PROOF MINING AND HIGH-LEVEL PROOF THEORETIC REASONING: A CASE STUDY ON GREEDY APPROXIMATION SCHEMES

THOMAS POWELL

Abstract. We carry out a logical analysis of a convergence proof for greedy approximation schemes in uniformly smooth Banach spaces. Though the proof is by contradiction, we are able to extract computable rates of convergence that depend on the corresponding modulus of uniform smoothness for the space. While our quantitative results represent a first proof theoretic study of greedy approximation schemes, we use this case study more generally as an opportunity to make explicit some of the high-level proof theoretic reasoning that enables us to transform a nonconstructive convergence proof to one where computable convergence rates are apparent, representing the proof using a series of formal derivations that are designed to capture core mathematical reasoning, as opposed to low-level proof theoretic bureaucracy. In this way we exemplify an approach to representing the process of program extraction that might, in particular, inform efforts to formalise proof mining in proof assistants.

§1. Introduction. Applied proof theory, also known as *proof mining*, is a subfield of logic that uses ideas and techniques from proof theory to produce new theorems in different areas of mainstream mathematics and computer science. This usually proceeds via a careful analysis of existing proofs in those areas, and can result in both quantitative and qualitative improvements of the corresponding theorems. While the use of proof theoretic techniques for this purpose was first proposed by Kreisel in the 1950s [29, 30], the field was brought to maturity through the work of Kohlenbach and others, an overview of which is documented in the monograph [17] and the more recent survey papers [18, 20].[1]

Progress in applied proof theory tends to assume one of two main forms: Concrete case studies in which proof theoretic techniques are used to obtain new results, often through the analysis of a single or a collection of related proofs; and so-called *logical metatheorems* which explain those case studies as instances of general logical phenomena, typically guaranteeing the extractability of (highly uniform) computational information within a specific setting. For example, early case studies in metric fixed

[1] An online proof mining bibliography, which at the time of writing lists over 250 papers, is currently maintained by N. Pischke at `https://nicholaspischke.github.io/bib/ref_date.html`.

point theory (e.g. [14, 15, 23]) were explained in [16] (later generalised in [10]), while the more recent metatheorems [26, 40, 44] cover a range of case studies on accretive and monotone set-valued operators (including but not limited to [7, 19, 21, 22, 27, 31]). Current efforts to bring proof mining to bear on probability theory (building on earlier work in measure theory [1, 2]) exemplify this pattern further, with initial case studies on martingale convergence [38, 37], laws of large numbers [32, 33] and stochastic optimization [35, 45] being developed in tandem with new logical metatheorems [34].

In the case studies, proof-theoretic 'workings' are often suppressed and streamlined, with everything presented in standard mathematical language and in a style suited to the area of application (which is indeed crucial, as in most cases this work is published in specialist journals within that area of application). Metatheorems, on the other hand, are expressed in the language of formal logic, and require sophisticated proof-theoretic machinery that includes proof interpretations (and above all Gödel's Dialectica [11]), logical relations based on majorizability [12], and abstract types [16]. The point here is that the connection between metatheorem and case study is subtle: Metatheorems are never used *directly* (and certainly not *bureaucractically*) to obtain new applications. It is rather the case that practitioners of proof mining carry in their mind a broad and constantly evolving proof-theoretic intuition, one that can in a given context be made rigorous and formal through logical metatheorems, but which can also be utilised without explicit reference to formal logic and in conjunction with ordinary mathematical insights to obtain concrete results. But as a consequence, it is not necessarily obvious to a non-specialist precisely how proof theory was used to obtain those results.

Our paper explores a space between the purely mathematical case studies, and the fully formal metatheorems, where we seek to capture in a dynamic manner the way in which proof-theoretic insights are used to produce concrete theorems in mathematics. We present a new case study in convex optimization, but offer an alternative style of presentation where we attempt to make explicit some of the underlying proof theoretic intuitions involved in "unwinding" the proof in question, making use in particular of an informal type of proof tree. This is certainly not standard in applied proof theory, and technically not necessary to prove the main results. However, it is hoped that this presentation will be of independent interest, hinting at ways in which proof mining could be effectively implemented (or even partially automated) in a proof assistant by focusing on 'high-level' inference rules that represent commonly occurring mathematical patterns, along with transformations that indicate how these patterns can be manipulated to obtain computational information.

Our starting point is an elegant proof on the convergence of greedy approximation schemes, given as Theorem 3.4 of [6]. Our analysis of this proof exhibits many of the characteristic features of proof mining, namely:

(i) The proof is at first glance non-constructive (establishing that a limit inferior must be zero by showing that it can't be positive), and yet one can nevertheless extract direct rates of convergence with very low computational complexity;
(ii) the theorem applies to arbitrary Banach spaces with certain geometric properties (in this case uniform smoothness), but the rates are highly uniform and only depend on the appropriate *modulus* of uniform smoothness along with some other basic data;
(iii) the initial quantitative analysis can be extended to produce several qualitative strengthenings of the original result, including a weakening of the convergence condition on the error terms and an extension to fixed step sizes.

We analyse this proof, and in doing so provide a first application of proof theory to greedy approximation schemes. We anticipate that further work in this direction would be of interest, given that there exist a wealth of related nonconstructive convergence proofs that depend geometric properties of Banach spaces (e.g. [47, 48]), which are in turn directly relevant to solving high-dimensional problems in machine learning and statistics.

However, we consider our presentation of the extraction process to be the central contribution of the paper: Firstly, in doing so we provide an expository account of the way in which proof theory can be used to generate new theorems in mathematics; Secondly, and more importantly, given the ever increasing capabilities and widespread use of proof assistants, we envisage that high-level descriptions of the underlying proof transformations of the kind studied here might provide insight into the sort of mathematical libraries and tactics that would be helpful in both formalising or even partially automating the proof mining process. Indeed, we conjecture that the development of domain-specific logical systems that, much like our informal proof rules, are designed to capture the high-level structures on which proof mining operates, will be essential for any *practical* automation of the proof mining process.

Structure of the paper. The paper has been written to appeal to a reader who is more interested in how an applied proof theorist might manipulate proofs in general, than in the specific mathematical details of the situation at hand. For that reason, the necessary mathematical background is presented in a self-contained manner in Section 2, and logic and analysis are subsequently separated as much as possible, so that the reader primarily interested in the former can skim over passages concerning the latter while still appreciating the main points.

§2. Mathematical background. We start by presenting the main subject of our case study, along with the convergence theorem whose proof we will analyse. The reader interested in the deeper mathematical context is encouraged to consult the references, particularly the original paper of Darken et al. [6] in which our chosen proof is just one among a series of results on convex approximation schemes in Banach spaces, and the more recent monograph on greedy algorithms by Temlyakov [47].

2.1. Greedy approximation schemes. The following definitions and notation are taken from [6]. Let X be a real Banach space, and suppose that $S \subseteq X$ is some arbitrary set. Let $\mathrm{co}(S)$ denote the set of all convex combinations of elements of S, that is, objects of the form

$$\lambda_1 y_1 + \ldots + \lambda_n y_n$$

for $n \geq 1$, $y_1, \ldots, y_n \in S$ and $\lambda_1, \ldots, \lambda_n \in [0,1]$ with $\lambda_1 + \ldots + \lambda_n = 1$. Now suppose that $x^* \in \overline{\mathrm{co}(S)}$, where $\overline{\mathrm{co}(S)}$ denotes the closure of $\mathrm{co}(S)$. A natural question arises:

> *Can we construct incremental approximates to x^* from elements of $\mathrm{co}(S)$ where each approximant is improved by forming a convex combination with a single new element of S?*

In other words, we consider incremental sequences (x_n) of the form

$$x_{n+1} = (1-\lambda_n)x_n + \lambda_n y_n$$

for $x_0 \in S$, $y_n \in S$ and $\lambda_n \in [0,1)$ (we assume that $\lambda_n < 1$ else we would have $x_{n+1} = y_n \in S$ and we are back where we started). The most natural way of choosing y_n and λ_n at each step is to require them to be optimal, which leads naturally to the following slightly more general notion:

Definition 2.1. A sequence (x_n) of elements of X is called *greedy* with respect to $x^* \in X$ and S if for all $n \in \mathbb{N}$:

$$\|x_{n+1} - x^*\| \leq \inf\{\|(1-\lambda)x_n + \lambda y - x^*\| \mid y \in S, \lambda \in [0,1)\}.$$

To account for the fact that S may not be compact, and thus the infimum in the above definition not attained, we loosen the notion in order to incorporate error terms (ϵ_n), leading to our main definition of greedy approximation scheme.

Definition 2.2. Let (ϵ_n) be a sequence of positive real numbers. A sequence (x_n) of elements of X is called (ϵ_n)-*greedy* with respect to $x^* \in X$ and S if for all $n \in \mathbb{N}$:

$$\|x_{n+1} - x^*\| \leq \inf\{\|(1-\lambda)x_n + \lambda y - x^*\| \mid y \in S, \lambda \in [0,1)\} + \epsilon_n.$$

Greedy approximation schemes in Hilbert spaces were first studied by Jones [13], where convergence to $x^* \in \overline{\mathrm{co}(S)}$ with rate $\mathcal{O}(1/\sqrt{n})$ is proven, and the connection of such schemes to artificial neural networks is highlighted where, roughly speaking, improving an approximant by combining with a new element of S can be seen as a generalisation of improving the accuracy of a neural network by adding an additional neuron (see [13, Section 4] or [6, Section 1.3] for further details of this connection).

When X is a general Banach space, on the other hand, convergence is no longer guaranteed. The example given in [6] is $\mathbb{R}^2$ under the L^1 norm: Here, if we take $S := \{(0,-1),(2,1/2),(-2,1/2)\}$ so that $(0,0) \in \mathrm{co}(S)$, the only greedy incremental scheme starting from $x_0 := (0,-1)$ is $x_0 = x_1 = x_2 = \ldots$, since there is no way to strictly decrease the distance to $(0,0)$ through convex combinations with $(2,1/2)$ or $(-2,1/2)$. Geometrically speaking, the problem here is that the unit ball in this space is a diamond, and the line segment between $(0,-1)$ and $(2,1/2)$ only intersects this unit ball at $(0,-1)$, which would not be the case with the Euclidean metric.

The critical issue is connected to the notion of *smoothness* (more informally, spaces where "balls don't have corners"). It turns out that one can establish convergence of greedy algorithms in general Banach spaces by assuming additional smoothness properties. However, the proofs are more difficult, and do not always come with corresponding rates. Establishing rates in such cases is one of the goals of this paper.

2.2. Geometric properties of Banach spaces. In what follows we present some basic facts about Banach spaces: Further details for this section can be found in e.g. [3, Chapter 2] and [47, Chapter 6]. Let X^* denote the dual of X, and $J : X \to 2^{X^*}$ the so-called normalized duality mapping function defined by

$$J(x) := \{y \in X^* \mid y(x) = \|x\|^2 = \|y\|^2\}.$$

A space X is defined to be *smooth* if J is single-valued, in which case we let $j : X \to X^*$ denote the corresponding unique duality map.

In the special case that X is a Hilbert space the duality map is just the inner product $j(x) = \langle x, - \rangle$, and as such the duality map often plays the role of mimicking an inner product in Banach spaces. Crudely speaking, the nicer the duality map, the more X behaves like a Hilbert space. In this respect, an important notion is the modulus of smoothness

$$\rho_X(t) := \sup\left\{ \frac{\|x+ty\| + \|x-ty\|}{2} - 1 \;\middle|\; \|x\| = \|y\| = 1 \right\}$$

for $t \in (0,\infty)$, which in a certain sense gives a quantitative measure of 'niceness' in this context. A Banach space X is *uniformly smooth* if

$$\lim_{t\to 0} \frac{\rho_X(t)}{t} = 0.$$

The following standard lemma (cf. [47, Lemma 6.1]) connects the duality map with the modulus of smoothness. We give its proof in full as this forms part of the overall proof that will be analysed.

Lemma 2.3. *Suppose that X is smooth. Take $x \neq 0$ and let $F_x := j(x)/\|x\|$. Then for any $y \in X$ and $t \in (0, \infty)$:*

$$\|x + ty\| \leq \|x\| \left(1 + 2\rho_X \left(t\|y\|/\|x\|\right)\right) + tF_x(y).$$

Proof. From the definition of ρ_X we obtain

$$\|x + ty\| + \|x - ty\| \leq 2\|x\| \left(1 + \rho_X \left(t\|y\|/\|x\|\right)\right)$$

and using in addition that from $\|F_x\| = 1$ we have

$$\|x - ty\| \geq F_x(x - ty) = F_x(x) - tF_x(y) = \|x\| - tF_x(y)$$

the result follows. $\dashv$

Remark 2.4. Using Lemma 2.3 in along with $\|x + ty\| \geq F_x(x + ty) = \|x\| + tF_x(y)$ proves

$$0 \leq \|x + ty\| - \|x\| - tF_x(y) \leq 2\rho_X \left(t\|y\|/\|x\|\right).$$

For $\|x\| = \|y\|$ it follows that

$$0 \leq \frac{\|x + ty\| - \|x\|}{t} - F_x(y) \leq \frac{2\rho_X(t)}{t}$$

and thus uniformly smooth spaces have the nice geometric property that

$$\lim_{t \to 0} \left(\frac{\|x + ty\| - \|x\|}{t}\right) = F_x(y)$$

and moreover the limit is attained uniformly in x and y with $\|y\| = \|x\|$. In other words, X has a *uniformly Gâteuax differentiable norm*, and in fact this is an equivalent characterisation of being uniformly smooth.

2.3. Convergence of greedy incremental sequences in Banach spaces. We now state and prove the main result that we will analyse: Roughly speaking, this says that if X is uniformly smooth, if (x_n) is an (ϵ_n)-greedy approximation schemes with respect to x^*, then $x_n \to x^*$ provided that $\sum_{i=0}^{\infty} \epsilon_i < \infty$. The material in this subsection is taken entirely from [6], with the proof only very slightly reformulated from its original presentation.

We first require a lemma that applies smoothness to greedy algorithms, essentially showning that sequences that are (ϵ_n)-greedy with respect to x^* are, in particular, *quasi-Fejér monotone* with respect to x^* in a certain sense (see [4, 5] for surveys of this fundamental concept and its role in convex optimization). The result below incorporates Lemma 3.3 of [6], along with part of the main proof of Theorem 3.4 from the same paper, and we have deliberately re-structured things in this way in order to separate out those parts of the proof that use functional analysis of some kind. This is convenient because it allows us, in the proof of Theorem 2.6, to focus on the main combinatorial structure of the overall proof, making it easier to organise the quantitative analysis that follow.

LEMMA 2.5 (Cf. Lemma 3.3 of [6]). *Let X be a Banach space with modulus of smoothness ρ_X. Suppose that $S \subseteq X$ and $x^* \in \overline{\mathrm{co}(S)}$, that (x_n) is (ϵ_n)-greedy with respect to x^* and S, and $K > 0$ is such that $\sup\{\|y - x^*\| \mid y \in S\} \leq K$. Then for any $n \in \mathbb{N}$ and $b > 0$, we have*

$$\|x_{n+1} - x^*\| \leq (1 - \alpha(b))\|x_n - x^*\| + \epsilon_n$$

whenever $\|x_n - x^*\| \geq b$, where

$$\alpha(b) := \sup\left\{ \lambda \left(1 - \frac{2K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)}\right) \;\middle|\; \lambda \in [0,1) \right\}$$

for $u(b,\lambda) := \lambda K/(1-\lambda)b$.

PROOF. Fix $n \in \mathbb{N}$ and assume that $\|x_n - x^*\| \geq b > 0$. For $y \in S$ and $\lambda \in [0,1)$, writing

$$\|(1-\lambda)x_n + \lambda y - x^*\| = (1-\lambda)\|(x_n - x^*) + \lambda(y - x^*)/(1-\lambda)\|$$

and applying Lemma 2.3 to the right hand side we have

$$\begin{aligned} &\|(1-\lambda)x_n + \lambda y - x^*\| \\ &\leq (1-\lambda)\left(1 + 2\rho_X\left(\frac{\lambda\|y - x^*\|}{(1-\lambda)\|x_n - x^*\|}\right)\right)\|x_n - x^*\| + \lambda F_n(y - x^*) \end{aligned}$$

where we write $F_n := j(x_n - x^*)/\|x_n - x^*\|$. Using the standard fact that the modulus of smoothness is monotone, it follows from $b \leq \|x_n - x^*\|$ and $\|y - x^*\| \leq K$ for $y \in S$ that

$$\begin{aligned} (1-\lambda)\rho_X\left(\frac{\lambda\|y - x^*\|}{(1-\lambda)\|x_n - x^*\|}\right) &\leq (1-\lambda)\rho_X\left(\frac{\lambda K}{(1-\lambda)b}\right) \\ &= \frac{\lambda K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)} =: \beta(\lambda, n) \end{aligned}$$

for $u(b,\lambda)$ as defined in the statement of the result. Putting these together and defining $\beta(\lambda, n)$ as indicated in the previous equation, we obtain

$$(*) \qquad \begin{aligned} &\|(1-\lambda)x_n + \lambda y - x^*\| \\ &\leq [1 - \lambda(1 - 2\beta(\lambda, n))]\,\|x_n - x^*\| + \lambda F_n(y - x^*). \end{aligned}$$

Now, since $x^* \in \overline{\mathrm{co}(S)}$, we can make the positive part of $F_n(y - x^*)$ arbitrary small, in the sense that for all $\varepsilon > 0$ there exists $y \in S$ such that $F_n(y - x^*) \leq \varepsilon$. To see this, pick some $z \in \mathrm{co}(S)$ such that $\|z - x^*\| \leq \varepsilon$. Writing $z = \sum_{i=1}^k \lambda_i y_i$ for $y_i \in S$, we must have $F_n(y_i - x^*) \leq \varepsilon$ for some $i = 1, \ldots, k$, else, using that F_n is linear and $\|F_n\| \leq 1$ we have

$$\varepsilon = \varepsilon \sum_{i=1}^k \lambda_i < \sum_{i=1}^k \lambda_i F_n(y_i - x^*) = F_n(z - x^*) \leq \|z - x^*\| \leq \varepsilon$$

Therefore $\inf\{F_n(y-x^*)\,|\,y\in S\}\leq 0$, and more generally the infimum over $y\in S$ and $\lambda\in[0,1)$ of the quantity

$$(1-\lambda\,(1-2\beta(\lambda,n)))\,\|x_n-x^*\|+\lambda F_n(y-x^*)$$

is bounded by

$$(1-\sup\{\,\lambda\,(1-2\beta(\lambda,n))\;|\;\lambda\in[0,1)\})\,\|x_n-x^*\|$$

which is just $(1-\alpha(b))\|x_n-x^*\|$. Thus combining the above with $(*)$ and the definition of being (ϵ_n)-greedy we have

$$\begin{aligned}\|x_{n+1}-x^*\|&\leq\inf\{\|(1-\lambda)x_n+\lambda y-x^*\|\,|\,y\in S,\,\lambda\in[0,1)\}+\epsilon_n\\&\leq(1-\alpha(b))\|x_n-x^*\|+\epsilon_n\end{aligned}$$

and the lemma is proven. ⊣

The main result can now be stated and proved using little more than elementary analysis.

Theorem 2.6 (Cf. Theorem 3.4 of [6]). *Let X be a Banach space with modulus of smoothness ρ_X and $S\subseteq X$ be bounded. Suppose that $x^*\in\overline{\mathrm{co}(S)}$, and that (x_n) is (ϵ_n)-greedy with respect to x^* and S for some sequence (ϵ_n) of positive reals with $\sum_{i=0}^{\infty}\epsilon_i<\infty$. Then $x_n\to x^*$ as $n\to\infty$.*

Proof. Define $a_n:=\|x_n-x^*\|$ and let $a_\infty:=\liminf_{n\to\infty}a_n$. Using the fact that (x_n) is (ϵ_n)-greedy, we have $a_{n+1}\leq a_n+\epsilon_n$ by definition (by simply setting $\lambda=0$), and thus more generally

$$a_{n+m}\leq a_n+\sum_{i=n}^{n+m-1}\epsilon_i\leq a_n+\sum_{i=n}^{\infty}\epsilon_i$$

for any $m,n\in\mathbb{N}$. But since $\sum_{i=n}^{\infty}\epsilon_i\to 0$ as $n\to\infty$, we must in fact have $a_n\to a_\infty$ as $n\to\infty$. It therefore suffices to show that $a_\infty=0$.

To this end, suppose for contradiction that $a_\infty>0$. Since S is bounded there exists $K>0$ is such that $\sup\{\|y-x^*\|\;|\;y\in S\}\leq K$, and since $a_\infty/2\leq a_n$ for n sufficiently large, we can apply Lemma 2.5 to show that

$$a_{n+1}\leq(1-\alpha(a_\infty/2))a_n+\epsilon_n$$

for n sufficiently large. Taking the limit as $n\to\infty$ and using that $a_n\to a_\infty$ and $\epsilon_n\to 0$ yields

$$a_\infty\leq(1-\alpha(a_\infty/2))a_\infty$$

and therefore $\alpha(a_\infty/2)\leq 0$. But we have $u(a_\infty/2,\lambda)\to 0$ as $\lambda\to 0$, and therefore by uniform smoothness of X it follows that

$$\frac{\rho_X(u(\lambda,a_\infty/2))}{u(\lambda,a_\infty/2)}\to 0\quad\text{as }\lambda\to 0$$

from which we see that $\alpha(a_\infty/2)>0$, a contradiction. Thus $a_\infty=0$ and we are done. ⊣

§3. A high-level analysis of the proof. In this section, we start to apply proof theoretic reasoning to the ideas presented so far. More specifically, we carry out a series of steps that apply to the high-level structure of the proof of Theorem 2.6, our end goal being a computational version of Theorem 2.6, namely a rate of convergence for $x_n \to x^*$ as $n \to \infty$, by which we mean a computable function $f : (0, \infty) \to \mathbb{N}$ satisfying

$$\forall \varepsilon > 0 \, \forall n \geq f(\varepsilon) \, (\|x_n - x^*\| < \varepsilon) \, .$$

We represent the proof via a series of proof trees in natural deduction style, where inferences typically represent a whole series of formal steps conflated into one. Our main goal in representing the proof this way is to identify its main features, and then carry out a series of transformations on the proof which pay special attention to the following questions:

1. If we have used an assumption in part of the proof, can we in fact replace it with a weaker assumption?
2. Can we phrase formulas in a more uniform way by expressing them in terms of bounds?
3. How does computational information flow through the proof?

Most of these questions can be tackled formally using proof theoretic methods, such as logical metatheorems, majorizability, and variants of the Dialectica interpretation ([17] is the standard reference). However, here we aim to show how one might transform a proof "by hand" and in a semi-formal manner, ignoring parts of the proof that are uninteresting and focusing on the key *mathematical* rather than *logical* steps. We envisage that the kind of high-level inference rules we sketch below, along with the manipulations performed on them, could become formally captured in a domain-specific logical system specially designed for the computational analysis of proofs of a particular shape, though the precise design of such a system is something we leave to future work!

One important thing to note is that, initially at least, we analyse the proof *as it is*, resulting in a preliminary rate of convergence that is a direct reflection of the convergence proof as found in [6] (and as presented in the previous section). We then introduce improvements of the proof in the following section, where the initial analysis helps in identifying parts that can be optimized, resulting in both an improved rates and qualitative strengthenings of the convergence theorem itself.

3.1. The overall structure of the proof of Theorem 2.6. For the remainder of this section, we will fix several things: X will be a Banach space with ρ its modulus of smoothness, $S \subseteq X$ and $x^* \in X$. We suppose that $K > 0$ is such that $\|y - x^*\| \leq K$ for all $y \in S$, which in particular exists whenever S is bounded. Finally, for now we let (x_n) be an arbitrary sequence in X and (ϵ_n) an arbitrary sequence of nonnegative real numbers. We treat these throughout as global parameters, and also for convenience fix the notation $a_n := \|x_n - x^*\|$. The goal is therefore to prove that $a_n \to 0$.

The main technical lemma we required – Lemma 2.5 – can then be represented as the single inference

$$\frac{(\epsilon_n)\text{-}x^*\text{-greedy} \quad x^* \in \overline{\mathrm{co}(S)}}{\forall n\, \forall b > 0(b \leq a_n \Longrightarrow P(b, a_n, a_{n+1}, \epsilon_n))}\,\text{L2.5}$$

where "(ϵ_n)-x^*-greedy" is shorthand for the statement "(x_n) is (ϵ_n)-greedy with respect to x^*" and the predicate P is defined by

$$P(b, a, a', \epsilon) := a' \leq (1 - \alpha(b))a + \epsilon$$

for $\alpha(b)$ defined as in Lemma 2.5. Rather than immediately analysing the somewhat intricate proof of that lemma, we will leave it for now and consider how it fits in to the main proof of Theorem 2.6. Letting $a_\infty := \liminf_{n\to\infty} a_n$, the first step in the main proof has the overall form

$$(\Gamma_1) \qquad \frac{\dfrac{(\epsilon_n)\text{-}x^*\text{-greedy}}{\forall n(a_{n+1} \leq a_n + \epsilon_n)} \quad \sum_{i=0}^\infty \epsilon_i < \infty}{a_n \to a_\infty}$$

Of course, there are a number of additional (elementary) steps involved in inferring $a_n \to a_\infty$ from the two premises, but we note that the property of being (ϵ_n)-greedy is used in a weak way, in that we only require $a_{n+1} \leq a_n + \epsilon_n$. Moving on to the second main step of the proof, we now include an open assumption $\{a_\infty > 0\}$ with the aim of reaching a contradiction:

(Γ_2)

$$\frac{\forall n \geq N_\exists(a_\infty/2 \leq a_n) \quad \dfrac{\dfrac{(\epsilon_n)\text{-}x^*\text{-greedy} \quad x^* \in \overline{\mathrm{co}(S)}}{\forall n\forall b > 0(b \leq a_n \Longrightarrow P(b, a_n, a_{n+1}, \epsilon_n))}\,\text{L2.5} \quad \{a_\infty > 0\}}{\forall n(a_\infty/2 \leq a_n \Longrightarrow P(a_\infty/2, a_n, a_{n+1}, \epsilon_n))}}{\forall n \geq N_\exists(P(a_\infty/2, a_n, a_{n+1}, \epsilon_n))}$$

and we label this derivation (Γ_2). Here, $N_\exists$ is simply some natural number that we know to exist by definition of a_∞ (where the notation is chosen to indicate that this is essentially an epsilon term for which we do not have an explicit value). The conclusion of (Γ_2) is simply the statement that

$$a_{n+1} \leq (1 - \alpha(a_\infty/2))a_n + \epsilon_n$$

for n sufficiently large. Continuing, we have the following crucial step:

$$(\Gamma_3) \qquad \frac{\begin{array}{c}\{a_\infty > 0\} \\ \Big\|\,\Gamma_2 \\ \forall n \geq N_\exists(P(a_\infty/2, a_n, a_{n+1}, \epsilon_n))\end{array} \quad \begin{array}{c}\Big\|\,\Gamma_1 \\ a_n \to a_\infty\end{array} \quad \dfrac{\sum_{i=0}^\infty \epsilon_i < \infty}{\epsilon_n \to 0}}{P(a_\infty/2, a_\infty, a_\infty, 0)}\,{n\to\infty}$$

where we use the shorthand

$$\begin{array}{c} \Big\Vert\,\Gamma_1 \\ a_n \to a_\infty \end{array}$$

to represent the proof tree (Γ_1), and similarly for (Γ_2). The final inference in (Γ_3) represents taking the limit as $n \to \infty$ to establish

$$a_\infty \leq (1 - \alpha(a_\infty/2))a_\infty.$$

We reach our contradiction by using uniform smoothness of the space, with an inference we mark as $(\star)$ below, and from this can therefore derive $a_\infty = 0$ using classical logic, eliminating the open assumption $\{a_\infty > 0\}$:

$$(\Gamma_4) \qquad \cfrac{\cfrac{\begin{array}{c}\{a_\infty > 0\} \\ \Big\Vert\,\Gamma_3 \\ P(a_\infty/2, a_\infty, a_\infty, 0)\end{array} \qquad \cfrac{\cfrac{X \text{ is U. S.}}{\forall b > 0(\alpha(b) > 0)}\,(\star) \qquad \{a_\infty > 0\}}{\alpha(a_\infty/2) > 0}}{\perp}}{a_\infty = 0}\,\Longrightarrow_I$$

Formally, there is now one final step in the proof, namely:

$$(\Gamma_5) \qquad \cfrac{\begin{array}{c} \Big\Vert\,\Gamma_1 \\ a_n \to a_\infty \end{array} \quad \begin{array}{c} \Big\Vert\,\Gamma_4 \\ a_\infty = 0 \end{array}}{a_n \to 0}$$

and thus expanding the definition of (Γ_5) in full would give us a complete (high-level) representation of the proof of Theorem 2.6. We now set out to analyse this proof with the three main questions posed at the beginning of the section in mind. We start at the bottom and work back up.

3.2. Using $a_\infty = 0$ in the final step. We start by considering precisely how the final step is proven, and asking whether we can extract any computational information at this stage. We first note that in the special case that (a_n) is a sequence of nonnegative reals, the otherwise more complex statement $a_\infty = 0$ can be expressed as

$$(*) \qquad \forall b > 0, m \in \mathbb{N}\, \exists n \geq m(a_n < b).$$

So assuming that we have proven $(*)$, how exactly do we derive $a_n \to 0$ from $a_n \to a_\infty$? For a start, we observe that we do not require the full statement $a_n \to a_\infty$ at all: It is sufficient to notice that two premises used to prove this in (Γ_1) – which is in general a more complex argument – can be combined with $(*)$ in a simple way to establish $a_n \to 0$. To be more specific, suppose that $f : (0,\infty) \to \mathbb{N}$ is a rate of convergence for $\sum_{i=0}^\infty \epsilon_i < \infty$ in the sense that

$$\forall b > 0 \left(\sum_{i=f(b)}^{\infty} \epsilon_i < b \right).$$

Then for any $n \geq f(b)$, using that $\forall n(a_{n+1} \leq a_n + \epsilon_n)$ we must have

$$a_{n+k} \leq a_n + \sum_{i=n}^{n+k-1} \epsilon_i \leq a_n + \sum_{i=f(b)}^{\infty} \epsilon_i < a_n + b$$

for any $n, k \in \mathbb{N}$ and $b > 0$. Now suppose that we have a computable bound for n in $(*)$ for $m := f(b)$ i.e. a function $\Phi : (0,\infty) \to \mathbb{N}$ such that

$$\forall b > 0 \, \exists n \in [f(b); \Phi(b)](a_n < b)$$

where we use the notation $[f(b); \Phi(b)] := [f(b), \Phi(b)] \cap \mathbb{N}$. Then it is not hard to see that $\Phi(b/2)$ must be a rate of convergence for $a_n \to 0$, and thus we have not only proved that $a_n \to 0$ but shown exactly what quantitative information we need from the assumption $a_\infty = 0$ to obtain a rate. In terms of our proof trees, we have essentially transformed (Γ_5) into the following *computational* derivation, eliminating the original (Γ_1), where $b > 0$ is now some global free variable:

$$(\Gamma_5^c) \qquad \cfrac{\cfrac{(\epsilon_n)\text{-}x^*\text{-greedy}}{\forall n(a_{n+1} \leq a_n + \epsilon_n)} \quad \sum_{i=f(b)}^\infty \epsilon_i < b \quad \begin{array}{c} \Big\| \Gamma_4^c \\ \exists n \in [f(b); \Phi(b)](a_n < b) \end{array}}{\forall n \geq \Phi(b)(a_n < 2b)}$$

where now (Γ_4^c) is some (currently hypothetical) computation transformation of (Γ_4) that allows us to replace $a_\infty = 0$ with a derivation of

$$\exists n \in [f(b); \Phi(b)](a_n < n)$$

for some explicit function Φ. The challenge is now has shifted to transforming the original derivation (Γ_4) to a computational one (Γ_4^c).

3.3. Simplifying (Γ_4). A crucial observation at this stage is that (Γ_4) simplifies: Because we have only used $a_\infty = 0$ in the weakened form,

$$(**) \qquad \exists n \geq f(b)(a_n < b)$$

where $b > 0$ is now some ambient free variable, we can try to substitute this in the conclusion of (Γ_4) and then replace the open assumption $\{a_\infty > 0\}$ with the stronger *negation* of $(**)$ i.e. the more concrete assumption

$$\{\forall n \geq f(b)(a_n \geq b)\},$$

and simplify the proof tree accordingly. This process involves a series of straightforward heuristic steps. Starting with (Γ_2), we observe that here $\{a_\infty > 0\}$ is crucially used to establish that $\forall n \geq N_\exists (a_\infty/2 \leq a_n)$ for sufficiently large $N_\exists$ where $a_\infty/2 > 0$. For now we see if we can just replace $a_\infty/2$ with b, and set $N_\exists := f(b)$ as follows:

$$(\Gamma_2^s) \qquad \dfrac{\{\forall n \geq f(b)(a_n \geq b)\} \quad \dfrac{(\epsilon_n)\text{-}x^*\text{-greedy} \quad x^* \in \overline{\mathrm{co}(S)}}{\forall n(b \leq a_n \Longrightarrow P(b, a_n, a_{n+1}, \epsilon_n))}\,\mathrm{L2.5}}{\forall n \geq f(b)\, P(b, a_n, a_{n+1}, \epsilon_n)}$$

By a simple substitution, (Γ_3) then becomes

$$(\Gamma_3^s) \qquad \dfrac{\begin{array}{c}\{\forall n \geq f(b)(a_n \geq b)\}\\ \Big\Vert\,\Gamma_2^s\\ \forall n \geq f(b)\, P(b, a_n, a_{n+1}, \epsilon_n)\end{array} \quad \begin{array}{c}\Big\Vert\,\Gamma_1\\ a_n \to a_\infty\end{array} \quad \dfrac{\sum_{i=0}^\infty \epsilon_i < \infty}{\epsilon_n \to 0}}{P(b, a_\infty, a_\infty, 0)}\,{\scriptstyle n\to\infty}$$

and then the entire modified version of (Γ_4) would become

$$(\Gamma_4^s) \qquad \dfrac{\dfrac{\begin{array}{c}\{\forall n \geq f(b)(a_n \geq b)\}\\ \Big\Vert\,\Gamma_3^s\\ P(b, a_\infty, a_\infty, 0)\end{array} \quad \dfrac{X \text{ is U. S.}}{\alpha(b) > 0}\,(\star)}{\bot}}{\exists n \geq f(b)(a_n < b)}\,\Longrightarrow_I$$

where here replacing $a_\infty/2$ with b makes no difference to the way in which we derive a contradiction. Though not yet fully computational, (Γ_4^s) represents a simplified and more explicit version of (Γ_4), reflecting throughout our weakened use of $a_\infty = 0$.

3.4. Analysing (Γ_4^s). We have demonstrated that in order to obtain a computable a rate of convergence for $a_n \to 0$ it suffices find a bound $\Phi(b)$ for the existential quantifier in $\exists n \geq f(b)(a_n < b)$, where f is a rate of convergence for $\sum_{i=0}^\infty \epsilon_i < \infty$. We propose to do this by analysing the simplified version (Γ_4^s) of (Γ_4) arrived at above.

Here a key observation is that if we can weaken the open assumption with a bound on how many $n \geq f(b)$ are needed in order to be able to reach a contradiction as before, then this will be exactly the bound we are looking for. In other words, we want to produce $\Phi(b)$ satisfying

$$\dfrac{\begin{array}{c}\{\forall n \in [f(b); \Phi(b)](a_n \geq b)\}\\ \Big\Vert\,\Gamma_3^s\\ P(b, a_\infty, a_\infty, 0)\end{array} \quad \dfrac{X \text{ is U. S.}}{\alpha(b) > 0}\,(\star)}{\bot}$$

Focusing first on the final derivation above, one immediate observation is that $\alpha(b)$ is not necessarily computable, which might pose a problem in the likely case that we want to use property $\alpha(b) > 0$ in a computational way. Here we make another assumption that we hope to verify later on, namely the existence of a computable function $\xi : (0,\infty) \to (0,1)$ witnessing that $\alpha(b) > 0$ i.e. such that $\alpha(b) \geq \xi(b) > 0$ for any $b > 0$.

We now shift our up the proof tree to the final inference of (Γ_3^s), where we take the limit as $n \to \infty$ to establish $P(b, a_\infty, a_\infty, 0)$ i.e.

$$a_\infty \leq (1 - \alpha(b))a_\infty. \tag{†}$$

Given the infinitary nature of (†) it is not clear at first glance how we could weaken our open assumption $\{\forall n \geq f(b)(a_n \geq b)\}$ to being true only in a finite range. However, a natural question to ask here is the following: If $P(b, a_n, a_{n+1}, \epsilon_n)$ fails to be true in the limit – in the sense of (†) – can we show that it also fails to hold for n sufficiently large? In particular, here the only property of $a_n \to a_\infty$ that is important is that a_n and a_{n+1} converge to the *same value*, so could we replace this with a_n and a_{n+1} being sufficiently close together?

To this end, let us take some $\delta > 0$. Then by $a_n \to a_\infty$ and $\epsilon_n \to 0$ there exists, more concretely, some $k \geq f(b)$ such that $a_k - a_{k+1} < \delta$ and $\epsilon_k < \delta$. Then from $P(b, a_k, a_{k+1}, \epsilon_k)$ and our assumption $\alpha(b) \geq \xi(b)$ it follows that

$$a_{k+1} \leq (1 - \xi(b))(a_{k+1} + \delta) + \delta.$$

which can be rearranged as

$$\frac{\xi(b) \cdot a_{k+1}}{2 - \xi(b)} \leq \delta$$

But using also that $b \leq a_{k+1}$, the above fails for

$$\delta_{\xi,b} := \tfrac{1}{2}\xi(b) \cdot b,$$

so we have reached a contradiction, and we can replace the final inference of (Γ_3^s) with a finitary version which does not use the whole the limit as $n \to \infty$. We now formulate this finitisation of (Γ_3^s) as the following proof tree (Γ_3^f), where $\delta_{\xi,b}$ is defined as above:

(Γ_3^f)

$$\dfrac{\dfrac{\{\forall n \geq f(b)(a_n \geq b)\}}{\forall n \geq f(b)\,(b \leq a_{n+1} \wedge P(b, a_n, a_{n+1}, \epsilon_n))}\,\Gamma_2^s \qquad \dfrac{\dfrac{}{a_n \to a_\infty}\,\Gamma_1 \qquad \dfrac{\sum_{i=0}^\infty \epsilon_i < \infty}{\epsilon_n \to 0}}{\exists k \geq f(b)(a_k - a_{k+1}, \epsilon_k < \delta_{\xi,b})}}{\exists k \geq f(b)(b \leq a_{k+1} \wedge P(b, a_k, a_{k+1}, \epsilon_k) \wedge a_k - a_{k+1}, \epsilon_k < \delta_{\xi,b})}$$

Then we have obtained our contradiction as follows:

$$\dfrac{\dfrac{\{\forall n \geq f(b)(a_n \geq b)\}}{\exists k \geq f(b)(b \leq a_{k+1} \wedge P(b, a_k, a_{k+1}, \epsilon_k) \wedge a_k - a_{k+1}, \epsilon_k < \delta_{\xi,b})}\,\Gamma_3^f \qquad \dfrac{X \text{ is U. S.}}{\alpha(b) \geq \xi(b) > 0}}{\bot}$$

We now ask: How much of the assumption $\{\forall n \geq f(b)(a_n \geq b)\}$ do we need to obtain this contradiction? Inspecting (Γ_5) it is readily apparent that if $\Psi(b)$ is a bound on a witness for $\exists k \geq f(b)(a_k - a_{k+1}, \epsilon_k < \delta_{\xi,b})$, then we can replace the open assumption with

$$\{\forall n \in [f(b); \Psi(b) + 1](a_n \geq b)\}$$

(here the -1 is necessary from our additional use of the assumption for $b \leq a_{k+1}$), and thus $\Psi(b/2) + 1$ is a rate of convergence for $a_n \to 0$.

3.5. The final step. All that remains in order to obtain our rate of convergence – aside from some assumptions involving uniform smoothness that we have postponed to later – is to analyse the following fragment of our modified proof tree:

$$\dfrac{\dfrac{}{a_n \to a_\infty}\,\Gamma_1 \qquad \dfrac{\sum_{i=0}^{\infty} \epsilon_i < \infty}{\epsilon_n \to 0}}{\exists k \geq f(b)(a_k - a_{k+1}, \epsilon_k < \delta_{\xi,b})}$$

We will actually provide a bound $\Psi(N, \delta)$ for the more general statement

$$\forall N, \delta \exists k \geq N(a_k - a_{k+1}, \epsilon_k < \delta)$$

and then

$$(+) \qquad \Phi(b) := \Psi(f(b/2), \delta_{\xi,b/2}) + 1$$

would be our rate of convergence for $a_n \to 0$. Witnessing the above turns out to be simpler that it might look: From $\sum_{i=0}^{\infty} \epsilon_i < \infty$ we clearly have $\epsilon_k < \delta$ for all $k \geq f(\delta)$ sufficiently large, and then finding k such that we also have $a_k - a_{k+1} < \delta$ does not in any way require the assumption $a_n \to a_\infty$, rather we simply need a bound on a single element of (a_n). We now present this construction in general terms as follows:

Lemma 3.1. *Let (a_n), (K_n) and (ϵ_n) be sequences of nonnegative reals with $K_n \geq a_n$ for all $n \in \mathbb{N}$ and $\epsilon_n \to 0$ as $n \to \infty$ with rate f. Let*

$$\Psi(N, \delta) := M_{N,\delta} + \left\lceil \frac{K_{M_{N,\delta}}}{\delta} \right\rceil \quad \textit{for} \quad M_{N,\delta} := \max\{N, f(\delta)\}.$$

Then for any $N \in \mathbb{N}$ and $\delta > 0$ there exists some $k \in [N; \Psi(N, \delta)]$ such that $a_k - a_{k+1} < \delta$ and $\epsilon_k < \delta$.

Proof. Suppose that $a_k - a_{k+1} \geq \delta$ for all $k \in [M_{N,\delta}; \Theta(N, \delta)]$. Then we have

$$\begin{aligned} K_{M_{N,\delta}} \geq a_{M_{N,\delta}} &\geq a_{M_{N,\delta}+1} + \delta \\ &\geq \dots \\ &\geq a_{M_{N,\delta}+\lceil K_{M_{N,\delta}}/\delta \rceil+1} + (\lceil K_{M_{N,\delta}}/\delta \rceil + 1)\delta \\ &\geq (\lceil K_{M_{N,\delta}}/\delta \rceil + 1)\delta \\ &> K_{M_{N,\delta}} \end{aligned}$$

which is a contradiction. Therefore there exists some

$$k \in [M_{N,\delta}; \Theta(N, \delta)] \subseteq [N; \Theta(N, \delta)]$$

such that $a_k - a_{k+1} < \delta$, and since $k \geq M_{N,\delta} \geq f(\delta)$ it also follows that $\epsilon_k < \delta$ by definition of f. ⊣

Assuming for simplicity that f is monotone, so that we can simplify $\max\{f(b/2), f(\delta_{\xi,b/2})\} = f(\delta_{\xi,b/2})$, and supposing that $\sum_{i=0}^{\infty} \epsilon_i \leq L$, by which we have $a_n \leq a_0 + \sum_{i=0}^{\infty} \epsilon_i \leq K + L$ for all $n \in \mathbb{N}$, we then obtain a concrete rate of convergence for $a_n \to 0$ from (+), namely

$$\text{(R)} \qquad \Phi(b) := f(\mu_{\xi,b}) + \left\lceil \frac{K+L}{\mu_{\xi,b}} \right\rceil + 1 \quad \text{for} \quad \mu_{\xi,b} := \frac{b}{4} \cdot \xi\left(\frac{b}{2}\right).$$

Equipped with this high-level quantitative understanding of the main combinatorial structure of the proof, we now start to put things together in an ordinary mathematical style.

§4. The main result.

We complete our quantitative analysis of the proof of Theorem 2.6 by first formalising and optimizing the ideas of the previous section into abstract quantitative convergence lemmas, and second dealing with the hitherto postponed treatment of uniform smoothness. It is at this stage that we start to make greater use of ideas and techniques from the broader proof mining literature.

4.1. Abstract convergence lemmas. The result of Section 3 was a preliminary rate of convergence for $\|x_n - x^*\| \to 0$ in some (as yet hypothetical) computable function ξ satisfying $0 < \xi(b) \leq \alpha(b)$ for all $b > 0$, with $\alpha(b)$ defined in terms of the modulus of smoothness of the space as in Lemma 2.5. While there is nothing preventing us from just using this rate directly, it is sensible to first ask whether we can optimize the construction – in particular, whether there are local modifications that either improve our initial convergence rate or weaken any assumptions used. Here we now depart from our rigid analysis of the original proof of Theorem 2.6 and inject into it new insights.

Let us consider our analysis of (Γ_4^s) in Section 3.4. The core of this part of the proof, which allows us to derive our contradiction, is the limiting step in (Γ_3^s), which uses $\sum_{i=0}^{\infty} \epsilon_i < \infty$ in the weaker form of $\epsilon_n \to 0$ as $n \to \infty$. A natural question arises: Can we either exploit the strong property $\sum_{i=0}^{\infty} \epsilon_i < \infty$ in this step to gain a better bound $\Phi(b)$, or replace $\sum_{i=0}^{\infty} \epsilon_i < \infty$ with the weaker property $\epsilon_n \to 0$ and $n \to \infty$ throughout the proof? We now show that both are possible.

Our first result captures a refined analysis of (Γ_4^s) that makes full use of the property $\sum_{i=0}^{\infty} \epsilon_i < \infty$. We no longer spell out the details using proof trees, now instead adopting a standard presentation.

LEMMA 4.1. *Let (a_n) and (ϵ_n) be sequences of nonnegative reals, and $\xi : [0, \infty) \to [0, 1)$ be a function with $\xi(0) = 0$ and $\xi(b) > 0$ for $b > 0$, such that for all $n \in \mathbb{N}$ and $b \geq 0$*

$$a_{n+1} \leq (1 - \xi(b))a_n + \epsilon_n$$

whenever $a_n \geq b$. Suppose that $a_0 \leq K$ and $\sum_{i=0}^{\infty} \epsilon_i \leq L$ with rate f. Then $a_n \to 0$ as $n \to \infty$ with rate

$$\Phi(b) := f\left(\frac{b}{2}\right) + \left\lceil \frac{K+L}{\mu_{\xi,b}} \right\rceil \quad \text{for} \quad \mu_{\xi,b} := \frac{b}{2} \cdot \xi\left(\frac{b}{2}\right).$$

PROOF. Fix $b > 0$. We show that there exists some $n \in [f(b/2); \Phi(b)]$ such that $a_n < b/2$. If this were not the case, then for all n it that range we would have

$$\mu_{\xi,b} \leq a_n \cdot \xi\left(\frac{b}{2}\right) \leq a_n - a_{n+1} + \epsilon_n,$$

and therefore summing both sides

$$K + L < \sum_{i=f(b/2)}^{\Phi(b)} \mu_{\xi,b} \leq \sum_{i=f(b/2)}^{\Phi(b)} (a_i - a_{i+1} + \epsilon_i) \leq a_{f(b/2)} + \sum_{i=f(b/2)}^{\Phi(b)} \epsilon_i$$

and now also using that $a_{n+1} \leq a_n + \epsilon_n$ for all $n \in \mathbb{N}$, we have

$$a_{f(b/2)} + \sum_{i=f(b/2)}^{\Phi(b)} \epsilon_i \leq a_0 + \sum_{i=0}^{f(b/2)-1} \epsilon_i + \sum_{i=f(b/2)}^{\Phi(b)} \epsilon_i \leq a_0 + \sum_{i=0}^{\infty} \epsilon_i$$

and thus we have shown that

$$K + L < a_0 + \sum_{i=0}^{\infty} \epsilon_i \leq K + L$$

a contradiction. Therefore let $n \in [f(b/2); \Phi(b)]$ be such that $a_n < b/2$. But then just as in Section 3.2 we have

$$a_{n+k} \leq a_n + \sum_{i=f(b/2)}^{\infty} \epsilon_i < a_n + \frac{b}{2} < b$$

for any $k \in \mathbb{N}$, and so the result follows. $\dashv$

It is instructive to compare the rate given in Lemma 4.1 above to the initial rate (R): The former is an improvement of the latter, but shares the same basic structure. Generally speaking, Lemma 4.1 is more broadly applicable to algorithms that are *quasi-Fejér monotone* and enjoy a certain type of regularity property, here represented by the function ξ. The general quantitative study of quasi-Fejér monotone sequences from the perspective of proof mining began explicitly in [25], was further developed in [42] (notably under regularity assumptions), and recently lifted to a stochastic setting in [35], where, for example, Theorem 3.4 of [35] is a direct generalisation of Lemma 4.1 applicable to stochastic algorithms.

Our next result now keeps the analysis of (Γ_4^s) intact, but instead adjusts (Γ_1), replacing the weaker property $a_{n+1} \leq a_n + \epsilon_n$ with the stronger recurrence inequality satisfied by ϵ_n-greedy algorithms in uniformly smooth Banach spaces, conversely allowing us to weaken the assumption $\sum_{i=0}^{\infty} \epsilon_i < \infty$ to $\epsilon_n \to 0$ as $n \to \infty$ throughout the proof. To obtain a rate of convergence that is *uniform* in (a_n) we now require an explicit bound on the whole sequence (a_n), though boundedness of (a_n) is not required as an assumption in order to establish convergence. The result is inspired by a similar recurrence inequality used in [46] (cf. Lemma 3.1 of that paper) in the somewhat different context of Krasnoselskii-Mann type algorithms for computing fixpoints of weakly contractive mappings.

Lemma 4.2. *Let (a_n) and (ϵ_n) be sequences of nonnegative reals, and $\xi : [0, \infty) \to [0, 1)$ be a function with $\xi(0) = 0$ and $\xi(b) > 0$ for $b > 0$, such that for all $n \in \mathbb{N}$ and $b \geq 0$*

$$a_{n+1} \leq (1 - \xi(b))a_n + \epsilon_n$$

whenever $a_n \geq b$. Suppose that $\epsilon_n \to 0$ as $n \to \infty$ with rate f. Then $a_n \to 0$ as $n \to \infty$ with rate

$$\Phi(b) := f(\mu_{\xi,b}) + \left\lceil \frac{L}{\mu_{\xi,b}} \right\rceil + 1 \quad \text{for} \quad \mu_{\xi,b} := \frac{b}{4} \cdot \xi\left(\frac{b}{2}\right)$$

where $L > 0$ is any bound on (a_n).

Proof. Fix $b > 0$ and apply Lemma 3.1 for $N := f(\mu_{\xi,b})$ and $\delta := \mu_{\xi,b}$, by which there exists $k \in [f(\mu_{\xi,b}); \Theta(b)]$ such that $a_k - a_{k+1}, \epsilon < \mu_{\xi,b}$ for

$$\Theta(b) := f(\mu_{\xi,b}) + \left\lceil \frac{K_b}{\mu_{\xi,b}} \right\rceil$$

where K_b is any bound for $a_{f(\mu_{\xi,b})}$. Suppose for contradiction that $a_l \geq b/2$ for all $l \in [f(\mu_{\xi,b}); \Theta(b) + 1]$. Then in particular we have

$$a_{k+1} \leq (1 - \xi(b/2))(a_{k+1} + \mu_{\xi,b}) + \mu_{\xi,b}$$

and therefore

$$\mu_{\xi,b} \leq \frac{a_{k+1}}{2} \cdot \xi\left(\frac{b}{2}\right) < \frac{\xi(b/2) \cdot a_{k+1}}{2 - \xi(b/2)} \leq \mu_{\xi,b}.$$

Thus there exists $l \in [f(\mu_{\xi,b}); \Theta(b)+1]$ such that $a_l < b/2$, and we now show by induction that $a_{l+m} < b$ for all $m \in \mathbb{N}$. For the induction step, we first note that since $l+m \geq l \geq f(\mu_{\xi,b})$ we have $\epsilon_{l+m} < \mu_{\xi,b}$. We then consider two cases: Either $a_{l+m} < b/2$, and then

$$a_{l+m+1} \leq a_{l+m} + \epsilon_{l+m} < \frac{b}{2} + \mu_{\xi,b} < b,$$

or otherwise we have $b/2 \leq a_{l+m} < b$, and then

$$a_{l+m+1} \leq (1-\xi(b/2))a_{l+m} + \mu_{\xi,b} \leq a_{l+m} - 2\mu_{\xi,b} + \mu_{\xi,b} < b.$$

Therefore $a_n \to 0$ as $n \to \infty$ with convergence rate $\Theta(b)+1$, and setting $K_b := L$ for some uniform bound L on (a_n) gives the result. ⊣

4.2. Handling uniform smoothness. Up until now, we have postponed the fact that we need to deal with uniform smoothness in a computational way: So far, our computational analysis has dealt with nothing beyond sequences of real numbers. Interestingly, and as is very often the case in applied proof theory, these results on the convergence of sequence of reals contain the core computational content of the original proof (see [9] for a comprehensive survey on the general importance of such results in analysis, and [36] for a recent discussion of the role they play in the context of applied proof theory, which also contains references to some the many places in which quantitative version of such results have played a role). Indeed, the only computational role that uniform smoothness plays in the proof of Theorem 2.6 is in establishing that $\alpha(b) > 0$ for α defined as in Lemma 2.5, which explicitly involves the modulus of smoothness ρ_X, or more precisely, a rate of convergence for $\rho_X(t)/t \to 0$ as $t \to 0$.

For the vast majority of case studies in applied proof theory that take place in uniformly smooth Banach spaces, such a rate of convergence is essentially all that is required (though see [8] for an example where additional properties of the modulus of smoothness are needed). In such cases, it is often convenient reformulate the proof using the following alternative definition of uniform smoothness: A Banach space is uniformly smooth if and only if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\text{(US)} \qquad \|x\| = 1 \wedge \|y\| \leq \delta \Longrightarrow \|x+y\| + \|x-y\| \leq 2 + \varepsilon\|y\|$$

for any $x, y \in X$. This characterisation of uniform smoothness is simpler from a logical perspective, and admits a direct computational interpretation in the form of a so-called *modulus of uniform smoothness* $\omega : (0,\infty) \to (0,\infty)$, which is defined to be any function that for any $\varepsilon > 0$ returns some witness for δ in (US). Moduli of uniform smoothness, which are distinct from the (uniquely defined) modulus of smoothness, were first used in [24] and appear in many other places in proof mining as a quantitative analogue of uniform smoothness.

For the purpose of our case study, our main task is therefore to reformulate the main lemmas on uniform smoothness in terms of the simpler logical modulus. We start with Lemma 2.3:

LEMMA 4.3. *Let (X, ω) be a uniformly smooth Banach space with modulus of uniform smoothness ω. Take $x \neq 0$ and let $F_x := j(x)/\|x\|$. Then*

$$\frac{t\|y\|}{\|x\|} \leq \omega(\varepsilon) \Longrightarrow \|x + ty\| \leq \|x\| \left(1 + \frac{\varepsilon t\|y\|}{\|x\|}\right) + tF_x(y)$$

for any $y \in X$ and $t \in (0, \infty)$.

PROOF. Analogous to the proof of Lemma 2.3. ⊣

We now give a computational version of Lemma 2.5, which is formulated in terms of the modulus ω instead of ρ_X, and implicitly replaces the function $\alpha(b)$ defined there with a function $\xi(b)$ computable in ω.

LEMMA 4.4. *Let (X, ω) be a uniformly smooth Banach space. Suppose that $S \subseteq X$ and $x^* \in \overline{\mathrm{co}(S)}$, that (x_n) is (ϵ_n)-greedy with respect to x^* and S, and $K > 0$ is such that $\sup\{\|y - x^*\| \mid y \in S\} \leq K$. Using our notation $a_n := \|x_n - x^*\|$, for any $n \in \mathbb{N}$ and $b > 0$, if $b \leq a_n$ then*

$$a_{n+1} \leq (1 - \xi(b))a_n + \epsilon_n$$

for

$$\xi(b) := \frac{1}{4} \min \left\{ \frac{b}{K} \cdot \omega\left(\frac{b}{2K}\right), 1 \right\}$$

PROOF. Fix $n \in \mathbb{N}$, $b, \varepsilon > 0$ and $\lambda \in (0, 1)$. Suppose that $b \leq a_n$. If

$$(*) \qquad \frac{\lambda K}{(1 - \lambda)b} \leq \omega(\varepsilon b)$$

then analogously to the proof of Lemma 2.5, by Lemma 4.3 we have

$$\begin{aligned} &\|(x_n - x^*) + \lambda(y - x^*)/(1 - \lambda)\| \\ &\leq \|x_n - x^*\| \left(1 + \frac{\varepsilon b\lambda\|y - x^*\|}{(1 - \lambda)\|x_n - x^*\|}\right) + \lambda F_n\left(\frac{y - x^*}{1 - \lambda}\right) \end{aligned}$$

for any $y \in S$ and thus

$$\begin{aligned} \|(1 - \lambda)x_n + \lambda y - x^*\| &\leq (1 - \lambda)a_n \left(1 + \frac{\varepsilon\lambda K}{1 - \lambda}\right) + \lambda F_n(y - x^*) \\ &\leq (1 - \lambda(1 - \varepsilon K))a_n + \lambda F_n(y - x^*) \end{aligned}$$

where as before we write $F_n := F_{x_n - x^*}$. Now, if we define

$$\lambda_\varepsilon := \frac{1}{2} \min \left\{ \frac{b \cdot \omega(\varepsilon b)}{K}, 1 \right\}$$

then $(*)$ holds for $\lambda := \lambda_\varepsilon$, and so using again the fact that

$$\inf\{F_n(y - x^*) \mid y \in S\} \leq 0$$

we have

$$
\begin{aligned}
&\inf\{\|(1-\lambda)x_n+\lambda y-x^*\| \mid y\in S, \lambda\in[0,1)\}\\
&\leq \inf\{(1-\lambda_\varepsilon(1-\varepsilon K))a_n+\lambda_\varepsilon F_n(y-x^*) \mid y\in S\}\\
&\leq (1-\lambda_\varepsilon(1-\varepsilon K))a_n
\end{aligned}
$$

and so finally, setting $\varepsilon := 1/2K$ we have

$$a_{n+1}\leq (1-\lambda_\varepsilon/2)a_n+\epsilon_n=(1-\xi(b))a_n+\epsilon_n$$

for $\xi(b)$ defined as in the statement of the lemma. $\dashv$

We can now simply combine Lemma 4.3 with Lemmas 4.1 and 4.2 to get our main result - a computational interpretation (and qualitative strengthening) of Theorem 2.6.

THEOREM 4.5. *Let (X,ω) be a uniformly smooth Banach space. Suppose that $S\subseteq X$ and $x^*\in\overline{\mathrm{co}(S)}$, that (x_n) is (ϵ_n)-greedy with respect to x^* and S, and $K>0$ is such that $\sup\{\|y-x^*\| \mid y\in S\}\leq K$.*

(a) Suppose that $\epsilon_n\to 0$ as $n\to\infty$ with rate f. Then $x_n\to x^*$ as $n\to\infty$ with rate

$$\Phi(b):=f(\mu_{\omega,b,K})+\left\lceil\frac{L}{\mu_{\omega,b,K}}\right\rceil+1$$

where

$$\mu_{\omega,b,K}:=\frac{b}{16}\min\left\{\frac{b}{2K}\cdot\omega\left(\frac{b}{4K}\right),1\right\}.$$

and $L>0$ is any bound on $(\|x_n-x^*\|)$.

(b) In the special case that $\sum_{i=0}^\infty \epsilon_i\leq L$ with rate f, the rate of convergence for $x_n\to x^*$ as $n\to\infty$ can be simplified to

$$\Phi(b):=f\left(\frac{b}{2}\right)+\left\lceil\frac{K+L}{2\mu_{\omega,b,K}}\right\rceil$$

for $\mu_{\omega,b,K}$ as defined in part (a).

PROOF. We apply Lemmas 4.2 and 4.1 respectively to obtain parts (a) and (b), with $a_n:=\|x_n-x^*\|$ and $\xi(b)$ defined as in Lemma 4.3 for $b>0$, and extended to $b=0$ by setting $\xi(0):=0$, noting that (a_n) and $\xi(b)$ satisfy the premises of Lemmas 4.1 and 4.2 by Lemma 4.3, where for the case $b=0$ we have $a_{n+1}\leq a_n+\epsilon_n$ by definition of (x_n) being (ϵ_n)-greedy with respect to x^*. $\dashv$

§5. Extensions. There are many different ways in which our analysis can be further generalised by exploiting our careful examination of the proof. We briefly discuss two such extensions, highlighting the fact that much of the power of proof theoretic reasoning lies not merely in the ability to directly unwind a particular proof, but in deriving qualitative strengthenings of results by analysing those proofs further.

5.1. Rates of convergence for fixed step sizes. A natural question to ask once we have established convergence of greedy approximations schemes for *optimal* step sizes is whether we can establish an analogous result when the step sizes are fixed in advance. A number of results of this kind, for $\lambda_n := 1/(n+1)$, are provided in [6] in the special case that X has modulus of smoothness of the form $\rho(u) \leq \gamma u^t$ for $t > 1$, though no such result is given in the general case, the authors simply noting in reference to the proof analysed in this paper that *"The stepwise selection of $\lambda = \lambda_n$ in the above proof apparently depends upon the modulus of smoothness $\rho(u)$"*. We now provide a general convergence theorem for fixed step sizes, in terms of the alternative modulus of smoothness ω, exploiting our quantitative analysis. We label a sequence (x_n) (λ_n)-(ε_n)-greedy with respect to x^* and S if for all $n \in \mathbb{N}$:

$$\|x_{n+1} - x^*\| \leq \inf \{\|(1-\lambda_n)x_n + \lambda_n y - x^*\| \mid y \in S\} + \epsilon_n$$

where now $(\lambda_n) \subset [0,1)$ is some fixed sequence. We also assume that

$$\inf \{\|(1-\lambda_n)x_n + \lambda_n y - x^*\| \mid y \in S\} \leq \|x_n - x^*\|$$

for all $n \in \mathbb{N}$ i.e. step sizes in $[0,1)$ are meaningful in that they do not worsen our estimate. By a simple adaptation of the results of the previous section we obtain the following:

THEOREM 5.1. *Let (X, ω) be a uniformly smooth Banach space. Suppose that $S \subseteq X$ and $x^* \in \text{co}(S)$, that (x_n) is (λ_n)-(ε_n)-greedy with respect to x^* and S for some fixed step sizes $(\lambda_n) \subset [0,1)$, $K > 0$ is such that $\sup\{\|y - x^*\| \mid y \in S\} \leq K$, and $\sum_{i=0}^{\infty} \epsilon_i \leq L$ with rate f. Finally, suppose that $\xi : (0,\infty) \to (0,1)$ is such that for all $b > 0$*

$$2\xi(b) \leq \lambda_n \leq \frac{1}{2} \min \left\{ \frac{b}{K} \cdot \omega\left(\frac{b}{2K}\right), 1 \right\}$$

for all $n \in [f(b/2); \Phi_\xi(b)]$, where Φ_ξ is defined in terms of ξ as in Lemma 4.1. Then $x_n \to x^*$ as $n \to \infty$ with rate Φ_ξ.

PROOF. Fix $n \in [f(b/2); \Phi(b)]$ and $b > 0$. Suppose that $b \leq a_n$. Then using the upper bound on λ_n we have

$$\frac{\lambda_n K}{(1-\lambda_n) b} \leq \omega(b/2K)$$

and so just as in the proof of Lemma 4.4 (for $\varepsilon := 1/2K$) it follows that

$$\|(1-\lambda_n)x_n + \lambda_n y - x^*\| \leq (1-\lambda_n/2)a_n + \lambda_n F_n(y-x^*)$$

for $a_n := \|x_n - x^*\|$ and $F_n := F_{x_n - x^*}$ as usual, and thus also

$$a_{n+1} \leq (1-\lambda_n/2)a_n + \epsilon_n \leq (1-\xi(b))a_n + \epsilon_n$$

where for the second inequality we now use the lower bound on λ_n. Since we also have $a_{n+1} \leq a_n + \epsilon_n$ (for any $n \in \mathbb{N}$), we have established that the recurrence inequality of Lemma 4.1 is satisfied for ξ (extended by $\xi(0) := 0$) but now only for $n \in [f(b/2); \Phi(b)]$. But a simple inspection of the proof of Lemma 4.1 reveals that we only require the full recurrence inequality to hold in that range to establish the existence of some $n \in [f(b/2); \Phi(b)]$ such that $a_n < b/2$ (though we also need $a_{n+1} \leq a_n + \epsilon_n$ for any $n \in \mathbb{N}$), and then $a_{n+k} < b$ for any $k \geq \Phi(b)$ follows immediately. $\dashv$

For example, by setting $\omega(\varepsilon) := \varepsilon^r$, $\xi(b) := b^s$ and $\lambda_n := n^{-t}$ we can use Theorem 5.1 to obtain conditions on r, s, t such that convergence of (x_n) is guaranteed with a simple rate of the form $O(n^{-k})$, a result similar in spirit to e.g. [6, Theorem 3.5], but now for more general step-sizes. However, we stress that Theorem 5.1 is not proposed as an improvement to the more refined convergence rates of [6], but rather a simple illustration of how our proof-theoretic analysis can be readily extended.

5.2. Greedy algorithms in hyperbolic spaces. A far deeper extension of our analysis would involve working over a more general class of spaces. One candidate here is represented by *hyperbolic spaces*, which informally speaking are metric spaces (X, d) that come equipped with a function $\oplus$ and associated axioms that allow one to reason abstractly about convex combinations $(1-\lambda)x \oplus \lambda y$ of elements $x, y \in X$ (see [16] for a formal definition widely used in the context of proof mining, and a discussion on the relationship of this definition to various others in the literature). As such, the notion of an (ϵ_n)-greedy algorithm can be readily extended from normed spaces to this more general setting, as a sequence (x_n) of elements satisfying

$$d(x_{n+1}, x^*) \leq \inf\{d((1-\lambda)x \oplus \lambda y, x^*) \mid y \in S, \lambda \in [0,1)\} + \epsilon_n.$$

However, in order to establish *convergence* we would need a suitable notion of uniform smoothness for hyperbolic spaces, which would in turn presumably require an appropriate notion of duality. Only recently have such generalisations of concepts from normed spaces been explored, notably via the smooth hyperbolic spaces of Pinto [39] and the dual systems of Pischke [41]. We therefore propose that the convergence of greedy algorithms in hyperbolic spaces represents an interesting case study that could build on this recent work, though we leave this open for now.

§6. Outlook: Formalising and automating proof mining. As stated at the beginning of the paper, both our choice of case study and the manner in which we have presented our proof-theoretic analaysis were intended to prove insights into how the proof mining process could potentially be formalised and partially automated. We conclude in this spirit with a short list of ideas and conjectures in that direction. While these have some parallels with the general programme of *formalising proofs with explicit computational content* articulated by Koutsoukou-Argyraki in [28], rather than focusing on toy examples, everything that we propose is firmly tethered to a real and representative case study in proof mining.

Firstly, can we develop domain specific logical systems specifically designed for proof mining in particular areas? We envisage that these might contain high-level, sound inference rules of the kind we considered in Section 3, which when embedded into a proof assistant would allow us to automate various operations on derivations in those systems. In particular, we would hope that the use of high-level rules would facilitate the automatic extraction of *readable* bounds by implicitly suppressing unnecessary bureaucracy.

More generally, it is natural to ask whether are there useful procedures and tactics that can be developed organically within a proof assistant that allow for proof mining *as done in practice* to be more easily formalised. Here the aim is to circumvent any *formal* application of the Dialectica interpretation or associated logical metatheorems (many implementations of which already exist, most recently in Lean by Cheval[2]), which is of course not how proof mining is done in reality. Ideas in this direction include: tactics that draw quantifiers out of formulas in the proof assistant's ambient type theory; the automatic identification of assumptions that can potentially be weakened (as done in an ad-hoc way in Section 4.1); tactics that dynamically inform us how much of a quantified statement is currently being used (as illustrated in our manipulations on proof trees in Section 3, and in particular our extension to fixed step-sizes in Section 5.1); or procedures on a meta-level that automatically lift proofs to more general spaces (as discussed in the context of the present case study in Section 5.2).[3]

Finally, what libraries would be useful for practical proof mining? One simple suggestion here are libraries that translate mathematical moduli to their logical counterpart. A concrete example would be a library that formulates standard lemmas on uniformly smooth Banach spaces phrased in terms of the modulus ρ_X to equivalent results in terms of the *proof*

---

[2] `https://github.com/hcheval/formalized-proof-mining`

[3] The idea that proof assistants could be utilised to generalise proofs to more abstract spaces was communicated to the author by N. Pischke.

theoretic modulus ω (as done in Section 4.2), where here one could potentially consider a partially automated translation from the former to the latter. A more far reaching effort would be the systematic formalisation of abstract convergence lemmas (two examples of which are given in Section 4.1) along with their associated computational content:[4] Convergence lemmas of this kind are used throughout optimization (as recently surveyed in [9]), and quantitative variants thereof play a central role in proof mining, as discussed in [36]. In particular, the quantitative analysis of *stochastic* recurrence inequalities via martingale methods lies at the heart of recent applications of proof mining in stochastic optimization, including [35, 38, 37, 43, 45]. The potential use of such a library in conjunction with automated reasoning techniques to generate computer-checked, quantitative convergence proofs for both deterministic and stochastic algorithms is extremely exciting prospect, but one we firmly postpone to future work!

Acknowledgements. The author is grateful to the anonymous referee for their comments, and to Sam Sanders and particularly Nicholas Pischke for helpful suggestions and discussion.

Funding. The author was partially supported by the Engineering and Physical Sciences Research Council grant EP/W035847/1. No new data were created during this study.

REFERENCES

[1] Jeremy Avigad, Edward T. Dean, and Jason Rute, *A metastable dominated convergence theorem*, ***Journal of Logic & Analysis***, vol. 4 (2012), no. 3, pp. 1–19.

[2] Jeremy Avigad, Philipp Gerhardy, and Henry Towsner, *Local stability of ergodic averages*, ***Transactions of the American Mathematical Society***, vol. 362 (2010), no. 1, pp. 261–288.

[3] Charles Chidume, ***Geometric properties of Banach spaces and nonlinear iterations***, Springer, 2009.

[4] Patrick L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, ***Inherently parallel algorithms in feasibility and optimization and their applications***, Studies in Computational Mathematics, vol. 8, Elsevier, 2001, pp. 115–152.

[5] ———, *Fejér monotonicity in convex optimization*, ***Encyclopedia of Optimization***, Springer, New York, 2nd ed., 2009, pp. 1016–1024.

[6] C. Darken, M. J. Donahue, L. Gurvits, and E. Sontag, *Rates of convex approximation in non-Hilbert spaces*, ***Constructive Approximation***, vol. 13 (1997), pp. 187–220.

[7] Bruno Dinis and Pedro Pinto, *Quantitative results on the multi-parameters proximal point algorithm*, ***Journal of Convex Analysis***, vol. 28 (2021), no. 3, pp. 729–750.

[4]Some initial, ongoing work in this direction can be found at `https://github.com/t-powell/quantitative-recurrence-inequalities`.

[8] RICHARD FINDLING and ULRICH KOHLENBACH, *Rates of convergence and metastability for Chidume's algorithm for the approximation of zeros of accretive operators in Banach spaces*, ***Numerical Functional Analysis and Optimization***, vol. 45 (2024), pp. 216–233.

[9] BARBARA FRANCI and SERGIO GRAMMATICO, *Convergence of sequences: A survey*, ***Annual Reviews in Control***, vol. 53 (2022), pp. 161–186.

[10] PHILIPP GERHARDY and ULRICH KOHLENBACH, *General logical metatheorems for functional analysis*, ***Transactions of the American Mathematical Society***, vol. 360 (2008), pp. 2615–2660.

[11] KURT GÖDEL, *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*, ***dialectica***, vol. 12 (1958), no. 3–4, pp. 280–287.

[12] WILLIAM A. HOWARD, *Hereditarily majorizable functionals of finite type*, ***Metamathematical investigation of intuitionistic arithmetic and analysis*** (Anne S. Troelstra, editor), Lecture Notes in Mathematics, vol. 344, Springer-Verlag, 1973.

[13] LEE JONES, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, ***The Annals of Statistics***, vol. 20 (1992), no. 1, pp. 608–613.

[14] ULRICH KOHLENBACH, *A quantitative version of a theorem due to Borwein-Reich-Shafrir*, ***Numerical Functional Analysis and Optimization***, vol. 22 (2001), pp. 641–656.

[15] ———, *Uniform asymptotic regularity for Mann iterates*, ***Journal of Mathematical Analysis and Applications***, vol. 279 (2003), pp. 531–544.

[16] ———, *Some logical metatheorems with applications in functional analysis*, ***Transactions of the American Mathematical Society***, vol. 357 (2005), no. 1, pp. 89–128.

[17] ———, ***Applied Proof Theory: Proof Interpretations and their Use in Mathematics***, Springer Monographs in Mathematics, Springer, 2008.

[18] ———, *Recent progress in proof mining in nonlinear analysis*, ***IFCoLoG Journal of Logics and their Applications***, vol. 10 (2017), no. 4, pp. 3361–3410.

[19] ———, *A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space*, ***Foundations of Computational Mathematics***, vol. 19 (2019), pp. 83–99.

[20] ———, *Proof-theoretic methods in nonlinear analysis*, ***Proceedings of the International Congress of Mathematicians 2018***, vol. 2, World Scientific, 2019, pp. 61–82.

[21] ———, *Quantitative analysis of a Halpern-type Proximal Point Algorithm for accretive operators in Banach spaces*, ***Journal of Nonlinear and Convex Analysis***, vol. 21 (2020), no. 9, pp. 2125–2138.

[22] ULRICH KOHLENBACH and ANGELIKI KOUTSOUKOU-ARGYRAKI, *Rates of convergence and metastability for abstract cauchy problems generated by accretive operators*, ***Journal of Mathematical Analysis and Applications***, vol. 423 (2015), pp. 1089–1112.

[23] ULRICH KOHLENBACH and LAURENȚIU LEUŞTEAN, *Mann iterates of directionally nonexpansive mappings in hyperbolic spaces*, ***Abstract and Applied Analysis***, vol. 2003 (2003), no. 8, pp. 449–477.

[24] ———, *On the computational content of convergence proofs via Banach limits*, ***Philosophical Transactions of the Royal Society A***, vol. 370 (2012), pp. 3449–3463.

[25] ULRICH KOHLENBACH, LAURENȚIU LEUŞTEAN, and ADRIANA NICOLAE, *Quantitative results on Fejér monotone sequences*, ***Communications in Contemporary Mathematics***, vol. 20 (2018), p. 42pp.

[26] Ulrich Kohlenbach and Nicholas Pischke, *Proof theory and nonsmooth analysis*, ***Philosophical Transactions of the Royal Society A***, vol. 381 (2023), no. 2248, p. 20220015.

[27] Ulrich Kohlenbach and Thomas Powell, *Rates of convergence for iterative solutions of equations involving set-valued accretive operators*, ***Computers and Mathematics with Applications***, vol. 80 (2020), pp. 490–503.

[28] Angeliki Koutsoukou-Argyraki, *On preserving the computational content of mathematical proofs: Toy examples for a formalising strategy*, ***Connecting with Computability. CiE 2021***, LNCS, vol. 12813, Springer, 2021, pp. 285–296.

[29] Georg Kreisel, *On the interpretation of non-finitist proofs, Part I*, ***The Journal of Symbolic Logic***, vol. 16 (1951), pp. 241–267.

[30] ———, *On the interpretation of non-finitist proofs, Part II: Interpretation of number theory. applications.*, ***The Journal of Symbolic Logic***, vol. 17 (1952), pp. 43–58.

[31] Laurențiu Leuştean, Adriana Nicolae, and Andrei Sipoş, *An abstract proximal point algorithm*, ***Journal of Global Optimization***, vol. 72 (2018), no. 3, pp. 553–577.

[32] Morenikeji Neri, *Quantitative strong laws of large numbers*, ***Electronic Journal of Probability***, vol. 30 (2024), pp. 1–22.

[33] ———, *A finitary Kronecker's lemma and large deviations in the strong law of large numbers on Banach spaces*, ***Annals of Pure and Applied Logic***, vol. 176 (2025), no. 6, p. 103569.

[34] Morenikeji Neri and Nicholas Pischke, *Proof mining and probability theory*, ***Forum of Mathematics, Sigma***, (2025), To appear.

[35] Morenikeji Neri, Nicholas Pischke, and Thomas Powell, *On the asymptotic behaviour of stochastic processes, with applications to supermartingale convergence, Dvoretzky's approximation theorem, and stochastic quasi-Fejér monotonicity*, ***ArXiV e-prints***, (2025), arXiv, math.OC, 2504.12922.

[36] Morenikeji Neri and Thomas Powell, *A computational study of a class of recursive inequalities*, ***Journal of Logic and Analysis***, vol. 15 (2023), no. 3, pp. 1–48.

[37] ———, *On quantitative convergence for stochastic processes: Crossings, fluctuations and martingales*, ***Transactions of the American Mathematical Society, Series B***, vol. 12 (2025), pp. 974–1019.

[38] ———, *A quantitative Robbins-Siegmund theorem*, ***Annals of Applied Probability***, (2025), To appear.

[39] Pedro Pinto, *Nonexpansive maps in nonlinear smooth spaces*, ***Transactions of the American Mathematical Society***, vol. 377 (2024), no. 9, pp. 6379–6426.

[40] Nicholas Pischke, *Logical metatheorems for accretive and (generalized) monotone set-valued operators*, ***Journal of Mathematical Logic***, vol. 24 (2024), no. 2, p. 59pp.

[41] ———, *Duality, Fréchet differentiability and Bregman distances in hyperbolic spaces*, ***Israel Journal of Mathematics***, (2025), To appear.

[42] ———, *Generalized Fejér monotone sequences and their finitary content*, ***Optimization***, vol. 74 (2025), no. 14, pp. 3771–3838.

[43] ———, *Mean-square and linear convergence of a stochastic proximal point algorithm in metric spaces of nonpositive curvature*, ***ArXiV e-prints***, (2025), arXiv, math.OC, 2510.10697.

[44] ———, *A proof-theoretic metatheorem for nonlinear semigroups generated by an accretive operator and applications*, ***Selecta Mathematica***, vol. 31 (2025), no. 32.

[45] Nicholas Pischke and Thomas Powell, *Asymptotic regularity of a generalised stochastic Halpern scheme with applications*, ***ArXiV e-prints***, (2024), arXiv,

math.OC, 2411.04845.

[46] Thomas Powell and Franziskus Wiesnet, *Rates of convergence for asymptotically weakly contractive mappings in normed spaces*, ***Numerical Functional Analysis and Optimization***, vol. 42 (2021), no. 15, pp. 1802–1838.

[47] Vladimir N. Temlyakov, ***Greedy approximation***, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2011.

[48] Ambuj Tewari, Pradeep Ravikumar, and Inderjit Dhillon, *Greedy algorithms for structurally constrained high dimensional problems*, ***Advances in Neural Information Processing Systems 24 (NIPS 2011)***, 2011.

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF BATH
BATH, BA2 7AY, UNITED KINGDOM
E-mail: trjp20@bath.ac.uk