

A computational study of a class of recursive inequalities

Morenikeji Neri and Thomas Powell

July 29, 2022

Abstract

We examine the convergence properties of sequences of nonnegative real numbers that satisfy a particular class of recursive inequalities, from the perspective of proof theory and computability theory. We first establish a number of results concerning rates of convergence, setting out conditions under which computable rates are possible, and when not, providing corresponding rates of metastability. We then demonstrate how the aforementioned quantitative results can be applied to extract computational information from a range of proofs in nonlinear analysis. Here we provide both a new case study on subgradient algorithms, and survey a number of recent results which each involve an instance of our main recursive inequality. All of the relevant concepts from both proof theory and mathematical analysis are defined and motivated within the paper itself, and as such, we hope that this work also forms an accessible overview of aspects of current research in applied proof theory.

1 Introduction

Recursive inequalities on sequences of nonnegative real numbers play an important role in functional analysis. They can be used to establish convergence properties of algorithms in a very general setting, and often yield explicit *rates* of convergence in addition. A simple example of this phenomenon is represented by the inequality

$$\mu_{n+1} \leq c\mu_n \tag{1}$$

for $c \in [0, 1)$. Here, any sequence $\{\mu_n\}$ of nonnegative reals that satisfies (1) converges to zero, and moreover an effective rate of convergence for $\mu_n \rightarrow 0$ is given by $\mu_n \leq c^n \mu_0$. The inequality (1) is associated most famously with the Banach fixed point theorem: Suppose that (X, d) is a metric space and $T : X \rightarrow X$ a contractive mapping with constant c i.e.

$$d(T(x), T(y)) \leq cd(x, y)$$

for all $x, y \in X$. If x^* is a fixpoint of T , and $\{x_n\}$ the algorithm defined by $x_{n+1} := Tx_n$ for some starting value $x_0 \in X$, it is easy to see that $\mu_n :=$

$d(x_n, x^*)$ satisfies (1), since

$$d(x_{n+1}, x^*) = d(T(x_n), T(x^*)) \leq cd(x_n, x^*)$$

Therefore $x_n \rightarrow x^*$ follows from $\mu_n \rightarrow 0$, and moreover, the same rate of convergence applies, with $d(x_n, x^*) \leq c^n d(x_0, x^*)$.

What we have just sketched is a simple but representative example of a much broader strategy for proving that an algorithm $\{x_n\}$ in some space X converges to a point x^* (where the latter could be, for example, a minimiser of a function, a fixpoint of an operator, or a solution of an equation), namely:

1. Show that $\mu_n := d(x_n, x^*)$ satisfies some recursive inequality,
2. Prove that any sequence $\{\mu_n\}$ of nonnegative reals satisfying that recursive inequality converges to zero.

An important advantage of this strategy is that individual classes of recursive inequalities can typically be used to prove the convergence of a wide range of different algorithms. Moreover, as a happy byproduct, additional properties pertaining to the recursive inequality that we are able to uncover, such as a rate of convergence, often then apply uniformly to any algorithm that can be reduced to that inequality.

The purpose of this article is to study several classes of sequences that all satisfy the following recursive inequality:

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \gamma_n \tag{*}$$

where we assume throughout that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers, with the additional conditions that:

- $\sum_{i=0}^{\infty} \alpha_i = \infty$, and
- $\gamma_n \rightarrow 0$ as $n \rightarrow \infty$.

The significance of (*) as a basic characterising property for a range of interrelated convergent sequences has been exploited in several places in the mathematics literature – see for example Section 2 of [3]. We give a detailed description of the mathematical relevance of (*) in Section 2.1 below, but informally speaking, the $\{\alpha_n\}$ are typically step sizes for some algorithm, while $\{\gamma_n\}$ represents an “error-term” that eventually vanishes. We are then interested in the convergence properties of $\{\mu_n\}$ and $\{\beta_n\}$. As we will see, the aforementioned conditions on $\{\alpha_n\}$ and $\{\gamma_n\}$ alone are not sufficient to establish convergence of either $\{\mu_n\}$ or $\{\beta_n\}$, but their behaviour can be controlled by imposing a range of additional conditions, which then divide (*) into a number of interesting subclasses that each have distinct properties and applications. While (1) represents a very simple case of (*) for $\beta_n := \mu_n$, $\alpha_n := 1 - c$ and $\gamma_n := 0$, in general, proving convergence of $\{\mu_n\}$ and $\{\beta_n\}$ can be extremely subtle.

Aims and overview of the paper

For us, the relevance of (\star) lies primarily in the fact that it has implicitly appeared in several recent papers in applied proof theory, where classes of sequences satisfying (\star) have been analysed from a quantitative perspective in order to extract rates of convergence for concrete problems in nonlinear analysis. In this paper, we generalise many of these results and classify them within a overarching framework, establish new abstract quantitative lemmas for sequences of nonnegative reals, and present new applications.

Our paper is divided into two main parts. In Section 3 we focus exclusively on sequences of real numbers satisfying (\star) , and address the following questions:

- (a) Under which conditions on $\{\alpha_n\}$ and $\{\gamma_n\}$ can we prove that $\{\mu_n\}$ and $\{\beta_n\}$ converge, and when do we obtain computable rates of convergence?
- (b) Where computable rates of convergence are not possible, can we obtain computable bounds on the corresponding rates of *metastable* convergence for these sequences?

Answers to these questions are both interesting and useful. Where (a) can be answered positively, we are then in a position extract concrete rates of convergence from more complex proofs which reduce to that particular instance of (\star) as a lemma. When the answer is negative, we have demonstrated that there is no *uniform* method for obtaining computable rates in this way, and this can also be revealing: Quite often, analysts who use recursive inequalities of the latter type simply observe that their proof does not yield convergence rates. Our results then provide an explanation as to why that is the case, using tools from computable analysis such as Specker sequences. On the other hand, for all cases that we consider we are able to answer (b) positively, and provide a corresponding quantitative lemma that yields rates of metastability, given appropriate metastable rates for assumptions. This allows us to extract quantitative information from proofs where computable rates of convergence do not seem to be possible.

Section 4 then focuses on applications, giving concrete examples of the phenomena we have just described. We begin with a brief overview of recursive inequalities in the applied proof theory literature, and then present a new case study which uses some of our new results on recursive inequalities in the context of subgradient methods, to firstly explain why rates of convergence cannot be extracted from the kind of convergence proofs given in the literature, and secondly providing instead a rate of metastability for a class of subgradient algorithms. This is followed by a high-level survey of some recent work, where we demonstrate how our abstract numerical results can be more broadly utilised in a wide range of different settings.

Our main technical contributions include the following:

- (Real analysis) Strengthenings of existing convergence results on sequences of real numbers, with necessary and sufficient conditions for convergence for both main classes studied (Proposition 3.8 and Proposition 3.13);

- (Computability theory) Negative results showing that for sequences of reals satisfying (\star) , computable rates of convergence are in general not possible without additional assumptions (Theorem 3.6 and Proposition 3.16);
- (Proof theory) The extraction of computable rates of metastability for all convergence results considered (Theorems 3.9, 3.17 and 3.19), along with rates of convergence in special cases (Corollaries 3.11, 3.18 and 3.20).
- (Applied proof theory) An abstract characterisation of a class of gradient descent methods along with a general quantitative convergence result for these algorithms (Theorem 4.2), and application to a projective subgradient method of Alber et al. [4] (Theorem 4.5).

In addition to this technical work, our secondary objective has been to write a paper that serves as a useful and accessible introduction to recent research in applied proof theory. Quantitative versions of lemmas involving recursive inequalities have become increasingly common over the last 5–10 years, and while this article is by no means intended as a comprehensive overview of all of these, by focusing on a relevant class of inequalities we hope to provide some general insight into the role they play in proof theoretic approaches to analysis. Our work coincides nicely with a recent survey paper on convergent sequences [10], which provides an extensive overview of their many applications over different areas of analysis and probability theory. Our paper is similar in spirit (though here with a focus on proof theory and computability theory) in presenting an overview of several variations a type of recursive inequality, and illustrating how these are applied in concrete areas.

With all this in mind, our aim has been to write for as general an audience as possible. All of the core definitions and concepts from both proof theory and computability theory are introduced in Section 2.2, and our main abstract results in Section 3 do not require any knowledge of analysis beyond basic facts about sequences and series. The more advanced areas of mathematical analysis covered in Section 4 are carefully motivated, with simple illustrative examples provided before each of the main results, and it should be possible for readers who are not experts in the relevant subfields of analysis to follow these with no prior knowledge beyond some elementary facts about inner products and normed spaces.

Some of this research has been formalized in the Lean proof assistant, as part of a broader strategy of building a library for key definitions and lemmas from applied proof theory. We conclude the paper with a brief account of this effort, arguing why recursive inequalities form such a good starting point for an applied proof theory library, and by outlining some potentially interesting directions for future work.

2 Preliminaries

This paper is about both mathematical analysis and logic, and we divide this background section into two distinct parts, focusing on each in turn. We begin by motivating the recursive inequality (\star) in a little more detail, justifying why we have chosen it as an object of study and preparing for the applications that will be presented in Section 4. We then collect together some key facts and definitions from proof theory and computability theory that will be used in what follows.

2.1 Mathematical context

At the very beginning of the paper we gave, via Banach's fixed point theorem, a simple example of how recursive inequalities underlie convergence proofs. We now explain how the more complex recursive inequality (\star) appears in nonlinear analysis, first through a simple example and then with a high-level overview, detailing the role that each of the individual components usually play.

2.1.1 A simple example: weakly contractive mappings and the sine function

For purely illustrative purposes, let us consider a simple class of mappings known as ψ -weakly contractive mappings, introduced in [1] and widely studied since. Letting X be some normed space and $\psi : [0, \infty) \rightarrow [0, \infty)$ a nondecreasing function with $\psi(0) = 0$ and $\psi(t) > 0$ for $t > 0$, a mapping $T : X \rightarrow X$ is called ψ -weakly contractive on some $C \subseteq X$ if it satisfies

$$\|Tx - Ty\| \leq \|x - y\| - \psi(\|x - y\|) \quad (2)$$

for all $x, y \in C$. Any contractive mapping on C is also ψ -weakly contractive for $\psi(t) = (1 - c)t$, but the ψ -weakly contractive mappings are much more general. For example, setting $X = \mathbb{R}$, we can show, by considering its Taylor expansion, that

$$|\sin(x) - \sin(y)| \leq |x - y| - \frac{|x - y|^3}{8}$$

for $x, y \in [0, 1]$, and so the sine function is ψ -weakly contractive on $[0, 1]$ for $\psi(t) = t^3/8$ (cf. [1, p.14]).

Now suppose that x^* is a fixed point of a ψ -weakly contractive mapping T . A classic algorithm for approximating fixed points is the so-called Krasnoselskii-Mann (KM) scheme, defined recursively from some starting point x_0 by

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_nTx_n \quad (3)$$

where $\{\alpha_n\}$ is a sequence of nonnegative real numbers with $\sum_{i=0}^{\infty} \alpha_i = \infty$. Again, this is a generalisation of the scenario considered in the introduction, where the algorithm $x_{n+1} = Tx_n$ corresponds to the special case $\alpha_n = 1$. Using

(2) and (3) along with the triangle inequality, we can calculate that

$$\begin{aligned}
\|x_{n+1} - x^*\| &= \|(1 - \alpha_n)x_n + \alpha_n T x_n - x^*\| \\
&= \|(1 - \alpha_n)(x_n - x^*) + \alpha_n(T x_n - x^*)\| \\
&\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n\|T x_n - T x^*\| \\
&\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n(\|x_n - x^*\| - \psi(\|x_n - x^*\|)) \\
&= \|x_n - x^*\| - \alpha_n\psi(\|x_n - x^*\|)
\end{aligned}$$

and thus $\mu_n := \|x_n - x^*\|$ satisfies the following special case of (\star) for $\beta_n := \psi(\mu_n)$ and $\gamma_n := 0$:

$$\mu_{n+1} \leq \mu_n - \alpha_n\psi(\mu_n) \quad (4)$$

As we will see in Section 3, it can be shown that if $\sum_{i=0}^{\infty} \alpha_i = \infty$, then whenever $\{\mu_n\}$ satisfies (4) we have $\mu_n \rightarrow 0$. There also exist effective rates of convergence for $\mu_n \rightarrow 0$ in terms of the speed of divergence of $\sum_{i=0}^{\infty} \alpha_i$. These will be given in a general setting in Section 3 and are also formulated in a slightly different way in [1].

As a consequence, by focusing on the convergence behaviour of sequences of reals satisfying (4), we can show that whenever x^* is a fixed point of some ψ -weakly contractive mapping T , the corresponding KM algorithm converges to this fixed point with a known rate of convergence. In particular, setting $T = \sin$ and $\alpha_n = 1$, we can show that $x_{n+1} = \sin x_n$ converges to zero for any $x_0 \in [0, 1]$ and provide an explicit convergence speed.

In Section 4.3.1 we will give a slightly more intricate asymptotic variant of being ψ -weakly contractive, which leads to a version of (4) with a nonzero error terms $\{\gamma_n\}$.

2.1.2 The general setting

Moving away from our concrete example to the general setting, the recursive inequality (\star) is applied in a broadly similar way across each of the different setting we explore in this paper. We typically have a mapping T on some normed space X , a special value x^* (such as a minimizer or a fixed point), and define an algorithm $\{x_n\}$ in terms of some step size $\{\alpha_n\}$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$, and we show that

$$\|x_{n+1} - x^*\| \leq \|x_n - x^*\| - \alpha_n F(x_n) + \gamma_n$$

where $F : X \rightarrow \mathbb{R}$ is some function, $\{\gamma_n\}$ a sequence of error terms with $\gamma_n \rightarrow 0$, and both F and $\{\gamma_n\}$ are defined in terms of the mapping T or the space X . We are then usually interested in the convergence behaviour of $\|x_n - x^*\|$ or $F(x_n)$ (i.e. μ_n or β_n), and this leads naturally to the two questions (a) and (b) posed in the introduction.

The first part of this paper focuses exclusively on convergence properties of sequences of real numbers, but in Section 4 we outline several concrete applications of these convergence results, which each adhere to the above phenomenon, namely:

- $\{x_n\}$ is a generalised gradient descent algorithm, $T : X \rightarrow \mathbb{R}$ is a convex, continuous function and x^* a minimiser of T (a new application discussed in Section 4.2.1).
- $\{x_n\}$ a modified Krasnoselkii-Mann scheme, $T : X \rightarrow X$ is an asymptotically ψ -weakly contractive mapping and x^* is a fixed point of T (surveyed in Section 4.3.1).
- $\{x_n\}$ is an implicit scheme, $T : X \rightarrow 2^X$ a set-valued accretive operator and x^* a zero of T (surveyed in Section 4.3.2).

These in turn are just representative examples of a much wider variety of settings in which recursive inequalities of the form (\star) have been applied.

2.2 Preliminaries from proof theory and computable analysis

We now turn our attention to the relevant background from logic, outlining some key definitions and results which will be important in what follows. We let \mathbb{Q}_+ denote the set of strictly positive rational numbers.

Definition 2.1. Let $\{a_n\}$ be a sequence of real number that converges to some limit a . A rate of convergence for $a_n \rightarrow a$ is any function $\phi : \mathbb{Q}_+ \rightarrow \mathbb{N}$ satisfying

$$\forall \varepsilon \in \mathbb{Q}_+ \forall n \geq \phi(\varepsilon) (|a_n - a| \leq \varepsilon)$$

Similarly, ϕ is a rate of Cauchy convergence for $\{a_n\}$ if

$$\forall \varepsilon \in \mathbb{Q}_+ \forall m, n \geq \phi(\varepsilon) (|a_n - a_m| \leq \varepsilon)$$

We say that $\{a_n\}$ has a computable rate of (Cauchy) convergence if it possesses a rate of (Cauchy) convergence ϕ that is computable in the usual sense. Any computable rate of convergence can be converted to a computable rate of Cauchy convergence, and vice-versa.

While the existence of a (not necessarily computable) rate of convergence is logically equivalent to convergence itself, it is well known that there exist convergent sequences of computable numbers that do not have a computable rate of convergence. The canonical examples of this phenomenon are the so-called Specker sequences:

Proposition 2.2 (E. Specker 1949 [32]). *There exists a computable, monotonically increasing, bounded sequence $\{a_n\}$ of rational numbers whose limit a is not a computable real number, and thus no computable rate of convergence for $a_n \rightarrow a$ exists.*

It is not necessarily the case that having a computable limit guarantees the existence of a rate of convergence either: Provided we no longer insist on monotonicity, we can construct a computable sequence of rationals that converges to 0 but with no computable rate. We now state and prove a slightly generalized version of what is essentially a folklore construction, which will be important in Section 3.

Proposition 2.3. *Let $\{a_n\}$ be any strictly decreasing computable sequence of positive rationals that converges to 0. Letting T_m denote the Turing machine with index m , define*

$$s_n := \begin{cases} a_m & \text{for the minimum } m \leq n \text{ such that } T_m \text{ halts on} \\ & \text{input } m \text{ in exactly } n \text{ steps} \\ 0 & \text{if no such } m \text{ exists} \end{cases}$$

Then $s_n \rightarrow 0$ but has no computable rate of convergence.

Proof. Fix $n \in \mathbb{N}$ and let N be such that any of the machines T_i which terminate on input i for $i = 0, \dots, n-1$ do so in at most N steps. Then for any $k \geq N+1$ we have $s_k \leq a_n$: Were this not the case then $s_k = a_m$ where $m \leq k$ is the least such that T_m halts on input m in exactly k steps. But since $\{a_n\}$ is strictly decreasing, we must have $m < n$ and therefore $k \leq N$. Since $a_n \rightarrow 0$ this therefore implies that $s_n \rightarrow 0$.

Now suppose for contradiction that $s_n \rightarrow 0$ with some computable rate of convergence ϕ . We argue that for any $k \in \mathbb{N}$, if T_k halts on input k , then it does so in less than $\psi(k) := \max\{k, \phi(a_{k+1})\}$ steps: If not and $T_k(k)$ halts in n steps for $\psi(k) \leq n$, then since $k \leq n$ we must have $s_n = a_m$ for some $m \leq k$ (i.e. the minimal $m \leq n$ such that $T_m(m)$ halts in exactly n steps). Thus $s_n = a_m \geq a_k > a_{k+1}$. But since ϕ is a rate of convergence for $s_n \rightarrow 0$ and $\phi(a_{k+1}) \leq n$ then $s_n \leq a_{k+1}$, a contradiction. Therefore if ϕ were computable then $\psi(k)$ forms a computable upper bound on the number of steps it takes T_k to halt on input k , contradicting the unsolvability of the halting problem. \square

In scenarios where we are, in general, unable to provide a computable rate of convergence, we can often nevertheless produce computable bounds for the following classically equivalent reformulation, known as *metastable* convergence.

$$\forall \varepsilon \in \mathbb{Q}_+ \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists n \forall k \in [n, n + g(n)] (|a_k - a| \leq \varepsilon) \quad (5)$$

where $[a, a+b] := \{a, a+1, \dots, a+b\}$. Note that (5) easily follows from $a_n \rightarrow a$ since if n is such that $|a_k - a| \leq \varepsilon$ for all $k \geq n$, then we also have $|a_k - a| \leq \varepsilon$ for $k \in [n, n + g(n)]$. The other direction, on the other hand, requires us to reason by contradiction: If $a_n \not\rightarrow a$ then there exists some $\varepsilon > 0$ such that

$$\forall n \exists k \geq n (|a_k - a| > \varepsilon)$$

and therefore there exists some function $g : \mathbb{N} \rightarrow \mathbb{N}$ bounding k in n i.e.

$$\forall n \exists k \in [n, n + g(n)] (|a_k - a| > \varepsilon)$$

which contradicts (5).

Definition 2.4. A rate of metastability is a functional $\Phi : \mathbb{Q}_+ \times (\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N}$ bounding n in (5) i.e.

$$\forall \varepsilon \in \mathbb{Q}_+ \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists n \leq \Phi(\varepsilon, g) \forall k \in [n, n + g(n)] (|a_k - a| \leq \varepsilon)$$

There is an analogous notion of metastability for Cauchy convergence, namely

$$\forall \varepsilon \in \mathbb{Q}_+ \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists n \forall i, j \in [n, n + g(n)] (|a_i - a_j| \leq \varepsilon)$$

and rates of metastability are defined similarly:

$$\forall \varepsilon \in \mathbb{Q}_+ \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists n \leq \Phi(\varepsilon, g) \forall i, j \in [n, n + g(n)] (|a_i - a_j| \leq \varepsilon)$$

From the perspective of logic, metastability (5) is just the Herbrand normal form of convergence, and a rate of metastability a solution to the so-called ‘no-counterexample interpretation’ of the convergence statement [22, 23]. However, metastable convergence theorems have been independently studied in mathematics: In particular, metastable versions of the monotone convergence principle and other ‘infinitary’ statements have been discussed by T. Tao in [33, 34], and have found nontrivial applications in several areas of mathematics. Highly uniform and computable rates of metastability can often be obtained convergence proofs, even when those proofs are nonconstructive and computable rates of convergence are not possible. Indeed, the extraction of rates of metastability using proof theoretic techniques is a standard result in applied proof theory (e.g. [5, 14, 19, 28]).

A number of the results that we study involve convergence or divergence properties of infinite series. We note that in the case that $\sum_{i=0}^{\infty} b_i < \infty$ where $\{b_i\}$ a sequence of nonnegative reals, a rate ϕ of Cauchy convergence for $\sum_{i=0}^{\infty} b_i$ satisfies

$$\forall \varepsilon \in \mathbb{Q}_+ \forall n, m \in \mathbb{N} \left(n \geq \phi(\varepsilon) \implies \sum_{i=n}^{n+m} b_i \leq \varepsilon \right)$$

while a corresponding rate of metastability Φ satisfies

$$\forall \varepsilon \in \mathbb{Q}_+ \forall g : \mathbb{N} \rightarrow \mathbb{N} \exists n \leq \Phi(\varepsilon, g) \left(\sum_{i=n}^{n+g(n)} b_i \leq \varepsilon \right)$$

Finally, for infinite series that diverge, we will need to introduce a computational analogue of this property.

Definition 2.5. Suppose that $\{b_n\}$ is a sequence of nonnegative real numbers such that $\sum_{i=0}^{\infty} b_i = \infty$. A rate of divergence for $\sum_{i=0}^{\infty} b_i = \infty$ is any function $r : \mathbb{N} \times \mathbb{Q}_+ \rightarrow \mathbb{N}$ such that

$$\forall n \in \mathbb{N} \forall x \in \mathbb{Q}_+ \left(\sum_{i=n}^{r(n,x)} b_i \geq x \right)$$

For simplicity in formulating our bounds, we also assume that rates of divergence are monotone in their first argument i.e.

$$m \leq n \implies r(m, x) \leq r(n, x)$$

for all $m, n \in \mathbb{N}$ and $x \in \mathbb{Q}_+$.

Remark 2.6. Our monotonicity assumption is harmless: Any (potentially non-monotone) rate of divergence r can be computably converted to one that is monotone by setting

$$\tilde{r}(n, x) := \max\{r(k, x) \mid k \leq n\}$$

Then \tilde{r} is monotone in its first argument, to see that it is also a rate of divergence, note that $\tilde{r}(n, x) = r(k, x) \geq r(n, x)$ for some $k \leq n$, and so

$$\sum_{i=n}^{\tilde{r}(n, x)} b_i \geq \sum_{i=n}^{r(n, x)} b_i \geq x$$

3 Abstract quantitative results

In this first main section, we restrict our attention to convergence properties of sequences of reals satisfying our main recursive inequality. To recall: for the remainder of the section, $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ will be sequences of nonnegative real numbers such that

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \gamma_n \tag{*}$$

for all $n \in \mathbb{N}$. Moreover, we will always assume that

- $\sum_{i=0}^{\infty} \alpha_i = \infty$, and
- $\gamma_n \rightarrow 0$ as $n \rightarrow \infty$.

Our main question will concern the convergence behaviour of $\{\mu_n\}$ and $\{\beta_n\}$, and in particular under which conditions rates of convergence are possible.

We begin with the observation that the above conditions alone do not guarantee convergence of either $\{\mu_n\}$ or $\{\beta_n\}$: After all, setting

$$\alpha_n = \frac{1}{n+1}, \quad \beta_n = \sqrt{n+1}, \quad \gamma_n = \frac{2}{\sqrt{n+1}}, \quad \mu_0 = 0 \quad \text{and} \quad \mu_{n+1} = \sum_{i=0}^n \frac{1}{\sqrt{i+1}}$$

we see that $(*)$ is satisfied with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\gamma_n \rightarrow 0$, but both $\{\mu_n\}$ and $\{\beta_n\}$ diverge. However, there are two key ways of strengthening $\gamma_n \rightarrow 0$ so that we start to see better behaviour, and these will in turn form the two main subclasses that will be studied below. The following facts are well-known (cf. [3]):

Proposition 3.1 ([3]). *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers satisfying $(*)$, and that $\sum_{i=0}^{\infty} \alpha_i = \infty$. Then either one of the following conditions implies the existence of a subsequence $\{\beta_{l_n}\}$ with $\beta_{l_n} \rightarrow 0$:*

$$(I) \quad \sum_{i=0}^{\infty} \gamma_i < \infty$$

(II) $\gamma_n/\alpha_n \rightarrow 0$ as $n \rightarrow \infty$.

Moreover, in the case (I) we also have $\mu_n \rightarrow \mu$ for some limit $\mu \geq 0$.

Proof. We suppose that there exists $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $\beta_n > \varepsilon$ for all $n \geq N$, and show that each of (I) and (II) lead to a contradiction, from which we can infer the existence of a subsequence $\{\beta_{l_n}\}$ converging to zero.

(I) Rearranging (\star) we obtain

$$\alpha_n \varepsilon < \alpha_n \beta_n \leq \mu_n - \mu_{n+1} + \gamma_n$$

for all $n \geq N$, and thus

$$\varepsilon \sum_{i=N}^n \alpha_i \leq \mu_N - \mu_{n+1} + \sum_{i=N}^n \gamma_i \leq \mu_N + \sum_{i=0}^{\infty} \gamma_i$$

From $\sum_{i=0}^{\infty} \gamma_i < \infty$ we therefore infer $\sum_{i=0}^{\infty} \alpha_i < \infty$, a contradiction.

(II) Pick $N_0 \geq N$ such that $\gamma_n/\alpha_n \leq \varepsilon/2$ for all $n \geq N_0$. Then from (\star) and $\beta_n > \varepsilon$ we have

$$\mu_{n+1} < \mu_n - \alpha_n \varepsilon + \frac{\alpha_n \varepsilon}{2} = \mu_n - \frac{\alpha_n \varepsilon}{2}$$

for all $n \geq N_0$, and therefore rearranging:

$$\frac{\varepsilon}{2} \sum_{i=N_0}^n \alpha_i \leq \mu_{N_0} - \mu_{n+1} \leq \mu_{N_0}$$

and so again we have $\sum_{i=0}^{\infty} \alpha_i < \infty$, a contradiction.

The additional fact that $\{\mu_n\}$ converges to some limit in case (I) follows from the a general result that whenever

$$\mu_{n+1} \leq \mu_n + \gamma_n$$

for $\sum_{i=0}^{\infty} \gamma_i < \infty$, then $\{\mu_n\}$ converges, noting that since $\alpha_n \beta_n \geq 0$ then

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \gamma_n \leq \mu_n + \gamma_n$$

We omit details here, but the proof can be found in e.g. [2, Proposition 1]. \square

The two cases (I) and (II) outlined above will be considered in Sections 3.1 and 3.2 respectively, where in each case further conditions are added to establish stronger convergence results. Before proceeding, we observe that the two cases are disjoint.

Example 3.2. (a) Define $\alpha_n = 1/(n+1)$, $\beta_n = 0$ and $\mu_{n+1} = \mu_n + \gamma_n$ for some $\mu_0 \geq 0$, where

$$\gamma_n = \begin{cases} 1/(n+1) & \text{if } n+1 = m^2 \text{ for some } m \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

Then these sequences satisfy case (I) of Proposition 3.1 but not case (II), since

$$\sum_{i=0}^{\infty} \gamma_i = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

but whenever $n+1 = m^2$ we have $\gamma_n/\alpha_n = 1$, therefore we cannot have $\gamma_n/\alpha_n \rightarrow 0$.

(b) Define $\alpha_n = \beta_n = 1/\sqrt{n+1}$, $\gamma_n = 1/(n+1)$ and $\mu_n = \mu_0$ for some $\mu_0 \geq 0$. Then these sequences satisfy case (II) of Proposition 3.1 but not case (I).

3.1 Case I: Convergence with $\sum_{i=0}^{\infty} \gamma_i < \infty$

As stated in Proposition 3.1 above, if $\sum_{i=0}^{\infty} \gamma_i < \infty$, then $\{\mu_n\}$ converges to some limit independently of any properties of $\{\alpha_n\}$ and $\{\beta_n\}$. Indeed, more general results of this kind are possible: For instance, Qihou [30] proves that if

$$\mu_{n+1} \leq (1 + \delta_n)\mu_n + \gamma_n \quad (6)$$

where $\{\delta_n\}$ is also sequence of nonnegative reals with $\sum_{i=0}^{\infty} \delta_n < \infty$, then $\{\mu_n\}$ converges. Convergence of $\{\beta_n\}$, on the other hand, is more subtle, and requires additional assumptions beyond $\sum_{i=0}^{\infty} \gamma_i < \infty$, as the following example demonstrates:

Example 3.3. Let $\alpha_n := 1/(n+1)$, $\gamma_n := 0$ and $\beta_n := 1$ if $n+1 = m^2$ else 0. Then

$$\sum_{i=0}^{\infty} \alpha_i \beta_i = \sum_{i=m^2}^{\infty} \frac{1}{i} = \sum_{i=0}^{\infty} \frac{1}{i^2} \leq L$$

for sufficiently large L . Now let $\mu_0 := L$ and $\mu_{n+1} := L - \sum_{i=0}^n \alpha_i \beta_i$. Then $\mu_n \geq 0$ for all $n \in \mathbb{N}$ and $\mu_{n+1} \leq \mu_n - \alpha_n \beta_n$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$, but $\{\beta_n\}$ does not converge.

A key result in this direction is due to Alber et al., which adds a regularity condition to the $\{\beta_n\}$ to establish convergence to zero:

Proposition 3.4 (cf. Lemma 2.2 of [3]). *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers satisfying (\star) with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \gamma_i < \infty$. Then whenever there exists $\theta > 0$ such that the following conditions holds:*

$$\beta_n - \beta_{n+1} \leq \theta \alpha_n \quad \text{for all } n \in \mathbb{N}$$

then $\mu_n \rightarrow \mu$ for some $\mu \geq 0$, and $\beta_n \rightarrow 0$.

The above result and its variants, proofs of which are typically by contradiction, plays a particularly important role in establishing convergence properties of subgradient algorithms (cf. [3, 4, 12, 26] as examples), and we will explore this domain of application later in Section 4.2. Therefore the question of whether computable rates of convergence for $\mu_n \rightarrow \mu$ and particularly $\beta_n \rightarrow 0$ can be produced is important, as these could be used to obtain computable rates of convergence for a very general class of subgradient methods. Our first main result gives a negative answer, and we also conjecture that the main construction here could be easily adapted to prove similar results for the numerous variants on Proposition 3.4 found in the literature. We first need a simple lemma which adapts the standard construction given in Proposition 2.3.

Lemma 3.5. *There exists a computable sequence $\{s_n\}$ of rational numbers with $\sum_{i=0}^{\infty} s_i < \infty$ such that $s_n \rightarrow 0$ with no computable rate of convergence.*

Proof. Let $\{a_n\}$ and $\{s_n\}$ be as in Proposition 2.3 but with $\sum_{i=0}^{\infty} a_i < \infty$ (e.g. $a_n = 1/(n+1)^2$). Define $b_n := a_n$ if there exists some $k \in \mathbb{N}$ such that $s_k = a_n$, and 0 otherwise, and note that $\sum_{i=0}^{\infty} b_i \leq \sum_{i=0}^{\infty} a_i < \infty$. Let $\{s_{n_i}\}$ (respectively $\{b_{m_j}\}$), be the subsequence of $\{s_n\}$ (respectively $\{b_n\}$) consisting of the sequence's nonzero elements. Then for each index i there is exactly one j such that $s_{n_i} = b_{m_j}$, and vice-versa, where for uniqueness we note that if $s_{n_i} = s_{n_{i'}} = b_{m_j}$, then $T_{m_j}(m_j)$ halts in exactly n_i and $n_{i'}$ steps, and thus $i = i'$. This means there is a bijection between $\{s_{n_i}\}$ and $\{b_{m_j}\}$, and therefore

$$\sum_{i=0}^{\infty} s_i = \sum_{i=0}^{\infty} s_{n_i} = \sum_{j=0}^{\infty} b_{m_j} = \sum_{j=0}^{\infty} b_j < \infty$$

where the first equality follows since $\{s_i\}$ is just $\{s_{n_i}\}$ padded out with zero elements (and similarly for the third), while the second equality follows from the fact that we can reorder the terms in series where all terms are positive. \square

Theorem 3.6. *For any sequence of positive reals $\{\alpha_n\}$ bounded above by $\alpha > 0$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$, together with $\theta > 0$, there exist sequences of positive reals $\{\mu_n\}$ and $\{\beta_n\}$, computable in $\{\alpha_n\}$ and θ and satisfying*

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n \quad \text{and} \quad \beta_n - \beta_{n+1} \leq \theta \alpha_n$$

such that $\mu_n \rightarrow \mu$ and $\beta_n \rightarrow 0$, but neither with a computable rate of convergence.

Proof. Take $\{s_n\}$ as in Proposition 3.5 and define $l : \mathbb{N} \rightarrow \mathbb{N}$ recursively by $l(0) := 0$, and $l(n+1) := k+1$ where $k \geq l(n)$ is the least number satisfying

$$\sum_{i=l(n)}^k \alpha_i \geq \frac{|s_n - s_{n+1}|}{\theta} \tag{7}$$

which is well defined by $\sum_{i=0}^{\infty} \alpha_i = \infty$. Now define $\{\beta_k\}$ as follows: $\beta_{l(n)} := s_n$, and if $l(n) < k < l(n+1)$ then

$$\beta_k := s_n + \text{sgn}(s_{n+1} - s_n) \cdot \theta \sum_{i=l(n)}^{k-1} \alpha_i$$

The intuition here is that $\{\beta_n\}$ represents a ‘padding out’ of $\{s_n\}$, designed in such a way as to ensure that the distance between consecutive elements is bounded by $\theta\alpha_n$. Since $l(n)$ is strictly increasing with $l(0) = 0$, $\{\beta_k\}$ is thereby defined for all $k \in \mathbb{N}$. We now show that $|\beta_k - \beta_{k+1}| \leq \theta\alpha_k$. There are two cases to deal with: If $l(n) \leq k < l(n+1) - 1$ then

$$|\beta_k - \beta_{k+1}| = |\text{sgn}(s_{n+1} - s_n) \cdot \theta\alpha_k| \leq \theta\alpha_k$$

and if $k = l(n+1) - 1$ then (assuming w.l.o.g. that $|s_n - s_{n+1}| > 0$, otherwise the claim is trivially true):

$$\begin{aligned} |\beta_{l(n+1)-1} - \beta_{l(n+1)}| &= \left| s_n + \text{sgn}(s_{n+1} - s_n) \cdot \theta \sum_{i=l(n)}^{l(n+1)-2} \alpha_i - s_{n+1} \right| \\ &= |\text{sgn}(s_{n+1} - s_n)| \cdot \left| \theta \sum_{i=l(n)}^{l(n+1)-2} \alpha_i - |s_n - s_{n+1}| \right| \\ &= |s_n - s_{n+1}| - \theta \sum_{i=l(n)}^{l(n+1)-2} \alpha_i \\ &\leq \theta \sum_{i=l(n)}^{l(n+1)-1} \alpha_i - \theta \sum_{i=l(n)}^{l(n+1)-2} \alpha_i \\ &= \theta\alpha_{l(n+1)-1} \end{aligned}$$

where if $l(n+1) - 1 = l(n)$ we use the convention that $\sum_{i=l(n)}^{l(n+1)-2} \alpha_i = 0$. For the third equality we use that $|\text{sgn}(s_n - s_{n+1})| = 1$ and $\theta \sum_{i=l(n)}^{l(n+1)-2} \alpha_i \leq |s_n - s_{n+1}|$, since otherwise this would contradict the construction of $l(n+1)$.

From the defining property of $l(n+1)$ we can also show that for $l(n) \leq k < l(n+1)$ we have either $s_n \leq \beta_k \leq s_{n+1}$ or $s_{n+1} \leq \beta_k \leq s_n$, and from this and the fact that $s_n \rightarrow 0$ it follows that $\beta_n \rightarrow 0$.

To show that there is no computable rate of convergence for $\beta_n \rightarrow 0$, suppose for contradiction that ϕ is such a rate. Fixing $\varepsilon > 0$, we have $\beta_n \leq \varepsilon$ for all $n \geq \phi(\varepsilon)$. However, since $l(n)$ is strictly monotone, we have $l(n) \geq n$ for all $n \in \mathbb{N}$, and thus $s_n = \beta_{l(n)} \leq \varepsilon$ for all $n \geq \phi(\varepsilon)$. Therefore ϕ is also a computable rate of convergence for $s_n \rightarrow 0$, which is not possible.

Finally, to define $\{\mu_n\}$, we first show that $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$. Let $c > 0$ be any upper bound on $\{s_n\}$. Using again that $s_n \leq \beta_k \leq s_{n+1}$ or $s_{n+1} \leq \beta_k \leq s_n$ for $l(n) \leq k < l(n+1)$, we have $\beta_k \leq s_n + s_{n+1}$ for k in this range, and therefore

$$\begin{aligned} \sum_{i=l(n)}^{l(n+1)-1} \alpha_i \beta_i &\leq (s_n + s_{n+1}) \sum_{i=l(n)}^{l(n+1)-1} \alpha_i \\ &\leq (s_n + s_{n+1}) \left(\frac{|s_n - s_{n+1}|}{\theta} + \alpha_{l(n+1)-1} \right) \\ &\leq (s_n + s_{n+1})(c/\theta + \alpha) \end{aligned}$$

where for the second inequality we use the defining property of $l(n+1)$, noting that since $l(n+1) - 1 = k$ for the *least* k satisfying 7, we must have

$$\sum_{i=l(n)}^{l(n+1)-1} \alpha_i = \sum_{i=l(n)}^k \alpha_i \leq \frac{|s_n - s_{n+1}|}{\theta} + \alpha_k$$

since if this were not the case then

$$\sum_{i=l(n)}^{k-1} \alpha_i > \frac{|s_n - s_{n+1}|}{\theta}$$

for $k-1 \geq l(n)$, contradicting minimality of k . Therefore summing over the whole sequence:

$$\begin{aligned} \sum_{i=0}^{\infty} \alpha_i \beta_i &= \sum_{n=0}^{\infty} \sum_{i=l(n)}^{l(n+1)-1} \alpha_i \beta_i \\ &\leq (c/\theta + \alpha) \sum_{n=0}^{\infty} (s_n + s_{n+1}) \\ &\leq 2(c/\theta + \alpha) \sum_{n=0}^{\infty} s_n < \infty \end{aligned}$$

Now to construct $\{\mu_n\}$, first of all, let $L > 0$ be any bound with $\sum_{i=0}^{\infty} \alpha_i \beta_i \leq L$, and let $\{\nu_n\}$ be some computable monotonically decreasing sequence converging to some limit $\nu \geq 0$. Define

$$\mu_0 := L + \nu_0 \quad \text{and} \quad \mu_{n+1} := L + \nu_{n+1} - \sum_{i=0}^n \alpha_i \beta_i$$

Then $\mu_{n+1} \leq \mu_n - \alpha_n \beta_n$, and we now claim that if exactly one of $\{\nu_n\}$ and $\sum_{i=0}^{\infty} \alpha_i \beta_i$ have a computable rate of Cauchy convergence, then $\{\mu_n\}$ cannot have a computable rate of Cauchy convergence. To see this we observe that for any $m \geq n$:

$$\mu_m - \mu_n = (\nu_m - \nu_n) - \sum_{i=n-1}^{m-1} \alpha_i \beta_i$$

and so if $\{\mu_n\}$ has a computable rate of Cauchy convergence, then a computable rate of convergence for either one of $\{\nu_n\}$ or $\sum_{i=0}^{\infty} \alpha_i \beta_i$ would imply a computable rate for the other.

It remains to show that we can always define $\{\nu_n\}$ such that exactly one of $\{\nu_n\}$ or $\sum_{i=0}^{\infty} \alpha_i \beta_i$ have a computable rate of convergence: If $\sum_{i=0}^{\infty} \alpha_i \beta_i$ has a computable rate, then let $\{\nu_n\}$ be some computable, monotonically decreasing sequence of rationals that converges to some noncomputable $\nu > 0$ (which exists by a simple adaptation of Proposition 2.2). On the other hand, if $\sum_{i=0}^{\infty} \alpha_i \beta_i$ does not have a computable rate, just set $\nu_n = 0$ for all $n \in \mathbb{N}$.

The final claim that $\{\mu_n\}$ and $\{\beta_n\}$ are computable sequences of rationals when $\{\alpha_n\}$ is follows by observing that all of the above constructions are computable, including the construction $\{\nu_n\}$ above, along with $\{s_n\}$ as taken from Proposition 2.3. We note that our proof gives two possible constructions for $\{\mu_n\}$, both of which are computable in $\{\alpha_n\}$ and θ , according to whether or not $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with a computable rate of convergence, and thus the theorem holds, even if we cannot read off from the proof which of these two constructions we would use. \square

As a consequence of Theorem 3.6, we cannot in general obtain rates of convergence for $\mu_n \rightarrow \mu$ and $\beta_n \rightarrow 0$ in Proposition 3.4, since even in the special case that $\gamma_n = 0$ and $\{\alpha_n\}$ is computable such that $\sum_{i=0}^{\infty} \alpha_i = \infty$ with a computable rate of divergence (e.g. $\alpha_n = 1$), there exist computable $\{\mu_n\}$ and $\{\beta_n\}$ such that all of the conditions of the proposition are satisfied, but neither $\{\mu_n\}$ nor $\{\beta_n\}$ converge with a computable rate. In the special case that $\alpha_n := 1$, we have $\sum_{i=0}^{\infty} \alpha_i \beta_i = \sum_{i=0}^{\infty} \beta_i < \infty$, and a rate of convergence for the latter would also imply a rate of convergence for $\beta_n \rightarrow 0$, which is not possible, therefore in this case we can also decide which of the two constructions for $\{\mu_n\}$ we would use, defining $\mu_n := L - \sum_{i=0}^n \beta_i$.

Having shown that computable rates of convergence are not possible in general, we can nevertheless give a quantitative version of Proposition 3.4 in terms of rates of metastability. We first observe that quantitative versions of recursive inequalities of the form (6) have been widely used in the applied proof theory literature, and rates of metastability for $\{\mu_n\}$ which are given in e.g. [17] would apply directly to the case of Proposition 3.4. As such, we focus here on completing the picture by giving a quantitative version of the property that $\beta_n \rightarrow 0$. In this case, the core of Proposition 3.4 is contained in the following result:

Proposition 3.7 (cf. Proposition 2 of [4]). *Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$. Then whenever there exists $\theta > 0$ such that the following condition holds:*

$$\beta_n - \beta_{n+1} \leq \theta \alpha_n \quad \text{for all } n \in \mathbb{N}$$

then $\beta_n \rightarrow 0$.

To derive $\beta_n \rightarrow 0$ under the conditions of Proposition 3.4, we simply observe that by (\star) and condition (i) of Proposition 3.4 we have

$$\sum_{i=0}^n \alpha_i \beta_i \leq \mu_0 - \mu_{n+1} + \sum_{i=0}^n \gamma_i \leq \mu_0 + \sum_{i=0}^{\infty} \gamma_i$$

and thus $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$, and so Proposition 3.7 applies directly.

The original proof of Proposition 3.7 uses a somewhat intricate and highly ineffective subsequence construction (cf. Proposition 2 of [4]). We now provide a strengthening of the result (and hence also of Proposition 3.4) which establishes a necessary and sufficient condition for $\beta_n \rightarrow 0$ under $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$, or in

the case of Proposition 3.7, under $\sum_{i=0}^{\infty} \gamma_i < \infty$. We then follow this with a quantitative variant, giving a simple constructive formulation of the standard nonconstructive proof strategy exhibited in e.g. [4].

Proposition 3.8. *Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers with $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$. Then $\beta_n \rightarrow 0$ if and only if there exists $\theta > 0$ such that*

$$\limsup_{N \rightarrow \infty} \left\{ \beta_n - \beta_m - \theta \sum_{i=n}^{m-1} \alpha_i \mid N \leq n < m \right\} \leq 0$$

We initially leave Proposition 3.8 without proof, as one direction will follow from the quantitative version of the result below.

Theorem 3.9 (Quantitative version of Proposition 3.8). *Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers and r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$. Let $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ be arbitrary. Then if $N_1, N_2 \in \mathbb{N}$ and $\theta > 0$ are such that, setting $N := \max\{N_1, N_2\}$, we have*

$$\sum_{i=N_1}^{r(N+g(N), \varepsilon/4\theta)} \alpha_i \beta_i \leq \frac{\varepsilon^2}{8\theta}$$

and

$$\beta_n - \beta_m \leq \theta \sum_{i=n}^{m-1} \alpha_i + \frac{\varepsilon}{4} \quad \text{for all } N_2 \leq n < m \leq r\left(N + g(N), \frac{\varepsilon}{4\theta}\right)$$

then we can conclude that $\beta_n \leq \varepsilon$ for all $n \in [N, N + g(N)]$.

Proof. Fix $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$, and suppose for contradiction that $\beta_n > \varepsilon$ for some $n \in [N, N + g(N)]$. We first show that there exists some $m \in [n, r(n, \varepsilon/4\theta)]$ with $\beta_m \leq \varepsilon/2$: If this were not the case then using monotonicity of r in its first component (cf. Remark 2.6) we would have

$$\frac{\varepsilon}{4\theta} \leq \sum_{i=n}^{r(n, \varepsilon/4\theta)} \alpha_i < \frac{2}{\varepsilon} \sum_{i=n}^{r(n, \varepsilon/4\theta)} \alpha_i \beta_i \leq \frac{2}{\varepsilon} \sum_{i=N_1}^{r(N+g(N), \varepsilon/4\theta)} \alpha_i \beta_i \leq \frac{\varepsilon}{4\theta}$$

which contradicts the claim, where for the third inequality we use that $N_1 \leq N \leq n$, and $n \leq N + g(N)$ together with our assumption that rates of divergence r are monotone in their first argument. Now let $n < m \leq r(n, \varepsilon/4\theta)$ be the least such index such that $\beta_n > \varepsilon$, $\beta_i > \varepsilon/2$ for $i = n, \dots, m-1$ and $\beta_m \leq \varepsilon/2$. Then since $N_2 \leq N \leq n$ we have

$$\begin{aligned} \frac{\varepsilon}{2} < \beta_n - \beta_m &\leq \theta \sum_{i=n}^{m-1} \alpha_i + \frac{\varepsilon}{4} \leq \frac{2\theta}{\varepsilon} \sum_{i=n}^{m-1} \alpha_i \beta_i + \frac{\varepsilon}{4} \\ &\leq \frac{2\theta}{\varepsilon} \sum_{i=N_1}^{r(N+g(N), \varepsilon/4\theta)} \alpha_i \beta_i + \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2} \end{aligned}$$

and so we have our contradiction. \square

Proof of Proposition 3.8. To establish one direction we use Theorem 3.9. If $\sum_{i=0}^{\infty} \alpha_i = \infty$ then it possesses some rate of divergence, and for any $\varepsilon > 0$ there must exist some N_1 and N_2 such that

$$\sum_{i=N_1}^{\infty} \alpha_i \beta_i \leq \frac{\varepsilon^2}{8\theta}$$

and

$$\beta_n - \beta_m - \theta \sum_{i=n}^{m-1} \alpha_i \leq \frac{\varepsilon}{4}$$

for all $N_2 \leq n < m$. Therefore the assumptions of Proposition 3.9 are satisfied for arbitrary $g : \mathbb{N} \rightarrow \mathbb{N}$, thus $\beta_n \leq \varepsilon$ for all $n \in [N, N + g(N)]$ also for arbitrary g . But this just means that $\beta_n \leq \varepsilon$ for all $n \geq N$, and so $\beta_n \rightarrow 0$.

In the other direction, if $\beta_n \rightarrow 0$ then there exists $N \in \mathbb{N}$ such that $|\beta_n - \beta_m| \leq \varepsilon$ for all $m, n \geq N$. But then for $m > n \geq N$ we have

$$\beta_n - \beta_m \leq |\beta_n - \beta_m| \leq \varepsilon \leq \sum_{i=n}^{m-1} \alpha_i + \varepsilon$$

and so $\limsup_{N \rightarrow \infty} \left\{ \beta_n - \beta_m - \sum_{i=n}^{m-1} \alpha_i \mid N \leq n < m \right\} \leq 0$. \square

Proposition 3.7 follows directly from Proposition 3.8 because if $\beta_n - \beta_{n+1} \leq \theta \alpha_n$ for all $n \in \mathbb{N}$, we have

$$\beta_n - \beta_m - \theta \sum_{i=n}^m \alpha_i \leq 0$$

for any $n < m$. In particular, we can set $N_2 := 0$ in Theorem 3.9 and replace $\varepsilon^2/8\theta$ with $\varepsilon^2/4\theta$ for the property of N_1 , so that Theorem 3.9 simplifies to the following:

Corollary 3.10. *Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequence of nonnegative real numbers and r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$. Let $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ be arbitrary. Then if $N \in \mathbb{N}$ and $\theta > 0$ are such that*

$$\sum_{i=N}^{r(N+g(N), \varepsilon/2\theta)} \alpha_i \beta_i \leq \frac{\varepsilon^2}{4\theta} \quad \text{and} \quad \beta_n - \beta_{n+1} \leq \theta \alpha_n \quad \text{for all } n \in \mathbb{N}$$

then $\beta_n \leq \varepsilon$ for all $n \in [N, N + g(N)]$.

As an immediate corollary we obtain the following:

Corollary 3.11. *Suppose that $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of nonnegative real numbers and r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$, and that there is some $\theta > 0$ such that $\beta_n - \beta_{n+1} \leq \theta \alpha_n$ for all $n \in \mathbb{N}$. Then*

- (a) If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of convergence ϕ , then $\beta_n \rightarrow 0$ with rate of convergence

$$\psi(\varepsilon) := \phi\left(\frac{\varepsilon^2}{4\theta}\right)$$

- (b) If $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ with rate of metastability Φ , then $\beta_n \rightarrow 0$ with rate of metastability

$$\Psi(\varepsilon, g) := \Phi\left(\frac{\varepsilon^2}{4\theta}, h\right) \text{ for } h(n) := r\left(n + g(n), \frac{\varepsilon}{2\theta}\right) - n$$

3.2 Case II: Convergence with $\gamma_n/\alpha_n \rightarrow 0$

Having thoroughly analysed (\star) under the additional condition that $\sum_{i=0}^{\infty} \gamma_i < \infty$, we now move onto the second condition that allows us to obtain favourable convergence results on $\{\mu_n\}$ and $\{\beta_n\}$. This second case has a slightly different flavour: Here the condition on $\{\gamma_n\}$ is primarily used to ensure convergence of $\{\mu_n\}$ towards zero. The fact that $\{\beta_n\}$ possesses a subsequence converging to zero is crucial for this, together with a requirement that there exists some modulus linking this subsequence to a corresponding subsequence of $\{\mu_n\}$. We show that the existence of such a modulus is also necessary when our sequences are strictly positive. Our starting point is a standard convergence result from the literature.

Proposition 3.12 (cf. Lemma 2.5 of [3]). *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers with $\alpha_n > 0$ for all $n \in \mathbb{N}$, satisfying (\star) with $\sum_{i=0}^{\infty} \alpha_i = \infty$, $\{\alpha_n\}$ bounded and $\gamma_n/\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Then there exists some sequence of indices $\{l_n\}$ such that $\beta_{l_n} \rightarrow 0$ as $n \rightarrow \infty$, and whenever, in addition, we have $\mu_{l_n} \rightarrow 0$, then $\mu_n \rightarrow 0$ as $n \rightarrow \infty$.*

Typically, the above result is applied in cases where $\beta_n = \psi(\mu_n)$ for some continuous, nondecreasing function $\psi : [0, \infty) \rightarrow [0, \infty)$ which is strictly positive on $(0, \infty)$ and satisfies $\psi(0) = 0$. These properties allow us to infer $\mu_{l_n} \rightarrow 0$ from $\psi(\mu_{l_n}) \rightarrow 0$, and thus $\lim_{n \rightarrow \infty} \mu_n = 0$. We now give an alternative formulation of Proposition 3.12 above, where we replace $\beta_n = \psi(\mu_n)$ with a more general modulus ω . In the case that the elements $\{\beta_n\}$ are strictly positive, the existence of such a modulus is even necessary for $\mu_n \rightarrow 0$, which justifies it as a natural property.

Proposition 3.13. *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers with $\alpha_n > 0$ for all $n \in \mathbb{N}$, satisfying (\star) with $\sum_{i=0}^{\infty} \alpha_i = \infty$, $\{\alpha_n\}$ bounded and $\gamma_n/\alpha_n \rightarrow 0$. Then if there exists a function $\omega : \mathbb{Q}_+ \rightarrow \mathbb{Q}_+$ such that for every $\varepsilon > 0$ we have*

$$\beta_n \leq \omega(\varepsilon) \implies \mu_n \leq \varepsilon \tag{8}$$

for all $n \in \mathbb{N}$, then $\mu_n \rightarrow 0$ as $n \rightarrow \infty$. Conversely, if $\mu_n \rightarrow 0$ and $\mu_n > 0 \implies \beta_n > 0$ for all $n \in \mathbb{N}$, then there necessarily exists an ω satisfying (8) for all $\varepsilon > 0$ and $n \in \mathbb{N}$.

Proof. One direction is by Proposition 3.12. Here $\gamma_n/\alpha_n \rightarrow 0$ implies there exists some subsequence $\beta_{l_n} \rightarrow 0$ as outlined in Proposition 3.1 (II), and (8) implies that $\mu_{l_n} \rightarrow 0$. To see this, note that for any $\varepsilon > 0$ we can choose N such that $\beta_{l_n} \leq \omega(\varepsilon)$ for all $n \geq N$, and so $\mu_{l_n} \leq \varepsilon$ for all $n \geq N$. Therefore $\mu_n \rightarrow 0$ by Proposition 3.12. For the other direction, assume now that $\mu_n \rightarrow 0$ and $\mu_n > 0 \implies \beta_n > 0$. Given $\varepsilon > 0$ let $N \in \mathbb{N}$ be such that $\mu_n \leq \varepsilon$ for all $n \geq N$, and set

$$\omega(\varepsilon) := \frac{1}{2} \cdot \min\{\beta_n > 0 \mid n < N\}$$

or just $\omega(\varepsilon) := 1$ if $\beta_n = 0$ for all $n < N$. Then supposing for contradiction that there is $n \in \mathbb{N}$ such that $\beta_n \leq \omega(\varepsilon)$ and $\mu_n > \varepsilon$, then we must firstly have $\beta_n > 0$ (using $\mu_n > 0 \implies \beta_n > 0$) and secondly $n < N$ (using $\mu_n \leq \varepsilon$ for $n \geq N$). But then $\beta_n \leq \omega(\varepsilon) \leq \beta_n/2$, a contradiction for $\beta_n > 0$. \square

So far we have focused on convergence of $\{\mu_n\}$. In most concrete applications, where $\beta_n = \psi(\mu_n)$, we also get $\beta_n \rightarrow 0$ as $n \rightarrow \infty$ from some basic property of ψ , e.g. continuity or monotonicity. We can also make this intuition formal in terms of a modulus going in the other direction:

Proposition 3.14. *Under the conditions of Proposition 3.13, together with the existence of a modulus ω satisfying (8), suppose that there also exists some $\tilde{\omega} : \mathbb{Q}_+ \rightarrow \mathbb{Q}_+$ such that for every $\varepsilon > 0$ we have*

$$\mu_n \leq \tilde{\omega}(\varepsilon) \implies \beta_n \leq \varepsilon \tag{9}$$

for all $n \in \mathbb{N}$. Then $\beta_n \rightarrow 0$ as $n \rightarrow \infty$, and if ϕ is a rate of convergence for $\mu_n \rightarrow 0$ then $\phi \circ \tilde{\omega}$ is a rate of convergence for $\beta_n \rightarrow 0$. Conversely, if $\beta_n \rightarrow 0$ and $\beta_n > 0 \implies \mu_n > 0$ for all $n \in \mathbb{N}$, then there exists a $\tilde{\omega}$ satisfying (9) for all $\varepsilon > 0$ and $n \in \mathbb{N}$.

Proof. In one direction, we have $\mu_n \rightarrow 0$ by Proposition 3.13, and then $\beta_n \rightarrow 0$ follows easily from (9). The other direction is proved analogously to second part of the proof of Proposition 3.13. \square

Example 3.15. Suppose that the conditions of Proposition 3.13 hold and $\beta_n := \psi(\mu_n)$ for some nondecreasing ψ with $\psi(0) = 0$ and $\psi(t) > 0$ for $t > 0$. Then (8) holds for $\omega(\varepsilon) := \psi(\varepsilon)/2$, since if $\mu_n > \varepsilon$ then $\psi(\mu_n) \geq \psi(\varepsilon) > \omega(\varepsilon)$, and (9) holds for $\tilde{\omega}(\varepsilon) := \mu_N$ for the least N such that $\beta_N \leq \varepsilon$ (this is well defined by the existence of a subsequence $\beta_{l_n} \rightarrow 0$). To see the latter, if $\mu_n \leq \mu_N$ then $\beta_n = \psi(\mu_n) \leq \psi(\mu_N) = \beta_N \leq \varepsilon$. Alternatively, if ψ is also continuous at 0, then $\tilde{\omega}$ can be defined to be any modulus of continuity.

The convergence theorems above play a role in numerous different convergence proofs in nonlinear analysis (some examples of which will be given in Section 4). Just as in the previous section, we now present some negative results which demonstrate the limitations on obtaining computable rates of convergence for $\mu_n \rightarrow 0$.

Proposition 3.16. *We can construct computable sequences $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ satisfying (\star) , with $\alpha_n > 0$ for all $n \in \mathbb{N}$, $\sum_{i=0}^{\infty} \alpha_i = \infty$, $\{\alpha_n\}$ bounded and $\gamma_n/\alpha_n \rightarrow 0$, along with ω satisfying (8), such that each of the following scenarios holds:*

- (a) ω is computable, but $\mu_n \rightarrow 0$ with no computable rate of convergence;
- (b) $\gamma_n/\alpha_n \rightarrow 0$ with computable rate of convergence, but $\mu_n \rightarrow 0$ with no computable rate of convergence.

Proof. (a) Define $\mu_n = \beta_n := s_n$ as in Proposition 2.3 for any $\{a_n\}$, $\alpha_n := 1$ and $\gamma_n := s_{n+1}$. Then (\star) is satisfied since

$$\mu_n - \alpha_n \beta_n + \gamma_n = s_n - s_n + s_{n+1} = s_{n+1} = \mu_{n+1}$$

we have that (8) is satisfied by $\omega(\varepsilon) := \varepsilon$, and $\gamma_n/\alpha_n = s_{n+1} \rightarrow 0$. But $\mu_n \rightarrow 0$ with no computable rate of convergence by construction.

- (b) Define $\mu_n := s_n$ as in Proposition 2.3 for $a_n := 1/(n+1)$, so that $s_n \leq 1$ for any $n \in \mathbb{N}$. Define $\alpha_n := n+1$, $\beta_n := 1/(n+1)^2$ and $\gamma_n := 2$. Then (\star) is satisfied since

$$\mu_n - \alpha_n \beta_n + \gamma_n = s_n - \frac{1}{n+1} + 2 \geq s_n + 1 \geq s_{n+1} = \mu_{n+1}$$

we have $\gamma_n/\alpha_n = 2/(n+1) \rightarrow 0$ with computable rate of convergence and (8) is satisfied by

$$\omega(\varepsilon) := \frac{1}{(\phi(\varepsilon) + 1)^2}$$

where ϕ is a (necessarily noncomputable) rate of convergence for $s_n \rightarrow 0$. □

This result shows that, in general, there is no way of obtaining a computable rate of convergence for $\mu_n \rightarrow 0$, even in cases where ω is computable, or we have a computable rate of convergence for $\gamma_n/\alpha_n \rightarrow 0$. We now provide a quantitative version of Proposition 3.12, giving a rate of metastability for $\mu_n \rightarrow 0$ in the general case, which can be converted into a rate of convergence if *both* ω is computable *and* we have a computable rate of convergence for $\gamma_n/\alpha_n \rightarrow 0$. This is a metastable generalisation of Lemma 3.1 of [29].

Theorem 3.17. *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of positive real numbers with $\alpha_n > 0$ for all $n \in \mathbb{N}$, satisfying (\star) , and that*

- (i) $c > 0$ is such that $\mu_n \leq c$ for all $n \in \mathbb{N}$
- (ii) $\alpha > 0$ is such that $\alpha_n \leq \alpha$ for all $n \in \mathbb{N}$,
- (iii) $\sum_{i=0}^{\infty} \alpha_i = \infty$ with rate of divergence r ,

Fix $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$, and define $h : \mathbb{Q}_+ \times \mathbb{N} \rightarrow \mathbb{N}$ by

$$h(\delta, m) := \tilde{g} \left(r \left(m, \frac{2c}{\delta} \right) \right) \quad \text{for} \quad \tilde{g}(i) := \max\{j + g(j) \mid j \leq i\}$$

Then if $p \in \mathbb{N}$ and $\delta > 0$ are such that there exists $m \leq p$ such that

$$\gamma_n / \alpha_n \leq \min \left\{ \frac{\delta}{2}, \frac{\varepsilon}{2\alpha} \right\} \quad \text{for all} \quad n \in [m, h(\delta, m)] \quad (10)$$

and

$$\beta_n \leq \delta \implies \mu_n \leq \frac{\varepsilon}{2} \quad \text{for all} \quad n \leq h(\delta, p) \quad (11)$$

then there exists some $N \leq r(p, 2c/\delta)$ such that

$$\mu_n \leq \varepsilon \quad \text{for all} \quad n \in [N, N + g(N)]$$

Proof. We first observe that for all $n \in [m, h(\delta, m)]$ it follows from (10) that

$$\mu_{n+1} \leq \mu_n - \alpha_n(\beta_n - \delta/2) \quad (12)$$

and (using $\alpha_n \leq \alpha$)

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \frac{\varepsilon}{2} \quad (13)$$

Now suppose for contradiction that $\mu_n > \varepsilon/2$ for all $n \in [m, r(m, 2c/\delta)]$. Then since

$$n \leq r \left(m, \frac{2c}{\delta} \right) \leq h(\delta, m) \leq h(\delta, p)$$

where here we use the definition of h and \tilde{g} , and the last step monotonicity of rates of divergence in their first argument, then from (11) we also have $\beta_n > \delta$, and therefore using (12) we have:

$$\mu_{n+1} \leq \mu_n - \alpha_n \cdot \frac{\delta}{2} \quad \text{and thus} \quad \alpha_n \cdot \frac{\delta}{2} < \mu_n - \mu_{n+1}$$

for n in this range. Therefore

$$\frac{\delta}{2} \sum_{i=m}^{r(m, 2c/\delta)} \alpha_i < \mu_m - \mu_{r(m, 2c/\delta)+1} \leq \mu_m \leq c$$

a contradiction. Therefore we have shown that there exists some $N \in [m, r(m, 2c/\delta)]$ such that $\mu_N \leq \varepsilon/2$, and in particular, $N \leq r(m, 2c/\delta) \leq r(p, 2c/\delta)$. We now show that in fact $\mu_n \leq \varepsilon$ for all $n \in [N, N + g(N)]$. We use induction, where the base case $n = N$ follows from the assumption. For the induction step we deal with two cases, first noting that $[N, N + g(N)] \subseteq [m, h(\delta, m)]$ by the definition of h and \tilde{g} . If $\mu_n \leq \varepsilon/2$ then from (13) we have

$$\mu_{n+1} \leq \mu_n - \alpha_n \beta_n + \varepsilon/2 \leq \varepsilon$$

On the other hand, if $\varepsilon/2 < \mu_n \leq \varepsilon$ then we have $\beta_n > \delta$ by (11) and $h(\delta, m) \leq h(\delta, p)$, and from (12)

$$\mu_{n+1} \leq \mu_n - \alpha_n(\beta_n - \delta/2) \leq \mu_n - \alpha_n \delta/2 \leq \mu_n \leq \varepsilon$$

which completes the induction. \square

Corollary 3.18. *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of positive real numbers with $\alpha_n > 0$ for all $n \in \mathbb{N}$, satisfying (\star) and that*

(i) $c > 0$ is such that $\mu_n \leq c$ for all $n \in \mathbb{N}$,

(ii) $\alpha > 0$ is such that $\alpha_n \leq \alpha$ for all $n \in \mathbb{N}$,

(iii) $\sum_{i=0}^{\infty} \alpha_i = \infty$ with rate of divergence r ,

(iv) $\gamma_n/\alpha_n \rightarrow 0$ with rate of convergence ϕ .

Then whenever there exists a modulus $\omega : (0, \infty) \rightarrow (0, \infty)$ such that for all $\varepsilon > 0$ and $n \in \mathbb{N}$:

$$\beta_n \leq \omega(\varepsilon) \implies \mu_n \leq \varepsilon$$

then $\mu_n \rightarrow 0$ with rate of convergence

$$\Phi(\varepsilon) := r \left(\phi \left(\min \left\{ \frac{\omega(\varepsilon/2)}{2}, \frac{\varepsilon}{2\alpha} \right\} \right), \frac{2c}{\omega(\varepsilon/2)} \right)$$

Proof. Directly from Theorem 3.17, defining $\delta := \omega(\varepsilon/2)$ and

$$p := \phi \left(\min \left\{ \frac{\delta}{2}, \frac{\varepsilon}{2\alpha} \right\} \right)$$

and noting that these satisfy the premise of the theorem for any $g : \mathbb{N} \rightarrow \mathbb{N}$. \square

3.2.1 Second variant of Case II

We conclude the section by considering a variant of Proposition 3.12 whose only difference is a slight change in the modulus condition (replacing μ_n with μ_{n+1}), but which comes with a slightly simpler computational interpretation, and, as we will see in Section 4, has already led to interesting applications in a number of areas. The following is a metastable variant of Lemma 3.4 of [21], which generalises that result:

Theorem 3.19. *Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers with $\alpha_n > 0$ for $n \in \mathbb{N}$, satisfying (\star) , and that*

(i) $c > 0$ is such that $\mu_n \leq c$ for all $n \in \mathbb{N}$

(ii) $\alpha > 0$ is such that $\alpha_n \leq \alpha$ for all $n \in \mathbb{N}$,

(iii) $\sum_{i=0}^{\infty} \alpha_i = \infty$ with rate of divergence r ,

Fix $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$, and define $h : \mathbb{Q}_+ \times \mathbb{N} \rightarrow \mathbb{N}$ by

$$h(\delta, m) := \tilde{g} \left(r \left(m, \frac{2c}{\delta} \right) \right) \quad \text{with} \quad \tilde{g}(i) := \max\{j + g(j) \mid j \leq i\}$$

Then if $p \in \mathbb{N}$ and $\delta > 0$ are such that there exists $m \leq p$ such that

$$\gamma_n / \alpha_n \leq \frac{\delta}{2} \quad \text{for all} \quad n \in [m, h(\delta, m)]$$

and

$$\beta_n \leq \delta \implies \mu_{n+1} \leq \varepsilon \quad \text{for all} \quad n \leq h(\delta, p)$$

then there exists some $N \leq r(m, 2c/\delta) + 1$ such that

$$\mu_n \leq \varepsilon \quad \text{for all} \quad n \in [N, N + g(N)]$$

Proof. Suppose for contradiction that $\mu_n > \varepsilon$ for all $n \in [m, r(m, 2c/\delta) + 1] \subseteq [m, h(\delta, m)]$. Then for $n \leq r(m, 2c/\delta)$ we also have $n \leq h(\delta, p)$, and since $\mu_{n+1} > \varepsilon$ we have $\beta_n > \delta$ for $n \in [m, r(m, 2c/\delta)]$. It follows that

$$\mu_{n+1} \leq \mu_n - \alpha_n \delta + \alpha_n \frac{\delta}{2} = \mu_n - \alpha_n \frac{\delta}{2}$$

and so

$$\frac{\delta}{2} \sum_{i=m}^{r(m, 2c/\delta)} \alpha_i < \mu_m - \mu_{r(m, 2c/\delta)+1} \leq c$$

a contradiction. So we have shown that there exists some $N \in [m, r(m, 2c/\delta) + 1]$ such that $\mu_N \leq \varepsilon$. We now show that $\mu_n \leq \varepsilon$ for all $n \in [N, N + g(N)]$. Suppose for contradiction that there is some $n \in [N, N + g(N)]$ such that $\mu_n \leq \varepsilon$ but $\mu_{n+1} > \varepsilon$. Then since $n \leq h(\delta, p)$ we have $\beta_n > \delta$ and thus

$$\varepsilon < \mu_{n+1} \leq \mu_n - \alpha_n \frac{\delta}{2} \leq \varepsilon$$

a contradiction. □

Corollary 3.20. Suppose that $\{\mu_n\}$, $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ are sequences of nonnegative real numbers with $\alpha_n > 0$ for $n \in \mathbb{N}$, satisfying (\star) , and that

(i) $c > 0$ is such that $\mu_n \leq c$ for all $n \in \mathbb{N}$

(ii) $\alpha > 0$ is such that $\alpha_n \leq \alpha$ for all $n \in \mathbb{N}$,

(iii) $\sum_{i=0}^{\infty} \alpha_i = \infty$ with rate of divergence r ,

(iv) $\gamma_n / \alpha_n \rightarrow 0$ with rate of convergence ϕ .

Then whenever there exists a modulus $\omega : (0, \infty) \rightarrow (0, \infty)$ such that for all $\varepsilon > 0$ and $n \in \mathbb{N}$:

$$\beta_n \leq \omega(\varepsilon) \implies \mu_{n+1} \leq \varepsilon$$

then $\mu_n \rightarrow 0$ with rate of convergence

$$\Phi(\varepsilon) := r \left(\phi \left(\frac{\omega(\varepsilon)}{2} \right), \frac{2c}{\omega(\varepsilon)} \right) + 1$$

where $c > 0$ is any upper bound on $\{\mu_n\}$.

Proof. Directly from Theorem 3.19, defining $\delta := \omega(\varepsilon)$ and $p := \phi(\delta/2)$, and observing that these satisfy the premise of the theorem for any $g : \mathbb{N} \rightarrow \mathbb{N}$. \square

4 Applications

In this second main section we demonstrate how our quantitative study of recursive inequalities can be applied to obtain numerical results from theorems in nonlinear analysis. Our main contribution is a new case study involving the subclass explored in Section 3.1, where we apply our rate of metastability for $\beta_n \rightarrow 0$ to obtain quantitative information for a class of gradient descent algorithms. Instances of the results from Section 3.2 have already been used in a number of recent papers, and we also provide a high-level outline some of these. In all cases, we start by presenting a simplified form of the application, motivating how the recursive inequality and associated conditions emerge naturally from the concrete setting, before discussing more detailed results.

4.1 Recursive inequalities in applied proof theory

Before we start, we give a very brief survey of some of the different types of recursive inequality that have been recently studied from a quantitative perspective in applied proof theory. We make no attempt to give a detailed or comprehensive overview of this topic (which would constitute a paper in its own right) - rather we content ourselves with a few pointers to representative papers in the proof theory literature, and explain how the inequalities used there relate to the class we study here.

Some of the first quantitative recursive inequalities appearing in the proof theory literature involve the general inequality already discussed in Section 3.1:

$$\mu_{n+1} \leq (1 + \delta_n)\mu_n + \gamma_n$$

which we used to show that $\{\mu_n\}$ converges to some limit under the conditions $\sum_{i=0}^{\infty} \gamma_i < \infty$ and $\sum_{i=0}^{\infty} \delta_i < \infty$. Rates of metastability for $\{\mu_n\}$ have been extracted, and then applied to obtain, for instance, bounds on the computation of approximate fixed points of asymptotically quasi-nonexpansive mappings in [17], or for nonexpansive mappings in uniformly convex hyperbolic spaces in [18].

If we consider the special case of (\star) where $\beta_n = \mu_n$ i.e.

$$\mu_{n+1} \leq (1 - \alpha_n)\mu_n + \gamma_n \quad (14)$$

we can show that $\mu_n \rightarrow 0$ for both $\sum_{i=0}^{\infty} \gamma_i < \infty$ and $\gamma_n/\alpha_n \rightarrow 0$. In fact, in this case the proofs can be simplified and the two conditions combined, and quantitative results giving both direct and metastable rates of convergence have been crucial in numerous different contexts. An early use of kind of recursive inequality was to extract rates of asymptotic regularity for the Halpern iterations of nonexpansive self-mappings in [24], and a particularly detailed discussion of variants of (14) is given in [19]. There have been many further examples since: The last years have seen numerous instances of (14) with combined conditions, including [8, 9, 20, 25].

Less common but perhaps more relevant to us are examples of the subclass considered in Section 3.2 for $\beta_n = \psi(\mu_n)$ or $\beta_n = \psi(\mu_{n+1})$. Here rates of convergence closely related to Corollary 3.18 or 3.20 have been applied in various different contexts. The main variant $\beta_n = \psi(\mu_n)$ was crucial in a recent paper by the second author and Wiesnet on rates of convergence for generalized asymptotically weakly contractive mappings [29], while the second variant $\beta_n = \psi(\mu_{n+1})$ considered in Section 3.2.1 was initially used in [15] to extract rates of convergence for pseudocontractive mappings. More recently, it has featured in [21], where Corollary 3.20 is applied to obtain rates of convergence for algorithms involving set-valued accretive operators, and [31], where the same result is used in the context of jointly nonexpansive mappings. Both [21] and [29] will be discussed in more detail in Sections 4.3.1 and 4.3.2 below.

4.2 Application of Case I: A quantitative analysis of a class of gradient descent methods

We now give a new quantitative result that utilises Proposition 3.8 (in the form of Corollary 3.11) in the context of gradient descent algorithms. As such, the content of this section should be seen as a new contribution to applied proof theory in nonlinear analysis, but we aim to present our work without assuming any technical background in the area of application. To help motivate what follows, we first consider a simple example:

Example 4.1 (Gradient descent in \mathbb{R}^n). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex, differentiable function, and consider the usual steepest descent algorithm defined by

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n) \quad (15)$$

for some initial x_0 , where the step sizes $\{\alpha_n\}$ satisfy

$$\sum_{i=0}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=0}^{\infty} \alpha_i^2 < \infty$$

Assuming that $L > 0$ is such that $\|\nabla f(x_n)\| \leq L$ for all $n \in \mathbb{N}$, by (15) we have

$$\|x_{n+1} - x_n\| \leq L\alpha_n \quad (16)$$

Now, let x^* be a point where f attains its minimum i.e. $f(x^*) \leq f(y)$ for all $y \in \mathbb{R}^n$. A basic property of a differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ being convex is that

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \quad (17)$$

for all $x, y \in \mathbb{R}^n$, and therefore in particular

$$0 \leq f(x_n) - f(x^*) \leq \langle \nabla f(x_n), x_n - x^* \rangle \quad (18)$$

for all $n \in \mathbb{N}$. Setting $\beta_n := f(x_n) - f(x^*)$ we have

$$\begin{aligned} 2\alpha_n\beta_n &\leq 2\langle \alpha_n \nabla f(x_n), x_n - x^* \rangle \quad \text{by (18)} \\ &= 2\langle x_n - x_{n+1}, x_n - x^* \rangle \quad \text{by (15)} \\ &= \|x_{n+1} - x_n\|^2 + \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 \quad (\text{properties of inner product}) \\ &\leq L^2\alpha_n^2 + \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 \quad \text{by (16)} \end{aligned} \quad (19)$$

and therefore the main recursive inequality (\star) is satisfied with

$$\mu_n := \frac{1}{2}\|x_n - x^*\|^2 \quad \text{and} \quad \gamma_n := \frac{1}{2}L^2\alpha_n^2$$

Moreover, from $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$ we clearly have $\sum_{i=0}^{\infty} \gamma_i < \infty$ and, using convexity of f together with the Cauchy-Schwartz inequality and (16):

$$\begin{aligned} \beta_n - \beta_{n+1} &= f(x_n) - f(x_{n+1}) \leq \langle \nabla f(x_n), x_n - x_{n+1} \rangle \\ &\leq \|\nabla f(x_n)\| \cdot \|x_n - x_{n+1}\| \leq L^2\alpha_n \end{aligned}$$

and thus the assumptions of Proposition 3.7 are satisfied for $\theta := L^2$. Therefore by the conclusion of that proposition, it follows that $f(x_n) \rightarrow f(x^*)$, and moreover, using the quantitative results of Section 3.1 we can produce a computable rate of metastability in terms of suitable input data.

We stress that for the special case of gradient descent in \mathbb{R}^n sketched above, more direct proofs of convergence are possible, with better rates of convergence. However, there exist a number of much more general results whose proofs use Proposition 3.7 or its variants in a crucial way, and in these cases, our quantitative results give us a means to extract computational information from those proofs.

A concrete example of such a result will be considered in Section 4.2.1 below, but we first give a new, abstract convergence theorem that axiomatises some key properties of steepest descent used above, making precise what we need in order to be able to reduce convergence to Proposition 3.7. Note that this result applies to an arbitrary real inner product space, and does not explicitly introduce a gradient descent method. Instead, this is implicit in the abstract properties that we require.

We start by motivating these abstract properties. Let $f : X \rightarrow \mathbb{R}$ be a function for some real inner product space X , $\{x_n\}$ and $\{u_n\}$ be sequences of

vectors in X , and $\{\alpha_n\}$ a sequence of nonnegative reals, with the intuitive idea being that $\{x_n\}$ is generated via some kind of gradient descent algorithm, where $\{u_n\}$ denotes the gradients of f at points $\{x_n\}$, and $\{\alpha_n\}$ the step sizes i.e.

$$x_{n+1} = x_n - \alpha_n u_n \quad (20)$$

Looking in more detail at our simple example, the precise definition of the algorithm is used to establish two key properties, namely

$$(I) \quad \|x_{n+1} - x_n\| \leq L\alpha_n \text{ (cf. (16))}$$

$$(II) \quad \langle \alpha_n u_n, x_n - x^* \rangle = \langle x_n - x_{n+1}, x_n - x^* \rangle \text{ (cf. (19))}$$

As we will see, we can replace (20) with general versions of these two properties, namely

$$(I) \quad \|x_{n+1} - x_n\| \leq c_n$$

$$(II) \quad \langle \alpha_n u_n, x_n - x^* \rangle \leq a \langle x_n - x_{n+1}, x_n - x^* \rangle + d_n$$

for some constant $a > 0$ and sequences of nonnegative reals $\{c_n\}$, $\{d_n\}$ with $\sum_{i=0}^{\infty} c_i^2 < \infty$ and $\sum_{i=0}^{\infty} d_i < \infty$, and still reduce the problem to an instance of (\star) . In a similar way, the fact that u_n represents a gradient of f at x_n is only used to establish

$$f(x_n) - f(y) \leq \langle u_n, x_n - y \rangle$$

and again, we will show that this can be generalised to incorporate an error term i.e.

$$f(x_n) - f(y) \leq \langle u_n, x_n - y \rangle + b_n$$

for some sequence of nonnegative reals $\{b_n\}$ with $\sum_{i=0}^{\infty} \alpha_i b_i < \infty$. We now use these properties to present an abstract, quantitative convergence theorem for a class of gradient descent-like methods, where we use Proposition 3.7 to establish convergence, and Corollary 3.11 to provide a rate of metastability. We certainly do not claim that this abstract version provides any kind of definitive characterisation of gradient descent methods in inner product spaces: It simply gives an abstract presentation of a type of convergence proof that utilises a specific reduction to (\star) . We give a nontrivial example of how this can then be applied in a concrete setting in Section 4.2.1.

Theorem 4.2. *Let X be a real inner product space with $Y \subseteq X$, and suppose that $f : X \rightarrow \mathbb{R}$ is a function. Let $\{\alpha_n\}$ be a sequence of nonnegative reals with $\sum_{i=0}^{\infty} \alpha_i = \infty$, and $\{b_n\}$, $\{c_n\}$ and $\{d_n\}$ sequences of nonnegative reals with $\sum_{i=0}^{\infty} \alpha_i b_i < \infty$, $\sum_{i=0}^{\infty} c_i^2 < \infty$ and $\sum_{i=0}^{\infty} d_i < \infty$. Finally, suppose that $x^* \in Y$, $\{x_n\}$ and $\{u_n\}$ are sequences of vectors with $x_n \in Y$ for all $n \in \mathbb{N}$, and $a, p, \theta > 0$ are constants, which satisfy the following properties for all $n \in \mathbb{N}$:*

$$(i) \quad f(x^*) \leq f(x_n) \text{ (} x^* \text{ acts as a minimizer)}$$

$$(ii) \quad f(x_n) - f(y) \leq \langle u_n, x_n - y \rangle + b_n \text{ for all } y \in Y \text{ (} u_n \text{ acts as a gradient)}$$

- (iii) $\|x_{n+1} - x_n\| \leq c_n$ (x_n acts as a gradient descent method, property I)
- (iv) $\langle \alpha_n u_n, x_n - x^* \rangle \leq a \langle x_n - x_{n+1}, x_n - x^* \rangle + d_n$ (x_n acts as a gradient descent method, property II)
- (v) $\|u_n\| \leq p$ (gradients are bounded)
- (vi) $p c_n + b_n \leq \theta \alpha_n$

Then $f(x_n) \rightarrow f(x^*)$. Moreover, if r is a rate of divergence for $\sum_{i=0}^{\infty} \alpha_i = \infty$ and $b, c, d, K > 0$ are such that

$$\sum_{i=0}^{\infty} \alpha_i b_i \leq b, \quad \sum_{i=0}^{\infty} c_i^2 \leq c, \quad \sum_{i=0}^{\infty} d_i \leq d, \quad \text{and} \quad \|x_0 - x^*\|^2 \leq K$$

Then for all $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ we have

$$\exists n \leq \Phi(\varepsilon, g) \forall k \in [n, n + g(n)] (f(x_k) \leq f(x^*) + \varepsilon)$$

where

$$\begin{aligned} \Phi(\varepsilon, g) &:= \tilde{h}^{(\lceil 4\theta e / \varepsilon^2 \rceil)}(0) \\ \tilde{h}(n) &:= r\left(n + g(n), \frac{\varepsilon}{2\theta}\right) + 1 \\ e &:= \frac{a(c + K)}{2} + b + d \end{aligned}$$

To prove Theorem 4.2, we require the following lemma, which is a simple variant of a standard rate of metastability for monotone increasing sequences.

Lemma 4.3. *Let $\{b_n\}$ be a sequence of nonnegative numbers with $\sum_{i=0}^{\infty} b_i \leq b$. Then a rate of metastability for $\sum_{i=0}^{\infty} b_i < \infty$ is given by*

$$\Phi(\varepsilon, g) := \tilde{g}^{(\lceil b/\varepsilon \rceil)}(0) \quad \text{where} \quad \tilde{g}(n) := n + g(n) + 1$$

Proof. Assume that the statement is false i.e. there exist $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ such that

$$\sum_{j=n}^{n+g(n)} b_j > \varepsilon \quad \text{for all } n \leq \Phi(\varepsilon, g)$$

Then for any $i < \lceil b/\varepsilon \rceil$ we have $\tilde{g}^{(i)}(0) \leq \Phi(\varepsilon, g)$ and thus, using that $\tilde{g}^{(i)}(0) + g(\tilde{g}^{(i)}(0)) = \tilde{g}^{(i+1)}(0) - 1$, we have

$$\sum_{j=\tilde{g}^{(i)}(0)}^{\tilde{g}^{(i+1)}(0)-1} b_j > \varepsilon$$

and so summing over all $i < \lceil b/\varepsilon \rceil$:

$$\sum_{i=0}^{\Phi(\varepsilon, g)-1} b_i = \sum_{i=0}^{\lceil b/\varepsilon \rceil-1} \left(\sum_{j=\tilde{g}^{(i)}(0)}^{\tilde{g}^{(i+1)}(0)-1} b_j \right) > \sum_{i=0}^{\lceil b/\varepsilon \rceil-1} \varepsilon = \lceil b/\varepsilon \rceil \cdot \varepsilon \geq b$$

contradicting $\sum_{i=0}^{\infty} b_i \leq b$. □

Proof of Theorem 4.2. Setting $\beta_n := f(x_n) - f(x^*)$ and noting that $\beta_n \geq 0$ by (i), we have:

$$\begin{aligned}
2\alpha_n\beta_n &\leq 2\langle\alpha_n u_n, x_n - x^*\rangle + 2\alpha_n b_n \quad \text{by (ii)} \\
&\leq 2a\langle x_n - x_{n+1}, x_n - x^*\rangle + 2(\alpha_n b_n + d_n) \quad \text{by (iv)} \\
&= a(\|x_{n+1} - x_n\|^2 + \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2) + 2(\alpha_n b_n + d_n) \quad (21) \\
&\leq a(\|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2) + [ac_n^2 + 2(\alpha_n b_n + d_n)] \quad \text{by (iii)}
\end{aligned}$$

where the third line is by a routine calculation and standard properties of real inner products, and therefore (\star) is satisfied for

$$\mu_n := \frac{a}{2}\|x_n - x^*\|^2 \quad \text{and} \quad \gamma_n := \frac{ac_n^2}{2} + \alpha_n b_n + d_n$$

In particular, it follows that

$$\begin{aligned}
\sum_{i=0}^{\infty} \alpha_i \beta_i &\leq \mu_0 + \sum_{i=0}^{\infty} \gamma_i \\
&= \frac{a}{2}\|x_0 - x^*\|^2 + \sum_{i=0}^{\infty} \left(\frac{ac_i^2}{2} + \alpha_i b_i + d_i \right) \quad (22) \\
&\leq \frac{aK}{2} + \frac{ac}{2} + b + d =: e
\end{aligned}$$

We also have

$$\begin{aligned}
\beta_n - \beta_{n+1} &= f(x_n) - f(x_{n+1}) \\
&\leq \langle u_n, x_n - x_{n+1} \rangle + b_n \quad \text{by (ii)} \\
&\leq \|u_n\| \|x_n - x_{n+1}\| + b_n \quad \text{Cauchy-Schwartz} \\
&\leq pc_n + b_n \quad \text{by (ii) and (v)} \\
&\leq \theta \alpha_n \quad \text{by (vi)}
\end{aligned}$$

therefore the conditions of Proposition 3.7 are satisfied for $\{\beta_n\}$ defined as above, and thus $f(x_n) - f(x^*) \rightarrow 0$. To get the rate of metastability, we apply Corollary 3.11 (b). Since $\sum_{i=0}^{\infty} \alpha_i \beta_i \leq e$, by Lemma 4.3, rate of metastability for $\sum_{i=0}^{\infty} \alpha_i \beta_i < \infty$ is therefore given by

$$\Psi(\varepsilon, h) := \tilde{h}^{\lceil e/\varepsilon \rceil}(0) \quad \text{where} \quad \tilde{h}(n) := n + h(n) + 1$$

for e as defined in (22) above, and therefore by Corollary 3.11 a rate of metastability for $f(x_n) - f(x^*) \rightarrow 0$ is given by

$$\Phi(\varepsilon, g) := \Psi\left(\frac{\varepsilon^2}{4\theta}, h\right)$$

where $h(n) = r(n + g(n), \varepsilon/2\theta) - n$. Unwinding the definitions gives the result. \square

Remark 4.4. A rate of metastability for $\{f(x_n)\}$ in the special case of gradient descent in \mathbb{R}^n as sketched in Example 4.1 can be read off from Theorem 4.2 by setting $e := (L + K)/2$ and $\theta := L^2$, though as emphasised already, alternative proofs for this simple case will yield better quantitative information.

4.2.1 An application for projective subgradient methods

We now use our abstract theorem to give a quantitative version of a result by Alber et al. [4], where convergence $f(x_n) \rightarrow f(x^*)$ for a generalised subgradient method $\{x_n\}$ towards a minimizer x^* of f is established, and then used to prove weak convergence of $\{x_n\}$. We provide a rate of metastability for $\{f(x_n)\}$, and conjecture that similar results on projective subgradient methods e.g. [3, 4, 12, 26] could also be reduced to Theorem 4.2 (or a simple variant thereof), although we leave the details to future work.

Suppose that H is a real valued Hilbert space, $C \subseteq H$ a closed, convex subset of H , and $f : H \rightarrow \mathbb{R}$ a convex and continuous function. The ε -subdifferential of f at $x \in H$ is defined by

$$\partial_\varepsilon f(x) := \{u \in H \mid f(y) - f(x) \geq \langle u, y - x \rangle - \varepsilon \text{ for all } y \in H\}$$

with the case $\varepsilon = 0$ coinciding with the usual subdifferential $\partial f(x)$. Letting $P_C : H \rightarrow C$ be the orthogonal projection of H into C , Alber. et al consider the following algorithm:

$$x_{n+1} = P_C \left(x_n - \frac{\alpha_n}{\nu_n} u_n \right) \quad \text{for } u_n \in \partial_{\varepsilon_n} f(x_n) \quad \text{with } u_n \neq 0 \quad (23)$$

where $\{\alpha_n\}$ satisfies $\sum_{i=0}^\infty \alpha_i = \infty$ and $\sum_{i=0}^\infty \alpha_i^2 < \infty$, $\{\varepsilon_n\}$ is a sequence of nonnegative error terms with $\varepsilon_n \leq \mu \alpha_n$ for some $\mu > 0$ and $\nu_n := \max\{1, \|u_n\|\}$. The algorithm halts if $0 \in \partial_{\varepsilon_n} f(x_n)$ at any point.

Theorem 4.5 (Cf. Lemma 1 and Theorems 1 and 2 of [4]). *Let $x^* \in C$ be a minimizer of f on C , and suppose that $\{x_n\}$ is an infinite sequence generated by the algorithm (23), whose components satisfy all of the properties outlined above. Suppose that $\rho > 1$ is such that $\|u_n\| \leq \rho$ for all $n \in \mathbb{N}$. Then $f(x_n) \rightarrow f(x^*)$. Moreover, if r is a rate of divergence for $\sum_{i=0}^\infty \alpha_i = \infty$ and $L, K > 0$ are such that $\sum_{i=0}^\infty \alpha_i^2 \leq L$ and $\|x_0 - x^*\|^2 \leq K$, then for all $\varepsilon > 0$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ we have*

$$\exists n \leq \Phi(\varepsilon, g) \forall k \in [n, n + g(n)] (f(x_k) \leq f(x^*) + \varepsilon)$$

where

$$\begin{aligned} \Phi(\varepsilon, g) &:= \tilde{h}(\lceil 4\theta e / \varepsilon^2 \rceil)(0) \\ \tilde{h}(n) &:= r \left(n + g(n), \frac{\varepsilon}{2\theta} \right) + 1 \\ e &:= \frac{\rho(L + K)}{2} + (\mu + 2\rho)L \\ \theta &:= \rho + \mu \end{aligned}$$

Proof. We show that conditions (i)–(vi) of Theorem 4.2 are satisfied, using details from the proof of Lemma 1 of [4], and the result then follows. Condition (i) holds because x^* is a minimiser of f on C and $x_n \in C$ for all $n \in \mathbb{N}$. For (ii) we use the definition of the subgradient: since $u_n \in \partial_{\varepsilon_n} f(x_n)$, for any $y \in C$ we have

$$f(x_n) - f(y) \leq \langle u_n, x_n - y \rangle + \varepsilon_n \leq \langle u_n, x_n - y \rangle + \mu \alpha_n$$

using $\varepsilon_n \leq \mu \alpha_n$, and so the property holds for $b_n := \mu \alpha_n$ with $\sum_{i=0}^{\infty} \alpha_i b_i \leq \mu L$. To establish condition (iii), setting $z_n := x_n - (\alpha_n / \nu_n) u_n$ we have

$$\|x_{n+1} - x_n\| = \|P_C(z_n) - P_C(x_n)\| \leq \|z_n - x_n\| = \frac{\alpha_n}{\nu_n} \|u_n\| \leq \alpha_n \quad (24)$$

using properties of the orthogonal projection (cf. Proposition 3 (i) of [4]) and $\|u_n\| \leq \nu_n$, and so property (iii) holds for $c_n := \alpha_n$ and $\sum_{i=0}^{\infty} c_i^2 \leq L$. For (iv) we first observe that

$$\begin{aligned} \frac{\alpha_n}{\nu_n} \langle u_n, x_n - x^* \rangle &= \langle x_n - x^*, x_n - z_n \rangle \\ &= \langle x_n - x^*, x_n - x_{n+1} \rangle + \langle x_n - x^*, x_{n+1} - z_n \rangle \\ &= \langle x_n - x^*, x_n - x_{n+1} \rangle + \langle z_n - x_n, z_n - x_{n+1} \rangle + \langle x^* - z_n, z_n - x_{n+1} \rangle \end{aligned} \quad (25)$$

By properties of the projection (cf. Proposition 3 (ii) of [4]), we have

$$\langle z_n - x^*, z_n - P_C(z_n) \rangle \geq 0$$

and using $P_C(z_n) = x_{n+1}$ we have

$$\langle x^* - z_n, z_n - x_{n+1} \rangle = -\langle z_n - x^*, z_n - P_C(z_n) \rangle \leq 0$$

and therefore from (25), also using that $\nu_n \leq \max\{1, \rho\} \leq \rho$, we have

$$\langle \alpha_n u_n, x_n - x^* \rangle \leq \rho (\langle x_n - x^*, x_n - x_{n+1} \rangle + \langle z_n - x_n, z_n - x_{n+1} \rangle) \quad (26)$$

Finally, we see that

$$\begin{aligned} \langle z_n - x_n, z_n - x_{n+1} \rangle &= \langle z_n - x_n, z_n - x_n \rangle + \langle z_n - x_n, x_n - x_{n+1} \rangle \\ &\leq \|z_n - x_n\|^2 + \|z_n - x_n\| \|x_n - x_{n+1}\| \\ &\leq 2\|z_n - x_n\|^2 \quad \text{by (24)} \\ &\leq \frac{2\alpha_n^2}{\nu_n^2} \|u_n\|^2 \\ &\leq 2\alpha_n^2 \quad \text{since } \|u_n\| \leq \nu_n \end{aligned}$$

Substituting this into (26), we have

$$\langle \alpha_n u_n, x_n - x^* \rangle \leq \rho \langle x_n - x^*, x_n - x_{n+1} \rangle + 2\rho \alpha_n^2$$

and so (iv) is satisfied for $a = \rho$ and $d_n = 2\rho \alpha_n^2$ with $\sum_{i=0}^{\infty} d_i \leq 2\rho L$. Finally, (v) holds by assumption since $\|u_n\| \leq \rho$, and (vi) holds since $\rho c_n + b_n = (\rho + \mu) \alpha_n$. Applying Theorem 4.2 with the relevant parameters i.e. $(a, b, c, d, p, \theta) := (\rho, \mu L, L, 2\rho L, \rho, \rho + \mu)$, gives us the result. \square

Remark 4.6. The proof of Theorem 1 of [4] uses Proposition 3.7 for $\beta_n := f(x_n) - f(x^*)$, establishing that $\sum_{i=0}^{\infty} \alpha_i \beta_i$ is bounded and therefore $\beta_n \rightarrow 0$. By Theorem 3.6, there is in general no way of obtaining a computable rate of convergence from a bound on $\sum_{i=0}^{\infty} \alpha_i \beta_i$ (rather a computable rate of convergence on the series is needed). The difficulty of reading off rates of convergence from their proof of the whole sequence $\{x_n\}$ rather than just a subsequence $\{x_{n_k}\}$ is alluded to by the authors in [4], who write, in the final paragraph “*This result does not give any information on the asymptotic behaviour of $\{f(x_k)\}$ outside the subsequence $\{x^{l_k}\}$ ”.* Theorem 3.6 gives a formal explanation as to why reading off a quantitative information on the whole sequence is difficult, although we are able to provide information on the behaviour of $\{f(x_k)\}$ in the form of a rate of metastability. We stress that this does not mean that an *alternative* proof technique would not yield a computable rate of convergence, at least in special cases.

Remark 4.7. Note that in [4], technically the existence of some $\rho > 0$ satisfying $\|u_n\| \leq \rho$ for all $n \in \mathbb{N}$ follows by establishing that the $\{x_n\}$ are bounded, and then using an additional boundedness assumption for the subgradient, namely that $\partial_\varepsilon f$ is bounded on bounded sets.

Remark 4.8. A quantitative study of another subgradient-like algorithm was recently given by Kohlenbach and Pitschke [27], who extract computational information from a theorem of Iiduke and Yamada [13]. Our case study above is separate to theirs (which in particular applies in \mathbb{R}^n), but it would be interesting to further explore any parallels between the two approaches.

4.3 Application of Case II: A survey of recent applications of Section 3.2

Having given a new case study which uses, in a crucial way, results of Section 3.1, we conclude this section on applications of the results of Section 3 with a brief discussion of how quantitative results pertaining to the second condition $\gamma_n/\alpha_n \rightarrow 0$, discussed in Section 3.2, have been used recently in different contexts.

4.3.1 Mann schemes for asymptotically weakly contractive mappings

An illuminating example of how the main recursive inequality (\star) with the second convergence condition $\gamma_n/\alpha_n \rightarrow 0$ can be applied is given by considering a slightly generalised notion of the class of ψ -weakly contractive mappings already discussed in Section 2.1. Convergence becomes more interesting when one is confronted with asymptotic variants of these mappings, which become weakly contractive in the limit. A simple example is the following generalisation of the original scheme (2):

$$\|T^n x - T^n y\| \leq \|x - y\| - \psi(\|x - y\|) + l_n$$

where $\{l_n\}$ is some sequence of nonnegative reals converging to zero, and we recall that ψ is a nondecreasing function with $\psi(0) = 0$ which is positive and

$(0, \infty)$. One can show in a very general setting that whenever x^* is a fixpoint of T , the Krasnoselki-Mann scheme (3) converges to x^* for any step sizes $\{\alpha_n\}$ with $\sum_{i=0}^{\infty} \alpha_i = \infty$. Indeed, for the simple example above, we see that

$$\begin{aligned} \|x_{n+1} - x^*\| &= \|(1 - \alpha_n)x_n + \alpha_n T^n x_n - x^*\| \\ &= \|(1 - \alpha_n)(x_n - x^*) + \alpha_n(T^n x_n - x^*)\| \\ &\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n\|T^n x_n - T^n x^*\| \\ &\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n(\|x_n - x^*\| - \psi(\|x_n - x^*\|) + l_n) \\ &= \|x_n - x^*\| - \alpha_n\psi(\|x_n - x^*\|) + \alpha_n l_n \end{aligned}$$

which is an instance of (\star) for $\mu_n := \|x_n - x^*\|$, $\beta_n := \psi(\|x_n - x^*\|)$ and $\gamma_n := \alpha_n l_n$. Moreover, we have $\gamma_n/\alpha_n = l_n \rightarrow 0$ and

$$\beta_n = \psi(\mu_n) \leq \frac{\psi(\varepsilon)}{2} \implies \mu_n \leq \varepsilon$$

by monotonicity of ψ (cf. Example 3.15), and so Corollary 3.18 can be applied to produce a rate of convergence for $x_n \rightarrow x^*$, assuming we have one for $l_n \rightarrow 0$. This corollary can in fact be utilised in a much more general setting to produce abstract quantitative convergence results for classes of weakly contractive mappings. The following is an example from a recent paper by the second author and Wiesnet [29], where weakly contractive mappings are discussed in much more detail:

Theorem 4.9 (cf. Theorem 4.1 of [29]). *Let X be a normed space with $E \subseteq X$, and suppose that $\{A_n\}$ is a sequence of mappings $A_n : E \rightarrow X$, and $\psi : [0, \infty) \rightarrow [0, \infty)$ a nondecreasing function with $\psi(0) = 0$ which is positive on $(0, \infty)$. Suppose that $\{k_n\}$ is some sequence of nonnegative reals and $\sigma : (0, \infty) \times (0, \infty) \rightarrow \mathbb{N}$ a modulus such that $\{A_n\}$ satisfies*

$$\|x - x^*\| \leq b \implies \|A_n x - x^*\| \leq (1 + k_n)\|x - x^*\| - \psi(\|x - x^*\|) + \delta$$

for all $\delta, b > 0$ and $n \geq \sigma(\delta, b)$. Suppose in addition that $\{x_n\}$ is a sequence satisfying

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n A_n x_n \tag{27}$$

where $\{\alpha_n\}$ is some sequence in $(0, \alpha]$ such that $\sum_{n=0}^{\infty} \alpha_n = \infty$ with rate of divergence r and $d > 0$ is such that $\prod_{i=0}^n (1 + \alpha_i k_i) \leq d$ for all $n \in \mathbb{N}$. Then whenever there exists $c > 0$ such that $\|x_n - x^*\| \leq c$ for all $n \in \mathbb{N}$, we have $\|x_n - x^*\| \rightarrow 0$ with rate

$$\Phi(\varepsilon) := r \left(\sigma \left(\frac{1}{2d} \min \left\{ \psi \left(\frac{\varepsilon}{2d} \right), \frac{\varepsilon}{\alpha} \right\}, c \right), 2d \int_{\varepsilon/2d}^c \frac{dt}{\psi(t)} \right) + 1$$

The rate of convergence stated here uses Corollary 3.18, though in a slightly refined form using monotonicity of ψ to replace the second argument of r with an integral. This instance of Corollary 3.18 is given as Lemma 3.1 in [29]. Much more sophisticated results, also using Corollary 3.18 in this way, are also given in [29].

4.3.2 Computing zeroes of set-valued accretive operators

In our final illustration, we give an instance of how the modified conditions set out in Section 3.2.1 can be used to produce rates of convergence for algorithms computing zeroes of set-valued accretive operators. Let X be a Banach space and X^* its dual space. The normalized duality mapping $J : X \rightarrow 2^{X^*}$ is defined by

$$J(x) := \{j \in X^* \mid \langle x, j \rangle = \|x\|^2 = \|j\|^2\}$$

An operator $A : D \rightarrow 2^X$ is said to be accretive if for all $x, y \in D$, if $u \in Ax$ and $v \in Ay$ then there exists some $j \in J(x - y)$ such that $\langle u - v, j \rangle \geq 0$. A notion of uniform accretivity was introduced by García-Falset in [11], where A is said to be uniformly accretive if there exists some nondecreasing $\phi : [0, \infty) \rightarrow [0, \infty)$ with $\phi(0) = 0$ and ϕ positive on $(0, \infty)$ such that whenever $u \in Ax$ and $v \in Ay$ then there exists some $j \in J(x - y)$ such that

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

For such operators, if there exists $x^* \in D$ such that $0 \in Ax^*$, we can show that the following algorithm

$$x_{n+1} = x_n - \alpha_n u_n \quad \text{for } u_n \in Ax_{n+1}$$

for $\sum_{i=0}^{\infty} \alpha_i = \infty$ converges to x^* . Suppose that $K > 0$ be such that $\|x_n - x^*\| \leq K$ for all $n \in \mathbb{N}$. Since $0 \in Ax^*$ and $u_n \in Ax_{n+1}$, we know that there exists some $j \in J(x_{n+1} - x^*)$ be such that $\langle u_n, j \rangle \geq \phi(\|x_{n+1} - x^*\|)$. We then have

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &= \langle x_{n+1} - x^*, j \rangle \quad \text{by } j \in J(x_{n+1} - x^*) \\ &= \langle x_n - x^*, j \rangle - \alpha_n \langle u_n, j \rangle \quad \text{by } x_{n+1} = x_n - \alpha_n u_n \\ &\leq \|x_n - x^*\| \cdot \|x_{n+1} - x^*\| - \alpha_n \langle u_n, j \rangle \quad \text{since } \|j\| = \|x_{n+1} - x^*\| \\ &\leq \|x_n - x^*\| \cdot \|x_{n+1} - x^*\| - \alpha_n \phi(\|x_{n+1} - x^*\|) \quad \text{using } j \in J(x_{n+1} - x^*) \\ &\leq \|x_{n+1} - x^*\| (\|x_n - x^*\| - \alpha_n \phi(\|x_{n+1} - x^*\|)/K) \quad \text{by } \|x_{n+1} - x^*\| \leq K \end{aligned}$$

and therefore

$$\|x_{n+1} - x^*\| \leq \|x_n - x^*\| - \alpha_n \phi(\|x_{n+1} - x^*\|)/K$$

Defining $\mu_n := \|x_n - x^*\|$ and $\beta_n := \phi(\|x_{n+1} - x^*\|)/K$, these then satisfy (\star) for $\gamma_n = 0$. Moreover, similarly to the previous section, we have

$$\beta_n = \phi(\mu_{n+1})/K \leq \phi(\varepsilon)/2K \implies \mu_{n+1} \leq \varepsilon$$

and so we can apply Corollary 3.20 to obtain a rate of convergence. Much more sophisticated results are possible, which use the full power of Corollary 3.20 with γ_n defined as some nontrivial error term. For example, we could replace the algorithm above with a perturbed scheme of the form

$$x_{n+1} = x_n - \alpha_n u_n \quad \text{for } u_n \in A_n x_{n+1}$$

where $\{A_n\}$ is some sequence of accretive operators which approach A in the Hausdorff metric. Similar calculations along the lines of the above give rise to an instance of (\star) where γ_n represents some error between $A_n x_{n+1}$ and $A x_{n+1}$. A series of quantitative results along these lines are given in [21], using a so-called *modulus of uniform accretivity at zero* [16] in place of the function ϕ .

Remark 4.10. A variant of Corollary 3.20 was first used by Kohlenbach and Körnlein in [15], and more recently in [31]. Both of these provide additional examples of how quantitative lemmas based on the second variant of Case II can be used in applied proof theory.

5 A note on implementation in the Lean theorem prover

As we have emphasised throughout, numerous convergence theorems in nonlinear analysis rely in a crucial way on the convergence properties of sequences of real numbers satisfying recursive inequalities. A recent survey encompassing a multitude of such inequalities (including those studied here) has been recently published [10]. In line with this, numerous quantitative results in applied proof theory rely on a quantitative analysis of the relevant recursive inequality, a partial account of which is provided by this paper.

While the formal proof interpretations that underlie applied proof theory, such as the Dialectica interpretation, have been implemented in a number of theorem provers (most notably the Minlog system¹), the majority of implementations tend to focus on the use of such translations as a verification strategy. In comparison, very little has been done on formalizing concrete applications of proof theory in mathematics.

However, significant progress here has recently been made by Cheval, who has developed several libraries in the Lean theorem prover², which implement not only the core logical techniques like the Dialectica interpretation, but also some logical metatheorems and quantitative results from the paper [8], which includes the use of recursive inequalities to establish convergence.

In parallel, the authors of this paper also have started building a Lean library³, the aim of which is to focus on recursive inequalities and the associated quantitative results. Formalizing these results is straightforward in the sense that they do not rely on extensive libraries of formal mathematics – indeed, most of the ordinary convergence results require little more than the ability to reason about sequences and series of real numbers. At the same time, such a library is clearly useful for the formalization of both nonlinear analysis in general and the corresponding proof theoretic applications, as it takes care of a core component of the normally much more complex convergence proofs in concrete settings. Moreover, in addition to the obvious value of having a range of core lemmas implemented, one could conceivably develop automated methods which

¹<https://www.mathematik.uni-muenchen.de/~logik/minlog/>

²<https://github.com/hcheval>

³<https://github.com/mneri123/Proof-mining->

attempt to generate recursive inequalities from given assumptions, and then automatically infer convergence properties, along with numerical bounds.

Our own formalisation effort is currently in an early stage. Several convergence proofs for simple recursive inequalities have now been implemented, and some key constructions from computable analysis have been formalized (such as Proposition 2.3), making use of Lean’s computability libraries, and some well-known rates of metastability, such as that for monotone bounded sequences, have been verified.

6 Concluding remarks

It is hoped that this paper has provided not only a thorough quantitative study of an important class of convergent sequences, but has also demonstrated the applicability of such results in analysing proofs in different areas of nonlinear analysis. We conclude by outlining some open questions.

An obvious direction for future work is a more in-depth proof theoretic study of subgradient type methods. There are numerous extensions of [4], many of which use variants of Proposition 3.7, and so the results of Section 3.1 could potentially be used to produce rates of convergence or metastability in different settings. For example, [3] establishes convergence results for subgradient methods in uniformly convex and uniformly smooth Banach spaces, and we anticipate our quantitative lemmas could be applied to extract quantitative information which would depend on the corresponding moduli of convexity and uniform smoothness.

A much broader challenge would be to explore further types of recursive inequalities, which have been nicely organised and classified in [10]. Particularly interesting here are stochastic recursive inequalities, which establish almost sure convergence of real-valued random variables, and in addition to standard methods used to prove convergence of sequences of real numbers, make crucial use of Martingale convergence theorems. Here we anticipate we could use recent work on formulating metastable variants of almost everywhere convergence [6], along with numerical results on upcrossings as used in [7].

Finally, as already alluded to above, work on formalizing convergence results for recursive inequalities in theorem provers raises a natural question of whether one can devise algorithms for reducing concrete convergence statements to a particular recursive inequality, thereby generating a convergence proof automatically. In practice, the reduction often involves little more than the systematic application of basic algebraic operations along with standard facts about norms or inner products, and it is conceivable that for a given method $\{x_n\}$ converging towards some point x^* , a search algorithm could potentially generate recursive inequalities satisfied by $\|x_n - x^*\|$, thereby providing an automated method for proving the convergence of numerical algorithms, along with quantitative information in the form of a rate of convergence or metastability.

References

- [1] ALBER, Y., AND GUERRE-DELABRIERE, S. Principle of weakly contractive maps in Hilbert spaces. In *New Results in Operator Theory and its Applications* (1997), I. Gohberg and Y. Lyubich, Eds., vol. 98, pp. 7–22.
- [2] ALBER, Y., AND GUERRE-DELABRIERE, S. On the projection methods for fixed point problems. *Analysis: International mathematical journal of analysis and its applications* 21 (2001), 17–39.
- [3] ALBER, Y., AND IUSEM, A. N. Extension of subgradient techniques for nonsmooth optimization in Banach spaces. *Set-Valued Analysis* 9 (2001), 315–335.
- [4] ALBER, Y., IUSEM, A. N., AND SOLODOV, M. V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* 81, 1 (1998), 23–35.
- [5] AVIGAD, J. The metamathematics of ergodic theory. *Annals of Pure and Applied Logic* 157, 2–3 (2009), 64–76.
- [6] AVIGAD, J., DEAN, E. T., AND RUTE, J. A metastable dominated convergence theorem. *Journal of Logic & Analysis* 4, 3 (2012), 1–19.
- [7] AVIGAD, J., GERHARDY, P., AND TOWNSNER, H. Local stability of ergodic averages. *Transactions of the American Mathematical Society* 362, 1 (2010), 261–288.
- [8] CHEVAL, H., AND LEUȘTEAN, L. Quadratic rates of asymptotic regularity for the Tikhonov-Mann iteration. *Optimization Methods and Software* (2022). Electronic publication ahead of print.
- [9] DINIS, B., AND PINTO, P. Quantitative results on the multi-parameters proximal point algorithm. *Journal of Convex Analysis* 28, 3 (2021), 729–750.
- [10] FRANCI, B., AND GRAMMATICO, S. Convergence of sequences: A survey. *Annual Reviews in Control* (2022). Article in press.
- [11] GARCÍA-FALSET, J. The asymptotic behavior of the solutions of the Cauchy problem generated by ϕ -accretive operators. *Journal of Mathematical Analysis and Applications* 310 (2005), 594–608.
- [12] GRAÑA DRUMMOND, L. M., AND IUSEM, A. N. A projected gradient method for vector optimization problems. *Optimization Problems Computational Optimization and Applications* 28 (2004), 5–29.
- [13] IIDUKA, H., AND YAMADA, I. A subgradient-type method for the equilibrium problem over the fixed point set and its applications. *Optimization* 58, 2 (2009), 251–261.

- [14] KOHLENBACH, U. Some computational aspects of metric fixed point theory. *Nonlinear Analysis* 61, 5 (2005), 823–837.
- [15] KOHLENBACH, U., AND KÖRNLEIN, D. Effective rates of convergence for Lipschitzian pseudocontractive mappings in general Banach spaces. *Nonlinear Analysis* 74 (2011), 5253–5267.
- [16] 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.
- [17] KOHLENBACH, U., AND LAMBOV, B. Bounds on iterations of asymptotically quasi-nonexpansive mappings. In *Proceedings of the International Conference on Fixed Point Theory and Applications* (2004), Yokohama Publishers, pp. 143–172.
- [18] KOHLENBACH, U., AND LEUŞTEAN, L. Asymptotically nonexpansive mappings in uniformly convex hyperbolic spaces. *Journal of the European Mathematical Society* 12, 1 (2010), 71–92.
- [19] KOHLENBACH, U., AND LEUŞTEAN, L. Effective metastability of Halpern iterates in CAT(0) spaces. *Advances in Mathematics* 321 (2012), 2526–2556.
- [20] KOHLENBACH, U., AND PINTO, P. Quantitative translations for viscosity approximation methods in hyperbolic spaces. Preprint, 2021.
- [21] 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.
- [22] KREISEL, G. On the interpretation of non-finitist proofs, Part I. *Journal of Symbolic Logic* 16 (1951), 241–267.
- [23] KREISEL, G. On the interpretation of non-finitist proofs, Part II: Interpretation of number theory. *Journal of Symbolic Logic* 17 (1952), 43–58.
- [24] LEUŞTEAN, L. Rates of asymptotic regularity for halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science* 13, 11 (2007), 1680–1691.
- [25] LEUŞTEAN, L., AND PINTO, P. Quantitative results on a Halpern-type proximal point algorithm. *Computational Optimization and Applications* 79, 1 (2021), 101–125.
- [26] MAINGÉ, P.-E. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set-valued analysis* 16 (2008), 899–912.

- [27] PITSCHKE, N., AND KOHLENBACH, U. Quantitative analysis of a subgradient-type method for equilibrium problems. *Numerical Algorithms* 90 (2022), 197–219.
- [28] POWELL, T. A new metastable convergence criterion and an application in the theory of uniformly convex Banach spaces. *Journal of Mathematical Analysis and Applications* 478, 2 (2019), 790–805.
- [29] POWELL, T., AND WIESNET, F. Rates of convergence for asymptotically weakly contractive mappings in normed spaces. *Numerical Functional Analysis and Optimization* 42, 15 (2021), 1802–1838.
- [30] QIHOU, L. Iteration sequences for asymptotically quasi-nonexpansive mappings with error member. *Journal of Mathematical Analysis and Applications* 259 (2001), 18–24.
- [31] SIPOŞ, A. Revisiting jointly firmly nonexpansive families of mappings. *Optimization* (2021). Electronic publication ahead of print.
- [32] SPECKER, E. Nicht konstruktiv beweisbare Sätze der Analysis. *Journal of Symbolic Logic* 14 (1949), 145–158.
- [33] TAO, T. Soft analysis, hard analysis, and the finite convergence principle. Essay posted 23 May 2007, 2007. Appeared in: ‘T. Tao, Structure and Randomness: Pages from Year One of a Mathematical Blog. AMS, 298pp., 2008’.
- [34] TAO, T. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems* 28 (2008), 657–688.