribution  $D_n$  over  $\Theta$ , and  $\zeta'_{n,1}, \dots, \zeta'_{n,l_n}$  independent samples drawn from some distribution  $D'_n$ . We can recognize this as an instance of (sHM) by defining the corresponding errors  $\xi_n := \frac{1}{k_n} \sum_{j=1}^{k_n} \tilde{T}(x_n, \zeta_{n,j}) - Tx_n$ , and analogously for  $\delta_n$ . In particular, the resulting number of oracle queries are in this case simply given by the bath sizes, i.e. we have  $\#\xi_n := k_n$  and  $\#\delta_n := l_n$ .

Now, suppose that we impose some standard assumptions on our oracle  $\tilde{T}$ :<sup>1</sup>

- (1)  $\tilde{T}$  is unbiased, i.e.  $\mathbb{E}[\tilde{T}(x_n, \zeta_{n,1})] = Tx_n$  for all  $n \in \mathbb{N}$ ,
- (2)  $\tilde{T}$  has controlled variance w.r.t.  $x_n$ , i.e.  $\mathbb{E}[\|\tilde{T}(x_n, \zeta_{n,1}) - Tx_n\|^2] \leq \sigma$  for all  $n \in \mathbb{N}$  and some constant  $\sigma$ .

Then whenever  $X$  is a separable Hilbert space it immediately follows (using Jensen's inequality and sample independence) that  $\mathbb{E}[\|\xi_n\|^2] \leq \sigma^2/k_n$ . Further, if  $X = (\mathbb{R}^d, \|\cdot\|)$  for some (not

---

<sup>1</sup>These variance assumptions on the oracles are indeed widely used, featuring in a number of recent works on stochastic Halpern schemes (see in particular [4, 5]).

necessarily Euclidean) norm  $\|\cdot\|$ , we have  $\mathbb{E}[\|\xi_n\|]^2 \leq c^2 \sigma^2 / k_n$  where  $c > 0$  is such that  $\|x\| \leq c \|x\|_E$  for all  $x \in \mathbb{R}^d$  (writing  $\|x\|_E$  for the Euclidean norm on  $\mathbb{R}^d$ ). In that case, we then obtain

$$\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|] \leq \sum_{n=0}^{\infty} \frac{c\sigma}{\sqrt{k_n}},$$

for some  $c > 0$ , and so the variance bounds on  $(\|\xi_n\|)$  that we require for convergence of (sHM) can be ensured by choosing the batchsizes  $(k_n)$  appropriately. Assuming analogous properties of the oracle  $\tilde{U}$ , the corresponding variance bounds for  $(\|\delta_n\|)$  are achieved similarly through appropriate choices for  $(l_n)$ . We now discuss one such choice in the following example for the instantiation (sH):

*Example 5.1.* Consider the special case of the stochastic Halpern iteration (sH), obtained from (sHM) by setting  $U := \text{Id}$  and  $\delta_n := 0$ . The corresponding minibatched scheme (sHM-mini) in that case to

$$(sH\text{-mini}) \quad x_{n+1} := (1 - \alpha_n) \frac{1}{k_n} \sum_{j=1}^{k_n} \tilde{T}(x_n, \zeta_{n,j}) + \alpha_n u,$$

as previously considered in [4]. If  $X$  is finite dimensional, then we have  $\mathbb{E}[\|\xi_n\|] \leq c\sigma/\sqrt{k_n}$  for some constant  $c > 0$  as outlined above. Choosing  $k_n = (n+1)^4$  hence yields  $\mathbb{E}[\|\xi_n\|] \leq K_1/(n+2)^2$  for  $K_1 := 4c\sigma$ . Under the additional conditions of Theorem 4.8 (recalling that (Hyp) is satisfied whenever we assume  $\text{Fix}T \neq \emptyset$ ) we would have

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

for  $L$  as defined in Theorem 4.8, corresponding to a rate of convergence  $\varphi(\varepsilon) := \varepsilon/2L$ . For this choice of  $(k_n)$ , a bound on the oracle growth is given by  $\psi(N) = N^5$ , and therefore we can compute an  $x_N$  such that

$$\mathbb{E}[\|Tx_N - x_N\|] < \varepsilon \text{ using at most } (\varepsilon/2L)^5 \text{ queries to the stochastic oracles.}$$

In this way we obtain a version of [4, Corollary 3.5] for linear rates *without* logarithmic factors, and this represents just an extremely simple case within our overall framework. Such complexity results extend to the instantiation (sKM-T) and a corresponding minibatched variant as will be discussed in detail in a forthcoming paper.

We conjecture that other methods of controlling variance beyond minibatching could also be expressed by instantiating  $\xi_n, \delta_n$  in a suitable way, which would come with alternative characterisations of oracle complexity. We anticipate that, in general, obtaining good bounds on the oracle complexity for instances of (sHM) will involve a tension between variance control and oracle growth, which in the simple case of minibatching corresponds, as detailed above, to a payoff between

$$\text{convergence speed of } \sigma/\sqrt{k_n} \rightarrow 0 \text{ and growth of } \sum_{n=0}^N k_n.$$

Because our explicit rates of convergence make completely precise how improved control on the variance (through e.g. faster convergence rates for  $\sum_{n=0}^{\infty} \mathbb{E}[\|\xi_n\|]$  and  $\sum_{n=0}^{\infty} \mathbb{E}[\|\delta_n\|]$ ) leads to improved rates of convergence for the algorithm itself, we anticipate that in any given model that allows us to explicitly describe  $\psi$  in terms of  $\mathbb{E}[\|\xi_n\|]$  and  $\mathbb{E}[\|\delta_n\|]$ , our abstract quantitative results would not only provide us with a general bound on the associated oracle complexity in terms of parameters representing variance control, but might even provide insights into specific choices of parameters that optimize oracle complexity.

**5.2. Applications of (sHM).** Our new stochastic iteration scheme (sHM) (along with the associated convergence results) immediately leads to new algorithms which solve problems that can be formulated in terms of fixed points of nonexpansive mappings. Two immediate examples here are monotone inclusion problems (where we could potentially generalise the results of [5] which are based on a stochastic Halpern scheme), and splitting methods, where this time our novel stochastic Krasnoselskii-Mann iteration with Tikhonov regularization terms is a direct generalisation of the nonstochastic method utilised for this purpose in [1], and so could be used to, for instance, compute zeroes of sums of maximally monotone operators that can only be accessed in a noisy way. Perhaps the most interesting application of our scheme lies in model-free reinforcement learning, where algorithms such as *Q-learning* [46] are naturally formulated as noisy methods for computing fixpoints. Only very recently has the special case (sH) of the scheme (sHM) corresponding to Halpern’s iteration been instantiated as a form of *Q-learning* [4], and we in that way also perceive our general method as an expanded class of learning algorithms.

To illustrate this latter application in a little more detail, suppose that  $(\mathcal{S}, \mathcal{A}, r, p)$  forms a Markov decision process (MDP) over some finite set of states  $\mathcal{S}$  and actions  $\mathcal{A}$ , 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  $t$  (see e.g. [42] for a standard reference). Instantiating not (sH) but (sKM-T) in the style of *Q-learning* results in the method

$$(KM-T-Q) \quad Q_{n+1}(s, a) := (1 - \beta_n) (U(\gamma_n Q_n)(s, a) + \delta_n(s, a)) + \beta_n (\gamma_n Q_n(s, a)),$$

where  $U$  is some suitable nonexpansive operator on  $\mathbb{R}^{\mathcal{S} \times \mathcal{A}}$  with respect to

$$\|Q\|_\infty := \max_{s \in \mathcal{S}} \max_{a \in \mathcal{A}} Q(s, a)$$

that captures the underlying Bellman equation  $UQ = Q$ , which in the case of discounted *Q-learning* would be given by

$$UQ(s, a) := r(s, a) + c \sum_{t \in \mathcal{S}} p(s, a, t) \max_{b \in \mathcal{A}} Q(t, b)$$

for some discount factor  $c \in [0, 1)$  (so that  $U$  in this case is even a strict contraction). The method (KM-T-Q), which can be perceived as an extension of ordinary *Q-learning* with Tikhonov regularization terms, is to the best of our knowledge already novel, even for the discounted case, though we consider it primarily of interest in the cases where  $U$  is properly nonexpansive. Notably, we claim that under suitable assumptions it could be used to compute optimal policies for MDPs in the *averaged reward* setting, as will be explored in a forthcoming work, just as Halpern’s scheme has been recently utilised to this end in [4] (under a minibatch strategy). The key aspect of those variants in that context, already motivating [4], 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, and the convergence results presented here would allow us to produce similar guarantees, which could then be translated to sample complexity bounds by introducing oracle complexity terms as discussed in Section 5.1. Concrete instantiations of the noise terms via minibatching are also possible in the context of reinforcement learning, in that they allow for sufficient variance control through suitable choices of batchsizes (see [4]), however the precise details are quite technical and we do not discuss them further here. In particular, this requires a more subtle version of variance control than that outlined in the previous section, where the associated oracle no longer has uniformly bounded variance in the sense of (2).

Ultimately, we envisage a broader use of our full framework in the context of reinforcement learning. Here our level of generality would allow us to consider *Q-learning* on a more abstract

level, where we could take as a starting point the generalised model of [34] and explore both different iterative methods and different forms of variance reduction, using the results given here to formulate abstract convergence theorems for learning algorithms. In particular we could consider *alternating*  $Q$ -learning algorithms based our main scheme (sHM), for example

$$\begin{aligned} \text{(sHM-Q)} \quad Q'_n(s, a) &:= (1 - \alpha_n) (TQ_n(s, a) + \xi_n(s, a)) + \alpha_n \hat{Q}(s, a) \\ Q_{n+1}(s, a) &:= (1 - \beta_n) (UQ'_n(s, a) + \delta_n(s, a)) + \beta_n Q'_n(s, a) \end{aligned}$$

where  $(\mathcal{S}, \mathcal{A}, r', p')$  is a second MDP over the same states and actions giving rise to the respective Bellman-type operator  $T$ . This iteration thus combines the traditional  $Q$ -learning procedure with a Halpern variant. This method bears a passing resemblance to double  $Q$ -learning [15], which incorporates a double estimator to reduce bias in ordinary  $Q$ -learning. Moreover, our scheme converges to a simultaneous fixpoint of two underlying nonexpansive operators, and thus might be relevant in situations where we are required to compute optimal policies concurrently across distinct environments. However, we leave an exploration of the potential merits of (sHM-Q) and similar algorithms to future work. In particular, a proper study of scenarios in which they could be exploited, together with an assessment of their performance against state-of-the-art reinforcement learning algorithms, would require substantial empirical work that is beyond the scope of the present paper.

## REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] X. Cai, C. Song, C. Guzmán, and J. Diakonikolas. Stochastic Halpern iteration with variance reduction for stochastic monotone inclusions. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Proceedings of NIPS'22*, pages 24766–24779. Curran Associates Inc., Red Hook, NY, USA, 2022.
- [6] H. Cheval, U. Kohlenbach, and L. Leuştean. On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications*, 197:233–251, 2023.
- [7] H. Cheval and L. Leuştean. Quadratic rates of asymptotic regularity for the Tikhonov-Mann iteration. *Optimization Methods & Software*, 37(6):2225–2240, 2022.
- [8] H. Cheval and L. Leuştean. Linear rates of asymptotic regularity for Halpern-type iterations. *Mathematics of Computation*, 94(353):1323–1333, 2025.
- [9] J.A. Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.
- [10] 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.
- [11] J. Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In J. Abernethy and S. Agarwal, editors, *Proceedings of COLT'20*, volume 125 of *Proceedings of Machine Learning Research*, pages 1428–1451, 2020.
- [12] B. Dinis and P. Pinto. On the convergence of algorithms with Tikhonov regularization terms. *Optimization Letters*, 15(4):1263–1276, 2021.
- [13] 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.
- [14] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73(6):957–961, 1967.

- [15] H.v. Hasselt. Double Q-learning. In J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Proceedings of NIPS'10*, volume 2, pages 2613–2621. Curran Associates Inc., Red Hook, NY, USA, 2010.
- [16] T.-H. Kim and H.-K. Xu. Strong convergence of modified Mann iterations. *Nonlinear Analysis. Theory, Methods & Applications*, 61(1–2):51–60, 2005.
- [17] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer Cham, 3rd edition, 2020.
- [18] U. Kohlenbach. Uniform asymptotic regularity for Mann iterates. *Journal of Mathematical Analysis and Applications*, 279(2):531–544, 2003.
- [19] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
- [20] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226(3):2764–2795, 2011.
- [21] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proceedings of ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.
- [22] U. Kohlenbach and L. Leuştean. Effective metastability of Halpern iterates in  $CAT(0)$  spaces. *Advances in Mathematics*, 231:2526–2556, 2012.
- [23] U. Kohlenbach and L. Leuştean. On the computational content of convergence proofs via Banach limits. *Philosophical Transactions of the Royal Society A*, 370(1971):3449–3463, 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, Providence, RI, USA, 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] L. Liang, Z. Xu, K.-C. Toh, and J.-J. Zhu. An Inexact Halpern Iteration with Application to Distributionally Robust Optimization, 2025. 58, 41pp.
- [32] F. Lieder. On the convergence rate of the Halpern-iteration. *Optimization Letters*, 15(2):405–418, 2021.
- [33] J. Lindenstrauss and L. Tzafriri. *Classical Banach Spaces II*, volume 97 of *Ergebnisse der Mathematik in ihrer Grenzgebiete*. Springer-Verlag Berlin Heidelberg, 1979.
- [34] M. Littman and C. Szepesvári. A generalized reinforcement-learning model: convergence and applications. In L. Saitta, editor, *Proceedings of ICML'96*, pages 310–318. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996.
- [35] W.R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510, 1953.
- [36] M. Neri and N. Pischke. Proof mining and probability theory. *Forum of Mathematics, Sigma*, 13, 2025. e187, 47pp.
- [37] 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.
- [38] M. Neri and T. Powell. A quantitative Robbins-Siegmund theorem. *Annals of Applied Probability*, 2025. to appear, arXiv e-print, math.OC, 2410.15986.
- [39] M. Neri and T. Powell. On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales. *Transactions of the American Mathematical Society, Series B*, 12:974–1019, 2025.
- [40] 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.
- [41] C. De La Vallée Poussin. Sur L'Intégrale de Lebesgue. *Transactions of the American Mathematical Society*, 16(4):435–501, 1915.

- [42] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, Hoboken, NJ, USA, 2014.
- [43] S. Sabach and S. Shtern. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27:640–660, 2017.
- [44] S. Saejung. Halpern iterations in CAT(0) spaces. *Fixed Point Theory and Applications*, 2010. Article ID 471781, 13pp.
- [45] 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.
- [46] C. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge, 1989.
- [47] D. Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.
- [48] R. Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58(5):486–491, 1992.
- [49] H.-K. Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66:240–256, 2002.
- [50] 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.
- [51] 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 M. Meila and T. Zhang, editors, *Proceedings of ICML’21*, volume 139 of *Proceedings of Machine Learning Research*, pages 12098–12109, 2021.

## APPENDIX

*Proof of Theorem 3.4.* Related to (a) – (d), we can immediately establish the following inequalities:

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

By taking the expectation, the rates immediately follow. □

*Proof of Theorem 3.5.* 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  $\alpha_n \|Tx_n - u\| \rightarrow 0$  almost surely: Using Markov’s inequality we have

$$\mathbb{P}(\exists n (\|Tx_n - u\| \geq K_0/\lambda)) \leq \mathbb{P}(Y \geq K_0/\lambda) \leq \lambda$$

for  $Y$  as in (Hyp’). Now noting that if  $\omega$  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  $n \geq \rho(\varepsilon\lambda/K_0)$ , and therefore we have

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

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

*Proof of Lemma 3.7.* Given  $n \in \mathbb{N}$ , we have

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

pointwise everywhere.  $\square$

*Proof of Theorem 3.8.* By the Lemma 3.7 above, we have

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

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

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

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

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

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

*Proof of Lemma 4.1.* 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

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

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

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

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

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

which completes the proof.  $\square$



*Proof of Theorem 4.6.* Using Lemma 3.1, (1), we have

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

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

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

From Theorem 4.4, we get  $\mathbb{E}[\|x_n - x_{n+1}\|] \leq 2L/(n+2)$  and similar as in the proof thereof, we have  $\mathbb{E}[\|\xi_{n+1} - \xi_n\|] \leq 2K_1/(n+2)^2 \leq 2K_1/(n+2)$ . Combined with the definition of  $\alpha_n$ , 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

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

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

*Proof of Theorem 4.8.* Using the inequalities listed in the proof of Theorem 3.4, we obtain

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

for all  $n \in \mathbb{N}$  pointwise everywhere, in this special case where  $U = \text{Id}$  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  $K_2 = 0$ ).  $\square$

*Proof of Theorem 4.9.* Using Lemma 3.7, we have

$$\|Uy_n - y_n\| \leq B(2\|y_{n-1} - y_n\| + \alpha_n \|Uy_n - u\| + \|\delta_{n-1}\|)$$

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

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

for  $n \geq 1$  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

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

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.  $\square$