# Proof mining and high-level proof theoretic reasoning: A case study on greedy approximation schemes

Thomas Powell

February 25, 2025

#### Abstract

We carry out a proof theoretic analysis of a proof by Darken et al. that establishes the convergence of 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. Our quantitative results represent a first proof theoretic study of greedy approximation schemes, whose applications include learning theory and neural networks. We use this case study 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 nonconstructive proofs in those areas, and can result in both quantitative and qualitative improvements of the original theorems. While the use of proof theoretic techniques for this purpose was first proposed by Kreisel in the 1950s [20, 21], the field was only brought to maturity in the last thirty years or so through the work of Kohlenbach and others, an overview of which is documented in the textbook [10] and the more recent survey papers [11, 13].

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 point theory (e.g. [7, 8, 16]) were explained by Kohlenbach in [9], while the more recent metatheorems of Pichke [24] cover a spate of case studies on accretive and monotone set-valued operators (including but not limited to [3, 12, 14, 15, 18, 22]).

In the case studies, proof-theoretic 'workings' are often suppressed so that the main results can be presented in standard mathematical language. As such, it is not necessarily obvious precisely how proof theory was used to obtain those results. Metatheorems, on the other hand, require complex logical machinery, and while they act as a guide in that they set out what kind of quantitative results are in principle possible within a given setting, new applications still require mathematical ingenuity, and are in any case often chosen to explore landscape hitherto unexplained by existing metatheorems. Thus the route from a metatheorem to a new mathematical theorem is rarely straightforward.

This paper aims to strike a balance between the purely mathematical case studies and the logically sophisticated metatheorems: 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 manipulations that an applied proof theorist might carry out implicitly. We represent these manipulations using a informal type of proof tree, which is certainly not standard in applied proof theory, and technically not necessary to prove the main results. But it is hoped that this presentation will be of independent interest, hinting at ways in which proof mining could be effectively implemented 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. In this way we capture, in a semi-formal manner, the important phenomenon that in proof mining, it is rarely necessary to fully formalise a proof to extract a meaningful program: It is the high-level structure that matters.

Our starting point is an elegant proof on the convergence of greedy approximation schemes, given as Theorem 3.4 of [2]. This theorem and its subsequent analysis exhibit 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;
- (iii) The initial quantitative analysis can be extended to produce several qualitative strengthenings of the original result, including a weakening of the condition on the error terms, and an exension to fixed step sizes.

Our work represents a first application of proof theory to greedy approximation schemes, and we anticipate that future case studies in this direction would be interest, given that there exist a wealth of related nonconstructive convergence proofs that depend geometric properties of Banach spaces (e.g. [28, 29]), which in turn are directly relevant to machine learning. However, we consider our presentation of the extraction process to be a 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, given the increasing importance and widespread use of proof assistants (whose potential role in connection to proof mining is discussed in [19]), we envisage that high-level descriptions of the underlying proof transformations might provide insight into the kind of mathematical libraries and tactics that would be helpful in both formalising or even partially automating the proof mining process.

#### Structure of the paper

This paper is intended to be of broad interest, presenting a representative case study in proof mining while giving insight into the underlying extraction process. It has been deliberately written to appeal to a reader who is more interested in how an applied proof theorist might manipulate nonconstructive proofs in general, than in the specific mathematical details of the situation at hand (in this case Banach spaces and their geometric properties). For that reason, all of the necessary mathematical background is provided in a detailed and self-contained manner in Section 2, and in the remainder of the paper, logic and analysis are 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.

We begin in Section 2 by introducing greedy approximation schemes and presenting, in full detail, the proof that we will analyse. In Section 3 we then formulate our task in proof-theoretic terms by setting out the overall logical structure of the main proof, before giving a detailed description of the extraction process. This 'informal analysis' is one of the most important parts of the paper, where we deliberately depart from the traditional presentation of proof mining case studies by looking in depth at how a nonconstructive mathematical proof can be transformed through examining the structure of that proof.

The two sections that follow present the resulting quantitative theorem: First, we deal with the overarching combinatorial structure of the proof itself and produce an abstract convergence result for sequences of reals that satisfy a particular recursive inequality. Then we take a closer look at how uniform smoothness is used in the proof and reformulate this in terms of the corresponding proof-theoretic modulus, before putting everything together and stating the main quantitative result. We conclude by discussing the various ways in which this main result can be extended in a *qualitative* way in Section 6.

# 2 Mathematical background

In this section we present 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. [2] in which our chosen proof is just one among a series of results on convex approximation schemes in Banach spaces, and the more recent textbook on greedy algorithms by Temlyakov [28]. For further background on Banach spaces, [1] is a standard textbook.

#### 2.1 Greedy approximation schemes

The following definitions and notation are taken from [2]. Let X be a real Banach space, and suppose that  $S \subseteq X$  is some arbitrary set. Let 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  $n \ge 1, y_1, \ldots, y_n \in S$  and  $\lambda_1, \ldots, \lambda_n \in [0, 1]$  with  $\lambda_1 + \ldots + \lambda_n = 1$ . Now suppose that  $x^* \in \overline{\operatorname{co}(S)}$ , where  $\overline{\operatorname{co}(X)}$  denotes the closure of  $\operatorname{co}(X)$ . A natural question arises:

Can we construct incremental approximates to  $x^*$  from elements of co(X) 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

$$x_0 \in S$$
 and  $x_{n+1} = (1 - \lambda_n)x_n + \lambda_n y_n$  for  $y_n \in S$  and  $\lambda_n \in [0, 1)$  (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  $\lambda_n$  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

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

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

To account for the fact that S may not be compact, and thus the infimum in the above definition not attained, we loosen the definition to incorporate error terms  $\{\epsilon_n\}$ . This is the main definition of greedy approximation scheme that will be used in what follows:

Definition 2.2. Let  $\{\epsilon_n\}$  be a sequence of positive real numbers. A sequence  $\{x_n\}$  of elements of X is called  $\{\epsilon_n\}$ -greedy with respect to  $x^* \in X$  and S if

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

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

Greedy approximation schemes in Hilbert spaces are studied by Jones [6], where convergence to  $x^* \in \overline{\operatorname{co}(S)}$  with rate  $\mathcal{O}(1/\sqrt{n})$  proven. Jones also highlights the connection of such schemes to artificial neural networks 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 [6, Section 4] or [2, 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 [2] is  $\mathbb{R}^2$  under the  $L^1$  norm: Here, if we take  $S := \{(0, -1), (2, 1/2), (-2, 1/2)\}$  then  $(0, 0) \in \operatorname{co}(S)$ , 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. [1, Chapter 2] and [28, 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 mapping function is just the inner product  $j(x) = \langle x, - \rangle$ , and as such the duality mapping often plays the role of mimicking an inner product in Banach spaces. Crudely speaking, the nicer the duality mapping, the more X behaves like a Hilbert space. In this respect, an important notion is the modulus of smoothness defined by:

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

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. [28, Lemma 6.1]) connects the duality mapping function 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.** Take  $x \neq 0$  and let  $F_x := j(x)/||x||$ . Then

$$||x + ty|| \le ||x|| (1 + 2\rho_X (t||y|| / ||x||)) + tF_x(y)$$

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

*Proof.* From the definition of  $\rho_X$  we obtain

$$||x + ty|| + ||x - ty|| \le 2||x|| (1 + \rho_X (t||y|| / ||x||))$$

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

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

the result follows.

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

$$0 \le ||x + ty|| - ||x|| - tF_x(y) \le 2\rho_X(t||y|| / ||x||)$$

For ||x|| = ||y|| = 1 it follows that

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

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, y. In other words, X has a *uniformly Fréchet 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  $\{\epsilon_n\}$ -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 [2], with the proof only very slightly reformulated from its original presentation.

We first require a lemma that applies smoothness to greedy algorithms. The result below incorporates Lemma 3.3 of [2], along with part of the main proof of Theorem 3.4 from the same paper, and we have deliberately re-organised 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 [2]). Let X be a Banach space with modulus of smoothness  $\rho_X$ . Suppose that  $S \subseteq X$  and  $x^* \in \overline{\operatorname{co}(X)}$ , that  $\{x_n\}$  is  $\{\epsilon_n\}$ -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 b > 0, if  $b \leq \|x_n - x^*\|$  then

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

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.* For  $y \in S$  and  $\lambda \in [0, 1)$ , writing

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

and applying Lemma 2.3 we have

$$\|(1-\lambda)x_{n} + \lambda y - x^{*}\| \leq (1-\lambda)\left(1 + 2\rho_{X}\left(\frac{\lambda\|y - x^{*}\|}{(1-\lambda)\|x_{n} - x^{*}\|}\right)\right)\|x_{n} - x^{*}\| + \lambda F_{n}(y - x^{*})$$
<sup>(2)</sup>

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

$$(1-\lambda)\rho_X\left(\frac{\lambda\|y-x^*\|}{(1-\lambda)\|x_n-x^*\|}\right) \le (1-\lambda)\rho_X\left(\frac{\lambda K}{(1-\lambda)b}\right) = \frac{\lambda K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)}$$
(3)

for  $u(b, \lambda)$  as defined in the statement of the result. Substituting (3) into (2) we obtain

$$\|(1-\lambda)x_n + \lambda y - x^*\| \le \left[1 - \lambda \left(1 - \frac{2K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)}\right)\right] \|x_n - x^*\| + \lambda F_n(y - x^*)$$
(4)

Now, since  $x^* \in \overline{\operatorname{co}(S)}$ , we can make  $F_n(y - x^*)$  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  $z \in \operatorname{co}(S)$  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

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

Therefore  $\inf\{F_n(y - x^*) \mid y \in S\} \le 0$ , and more generally

$$\inf\left\{ \left[ 1 - \lambda \left( 1 - \frac{2K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)} \right) \right] \|x_n - x^*\| + \lambda F_n(y - x^*) \mid y \in S, \lambda \in [0,1) \right\}$$
$$\leq \left[ 1 - \sup\left\{ \lambda \left( 1 - \frac{2K}{b} \cdot \frac{\rho_X(u(b,\lambda))}{u(b,\lambda)} \right) \mid \lambda \in [0,1) \right\} \right] \|x_n - x^*\|$$
$$= (1 - \alpha(b)) \|x_n - x^*\|$$

Thus combining the above with (4) and the definition of being  $\{\epsilon_n\}$ -greedy we have

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

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 [2]). Let X be a Banach space with modulus of smoothness  $\rho_X$  and  $S \subseteq X$  be bounded. Suppose that  $x^* \in \overline{\operatorname{co}(X)}$ , and that  $\{x_n\}$  is  $\{\epsilon_n\}$ -greedy with respect to  $x^*$  and S for some sequence  $\{\epsilon_n\}$  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  $\{\epsilon_n\}$ -greedy, we have  $a_{n+1} \leq a_n + \epsilon_n$ , and thus more generally

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

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  $\sup\{||y - x^*|| \mid y \in S\} \leq K$ . We must have  $a_{\infty}/2 \leq a_n$  for n sufficiently large, and so applying Lemma 2.5 we have

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

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

$$a_{\infty} \le (1 - \alpha(a_{\infty}/2))a_{\infty}$$

and therefore  $\alpha(a_{\infty}/2) \leq 0$ . Now consider the definition of  $\alpha(b)$ . We have  $u(b, \lambda) \to 0$  as  $\lambda \to 0$ , and therefore by uniform smoothness of X it follows that

$$\frac{\rho_X(u(\lambda,b))}{u(\lambda,b)} \to 0 \quad \text{as $\lambda \to 0$}$$

from which we see that  $\alpha(b) > 0$  for any b > 0, a contradiction for  $b = a_{\infty}/2$ . 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 highlevel structure of the proof of Theorem 2.6. 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 ([10] is the standard reference). However, here we aim to show how one might transform a proof "by hand" and in an informal manner, ignoring parts of the proof that are uninteresting and focusing on the key *mathematical* rather than *logical* steps. The end result presented in Sections 4 and 5 (and there re-proven in a conventional mathematical style) is a computational version of Theorem 2.6.

In relation to point 3 above, it will be helpful to give one basic definition:

Definition 3.1. For a sequence  $\{x_n\}$  of elements in some metric space (X, d)along with a point  $x^* \in X$ , a rate of convergence for  $x_n \to x^*$  as  $n \to \infty$  is a function  $f: (0, \infty) \to \mathbb{N}$  satisfying

$$\forall \varepsilon > 0 \,\forall n \ge f(\varepsilon)[d(x_n, x^*) < \varepsilon]$$

Our main task will be to find a computable rate of convergence for  $x_n \to x^*$  as  $n \to \infty$  in the context of Theorem 2.6.

## 3.1 The overall structure of the proof of Theorem 2.6

For the remainder of this section, we will fix several things. First, we let X denote a Banach space with  $\rho$  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  $\{\epsilon_n\}$  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 require – Lemma 2.5 – can then be represented as the single inference

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

where " $\{\epsilon_n\}$ -greedy" is shorthand for the statement " $\{x_n\}$  is  $\{\epsilon_n\}$ -greedy" and the predicate P is defined by

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

for  $\alpha(b)$  defined as in Lemma 2.5. Rather than going ahead and analysing the somewhat intricate proof of the 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

$$\frac{\{\epsilon_n\}\text{-greedy}}{\frac{\forall n[a_{n+1} \le a_n + \epsilon_n]}{a_n \to a_\infty}} \sum_{i=0}^{\infty} \epsilon_i < \infty$$
(\Gamma\_1)

Of course, there are a number of additional (elementary) steps involved in inferring  $a_n \to a_\infty$  from the two premisis, but for now the main thing we highlight is that the property of being  $\{\epsilon_n\}$ -greedy is used in a weak way, in that we only require  $a_{n+1} \leq a_n + \epsilon_n$ . We label this prooftree ( $\Gamma_1$ ) and will in future write it in shorthand as

$$\|_{\Gamma_1} \\ a_n \to a_\infty$$

We use a similar shorthand for other proof trees. 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. We start as follows:

$$\frac{\{\epsilon_n\}\text{-greedy} \quad x^* \in \overline{\operatorname{co}(S)}}{\frac{\forall b > 0, n[b \le a_n \implies P(b, a_n, a_{n+1}, \epsilon_n)]}{\frac{\forall n[a_{\infty}/2 \le a_n \implies P(a_{\infty}/2, a_n, a_{n+1}, \epsilon_n)]}{\forall n \ge N[P(a_{\infty}/2, a_n, a_{n+1}, \epsilon_n)]}} \quad \forall n \ge N[a_{\infty}/2 \le a_n]}$$

$$(\Gamma_2)$$

and we label this derivation ( $\Gamma_2$ ). Here, N is simply some natural number that we know to exist by definition of  $a_{\infty}$ , and the conclusion of ( $\Gamma_2$ ) is simply the statement that

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

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

$$\frac{\{a_{\infty} > 0\}}{\left\|\Gamma_{2}\right\|} \qquad \qquad \left\|\Gamma_{1} \qquad \frac{\sum_{i=0}^{\infty} \epsilon_{i} < \infty}{\epsilon_{n} \to 0}\right\|}{P(a_{\infty}/2, a_{n}, a_{n+1}, \epsilon_{n})} \qquad \qquad (\Gamma_{3})$$

where we now take the limit as  $n \to \infty$  to establish

$$a_{\infty} \le (1 - \alpha (a_{\infty}/2))a_{\infty}.$$

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  $\{a_{\infty} > 0\}$ :

$$\begin{cases}
\{a_{\infty} > 0\} & X \text{ is U. S.} \\
\|\Gamma_3 & \overline{\forall b > 0[\alpha(b) > 0]} \\
P(a_{\infty}/2, a_{\infty}, a_{\infty}, 0) & \overline{\alpha(a_{\infty}/2) > 0} \\
\hline
\frac{\bot}{a_{\infty} = 0} \implies_{I}
\end{cases} (\Gamma_4)$$

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

$$\frac{ \begin{aligned} & \|\Gamma_1 & \|\Gamma_4 \\ a_n \to a_\infty & a_\infty = 0 \\ \hline & a_n \to 0 \end{aligned}} \tag{($\Gamma_5$)}$$

and thus expanding the definition of  $(\Gamma_6)$  in full would give us a complete (informal) 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 of the derivation 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 unwrap the definition of  $a_{\infty} = 0$  and try to formulate it in the simplest possible logical terms. It is not difficult to show that in the special case that  $\{a_n\}$  is a sequence of nonnegative reals, the otherwise more complex statement the limit of  $\{a_n\}$  is equal to zero is equivalent to

$$\forall b > 0, m \in \mathbb{N} \,\exists n \ge m[a_n < b] \tag{5}$$

So assuming that we have proven (5), 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 general statement  $a_n \to a_\infty$  at all: It is sufficient to notice that two premises used to prove this in  $(\Gamma_1)$  – which is in general a more complex argument – can be combined with (5) 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  $n \ge f(b)$ , using that  $\forall n[a_{n+1} \le a_n + \epsilon_n]$  we must have

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

Now supposing that we have a computable bound for n in (5) for m := f(b) i.e. a function  $\Phi : (0, \infty) \to \mathbb{N}$  such that

$$\forall b > 0 \exists n [f(b) \le n \le \Phi(b) \land a_n < b]$$

then  $b \mapsto \Phi(b/2)$  must be a rate of convergence for  $a_n \to 0$  since  $a_n < b$ implies that  $a_{n+k} < 2b$  for all  $k \in \mathbb{N}$ . So we have not only proved that  $a_n \to 0$ but shown exactly what quantitative information we need from the assumption  $a_{\infty} = 0$  – in the simple form (5) – to obtain a rate. Essentially what we have done is transformed ( $\Gamma_6$ ) into the following *computational* derivation, which we accordingly label ( $\Gamma_5^c$ ):

$$\frac{\{\epsilon_n\}\text{-greedy}}{\forall n[a_{n+1} \le a_n + \epsilon_n]} \frac{\left\|\Gamma_4^c\right\|}{\sum_{i=f(b)}^{\infty} \epsilon_i < b} \exists n[f(b) \le n \le \Phi(b) \land a_n < n]}{\exists n \le \Phi(b) \forall k[a_{n+k} < 2b]}$$
( $\Gamma_5^c$ )

where now  $(\Gamma_4^c)$  is a hypothetical derivation of

$$\exists n [f(b) \le n \le \Phi(b) \land a_n < n]$$

for some function  $\Phi$ . The challenge is now has shifted to transforming the original derivation ( $\Gamma_4$ ) to a computational one ( $\Gamma_4^c$ ) so that we obtain such a  $\Phi$ .

## **3.3** Simplifying $(\Gamma_4)$

A crucial observation at this stage is that  $(\Gamma_4)$  simplifies: Because we have only used  $a_{\infty} = 0$  in the weakened form,

$$\exists n \ge f(b)[a_n < b] \tag{6}$$

where b is now some free variable, we can try to substitute this in the conclusion of ( $\Gamma_4$ ) and then replace the open assumption { $a_{\infty} > 0$ } with the stronger *negation* of (6). i.e. the more concrete assumption

$$\{\forall n \ge f(b)[a_n \ge b]\},\$$

and simplify the proof tree accordingly. This process involves a series of straightforward heuristic steps. Starting with  $(\Gamma_2)$ , we observe that here  $\{a_{\infty} > 0\}$  is simply used to establish that  $\forall n \geq N[a_{\infty}/2 \leq a_n]$  for sufficiently large N. For now we see if we can just replace  $a_{\infty}/2$  with b, and set N := f(b) as follows:

$$\frac{\{\epsilon_n\}\text{-greedy} \quad x^* \in \operatorname{co}(S)}{\frac{\forall b > 0, n[b \le a_n \implies P(b, a_n, a_{n+1}, \epsilon_n)]}{\forall n \ge f(b)[P(b, a_n, a_{n+1}, \epsilon_n)]}} \quad \{\forall n \ge f(b)[a_n \ge b]\} \quad (\Gamma_2^s)$$

Then  $(\Gamma_3)$  becomes

$$\begin{cases} \forall n \ge f(b)[a_n \ge b] \} \\ & \|\Gamma_2^s & \|\Gamma_1 & \underline{\sum_{i=0}^{\infty} \epsilon_i < \infty} \\ \frac{\forall n \ge f(b)[P(b, a_n, a_{n+1}, \epsilon_n)] & a_n \to a_\infty & \overline{\epsilon_n \to 0} \\ P(b, a_\infty, a_\infty, 0) & n \to \infty \end{cases}$$
 ( $\Gamma_3^s$ )

and then the entire modified version of  $(\Gamma_4)$  would be

$$\{ \forall n \ge f(b)[a_n \ge b] \}$$

$$\frac{ \left\| \Gamma_3^s - \frac{X \text{ is U. S.}}{\alpha(b) > 0} \right\|^{(\star)}}{\frac{1}{\exists n \ge f(b)[a_n < b]} \Longrightarrow r}$$

$$(\Gamma_4^s)$$

where here replacing  $a_{\infty}/2$  with b makes no difference to the way in which we derive a contradiction. We label this  $(\Gamma_4^s)$  i.e. if not a fully computational then at least a simplified version of  $(\Gamma_4)$ .

## **3.4** Analysing $(\Gamma_4^s)$

Let us now summarise our position: We demonstrated that the precise way that  $a_{\infty} = 0$  is used means 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 of  $(\Gamma_4)$  arrived at above.

We first note that if we can weaken the open assumption with a bound on how many  $n \ge f(b)$  it needs to hold in order to reach a contradiction, then this will be exactly the bound we are looking for. In other words, we want to produce  $\Phi(b)$  satisfying

$$\{ \forall n [\Phi(b) \ge n \ge f(b) \implies a_n \ge b] \}$$

$$\frac{ \left\| \Gamma_3^s \right\|_{\alpha(b) > 0}}{ \perp} \xrightarrow{X \text{ is U. S.}} (\star)$$

Looking at the final derivation of the above, one observation is that  $\alpha(b)$  is not necessarily computable, which might pose a problem if we want to use the fact that  $\alpha(b) > 0$  in a computational way. What we really need here is a computable function  $\xi : (0, \infty) \to (0, 1)$  witnessing the fact that  $\alpha(b) > 0$  i.e. such that  $\alpha(b) \ge \xi(b) > 0$  for any b > 0. Let us for now assume that we have such a  $\xi$ , and come back to the problem of finding it later.

The most obvious obstacles to our aim is that we have taken a limit as  $n \to \infty$  to establish  $P(b, a_{\infty}, a_{\infty}, 0)$  i.e.

$$a_{\infty} \le (1 - \alpha(b))a_{\infty} \tag{7}$$

and it is therefore 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 (7) – 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?

For argument's sake, let us take some  $\delta > 0$ . Then there exists some  $k \ge f(b)$  such that  $a_k - a_{k+1} < \delta$  and  $\epsilon_k < \delta$ . Then from  $P(b, a_k, a_{k+1}, \epsilon_k)$  and  $\alpha(b) \ge \xi(b)$  it follows that

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

which can be rearranged as

$$\frac{\xi(b) \cdot a_{k+1}}{2 - \xi(b)} \le \delta \tag{8}$$

assuming that  $\xi(b) < 2$ , which we can force to be the case if necessary. But (8) fails for e.g.

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

using also that  $b \leq a_{k+1}$ , we have reached a contradiction without using the whole the limit as  $n \to \infty$ . Let  $\Gamma_6$  be defined by

Then we have obtained our contradiction as follows:

$$\frac{\{\forall n \ge f(b)[a_n \ge b]\}}{ \left\| \Gamma_6 \right\|_{\mathcal{L}}} \frac{X \text{ is U. S.}}{\alpha(b) \ge \xi(b) > 0} (\star)$$

The question now becomes: How much of the open assumption  $\{\forall n \ge f(b)[a_n \ge b]\}$ do we need to obtain this contradiction? Inspecting  $(\Gamma_6)$  it is readily apparent that if  $\Psi(b)$  is a bound on a witness for  $\exists k \ge f(b)[a_k - a_{k+1} < \delta_{\xi,b}]$ , then we can replace the open assumption with

$$\{\forall n[\Psi(b) + 1 \ge n \ge f(b) \implies a_n \ge b]\}$$

(here the +1 coming from our additional use of the assumption for  $b \leq a_{k+1}$ ), and thus  $\Psi(b) + 1$  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 (modulo some assumptions involving uniform smoothness that we have delegated to later) is to analyse the following fragment of our modified proof tree:

$$\frac{\left\| \Gamma_{1} \quad \underbrace{\sum_{i=0}^{\infty} \epsilon_{i} < \infty}_{\epsilon_{n} \to 0} \right\|}{\exists k \ge f(b)[a_{k} - a_{k+1}, \epsilon_{k} < \delta_{\xi, b}]}$$

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

$$\forall N, \delta \exists k \ge N[a_k - a_{k+1}, \epsilon_k < \delta] \tag{9}$$

and then

$$\Phi(b) := \Gamma(f(b), \delta_{\xi, b}) + 1$$

would be our rate of convergence for  $a_n \to 0$ . Witnessing (9) turns out to be simpler that it might look. From  $\sum_{i=0}^{\infty} \epsilon_i < \infty$  we clearly have  $\epsilon_i < \delta$  for all *i* sufficiently large (where this can be made precise using the rate of convergence *f*). Finding *k* such that  $a_k - a_{k+1} < \delta$  doesn't require the assumption  $a_n \to a_{\infty}$ at all: It suffices that we know that  $\{a_n\}$  is bounded above by some K > 0. If it were the case that  $a_k - a_{k+1} \ge \delta$  for all  $k \ge N$ , we would have  $a_N \ge a_{N+i} - i\delta$ for all  $i \in \mathbb{N}$ , which is a contradiction for *i* sufficiently large since  $\{a_n\}$  is a sequence of nonnegative reals.

We have now processed the entire proof in an informal manner, and we are ready to put things back together in a formal way. The following section now presents everything we have done in an ordinary formal setting, where the machinery used to get there can be completely suppressed.

# 4 A preliminary quantitative result

We begin by presenting the content of Section 3.5 as a lemma, noting that the weaker condition  $\epsilon_n \to 0$  suffices. It is helpful to introduce some notation:

Definition 4.1. Given a pair of natural numbers m, n with  $m \leq n$ , we define  $[m, n] := \{m, m+1, \ldots, n-1, n\}$ , and thus  $k \in [m, n]$  is equivalent to  $m \leq k \leq n$ .

**Lemma 4.2.** Let  $\{a_n\}$  be a sequence of nonnegative reals bounded above by K > 0 and  $\{\epsilon_n\}$  a sequence of nonnegative reals with  $\epsilon_n \to 0$  as  $n \to \infty$  with rate f. Let

$$\Gamma(N,\delta) := \max\{N, f(\delta)\} + \lceil K/\delta \rceil$$

Then for any  $\delta > 0$  and  $N \in \mathbb{N}$  there exists some  $N \leq k \leq \Gamma(N, \delta)$  such that  $a_n - a_{n+1} < \delta$  and  $\epsilon_n < \delta$ .

*Proof.* Let  $N_1 := \max\{N, f(\delta)\}$  and suppose for contradiction that  $a_k - a_{k+1} \ge \delta$  for all  $k \in [N_1, N_1 + i]$ . Then for any  $i \in \mathbb{N}$  we have

$$K \ge a_{N_1} \ge a_{N_1+1} + \delta \ge a_{N_1+2} + 2\delta \ge \ldots \ge a_{N_1+i+1} + (i+1)\delta \ge (i+1)\delta$$

which is a contradiction for  $i := \lceil K/\delta \rceil$ . Thus there exists  $k \in [N_1, N_1 + i] \subseteq [N, N_1 + i]$  such that  $a_k - a_{k+1} < \delta$ . Since  $N_1 \ge f(\delta)$  we must also have  $\epsilon_k \le \delta$  since  $\sum_{i=N_1}^{\infty} \epsilon_i < \delta$  for this k.

The remaining work of Section 3 is now contained in the following result, which we carefully formulate to avoid any explicit mention of the underlying Banach space. This represents an abstract quantitative version of an underlying generalised *Fejér monotonicity* property with error terms enjoyed by  $||x_n - x^*||$  (see [25] for a detailed study of generalised Fejér monotonicity from the perspective of proof mining). Furthermore, we make the harmless assumption that our rate of convergence for  $\sum_{i=0}^{\infty} \epsilon_i < \infty$  is monotone in the sense that

$$\varepsilon \le \delta \implies f(\varepsilon) \ge f(\delta)$$

for the simple reason that this allows us to express our derived rate of convergence in a more concise way.

**Lemma 4.3.** Let  $\{a_n\}$  be a sequence of nonnegative reals bounded above by K > 0 and satisfying  $a_{n+1} \leq a_n + \epsilon_n$  where  $\{\epsilon_n\}$  is a sequence of nonnegative reals such that  $\sum_{i=0}^{\infty} \epsilon_i < \infty$  with (monotone) rate f. Suppose furthermore that these satisfy:

$$b \le a_n \implies a_{n+1} \le (1 - \alpha(b))a_n + \epsilon_n$$

for all b > 0 and  $n \in \mathbb{N}$ , where  $\alpha : (0, \infty) \to (0, \infty)$  is an arbitrary function. Then  $a_n \to 0$ , and a computable rate of convergence is given by

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

for any computable function  $\xi : (0, \infty) \to (0, 1)$  satisfying  $\alpha(b) \ge \xi(b)$  for all b > 0.

Proof. Fix  $b, \delta > 0$  and suppose that  $a_n \ge b$  for all  $n \in [f(b), \Gamma(f(b), \delta) + 1]$ where  $\Gamma$  is defined as in Lemma 4.2. By the same lemma, there exists some  $k \in [f(b), \Gamma(f(b), \delta)]$  such that  $a_k - a_{k+1} < \delta$  and  $\epsilon_k < \delta$ . In addition, it follows from our assumption that  $a_{k+1} \ge b$  and  $a_{k+1} \le (1 - \alpha(b))a_k + \epsilon_k$ , therefore reasoning exactly as in Section 3.4 we reach a contradiction for  $\delta_{\xi,b} := \frac{1}{2}\xi(b) \cdot b$ .

Therefore our assumption was false, and it follows that there exists some

$$f(b) \le n \le \Gamma(f(b), \delta_{\xi, b}) + 1 = \max\{f(b), f(\delta_{\xi, b})\} + \lceil K/\delta_{\xi, b}\rceil + 1$$
(10)

such that  $a_n < b$ . Using  $\delta_{\xi,b} < b$  together with monotonicity of f we can simplify the right hand side to

$$f(\delta_{\xi,b}) + \lceil K/\delta_{\xi,b} \rceil + 1 \tag{11}$$

Reasoning now as in Section 3.2, since  $n \ge f(b)$  we have  $a_{n+m} \le a_n + b < 2b$  for all  $m \in \mathbb{N}$ , and thus substituting  $b \mapsto b/2$  in (11) gives us a rate of convergence for  $a_n \to 0$ . Unwinding the definitions gives the result.

We will now see that the above lemma contains almost everything that we need to give a computational interpretation to Theorem 2.6.

# 5 The main result

Throughout Sections 3 and 4 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 [5] for a comprehensive survey on the general importance of such results in analysis, and [23] 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  $\alpha$  defined as in Lemma 2.5, which explicitly involves the modulus of smoothness  $\rho_X$ , 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 [4] 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

$$||x|| = 1 \land ||y|| \le \delta \implies ||x+y|| + ||x-y|| \le 2 + \varepsilon ||y||$$
(12)

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  $\delta$ in (12). Moduli of uniform smoothness, which are distinct from the (uniquely defined) modulus of smoothness, were first used in [17] and appear in many other places in the applied proof theory literature as a quantitative analogue of uniform smoothness, where they are equivalent to rates of convergence for  $\rho_X(t)/t \to 0$  as  $t \to 0$ .

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

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

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

Proof. Analogous to the proof of Lemma 2.3.

We now give a reformulation of the main lemma used.

**Lemma 5.2.** Let  $(X, \omega)$  be a uniformly smooth Banach space. Suppose that  $S \subseteq X$  and  $x^* \in co(X)$ , that  $\{x_n\}$  is  $\{\epsilon_n\}$ -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  $a_n := ||x_n - x^*||$ , any b > 0, if  $b \leq a_n$  then

$$a_{n+1} \le (1 - \alpha(b))a_n + \epsilon_n$$

for

$$\alpha(b) := \sup\left\{ \left. \frac{b \cdot \omega(\varepsilon \cdot b) \cdot (1 - \varepsilon K)}{2K} \right| \ \varepsilon \in (0, 1) \right\}$$

*Proof.* Applying Lemma 5.1 in an analogous way to the proof of Lemma 5.2 we see that for any  $\varepsilon > 0$ , if

$$\frac{\lambda K}{(1-\lambda)b} \le \omega(\varepsilon b)$$

then

$$(\|(1-\lambda)x_n + \lambda y - x^*\| \le (1-\lambda)a_n\left(1 + \frac{\varepsilon\lambda K}{(1-\lambda)}\right) + \lambda F_n(y - x^*)$$
$$\le (1-\lambda(1-\varepsilon K))a_n + \lambda F_n(y - x^*)$$

using the shorthand  $F_n := F_{x_n-x^*}$ . Now, if we define  $\lambda_{\varepsilon} := b \cdot \omega(\varepsilon \cdot b)/2K$  for arbitrary  $\varepsilon \in (0,1)$  then  $\lambda \in [0,1)$  and we can show that the premise of the above holds, and so in particular

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

from which the main result follows.

Now proving that  $\alpha(b) > 0$  for any b > 0 is extremely direct, since the expression inside the supremum must be positive for any  $\varepsilon < 1/K$ . We are now ready to present the computational version of Theorem 2.6.

**Theorem 5.3.** Let  $(X, \omega)$  be a uniformly smooth Banach space. Suppose that  $S \subseteq X$  and  $x^* \in \overline{\operatorname{co}(X)}$ , that  $\{x_n\}$  is  $\{\epsilon_n\}$ -greedy with respect to  $x^*$  and S for some sequence  $\{\epsilon_n\}$  of positive reals such that  $\sum_{i=0}^{\infty} \epsilon_n < \infty$  with (monotone) rate f. Suppose that K > 0 is such that  $\sup\{\|y - x^*\| \mid y \in S\} \leq K$ . Then  $x_n \to x^*$  as  $n \to \infty$  with rate of convergence given by

$$\Phi(b) := f(\psi(b)) + \left\lceil \frac{K}{\psi(b)} \right\rceil + 1 \quad for \quad \psi(b) := \frac{b^2}{32} \cdot \omega\left(\frac{b}{4K}\right)$$

*Proof.* We apply Lemma 4.3 for  $a_n := ||x_n - x^*||$ , where the main condition holds by Lemma 5.2 and the fact that  $a_{n+1} \leq a_n + \epsilon_n$ . For the rate of convergence, setting  $\varepsilon := 1/2K$  in the definition of  $\alpha(b)$  allows us to define computable  $\xi$  in a suitable way, namely

$$\alpha(b) \geq \frac{b}{4K} \cdot \omega\left(\frac{b}{2K}\right) =: \xi(b)$$

and substituting this into the  $\Phi$  as defined in Lemma 4.3 gives the rate.

## 6 Extensions

Having obtained the main result, we now discuss directions ways in which, by taking a closer look at the quantitative proofs, we can further generalise it. These typically revolve around looking at precisely how certain assumptions or hypotheses are used, and whether they can be weakened. In doing so we hope to illustrate 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.

#### 6.1 Weakening the convergence condition on $\{\epsilon_n\}$

One of the benefits of presenting the computational core of the original proof in such a simple and abstract way is that we can already start to make connections with existing results in the proof mining literature, with an eye to improving the result if possible. Lemma 4.3 identifies the main "recursive inequality" that plays a role in the original proof, namely

$$b \le a_n \implies a_{n+1} \le (1 - \alpha(b))a_n + \epsilon_n$$

Such recursive inequalities are widely used in proof mining, and the one above turns out to be very similar to an inequality already considered by the author in [27] (though in the very different setting of weakly contractive mappings), namely

$$a_{n+1} \le a_n - \psi(a_n) + \epsilon_n$$

for the case that  $\psi(b) := \alpha(b) \cdot b$ . There it is shown that  $a_n \to 0$  with computable rate, but under the weaker condition that  $\epsilon_n \to 0$ . We now show that we can adapt the above proof to establish the same result, also under this weaker condition.

**Lemma 6.1** (Strengthening of Lemma 4.3). Let  $\{a_n\}$  be a sequence of nonnegative reals bounded above by K > 0, and  $\{\epsilon_n\}$  a sequence of nonnegative reals such that  $\epsilon_n \to 0$  as  $n \to \infty$  with rate f. Suppose furthermore that these satisfy:

$$b \le a_n \implies a_{n+1} \le (1 - \alpha(b))a_n + \epsilon_n$$

for all b > 0 and  $n \in \mathbb{N}$ , where  $\alpha : [0, \infty) \to [0, \infty)$  is an arbitrary function with  $\alpha(0) = 0$  and  $\alpha(b) > 0$  for b > 0. Then  $a_n \to 0$ , with the same a computable rate of convergence as in Lemma 4.3.

*Proof.* By very similar reasoning to the proof of Lemma 4.3 we can show that there exists some

$$f(\psi(2b)) \le n \le \Phi(2b)$$

such that  $a_n < b$ . We now show directly that  $a_m < 2b$  for all  $m \ge n$ . For this we use induction. For the induction step, there are two cases to consider: Either  $a_m < b$ , in which case

$$a_{m+1} \le (1 - \alpha(b))a_m + \epsilon_m < b + b = 2b$$

where here we use that  $\epsilon_m < \psi(2b) < b$ . On the other hand, if  $b \leq a_m < 2b$  then

$$a_{m+1} \le a_m - \alpha(b) \cdot b + \epsilon_m \le a_m < 2b$$

where for the final inequality we use  $\epsilon_m < \psi(2b) = \frac{b}{2} \cdot \xi(b) < b \cdot \alpha(b)$ .

**Corollary 6.2.** Theorem 5.3 holds, with the same rate of convergence, but under the weaker assumption that  $\epsilon_n \to 0$  as  $n \to \infty$  with (monotone) rate f.

*Proof.* Exactly as in the proof of Theorem 5.3, but replacing Lemma 4.3 with Lemma 6.1 and extending the domain of  $\alpha$  to  $[0, \infty)$  by setting  $\alpha(0) = 0$ .

#### 6.2 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 ana 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 [2] in the special case that X has modulus of smoothness of the form  $\rho(u) \leq \gamma u^t$  for t > 1. These results are established by introducing new abstract convergence results for sequences of real numbers. Here, we instead ask whether results along these lines are already possible by analysing and refining the existing proof.

For Theorem 5.3, the main point where a particular value of  $\lambda$  is considered is the proof of Lemma 5.2, where we instantiate  $\lambda_{\varepsilon} := b \cdot \omega(\varepsilon \cdot b)/2K$  to establish  $a_{n+1} \leq (1 - \alpha(b))a_n + \epsilon_n$  for fixed b but arbitrary n. Thus at first glance, it doesn't seem possible to convert this into a local condition on  $\lambda_n$ , since the value is dependent only on b. However, referring to the proof of Theorem 4.3, we require this inequality to hold only in the case that  $n \in [f(\psi(b)), \Phi(b)]$ , so it might suffice to assume that  $\lambda_{\varepsilon} := b \cdot \omega(\varepsilon \cdot b)/2K$  for n in this range. We now prove this, first stating another modified form of Lemma 4.3. Interestingly, this version of the lemma doesn't seem to work under the weakened hypothesis  $\epsilon_n \to 0$  as  $n \to \infty$ , whose proof uses that the main recursive inequality to holds for all n to obtain convergence.

**Lemma 6.3.** Let  $\{a_n\}$  be a sequence of nonnegative reals bounded above by K > 0 and satisfying  $a_{n+1} \leq a_n + \epsilon_n$  where  $\{\epsilon_n\}$  is a sequence of nonnegative reals with  $\sum_{i=0}^{\infty} \epsilon_i < \infty$  with rate f. Let  $\alpha : (0, \infty) \to (0, \infty)$  be an arbitrary function and define

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

for any computable function  $\xi: (0,\infty) \to (0,1)$ . Then whenever

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

for all b > 0 and  $n \in [f(\psi(2b)), \Phi(2b)]$ , then  $a_n \to 0$  with rate of convergence  $\Phi$ .

Proof. Analogous to the proof of Lemma 4.3, but now keeping track of how the main inequality is used. Fix b > 0 and define  $\delta_{\xi,b} := \frac{1}{2}\xi(b) \cdot b = \psi(2b)$ . Suppose that  $a_n \geq b$  for all  $n \in [f(\psi(2b)), \Gamma(f(b), \delta_{\xi,b}) + 1] = [f(b), \Phi(2b)]$  where  $\Gamma$  is defined as in Lemma 4.2. Then there exists some  $k \in [f(\psi(2b)), \Gamma(f(b), \delta_{\xi,b})]$  such that  $a_k - a_{k+1} < \delta_{\xi,b}$  and  $\epsilon_k < \delta_{\xi,b}$ . In addition, it follows from our assumption that  $a_{k+1} \geq b$ , and since  $k \in [f(\psi(2b)), \Phi(2b)]$  we have  $a_{k+1} \leq (1 - \alpha(b))a_k + \epsilon_k$ , therefore reasoning exactly as in Section 3.4 we reach a contradiction. Therefore our assumption was false, and it follows that there exists some  $n \in [f(\psi(b)), \Phi(2b)]$  such that  $a_n < b$ . Reasoning now as in Section 3.2, since  $n \geq f(b)$  we have  $a_{n+m} \leq a_n + b < 2b$  for all  $m \in \mathbb{N}$ , and thus  $\Phi$  is a rate of convergence for  $a_n \to 0$ .

**Theorem 6.4.** Let  $(X, \omega)$  be a uniformly smooth Banach space. Suppose that  $S \subseteq X$  and  $x^* \in \overline{\operatorname{co}(X)}$ . Suppose that  $\{x_n\}$  is some sequence of approximants to  $x^*$  satisfying  $||x_{n+1} - x^*|| \leq ||x_n - x^*|| + \epsilon_n$  and

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

for all  $n \in \mathbb{N}$ , a sequence  $\{\epsilon_n\}$  of positive reals such that  $\sum_{i=0}^{\infty} \epsilon_i < \infty$  with rate f, and some fixed sequence  $\{\lambda_n\}$  of step sizes in [0,1). Let K > 0 be such that  $||y - x^*|| \leq K$  for all  $y \in S$ . Finally, suppose that  $\xi : (0,\infty) \to (0,1)$  is such that for all b > 0 there exists  $\varepsilon \in (0,1)$  such that

$$\xi(b) \le \lambda_n (1 - \varepsilon K) \quad and \quad \lambda_n \le \frac{b \cdot \omega(\varepsilon \cdot b)}{2K}$$

for all  $n \in [f(\psi(b)), \Phi(b)]$  for  $\psi$  and  $\Phi$  as defined in Lemma 6.3. Then  $x_n \to x^*$  as  $n \to \infty$  with rate  $\Phi$ .

*Proof.* We apply Lemma 6.3. Setting  $a_n := ||x_n - x^*||$ , all we need to do is show that  $a_{n+1} \leq (1 - \xi(b))a_n + \epsilon_n$  for all  $n \in [f(\psi(b)), \Phi(b)]$ . So picking such an n, arguing as in the proof of Lemma 5.2, if  $b \leq a_n$  then there exists  $\varepsilon \in (0, 1)$  such that  $\lambda_n \leq b \cdot \omega(\varepsilon \cdot b)/2K$  and therefore

$$\|(1-\lambda_n)x_n+\lambda_n y\| \le (1-\lambda_n(1-\varepsilon K))a_n+\lambda_n F_n(y-x^*)$$
  
$$\le (1-\xi(b))a_n+\lambda_n F_n(y-x^*)$$

and thus

$$a_{n+1} \leq \inf\{\|(1-\lambda_n)x_n + \lambda_n y\| \mid y \in S\} + \epsilon_n$$
  
$$\leq (1-\xi(b))a_n + \inf\{\lambda_n F_n(y-x^*) \mid y \in S\} + \epsilon_n$$
  
$$\leq (1-\xi(b))a_n + \epsilon_n$$

using that  $\inf \{\lambda_n F_n(y - x^*) \mid y \in S\} \leq 0$  as in Lemma 2.5. Thus all conditions of the Lemma 6.3 are satisfied and we obtain the result.

It is instructive to compare our Theorem 6.4 with the rates of convergence given in Section 3.2 of [2], notably their Theorem 3.5. The latter focuses on the special case that  $\rho(u) \leq \gamma u^t$  for t > 1 (which would correspond to a modulus of uniform smoothness of power type 1/(t-1)), and demonstrates that by fixing  $\lambda_n := 1/(n+1)$  we can obtain greedy approximation schemes that converge with  $||x_n - x^*|| = \mathcal{O}(1/n^{1-1/t})$ . This involves adapting their general proof idea to this special case, and using a different combinatorial result on convergence sequences of real numbers. Our result applies more generally to any modulus of uniform smoothness, and incorporates a range of conditions on the fixed step sizes  $\{\lambda_n\}$ .

# 6.3 Approximations to $\overline{\operatorname{co}(S)}$

Finally, we note that an alternative hypothesis we can consider weakening is  $x^* \in \overline{\operatorname{co}(S)}$ , which is used to establish that for any  $x_n$ , we can always find some  $y \in S$  such that  $F_{x_n-x^*}(y-x^*)$  is arbitrarily small. To do this, recall that we pick  $z := \sum_{i=1}^{n} a_i y_i$  with  $y_i \in S$  such that  $||z - x^*|| \leq \delta/2$ , and then it follows that  $F_{x_n-x^*}(y_i - x^*) \leq \delta$  for some i, else we would have

$$\delta \ge \|F_{x_n - x^*}\| \|z - x^*\| \ge F_{x_n - x^*}(z - x^*) = \sum_{i=1}^n F_{x_n - x^*}(y_i) > \delta \sum_{i=1}^n a_i = \delta$$

Fixing now some  $b, \delta > 0$ , if there exists  $z \in co(S)$  with  $||z - x^*|| \le \delta/2$ , then in Lemma 5.2 we would have

$$\begin{aligned} a_{n+1} &\leq \inf\{\left[1 + \lambda(\varepsilon K - 1)\right]a_n + \lambda F_n(y - x^*) \mid y \in S \lambda \in [0, 1]\} + \epsilon_n \\ &\leq \left[1 - \frac{b \cdot \omega(\varepsilon \cdot b) \cdot (1 - \varepsilon K)}{2K}\right]a_n + \frac{b \cdot \omega(\varepsilon \cdot b) \cdot F_n(y_i - x^*)}{2K} + \epsilon_n \\ &\leq \left[1 - \frac{b \cdot \omega(\varepsilon \cdot b) \cdot (1 - \varepsilon K)}{2K}\right]a_n + \delta + \epsilon_n \end{aligned}$$

where  $y_i$  is such that  $F_n(y_i - x^*) \leq \delta$ . In this way we could modify the main proof to obtain a rate of convergence for  $x_n \to x^*$  that can be verified up to b > 0provided we can show the existence of  $z \in co(S)$  such that  $||z - x^*|| \leq \delta$  for  $\delta$ sufficiently small depending on b. We do not discuss the details of this extension in any depth, but mention it simply to highlight that further generalisations are possible.

## 7 Conclusion

We have presented a new application of proof theory, where through the analysis of a nonconstructive convergence proof we are able to not only extract an explicit rate of convergence, but also provide several qualitative generalisations of the original theorem. The purpose of this work was twofold: (i) to apply proof theoretic techniques in a hitherto unexplored part of approximation theory, and (ii) to present the extraction process in a different way, emphasising the underlying proof theoretic steps that were performed in the analysis. Both of these points give rise, in turn, to directions for future work:

Firstly, It would be interesting to explore whether there are further applications of proof theory on the convergence of greedy approximation schemes in Hilbert and Banach spaces. This forms a rich area, covered in depth by the recent textbook [28] and also in many places in the machine learning literature (e.g. [29]). Here convergence proofs in general Banach spaces typically resort to geometric properties of those spaces, specifically variants of uniform smoothness, a property that can be universally axiomatised via the corresponding modulus and thus falls within existing metatheorems that guarantee the extractability of effective bounds [17]. Work in this direction represents a natural route through which ideas and techniques of proof mining ould be applied to machine learning, something that has been recently proposed by the author and Pischke in a quite different context in [26].

Secondly, an attempt to systematise the kind high-level manipulations on proof trees outlined in Section 3 and capture them within some specialised formal system could represent a fascinating project aimed at eventually automating aspects of applied proof theory. This idea would align well with the use of proof assistants, as discussed in [19], but would represent a significant challenge: While the straightforward extraction of computational content from formal proofs in a brute force manner can of course be done purely mechanically, by implementing the relevant proof interpretations, automating the high-level mathematical reasoning essential to the success of applied proof theory is another matter entirely, and may well require the development of sophisticated proof systems that are tailored to applications in a specific area.

Acknowledgments. The author was partially supported by the EPSRC grant EP/W035847/1. We thank Sam Sanders for comments on the presentation.

# References

- CHIDUME, C. Geometric Properties of Banach Spaces and Nonlinear Iterations. Springer, 2009.
- [2] DARKEN, C., DONAHUE, M. J., GURVITS, L., AND SONTAG, E. Rates of convex approximation in non-Hilbert spaces. *Constructive Approximation* 13 (1997), 187–220.
- [3] DINIS, B., AND PINTO, P. Quantitative results on the multi-parameters proximal point algorithm. *Journal of Convex Analysis 28*, 3 (2021), 729– 750.
- [4] FINDLING, R., AND KOHLENBACH, U. 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* 45 (2024), 216–233.

- [5] FRANCI, B., AND GRAMMATICO, S. Convergence of sequences: A survey. Annual Reviews in Control 53 (2022), 161–186.
- [6] JONES, L. A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *The Annals of Statistics 20*, 1 (1992), 608–613.
- [7] KOHLENBACH, U. A quantitative version of a theorem due to Borwein-Reich-Shafrir. Numerical Functional Analysis and Optimization 22 (2001), 641–656.
- [8] KOHLENBACH, U. Uniform asymptotic regularity for Mann iterates. Journal of Mathematical Analysis and Applications 279 (2003), 531–544.
- [9] KOHLENBACH, U. Some logical metatheorems with applications in functional analysis. Transactions of the American Mathematical Society 357, 1 (2005), 89–128.
- [10] KOHLENBACH, U. Applied Proof Theory: Proof Interpretations and their Use in Mathematics. Springer Monographs in Mathematics. Springer, 2008.
- [11] KOHLENBACH, U. Recent progress in proof mining in nonlinear analysis. IFCoLoG Journal of Logics and their Applications 10, 4 (2017), 3361–3410.
- [12] KOHLENBACH, U. A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space. Foundations of Computational Mathematics 19 (2019), 83–99.
- [13] KOHLENBACH, U. Proof-theoretic methods in nonlinear analysis. In Proceedings of the International Congress of Mathematicians 2018, vol. 2. World Scientific, 2019, pp. 61–82.
- [14] KOHLENBACH, U. Quantitative analysis of a Halpern-type Proximal Point Algorithm for accretive operators in Banach spaces. *Journal of Nonlinear* and Convex Analysis 21, 9 (2020), 2125–2138.
- [15] KOHLENBACH, U., AND KOUTSOUKOU-ARGYRAKI, A. Rates of convergence and metastability for abstract cauchy problems generated by accretive operators. *Journal of Mathematical Analysis and Applications* 423 (2015), 1089–1112.
- [16] KOHLENBACH, U., AND LEUŞTEAN, L. Mann iterates of directionally nonexpansive mappings in hyperbolic spaces. Abstract and Applied Analysis 2003, 8 (2003), 449–477.
- [17] KOHLENBACH, U., AND LEUŞTEAN, L. On the computational content of convergence proofs via Banach limits. *Philosophical Transactions of the Royal Society A 370* (2012), 3449–3463.

- [18] KOHLENBACH, U., AND POWELL, T. Rates of convergence for iterative solutions of equations involving set-valued accretive operators. *Computers and Mathematics with Applications 80* (2020), 490–503.
- [19] KOUTSOUKOU-ARGYRAKI, A. On preserving the computational content of mathematical proofs: Toy examples for a formalising strategy. In *Connect*ing with Computability. CiE 2021 (2021), vol. 12813 of LNCS, Springer, pp. 285–296.
- [20] KREISEL, G. On the interpretation of non-finitist proofs, Part I. Journal of Symbolic Logic 16 (1951), 241–267.
- [21] KREISEL, G. On the interpretation of non-finitist proofs, Part II: Interpretation of number theory. *Journal of Symbolic Logic* 17 (1952), 43–58.
- [22] LEUŞTEAN, L., NICOLAE, A., AND SIPOŞ, A. An abstract proximal point algorithm. Journal of Global Optimization 72, 3 (2018), 553–577.
- [23] NERI, M., AND POWELL, T. A computational study of a class of recursive inequalities. *Journal of Logic and Analysis* 15, 3 (2023), 1–48.
- [24] PISCHKE, N. Logical metatheorems for accretive and (generalized) monotone set-valued operators. *Journal of Mathematical Logic* 24, 2 (2024), 59pp.
- [25] PISCHKE, N. Generalized Fejér monotone sequences and their finitary content. *Optimization* (2025). To appear.
- [26] PISCHKE, N., AND POWELL, T. Asymptotic regularity of a generalised stochastic Halpern scheme with applications. Preprint, available at https: //arxiv.org/abs/2411.04845, 2024.
- [27] POWELL, T., AND WIESNET, F. Rates of convergence for asymptotically weakly contractive mappings in normed spaces. *Numerical Functional Anal*ysis and Optimization 42, 15 (2021), 1802–1838.
- [28] TEMLYAKOV, V. N. Greedy Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2011.
- [29] TEWARI, A., RAVIKUMAR, P., AND DHILLON, I. Greedy algorithms for structurally constrained high dimensional problems. In Advances in Neural Information Processing Systems 24 (NIPS 2011) (2011).