UNIVERSITATEA DIN BUCUREŞTI

Facultatea de Matematică şi Informatică

Teză de doctorat

Contributions to proof mining

Autor:
Andrei Sipoș

Coordonator ştiinţific:
Prof. dr. Laurențiu Leuștean

București, 2017

Contents

Preamble

The interpretative flavour of proof theory originally arose from the same motivations that drove, since David Hilbert, the very study of proofs as objects in themselves: showing, by acceptable means, the consistency of logical systems that can be thought to act as a workable foundation to modern mathematics. This goal had already been stifled in the 1930s by Kurt Gödel's second incompleteness theorem, which indicated that those means could not be a subset of the same logical system which was under investigation, and then somewhat salvaged by Gerhard Gentzen's work, which indicated that the non-finitary methods necessarily playing a role in the consistency proof of first-order Peano arithmetic could be limited to the schema of primitive recursive induction up to a large countable ordinal. Gödel himself later introduced [27] the alternative method, his functional or *Dialectica* interpretation (named after the journal it was published in), of proving the consistency of arithmetic, which works by translating the proofs of intuitionistic arithmetic into a quantifier-free, higher-typed calculus of his own devising (anticipated, though, by Hilbert in [32]), dubbed System T, whose consistency is *a priori* less doubtful, arguably, than that of the source theory.

Gentzen had identified in his papers three levels of the use of infinity in mathematics – arithmetic, analysis and set theory – and he had already started preliminary work on the consistency of the next level, analysis (in the form exhibited in the monograph of Hilbert and Bernays [33] and equivalent to its modern formulation as a two-sorted, first-order theory known as "second-order arithmetic"), before his untimely death in 1945. A non-trivial, interpretative solution to the problem was first proposed by another who came to meet an early death, namely Spector [84], who augmented Gödel's system with new constants denoting bar-recursive functionals (representing an extension to higher types of Brouwer's bar induction principle, which had already been suggested by Gödel as a possible way forward). However, these additional functionals were not nearly as intuitively sound as pure System T (and a second solution later conceived by Girard [25] using yet another extension called System F suffered from similar shortcomings). Georg Kreisel, who had recently introduced another proof interpretation for intuitionistic arithmetic called modified realizability [57], a variant of an earlier work of Kleene, convened a seminar on the foundations of analysis at Stanford in the summer of 1963 with the goal of finding a justification for Spector's functionals that would be acceptable on constructive grounds. Unfortunately, the seminar concluded by declaring the answer to be "negative by a wide margin" [58].

It was Kreisel, though, that foresaw an entirely different way of looking at these proof interpretations. He proposed that instead of considering them simply a destructive instrument that translates a hypothetical proof of contradiction inside a celebrated system to an almost impossible proof in a more reliable one, we should focus instead on translating existing proofs in mathematics, carrying the hope that the result of the translation will also contain more information (for example, witnesses on existentially quantified variables) than the original proof. That way, one could hope to answer the following question [56]:

"What more do we know if we have proved a theorem by restricted means than if we merely know that it is true?"

which became the driving force behind a research program proposed by Kreisel under the name "unwinding of proofs". In the following decades (which would, however, see some seminal research and expository work in "pure" interpretative proof theory [72, 87]) only sporadic advances would be made, one of the most significant being H. Luckhardt's 1989 analysis [73] of the proof of Roth's theorem on diophantine approximations.

In the early 1990s, Luckhardt's student Ulrich Kohlenbach devised the "monotone" variants of both modified realizability and Gödel's *Dialectica* [42, 43], new proof interpretations that could only extract bounds (which were, however, more uniform) instead of full witnesses, gaining instead the power of accepting more commonly used proof principles like the weak König lemma as additional axioms (or similarly bounded universal-existential sentences). This triggered a complete overhaul of Kreisel's program, which was renamed 'proof mining' (a name originally suggested by D. Scott), and which quickly led to quantitative results being obtained in the nonlinear analysis of separable spaces, particularly in approximation theory. The next major step was taken in the early 2000s, when Kohlenbach and his collaborators, including Laurenţiu Leuştean, started to analyze proofs in functional analysis in the context of abstract (metric) spaces [44, 48]. This culminated into a series of 'general logical metatheorems' (developed by Kohlenbach [45] and by Gerhardy and Kohlenbach [23, 24]) for both classical and semi-intuitionistic systems of higher-order arithmetic (appropriately modified in order to be able to tackle the abstract spaces used in nonlinear analysis), of proof-theoretic strength less than or comparable to classical analysis, which detail the circumstances in which bounds may be effectively extracted. Kohlenbach's monograph from 2008 [46] covers the major results in the field until then, while a survey of recent developments is [47].

This thesis positions itself firmly within this continuing mission of unwinding proofs in abstract nonlinear analysis. The first two chapters are fully expository: the first one presents the concepts in functional analysis that play a role in the theorems that are being examined, while the second details the abstract logical work of Kohlenbach that underlies the extraction of quantitative information out of proofs. Although some details have been inevitably omitted, we hope that the text is sufficiently self-contained in order to be comprehensible to the mathematically-minded reader.

The third chapter contains the most representative examples of the work usually done in proof mining. It focuses on pseudocontractions, a class of nonlinear mappings originally introduced by Browder and Petryshyn [13] in the context of Hilbert spaces, and which in recent years were partially generalized to Banach spaces. We first show that the class of Banach spaces where this generalization occurs actually coincides with the class of 2-uniformly smooth Banach spaces (with an assorted computation of the link between the significant constants involved in the two definitions) and that the relevant property of pseudocontractive mappings that allows for the generalization of the convergence theorems leads both to simplified proofs of those theorems and to an easier derivation of quantitative information previously obtained by other researchers in proof mining. Then, we proceed to analyze the convergence of the parallel algorithm for strict pseudocontractions, proven by López and Xu [71], where we (i) use the relevant property of before in order to get an easier proof which first focuses on nonexpansive mappings; (ii) relax a condition from the original formulation of the theorem by removing the square root; (iii) obtain individual rates of asymptotic regularity, as it has previously been done for the Kuhfittig iteration by Khan and Kohlenbach [40]. The last contribution of the chapter is to apply earlier results on Fejér monotonicity in totally bounded spaces [50] to the case of the Ishikawa iteration in the context for which it was originally introduced, i.e. Lipschitz pseudocontractions.

We see that, given the examples above about 2-uniformly smooth spaces and the relaxation of some assumptions, research in proof mining can easily spill over in related areas. This is the case of the main results of the fourth chapter, which essentially consist in the observation that a plethora of iterations from convex optimization known generically as the "proximal point algorithm" can be indeed considered as a single algorithm under highly general conditions which are in addition very natural. These conditions manifest themselves within two definitions that apply to families of mappings, isolating the classes of "jointly firmly nonexpansive" families and of "jointly (P_2)" families. The exact relations between the two definitions and between them and the aforementioned algorithms within the framework of CAT(0) and Hilbert spaces are carefully examined, sometimes leading to entirely new convergence proofs. In addition, full rates of convergence (and other quantitative information) are obtained for a special case of the algorithm dubbed "the uniform case", where techniques due to Kohlenbach and Oliva [41, 52] that use the modulus of uniqueness can be applied in an implicit way.

Finally, the last chapter is the most abstractly-minded of the set. It stems from a recent generalization of the metatheorems of proof mining of Günzel and Kohlenbach [28] that applies to all classes of metric structures which are formalizable in positive-bounded logic. A worked example is presented there in adapting the axioms obtained by Henson and Raynaud [31] for BL^pL^q Banach lattices. We treat here the case of L^p Banach spaces with no additional lattice structure, for which no concrete list of positive-bounded axioms has been previously presented in the literature (their axiomatization had been proved, though, in a non-constructive way using metric ultraproducts). We obtain such an axiomatization and we show that it can be used to obtain a logical metatheorem for this class of spaces. We also compute the modulus of uniform convexity for L^p spaces starting from our axiomatization.

Acknowledgements. I do not live in a social vacuum, and this thesis would probably not exist in its present form if not for certain people I wish now to direct my thanks to.

Firstly, to my advisor, Laurenţiu Leuştean, who displayed an enormous amount of patience with regard to me during these PhD years, continually suggesting possible avenues for research and more generally for my professional development, which included his projects and collaborations underway, eventually leading to me being able to conceive of myself as an emerging researcher.

To Ulrich Kohlenbach, for all the ideas that sprang forward from his papers, for the useful email exchanges that we have had, and for inviting me for a three-week stay in Darmstadt last November, during which we had a lot of illuminating discussions on the motivations and obstacles of the field.

To George Georgescu, who taught me the first notions of mathematical logic by lecturing on the subject and by supervising my bachelor's thesis, and who agreed to fulfill the role of interim advisor during my first year as a PhD student.

To Ioana Leuştean, who also supported me professionally by involving me into her own specialty of research.

To Genaro López-Acedo, Adriana Nicolae and the rest of the nonlinear analysis community of Universidad de Sevilla, who graciously received me for a two-week scientific stay in the summer of 2016.

To my parents – Odilia Roşianu and Ovidiu Sipoş – and to my closest friends – Alexandra Otiman, Claudia Chiriţă and Sergiu Nisioi – who have each made me feel at least once during these years that life is worth living, and who have done their best efforts to put up with me even during the bleakest of times.

Finally, there are two people that I want to mention – namely my maternal grandparents, Claudia and Nicolae Roşianu – who have been a constant source of support and encouragement ever since the beginning of my academic aspirations, but who unfortunately did not live enough to see this affair come to its end. I can only hope that whatever I have produced would have made them proud.

Andrei Sipoş
Mangalia, 28 June 2017

Addendum, 2019: This web version of my PhD thesis differs from the one I defended in September 2017 only by the correction of typos and by the updating of the bibliography. Further adjustments in this vein may follow.
For any result in this document, I consider its published form (see next page) as being the definitive one.

Map of original results

We present a general outline of the main, original results of this thesis, organized by the paper in which they were first obtained, so appropriate credit to co-authors can be given.

1. "A note on the Mann iteration for k-strict pseudocontractions in Banach spaces" [81] (sole author)
   - Lemma 3.1.1, detailing the equivalence between different conditions imposed on Banach spaces;
   - Equation (3.2), detailing the link between the two constants specific to 2-uniformly smooth Banach spaces;
   - Lemma 3.3.3 and Theorem 3.3.4, generalizing the convergence of the Mann iteration to strict pseudocontractions to 2-uniformly smooth Banach spaces;
   - Theorem 3.4.1, giving quantitative information to the above.
2. "Effective results on a fixed point algorithm for families of nonlinear mappings" [82] (sole author)
   - Theorem 3.6.6, giving a general rate of asymptotic regularity for the parallel algorithm for nonexpansive mappings;
   - Theorem 3.6.8, giving individual rates (i.e. for each mapping separately) of asymptotic regularity for the parallel algorithm for nonexpansive mappings;
   - Theorem 3.6.9, giving individual rates of asymptotic regularity for the parallel algorithm for strict pseudocontractions, showing that one may indeed eliminate the square root from the original condition of López and Xu.
3. "Quantitative results on the Ishikawa iteration of Lipschitz pseudo-contractions" [65] (together with L. Leuştean and V. Radu)
   - Theorem 3.8.10, giving a modulus of liminf for the Ishikawa iteration for Lipschitz pseudocontractions;
   - Theorem 3.8.17, giving a rate of metastability for the same iteration in the context of totally bounded metric spaces.
4. "Proof mining in L^p spaces" [83] (sole author)
   - Lemmas 5.1.4, 5.1.5, 5.1.6, giving a new abstract characterization of L^p Banach spaces, adapted for the higher-order situation in Theorem 5.1.7;

- Theorem 5.1.8, consecrating the axiomatization of this class into positive-bounded logic;
- Theorem 5.1.9, i.e. a general logical metatheorem for L^p spaces that uses the above axiomatization;
- Theorem 5.2.5, i.e. the derivation of the modulus of uniform convexity from this axiomatization.

5. "An abstract proximal point algorithm" [64] (together with L. Leuştean and A. Nicolae)

   - Theorem 4.1.1, showing the convergence of the proximal iteration in CAT(0) spaces from highly general conditions;
   - Definitions 4.1.12 and 4.1.14, which together with the associated study, isolate a natural class of families of mappings which can be subsumed into the general conditions above (Proposition 4.1.15, Proposition 4.1.18 and Proposition 4.1.19, linked together by Theorem 4.1.20);
   - Theorems 4.1.25, 4.1.27 and 4.1.35, giving new proofs for the convergence of the proximal point algorithm in classical, concrete situations;
   - Proposition 4.2.3, quantitatively analysing for the first time the "nonempty interior" argument in nonlinear analysis;
   - Theorem 4.2.9, giving a full convergence rate for the "uniform case" of the algorithm;
   - Section 4.2.1.1, showing that concrete uniform cases are subsumed in this general uniform case;
   - Theorem 4.2.17, giving a bound on the number of steps in which the algorithm sometimes converges.

6. "An application of proof mining to the proximal point algorithm in CAT(0) spaces" [66] (together with L. Leuştean)

   - Theorem 4.2.24, giving a rate of metastability for the algorithm in the context of totally bounded metric spaces.

Chapter 1

Preliminaries on analysis

The contributions to proof mining that give this thesis its title are situated mainly in the fields of nonlinear functional analysis and convex optimization. The results that are analyzed typically feature the asymptotic behaviour of some sequence in a metric space. The aim of this chapter is to survey the relevant classes of metric spaces and their associated phenomena and properties.

The first section presents the definitions and basic properties of the metric and normed spaces involved. The second section discusses weak convergence and its metric generalization. The third section focuses on the well-behaved subclasses of uniformly convex and uniformly smooth Banach spaces. The final section studies two relevant subclasses of mappings between spaces: firmly non-expansive and pseudocontractive ones.

No results in this chapter are original. For anything which is not ubiquitous in the literature, an effort was made to supply the needed references.

1.1 Classes of spaces

Let (X, d) be a metric space. A *geodesic* in X is a mapping $\gamma : [a, b] \to X$, where $a, b \in \mathbb{R}$, such that for all $s, t \in [a, b]$ we have that $d(\gamma(s), \gamma(t)) = |s - t|$. The space is called a *geodesic space* if for all $x, y \in X$, there is a geodesic $\gamma : [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$.

A typical example of a geodesic space is a normed space, defined as follows.

Definition 1.1.1. *We call a* **normed space** *a pair consisting of a real vector space E and a function $\|\cdot\| : E \to \mathbb{R}_+$, called its* **norm**, *satisfying:*

- *for all $x \in E$, $\|x\| = 0$ iff $x = 0$;*
- *for all $\alpha \in \mathbb{R}$ and $x \in E$, $\|\alpha x\| = |\alpha|\|x\|$;*
- *for all $x, y \in E$, $\|x + y\| \leq \|x\| + \|y\|$.*

A normed space E becomes a metric space by setting $d(x,y) := \|y-x\|$. (It is called a *Banach space* if it is complete as a metric space.) Then, for any $x, y \in E$, we may define a geodesic $\gamma : [0, \|y-x\|] \to E$, by putting, for all $t \in [0,1]$:

$$\gamma(t\|y-x\|) := (1-t)x + ty.$$

Then, obviously,

$$d(\gamma(s\|y-x\|), \gamma(t\|y-x\|)) = \|(1-t)x + ty - (1-s)x - sy\| = \|(t-s)(y-x)\| = |t-s|\|y-x\|,$$

so γ is a geodesic. We call it the *standard* one.

Note that in a normed space there may be multiple geodesics. Consider the l^1 norm (yielding the "Manhattan distance") on $\mathbb{R}^2$, given, for any $x, y \in \mathbb{R}$, by:

$$\|(x,y)\|_1 := |x| + |y|.$$

Then, it is a simple exercise to show that the concatenation of the two standard geodesics between the points $(0,0)$ and $(1,0)$ and between $(1,0)$ and $(1,1)$ is a geodesic from $(0,0)$ to $(1,1)$ which is different from the standard one.

However, in spaces that have some fixed class of geodesics (like the standard ones in Banach spaces) we may define a subset to be *convex* if along with any two points of it, it also contains the image of the unique geodesic (from the given class) which connects them.

A more restrictive kind of a normed space is an inner product space.

Definition 1.1.2. *We call a* **inner product space** *a pair consisting of a real vector space H and a function $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{R}$, called its* **inner product**, *satisfying:*

- *for all $x \in H$, $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ iff $x = 0$;*
- *for all $\alpha, \beta \in \mathbb{R}$ and $x, y, z \in H$, $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$;*
- *for all $x, y \in H$, $\langle x, y\rangle = \langle y, x\rangle$.*

An inner product space becomes a normed space by setting $\|x\| := \sqrt{\langle x, x\rangle}$, and it is called a *Hilbert space* if it is complete as a metric space.

We now discuss a nonlinear generalization of inner product spaces.

A geodesic space (X,d) is called a *CAT(0) space* if for all $z \in X$, all geodesics $\gamma : [a,b] \to X$ and all $t \in [0,1]$ we have that

$$d^2(z, \gamma((1-t)a + tb)) \leq (1-t)d^2(z, \gamma(a)) + td^2(z, \gamma(b)) - t(1-t)d^2(\gamma(a), \gamma(b)).$$

The above property is easily proven to be valid in inner product spaces. It may also be proven that any CAT(0) space is (unlike general Banach spaces as seen above) *uniquely geodesic* – that is, for any x, y in such a space X there is a unique geodesic $\gamma : [0, d(x,y)] \to X$ such that $\gamma(0) = x$ and $\gamma(d(x,y)) = y$ – and therefore we may now denote, for any $t \in [0,1]$, the point $\gamma(td(x,y))$ simply by

$(1-t)x + ty$. Given that, we might note that the metric of a CAT(0) space X is *convex* – i.e., for any $x_0, y_0, x_1, y_1 \in X$ and $t \in [0, 1]$, we have that

$$d((1-t)x_0 + tx_1, (1-t)y_0 + ty_1) \leq (1-t)d(x_0, y_0) + td(x_1, y_1). \tag{1.1}$$

In particular, an inner product space is a CAT(0) space, and the defining property is written in the following form.

Proposition 1.1.3. *Let H be an inner product space. Then for any $x, y, z \in H$ we have that:*

$$\|(1-t)x + ty - z\|^2 \leq (1-t)\|x - z\|^2 + t\|y - z\|^2 - t(1-t)\|x - y\|^2.$$

A way of dealing with the unique convex combinations in CAT(0) spaces is given by the following lemma.

Lemma 1.1.4 ([1, Lemma 2.4.(iii)])**.** *Let X be a CAT(0) space, $x, w \in X$ and $\lambda, \alpha \in (0, 1)$. Set $z := (1-\alpha)x + \alpha w$, $y := (1-\lambda)x + \lambda z$ and $\mu := \frac{(1-\lambda)\alpha}{1-\lambda\alpha}$. Then $z = (1-\mu)y + \mu w$.*

This class of spaces were originally studied by A. Alexandrov (and named "CAT(0)" by M. Gromov) as a way to generalize to abstract metric spaces the property of a Riemannian manifold that its sectional curvature on any plane is bounded above by 0. The original, more involved, definition involves the comparison of the side lengths of triangles in X with those of triangles in a "model space" (here, the ordinary Euclidean space).

A tool to exploit the similarity between "nonlinear" CAT(0) spaces and "linear" inner product spaces is the quasilinearization function introduced by Berg and Nikolaev [10]. Given any metric space (X, d), we may define this function $\langle \cdot, \cdot \rangle : X^2 \times X^2 \to \mathbb{R}$, for any $a, b, u, v \in X$, by the following (where we have denoted a pair $(w, w') \in X^2$ by $\overrightarrow{ww'}$):

$$\langle \overrightarrow{ab}, \overrightarrow{uv} \rangle := \frac{1}{2}(d^2(a, v) + d^2(b, u) - d^2(a, u) - d^2(b, v)).$$

Berg and Nikolaev gave the following characterization of this mapping.

Proposition 1.1.5 ([10, Proposition 14])**.** *In an arbitrary metric space X, the mapping $\langle \cdot, \cdot \rangle$ is the unique one that satisfies, for any $a, b, c, d, f \in X$, that:*

(i) $\langle \overrightarrow{ab}, \overrightarrow{ab} \rangle = d^2(a, b)$;

(ii) $\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle = \langle \overrightarrow{cd}, \overrightarrow{ab} \rangle$;

(iii) $\langle \overrightarrow{ba}, \overrightarrow{cd} \rangle = -\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle$;

(iv) $\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle + \langle \overrightarrow{ab}, \overrightarrow{df} \rangle = \langle \overrightarrow{ab}, \overrightarrow{cf} \rangle$.

One of their main results [10, Corollary 3] is that the "Cauchy-Schwarz" inequality for this "inner product" in a geodesic space (X, d) – i.e., that for all $a, b, c, d \in X$, $\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle \leq d(a, b)d(c, d)$ – is actually equivalent to (X, d) being CAT(0).

It can be easily checked (either by direct computation or by verification of the characterizing axioms) that given an inner product space H, we have, for any $a, b, c, d \in H$, that:

$$\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle = \langle a - b, c - d \rangle = \langle b - a, d - c \rangle.$$

In addition, there is another condition that Berg and Nikolaev proved [10, Theorem 6] to be equivalent with the one of a geodesic space X being a CAT(0) space – the *quadrilateral inequality condition*, i.e. that for any $x_1, x_2, y_1, y_2 \in X$, we have that:

$$d^2(x_1, x_2) + d^2(y_1, y_2) \leq d^2(x_1, y_1) + d^2(y_1, x_2) + d^2(x_2, y_2) + d^2(y_2, x_1) \tag{1.2}$$

1.2 Convergence

A linear mapping T between two Banach spaces E and E' is said to be *continuous* if it is continuous with respect to the topology induced by the metric. It is easy to show that such a linear map T is continuous iff it is *bounded* – that is, there is a $K \in \mathbb{R}$ such that for all $x \in E$, $\|T(x)\| \leq K\|x\|$. The least such K is called the norm of T. For a Banach space E, we denote by E^* the space of all continuous linear mappings $f : E \to \mathbb{R}$ and we call it the *continuous dual*, since it is also a Banach space w.r.t. the norm now defined. We have the following classical result.

Theorem 1.2.1 (Riesz's representation theorem)**.** *If H is a Hilbert space, then for any $f \in H^*$ there is an $x \in H$ such that for all $y \in H$, $f(y) = \langle x, y \rangle$.*

It is therefore immediate that the map $x \mapsto \langle x, \cdot \rangle$ is an isomorphism (even an isometry) of Hilbert spaces.

Beside the usual definition of convergence inherited from the metric structures, Banach spaces also feature a weaker version of it.

Definition 1.2.2. *Let E be a Banach space. A sequence $(x_n)_{n \in \mathbb{N}} \subseteq E$ is said to* **converge weakly** *to an $x \in E$ if for any $f \in E^*$ we have that $\lim_{n \to \infty} f(x_n) = f(x)$. We denote this situation by $x_n \rightharpoonup x$.*

By Riesz's representation theorem, we have the following corollary.

Corollary 1.2.3. *Let H be a Hilbert space. A sequence $(x_n)_{n \in \mathbb{N}} \subseteq H$ converges weakly to an $x \in H$ iff for any $y \in H$ we have that $\lim_{n \to \infty} \langle x_n, y \rangle = \langle x, y \rangle$.*

The analogue of weak convergence in "nonlinear" (here, CAT(0)) spaces has a somewhat convoluted history. The notion of Δ-convergence in general metric spaces, to be defined below, is due to Lim [67]. Equivalent notions of "weak" convergence (which reduce to ordinary weak convergence when we are restricting ourselves to Hilbert spaces) were studied in the realm of geodesic spaces e.g. by Jost [37] and by R. Espínola and A. Fernández-León [19] (who generalized a variant due to E. Sosov).

Definition 1.2.4. *Let X be a CAT(0) space. Let $(x_n)_{n \in \mathbb{N}} \subseteq X$ be a bounded sequence. We define:*

- *for any* $y \in X$, $r(y,(x_n)) := \limsup_{n\to\infty} d(y, x_n)$*;*
- $r((x_n)) := \inf\{r(y,(x_n)) \mid y \in X\}$*;*
- $\mathcal{A}((x_n)_{n\in\mathbb{N}}) := \{y \in X \mid r(y,(x_n)) = r((x_n))\}$*;*
- *an* **asymptotic center** *of* $(x_n)_{n\in\mathbb{N}}$ *to be an element of* $\mathcal{A}((x_n)_{n\in\mathbb{N}})$*.*

In addition, we say that $(x_n)_{n\in\mathbb{N}}$ **Δ-converges** *to a point* $x \in X$ *if for any subsequence* $(u_n)_{n\in\mathbb{N}}$ *of* $(x_n)_{n\in\mathbb{N}}$ *we have that* $\mathcal{A}((u_n)_{n\in\mathbb{N}}) = \{x\}$*.*

It is known that in a complete CAT(0) space the asymptotic center of any sequence exists and is unique [18, Proposition 7]. A complete CAT(0) space is also known as a *Hadamard space*.

Definition 1.2.5. *Let* X *be a CAT(0) space. Let* $(x_n)_{n\in\mathbb{N}} \subseteq X$ *and* $C \subseteq X$*. We say that* $(x_n)_{n\in\mathbb{N}}$ *is* **Fejér monotone** *with respect to* C *if for all* $c \in C$ *and* $n \in \mathbb{N}$ *we have that*

$$d(x_{n+1}, c) \leq d(x_n, c).$$

Proposition 1.2.6 ([6, Theorem 3.3])**.** *Let* X *be a complete CAT(0) space. Let* $(x_n)_{n\in\mathbb{N}} \subseteq X$ *and* $C \subseteq X$ *such that* $(x_n)_{n\in\mathbb{N}}$ *is Fejér monotone with respect to* C*. Suppose that for all* $w \in X$ *such that there is a subsequence* $(u_n)_{n\in\mathbb{N}}$ *of* $(x_n)_{n\in\mathbb{N}}$*, having* w *as its* Δ*-limit, we have that* $w \in C$*. Then* $(x_n)_{n\in\mathbb{N}}$ Δ*-converges to a point in* C*.*

1.3 Properties of Banach spaces

A canonical example of a Banach space is the space of p-integrable functions on a measure space $(\Omega, \mathcal{F}, \mu)$ – denoted by $L^p(\Omega, \mathcal{F}, \mu)$ or simply by $L^p(\mu)$, which is the Banach space built on the set of all real-valued measurable functions f on Ω having the property that

$$\int_\Omega |f|^p \mathrm{d}\mu < \infty,$$

a set then factored by the a.e.-equality relation (which makes the canonical seminorm into a norm).

We shall denote, for any Banach space E, by $S(E)$ the set of vectors of E of norm 1 (the "unit sphere" of E). The following notion was introduced by Clarkson in 1936.

Definition 1.3.1. *A Banach space is called* **uniformly convex** *if for any* $\varepsilon > 0$ *there is a* $\delta > 0$ *such that for any vectors* x, y *in the unit ball of the space such that* $\|x - y\| \geq \varepsilon$ *we have that* $\frac{\|x+y\|}{2} \leq 1 - \delta$*.*

If $\eta : \mathbb{R} \to \mathbb{R}$ is a function such that for any $\varepsilon > 0$, $\eta(\varepsilon)$ is a corresponding δ in the sense of the definition above, we call η a *modulus of uniform convexity* for the space. This clashes a bit with the notion in the mathematical analysis literature of "the" modulus of uniform convexity. This modulus has strictly positive values iff the space is uniformly convex, and in that case it is the "optimal" modulus of uniform convexity in our sense. We now give its definition.

Definition 1.3.2. *Let E be a Banach space. We define the* **modulus of convexity of** *E to be the map $\delta_E : [0,2] \to \mathbb{R}_+$, defined, for all $\varepsilon \in [0,2]$, by*

$$\delta_E(\varepsilon) := \inf\left\{1 - \frac{\|x+y\|}{2} \;\middle|\; x,y \in S(E), \|x-y\| \geq \varepsilon\right\}. \tag{1.3}$$

Typical examples of uniformly convex Banach spaces are Hilbert spaces, and also the L^p spaces defined above (the derivation of a valid modulus of uniform convexity for them will be the focus of the final section of the final chapter).

Definition 1.3.3. *Let E be a Banach space. We define the* **normalized duality mapping of** *E to be the map $J : E \to 2^{E^*}$, defined, for all $x \in E$, by*

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

A Banach space E is called *smooth* if for any $u \in S(E)$, we have that for any $v \in S(E)$, the limit

$$\lim_{h\to 0} \frac{\|u+hv\| - \|u\|}{h}$$

exists. This condition has been proven to be equivalent to the fact that the normalized duality mapping of the space, $J : E \to 2^{E^*}$, is single-valued – and we shall denote its unique section by $j : E \to E^*$. Therefore, for all $x \in E$, $j(x)(x) = \|x\|^2$ and $\|j(x)\| = \|x\|$. Hilbert spaces are smooth, and clearly $j(x)(y)$ is then simply $\langle x, y\rangle$, for any x, y in the space. This use of j will likely induce a remnant of the nice properties of the inner product.

Moreover, E has a *Fréchet differentiable norm* if, in addition, the limit above is attained uniformly in the variable $v \in S(E)$ and it is *uniformly smooth* (or has a *uniformly Fréchet differentiable norm*) if the limit is attained uniformly in the pair of variables $(u, v) \in S(E) \times S(E)$.

There are classical results stating that for any Banach space E, E is uniformly convex iff E^* is uniformly smooth and E is uniformly smooth iff E^* is uniformly convex.

One can also define (in analogy to the standard modulus of uniform convexity defined above) the *modulus of smoothness* of E to be the map $\rho_E : (0,\infty) \to \mathbb{R}$, defined, for all $\tau \in (0,\infty)$, by

$$\rho_E(\tau) := \sup\left\{\frac{\|u+\tau v\| + \|u-\tau v\|}{2} - 1 \;\middle|\; u,v \in S(E)\right\}.$$

It is known that a space E is uniformly smooth iff

$$\lim_{\tau\to 0} \frac{\rho_E(\tau)}{\tau} = 0.$$

This can happen, for example, if there are $c > 0$ and $q > 1$ such that for all τ, $\rho_E(\tau) \leq c\tau^q$. In that case, E is said to be *q-uniformly smooth*.

Such an "upper bound" for this modulus could be considered a generalized sort of modulus like in the definition of uniform convexity. In fact, such a modulus has already been considered for practical applications in [49, 54]. In addition, the following interesting relationship was proven for the first time in [54].

Theorem 1.3.4 ([54, Theorem 9.2.10])**.** *A Banach space E is uniformly smooth iff there is a section $j : E \to E^*$ of $J : E \to 2^{E^*}$ which is uniformly continuous.*

A great reference for various characteristic moduli and constants of normed linear spaces is [21].

1.4 Mappings

If X is a metric space and $T : X \to X$ is a mapping, we denote (throughout this whole thesis) by $Fix(T)$ the set of its fixed points.

The most basic form of a mapping between two metric spaces is that of a nonexpansive one. If X and Y are metric spaces, then a *nonexpansive mapping* between them is a function $T : X \to Y$ such that for all $x, y \in X$,

$$d_Y(Tx, Ty) \leq d_X(x, y).$$

The mapping type most used in convex optimization is a more particular form of nonexpansive mapping, namely the firmly nonexpansive one. This concept was originally defined in the context of Hilbert spaces, but here we shall use the following definition, introduced and studied extensively in [1] for a class of geodesic spaces which includes CAT(0) spaces.

Definition 1.4.1 (cf. [1, Section 3])**.** *Let X be a CAT(0) space. A mapping $T : X \to X$ is called* **firmly nonexpansive** *if for any $x, y \in X$ and any $t \in [0, 1]$ we have that*

$$d(Tx, Ty) \leq d((1-t)x + tTx, (1-t)y + tTy).$$

A strongly related property is the following one, which was dubbed the (P_2) **property** in [2, 51].

Definition 1.4.2. *Let X be a CAT(0) space. A mapping $T : X \to X$ is said to have the (P_2)* **property** *if for any $x, y \in X$ we have that*

$$2d^2(Tx, Ty) \leq d^2(x, Ty) + d^2(y, Tx) - d^2(x, Tx) - d^2(y, Ty).$$

Using Berg and Nikolaev's quasilinearization function, the (P_2) condition may be written in the following equivalent form:

$$0 \leq \langle \overrightarrow{TxTy}, \overrightarrow{xTx} \rangle - \langle \overrightarrow{TxTy}, \overrightarrow{yTy} \rangle.$$

The exact relationship between the two properties is given by the following result.

Theorem 1.4.3 (cf. [2, Section 2.2] and [8, Proposition 4.2])**.** *Let X be a CAT(0) space and $T : X \to X$. Then:*

1. *if T is firmly nonexpansive, then T is (P_2);*
2. *if X is a Hilbert space and T is (P_2), then T is firmly nonexpansive.*

Proof. 1. By applying the defining property of CAT(0) spaces twice (once for each argument) on the firm nonexpansiveness condition squared and with an arbitrary $t \in (0, 1)$, we obtain that:

$$\begin{aligned} d^2(Tx, Ty) \leq & (1-t)^2 d^2(x, y) + t(1-t)d^2(Tx, y) + t(1-t)d^2(x, Ty) \\ & + t^2 d^2(Tx, Ty) - t(1-t)d^2(x, Tx) - t(1-t)d^2(y, Ty). \end{aligned}$$

If we move the "$t^2 d^2(Tx, Ty)$" term to the left hand side, divide by $1 - t \neq 0$ and the let $t \to 1$, we obtain exactly the (P_2) condition.

2. Using the Berg-Nikolaev-style (P_2) condition from above and the special form using the inner product of the quasilinearization function in Hilbert spaces, what we claim is exactly the implication (v) ⇒ (vi) of [8, Proposition 4.2]. (Also, in the fourth chapter, we shall see proven a more general version of this.)

□

Let now X be a CAT(0) space and $T : X \to X$ be a mapping that satisfies the (P_2) property. Let $x, y \in X$. Applying (1.2) for $x_1 := x$, $x_2 := Ty$, $y_1 := Tx$, $y_2 := y$, we get that:

$$d^2(x, Ty) + d^2(y, Tx) \le d^2(x, Tx) + d^2(Tx, Ty) + d^2(y, Ty) + d^2(x, y).$$

By the above and the (P_2) property, we obtain that:

$$2d^2(Tx, Ty) \le d^2(Tx, Ty) + d^2(x, y)$$

or

$$d^2(Tx, Ty) \le d^2(x, y).$$

We have therefore proven that any mapping that satisfies the (P_2) property is nonexpansive. Another property of (P_2) mappings is the following. Let $x \in X$ and $p \in Fix(T)$. We have that:

$$2d^2(Tx, p) \le d^2(x, p) + d^2(Tx, p) - d^2(x, Tx)$$

or

$$d^2(Tx, p) \le d^2(x, p) - d^2(x, Tx). \tag{1.4}$$

In sharp contrast to firm nonexpansiveness, there are classes **more general** than the one of nonexpansive mappings that are used in nonlinear analysis. These classes, first defined by Browder and Petryshyn [13] and by Kato [39], will be used at first solely in Hilbert spaces.

Definition 1.4.4. *Let H be a Hilbert space, $C \subseteq H$ a convex subset, $k \in [0, 1)$. A mapping $T : C \to H$ is called a k-**strict pseudocontraction** if for all $x, y \in C$, we have that:*

$$\|Tx - Ty\|^2 \le \|x - y\|^2 + k\|(x - Tx) - (y - Ty)\|^2.$$

It is easy to see that the condition reduces to nonexpansiveness when $k = 0$. Also, it has been proven (see, e.g. [75, Proposition 2.1.(i)]) that any k-strict pseudocontraction is Lipschitzian of constant $\frac{1+k}{1-k}$.

Definition 1.4.5. *Let H be a Hilbert space, $C \subseteq H$ a convex subset. A mapping $T : C \to H$ is called a **pseudocontraction** if for all $x, y \in C$, we have that:*

$$\|Tx - Ty\|^2 \le \|x - y\|^2 + \|(x - Tx) - (y - Ty)\|^2.$$

The significance of the above concept lies in the following fact: an operator $T : C \to H$ is a pseudocontraction (by [13, Theorem 1.(1)]) if and only if its complement $U := Id - T$ is *monotone*, i.e. for all $x, y \in C$ we have that

$$\langle Ux - Uy, x - y\rangle \ge 0.$$

Monotone operators arise naturally in the study of partial differential equations: often such an equation can be written in the form $U(x) = 0$ (or $0 \in U(x)$ when considering multi-valued operators). Finding a zero of U is equivalent to finding a fixed point of its complement $T := Id - U$, hence the problem of finding fixed points of nonlinear operators is tightly linked to that of finding solutions to nonlinear differential equations.

There is, by [13, Theorem 1.(2)], a similar equivalence for k-strict pseudocontractions – an operator $T : C \to H$ is a pseudocontraction if and only if its complement $U := Id - T$ has the property that for any $x, y \in C$,

$$\langle Ux - Uy, x - y\rangle \geq \frac{1-k}{2}\|Ux - Uy\|^2.$$

We can write the above solely in terms of T, i.e.,

$$\langle (x - Tx) - (y - Ty), x - y\rangle \geq \frac{1-k}{2}\|(x - Tx) - (y - Ty)\|^2.$$

An issue that arises is that of the proper generalization of k-strict pseudocontractions to the case of Banach spaces. For smooth Banach spaces, given what we said above about their single-valued normalized duality mapping j, we may extend the condition above to the following one (used, for example, in [16, 90]), which we will take as our "official" definition of k-strict pseudocontractions in Banach spaces: for any x, y in the space, we have that

$$j(x - y)((x - Tx) - (y - Ty)) \geq \frac{1-k}{2}\|(x - Tx) - (y - Ty)\|^2.$$

In the third chapter, we shall convince ourselves that this is indeed a well-behaved extension.

Chapter 2

Preliminaries on logic

We now begin to examine the logical underpinnings of proof mining. The results of this thesis which pertain most to these fundamental assumptions will be seen in the final chapter. However, with regard to the theorems yielded by the practical application of the field in actual mathematics, such as those that we shall see in the two chapters following this one, it is a known and relevant truth that a thorough knowledge of proof theory is not necessary in order to understand them or to check them for correctness, as what is usually obtained is just another (more quantitative) theorem in the field of application (e.g. nonlinear analysis), which is proven using similar methods. Still, in order to get a faithful picture of the relevance and context for those results, it would be necessary to have a general exposition of proof mining in the abstract, even not taking into consideration the needs of the final chapter.

Now, it would be almost impossible to exhibit here a full presentation, with all the details, of the theorems which make up the bulk of Kohlenbach's monograph [46]. One must then compromise. And compromise is usually driven by goals. The goals that we have chosen and that we consider feasible are the following: (i) to make sure that almost all the concepts involved are defined, and so the reader can at least have a clear picture of *what* the theorems say; (ii) to put forward examples of tricky or significant parts of the proofs, so one could, if not get a full explanation, at least obtain a *feeling* of *why* the results should *morally* be true – in other words, to dispel any sensations that miracles are being performed.

The first section presents the purely arithmetical systems which form the basis of proof mining, with a focus on their higher-typed nature. The second section turns to the proof interpretations which may be applied to these systems, i.e. modified realizability and Gödel's *Dialectica*, focusing on their different roles with regard to the double negation translation. The last section presents Kohlenbach's approach to formalizing metric or normed spaces by augmenting these arithmetical systems, and the resulting extension of proof interpretations in the form of logical metatheorems.

As with the previous chapter, we do not make here any claim at originality. We generally use the presentation of [46], sometimes updating it by the conventions in [28].

2.1 The logical systems

The systems with regard to which the program of proof mining has been pursued so far comprise a tight family with some common, prominent features. Namely, they are all *systems of arithmetic* (capable of expressing basic facts about numbers, e.g. the first-order sentences), *higher-typed* (having terms representing functions of arbirary finite height) and possess a *Hilbert-style proof system*. Let us detail a bit these characteristics.

The set of types, denoted by $\mathbb{T}$, is given by the free algebra with a single binary operation $\to$ generated by the single constant 0, representing the type of natural numbers. If ρ and τ are types, then $\rho \to \tau$, also denoted by $\tau(\rho)$ or simply by $\tau\rho$, represents the type of functions from the objects of type ρ to the objects of type τ.

The *terms* of the system are assigned types at construction time and are given by the following rules:

1. we have a denumerable supply of variables of each type;
2. we have some fixed constants: 0^0 (by this we say that 0 is a constant of type 0), S^{00} (successor) and some schemas of constants: the classical combinators $\Pi^{\rho\tau\rho}_{\rho,\tau}$ and $\Sigma^{\tau\delta(\rho\delta)(\tau\rho\delta)}_{\delta,\rho,\tau}$, together with recursors R_ρ;
3. if t is a term of type $\tau\rho$ and s is a term of type ρ, then $t(s)$ is a term of type τ.

We now introduce some additional concepts related to types and terms.

By a *term of a type-tuple* $\vec{\rho} = (\rho_1, \ldots, \rho_n) \in \mathbb{T}^*$ we will mean a tuple of terms $\underline{t} = (t_1, ..., t_n)$ such that for each i, t_i is a term of type ρ_i. As we can see, we denote tuples of terms by underlined letters.

We define the operations $\to^1$: $\mathbb{T}^* \times \mathbb{T} \to \mathbb{T}^*$, $\to^2$: $\mathbb{T}^* \times \mathbb{T}^* \to \mathbb{T}^*$ recursively, by:

$$() \to^1 \rho := \rho$$

$$(\vec{\theta}, \tau) \to^1 \rho := \vec{\theta} \to^1 (\tau \to \rho)$$

$$\vec{\rho} \to^2 () := ()$$

$$\vec{\rho} \to^2 (\vec{\theta}, \tau) := (\vec{\rho} \to^2 \vec{\theta}, \vec{\rho} \to^1 \tau)$$

For any type ρ there exists an unique $\vec{\tau} = (\tau_1, \ldots, \tau_n) \in \mathbb{T}^*$ such that

$$\rho = \vec{\tau} \to^1 0 = \tau_1 \to (\tau_2 \to \ldots \to (\tau_n \to 0) \ldots)$$

– this is its *normal form*.

The application of terms can be similarly extended to tuples – $t()$ will be t; $t(\underline{t'}, u)$ will be $(t\underline{t'})u$; $()\underline{v}$ will be $()$ and finally $(\underline{t}, u)\underline{v}$ will be $(\underline{t}\underline{v}, u\underline{v})$.

The *formulas* of the system are constructed as follows:

1. if s and t are terms of type 0, then $s =_0 t$ is a 'prime formula';
2. if A and B are formulas, then $A \wedge B$, $A \vee B$ and $A \to B$ are formulas;
3. if A is a formula and x is a variable, then $\forall x A$ and $\exists x A$ are formulas.

We use the following abbreviations:

1. 1 for $S0$, etc.;
2. $\bot$ for $0 =_0 1$;
3. $\neg A$ for $A \to \bot$, $A \Leftrightarrow B$ for $(A \to B) \wedge (B \to A)$;
4. $s =_\rho t$ (where s and t are terms of type $\rho = 0\rho_k...\rho_1$) for
$$\forall y_1^{\rho_1}...\forall y_k^{\rho_k}(sy_1...y_k =_0 ty_1...y_k).$$

As we said before, the proof system will be Hilbert-style, given by axioms and rules of deduction. These are:

1. an axiomatization of intuitionistic logic used by Gödel in his *Dialectica* paper [27] for its simplicity;
2. axioms that state that $=_0$ is an equivalence relation;
3. axioms stating that the successor function is injective and that 0 is not a successor;
4. the induction schema:
$$A(0) \wedge \forall x^0(A(x) \to A(Sx)) \to \forall x A(x);$$
5. defining axioms for the three constant schemas.

We need to add a bit more if we are to define a system of intuitionistic arithmetic in all finite types. Higher-type equality, as seen above, is defined extensionally, and in order to make it well-behaved, the system is to be enriched. This can be done in the following ways – either we add what is called the quantifier-free rule of extensionality (where A_0 is a quantifier-free formula):

$$\frac{A_0 \to s =_\rho t}{A_0 \to r[x := s] =_\tau r[x := t]}$$

leading to the so-called weakly extensional Heyting arithmetic in all finite types ($\mathbf{WE\text{-}HA}^\omega$), or we add the axiom of extensionality (where $\rho = 0\rho_k...\rho_1$):

$$\forall z^\rho \forall x_1^{\rho_1} \forall y_1^{\rho_1} ... \forall x_k^{\rho_k} \forall y_k^{\rho_k} \left(\bigwedge_{i=1}^{k} (x_i =_{\rho_i} y_i) \to z x_1...x_k =_0 z y_1...y_k \right)$$

leading to fully extensional Heyting arithmetic in all finite types ($\mathbf{E\text{-}HA}^\omega$).

In addition, each variant of the system can be transformed into the corresponding 'Peano' version ($\mathbf{WE\text{-}PA}^\omega$, resp. $\mathbf{E\text{-}PA}^\omega$) by adding the law of excluded middle as an axiom schema,

$$A \vee \neg A.$$

The semantics for this logic is given by specifying a model which resembles a many-sorted first-order structure (e.g. we have an underlying set corresponding to each type). The standard model in all finite types $\mathcal{S}^\omega$ is constructed by putting $\mathcal{S}_0 := \mathbb{N}$ and for any types ρ, τ, $\mathcal{S}_{\tau(\rho)} := \mathcal{S}_\tau^{\mathcal{S}_\rho}$, assigning to any language constant its standard value. The satisfaction of sentences is then taken in the usual Tarskian sense.

What remains to be done is to find a way to express the real number system in this higher arithmetical framework. The key idea here is to take the ordinary definition of rational and later real numbers as equivalence classes of certain objects and then turn it into a workable *representation* by dealing with the representatives themselves instead of the classes.

For example, we take the usual Cantorian bijection $j : \mathbb{N}^2 \to \mathbb{N}$ and we postulate that $j(n,m)$ will represent the rational number $\frac{\frac{n}{2}}{m+1}$ if n is even and the number $-\frac{\frac{n+1}{2}}{m+1}$ otherwise. Equality and the other usual relations and operations on $\mathbb{Q}$ are easily seen to be decidable with respect to this representation.

Real numbers are to be conceived as functions f of type $1 := 0(0)$ satisfying the condition:

$$\forall n(|f(n) -_{\mathbb{Q}} f(n+1)|_{\mathbb{Q}} <_{\mathbb{Q}} 2^{-n-1}),$$

where the power of 2 is actually taken to be its rational representation as above. This condition actually certifies that f is a Cauchy sequence of rationals and hence "defines" a real. On the other hand, any function can be primitively recursively transformed into one that satisfies the condition above, by the operation $f \mapsto \widehat{f}$, defined as follows:

$$\widehat{f}(n) := \begin{cases} f(n), & \text{if, for all } k < n,\ |f(k) -_{\mathbb{Q}} f(k+1)|_{\mathbb{Q}} <_{\mathbb{Q}} 2^{-k-1}; \\ f(k), & \text{where } k < n \text{ is the least such that } |f(k) -_{\mathbb{Q}} f(k+1)|_{\mathbb{Q}} \geq_{\mathbb{Q}} 2^{-k-1}. \end{cases}$$

That way, any element of type 1 will be a representative of a real number. On the type 1, then, we may define equality of real numbers, $f_1 =_{\mathbb{R}} f_2$, by:

$$\forall n(|\widehat{f}_1(n+1) -_{\mathbb{Q}} \widehat{f}_2(n+1)|_{\mathbb{Q}} <_{\mathbb{Q}} 2^{-n}).$$

This time, equality is seen not to be decidable but rather Π^0_1 and hence co-recursively enumerable. Similarly, $<_{\mathbb{R}}$ is recursively enumerable (Σ^0_1) and $\leq_{\mathbb{R}}$ is again Π^0_1. This inevitable fact is considered to be one of the main impediments in working with effective formalizations of the reals. On the other hand, situations like the one happening in the usual definition of a convergent sequence, where "$< \varepsilon$" can be readily interchanged with "$\leq \varepsilon$", occur frequently, this giving us a leeway in minimizing the complexity of the formulas under discussion.

2.2 Proof interpretations

Proof interpretations are the basic ingredient of proof mining. They are highly varied in form, but frequently fall into the following schema. If T and S are two logical systems and $(\cdot)^I$ is a mapping from the formulas of the language of T to the ones of the language of S, then I is a *sound interpretation* if for any φ we have that:

$$T \vdash \varphi \text{ implies } S \vdash \varphi^I.$$

A basic example is given by Glivenko's theorem for propositional logic, which translates classical into intuitionistic logic. For any propositional formula φ, we have that

$$\vdash_c \varphi \text{ implies } \vdash_i \neg\neg\varphi.$$

Such a result already exhibits two important characteristics which will generally be present in all interpretations that we will present below. Firstly, the source system is a superset of the destination one, with some new proof principles added, which may be regarded as not particularly constructive. Then, the translated formula is equivalent to the original one in the stronger source system (it makes sense to compare them since the languages coincide). The "soundness" name in inspired by its standard sense of soundness with respect to a semantics, since the method of showing it is similar – induction over the structure of a proof (and here we might see why a Hilbert-style system is more suitable for this style of proof theory).

Nevertheless, Glivenko's theorem is just a "translation" and not a full-blown proof interpretation. For the latter, one might look at the modified realizability interpretation for intuitionistic logic, devised by Kreisel [57]. This applies to the language defined in the previous section. We shall now properly introduce it, along with the needed preliminary definitions.

We first define the function $t : Form \to \mathbb{T}^*$, where $Form$ is the set of all formulas in our higher-order language, by:

$$t(\text{a prime formula}) := ()$$
$$t(A \wedge B) := t(A) \circ t(B)$$
$$t(A \vee B) := (0) \circ t(A) \circ t(B)$$
$$t(A \to B) := t(A) \to^2 t(B)$$
$$t(\forall y^\rho A) := (\rho) \to^2 t(A)$$
$$t(\exists y^\rho A) := (\rho) \circ t(A)$$

The modified realizability "operator" accepts in the second argument a formula A and in the first argument a term of the type-tuple $t(A)$ – and must take care in its definition that this property is recursively maintained.

$$() \; mr \text{ a prime formula} := \text{the formula itself}$$
$$(\underline{t}, \underline{t'}) \; mr \; (A \wedge B) := (\underline{t} \; mr \; A) \wedge (\underline{t'} \; mr \; B)$$
$$(u, \underline{t}, \underline{t'}) \; mr \; (A \vee B) := (u = 0 \to (\underline{t} \; mr \; A)) \wedge (u \neq 0 \to (\underline{t'} \; mr \; B))$$
$$\underline{t} \; mr \; (A \to B) := \forall \underline{y}((\underline{y} \; mr \; A) \to (\underline{t}\underline{y} \; mr \; B))$$
$$\underline{t} \; mr \; (\forall y^\rho A) := \forall y(\underline{t}y \; mr \; A)$$
$$(u, \underline{t}) \; mr \; (\exists y^\rho A) := \underline{t} \; mr \; (A[y \leftarrow u])$$

A formula A is called **$\exists$-free** iff $\exists$ and $\vee$ do not appear in it. In that case, $t(A) = ()$. Also, note that for any formula A and any compatible term-tuple $\underline{t}$, we have that $\underline{t} \; mr \; A$ is an $\exists$-free formula of our language.

Another feature to which we draw attention is that a quantifier-free formula can be shown to be equivalent in $\mathbf{WE\text{-}HA}^\omega$ to a basic formula, and hence we may treat it practically as such.

Modified realizability possesses the first characteristic from above: the base system, considered to be the 'constructive' one, is $\mathbf{E\text{-}HA}^\omega$, while the additional principles are the 'Axiom of Choice', here defined by:

$$\mathbf{AC}^{\rho,\tau,A}\colon\ \forall x^\rho \exists y^\tau A(x,y) \to \exists Y^{\rho\to\tau} \forall x^\rho A(x, Yx)$$

and a restricted form of the 'Independence of Premise' schema (i.e. where we force A to be an $\exists$-free formula that does not have x as a free variable):

$$\mathbf{IP}^{\rho,A,B}_{\exists f}\colon\ (A \to \exists x^\rho B) \to \exists x^\rho (A \to B).$$

The soundness theorem for this interpretation is expressed as follows.

Theorem 2.2.1 ([87])**.** *Let A be a formula of our language and Δ be a set of $\exists$-free sentences. Suppose that:*

$$\mathbf{E\text{-}HA}^\omega + \mathbf{AC} + \mathbf{IP}^\omega_{\exists f} + \Delta \vdash A.$$

Then there exists a term-tuple $\underline{t}$ of type $t(A)$ such that $Var(\underline{t}) \subseteq FV(A)$ and:

$$\mathbf{E\text{-}HA}^\omega + \Delta \vdash \underline{t}\ mr\ A.$$

Moreover, this passage is computable, in the sense that there is an effective procedure which accepts as input a proof of A from the stated hypotheses and returns the corresponding proof of $\underline{t}$ mr A (and of course $\underline{t}$ as a byproduct).

Let us now return to the issue of this being a "proper" interpretation and also to bring forward the second desired characteristic. The translation of the formula A is actually the formula

$$\exists \underline{x} (\underline{x}\ mr\ A),$$

which can indeed be shown to be equivalent to A in the theory $\mathbf{E\text{-}HA}^\omega + \mathbf{AC} + \mathbf{IP}^\omega_{\exists f}$. However, the theory $\mathbf{E\text{-}HA}^\omega$ is by itself sufficiently constructive, so that the provability of $\exists \underline{x}(\underline{x}\ mr\ A)$ further implies the existence of a $\underline{t}$ such that $\underline{t}\ mr\ A$ is provable. In addition, we remark that $\exists$-free sentences may be added as lemmas without influencing the term extraction (because, as pointed out above, these sentences are "interpreted by themselves") – we shall usually see such a class of admissible sentences featured in all further proof interpretations, their complexity usually being traded off against extraction power. The fact that the universal sentences, for example, which shall be featured in the functional interpretation, play this role is in fact an old observation of Kreisel.

The next important question to ask ourselves is how do we deal with proofs in classical arithmetic. The answer is that we must first translate them into an intuitionistic system, using a massive generalization of Glivenko's procedure above, due to Kuroda (the first variant of it was developed by Gödel in 1933 [26]). It is defined as follows:

$$A^* := A, \text{ if } A \text{ is prime}$$

$$(A \wedge B)^* := A^* \wedge B^* \text{ (and the same for } \vee \text{ and } \to)$$

$$(\exists x A)^* := \exists x A^*$$

$$(\forall x A)^* := \forall x \neg\neg A^*$$

Finally, we put $A' := \neg\neg A^*$. We have the following soundness result:

Theorem 2.2.2 ([46, Proposition 10.6])**.** *Let A be a formula of our language and $\mathcal{P}$ be a set of purely universal sentences. Suppose that:*

$$\textbf{WE-PA}^\omega + \textbf{QF-AC} + \mathcal{P} \vdash A.$$

Then:

$$\textbf{WE-HA}^\omega + \textbf{QF-AC} + \textbf{M}^\omega + \mathcal{P} \vdash A'.$$

In the above, **QF-AC** is simply the restriction of the **AC** schema from before to quantifier-free formula. What is more relevant, here, is the meaning of $\textbf{M}^\omega$. This is *Markov's principle*, a principle which is generally accepted in the Russian school of constructivity, and which is formalized as follows:

$$\textbf{M}^{\underline{\rho}}\text{: } \neg\neg\exists \underline{x}^{\underline{\rho}} A_0(\underline{x}) \to \exists \underline{x}^{\underline{\rho}} A_0(\underline{x}),$$

where A_0 is a quantifier-free formula that has among its variables at least the tuple $\underline{x}$, which is of arbitrary type-tuple $\underline{\rho}$ (unlike Andrey Markov's original version of the schema, where only single variables representing natural numbers were used).

What remains is to translate the (semi-intuitionistic) system obtained above into a constructive one that may admit useful term extraction. One candidate is the modified realizability interpretation outlined before. However, that proof interpretation is not suitable because it cannot properly deal with Markov's principle. A better candidate is Gödel's *functional* or *Dialectica interpretation*, which we will now present.

In the sequel, we denote elements of $\mathbb{T}^* \times \mathbb{T}^*$ (pairs of tuples of types) like $[\vec{\rho}, \vec{\theta}]$, to avoid confusion with other kinds of brackets.

We define a new type-tuple-operator $t_D : Form \to \mathbb{T}^* \times \mathbb{T}^*$, in the following way. Assume that A and B are formulas such that $t_D(A) = [\vec{\rho}_1, \vec{\theta}_1]$ and $t_D(B) = [\vec{\rho}_2, \vec{\theta}_2]$.

$$\begin{aligned}
t_D(\text{a prime formula}) &:= [(), ()]\\
t_D(A \wedge B) &:= [\vec{\rho}_1 \circ \vec{\rho}_2, \vec{\theta}_1 \circ \vec{\theta}_2]\\
t_D(A \vee B) &:= [(0) \circ \vec{\rho}_1 \circ \vec{\rho}_2, \vec{\theta}_1 \circ \vec{\theta}_2]\\
t_D(A \to B) &:= [(\vec{\rho}_1 \to^2 \vec{\rho}_2) \circ (\vec{\rho}_1 \to^2 (\vec{\theta}_2 \to^2 \vec{\theta}_1)), \vec{\rho}_1 \circ \vec{\theta}_2]\\
t_D(\exists z^\tau A) &:= [(\tau) \circ \vec{\rho}_1, \vec{\theta}_1]\\
t_D(\forall z^\tau A) &:= [(\tau) \to^2 \vec{\rho}_1, \vec{\theta}_1]
\end{aligned}$$

For any formula A such that $t_D(A) = [\vec{\rho}, \vec{\theta}]$ we define a new formula A_D, whose free variables will be those of A together with new ones $\underline{x}$, $\underline{y}$ of types $\vec{\rho}$ and $\vec{\theta}$, respectively – for that reason, we will denote the formula by $A_D(\underline{x}; \underline{y})$ (occasionally, we will include in brackets the free variables of A). The definition is made inductively, as follows (we assume that we have already built $A_D(\underline{x}; \underline{y})$ and $B_D(\underline{u}; \underline{v})$):

$$
\begin{aligned}
(\text{a prime formula})_D &:= \text{the formula itself}\\
(A \wedge B)_D(\underline{x}, \underline{u}; \underline{y}, \underline{v}) &:= A_D(\underline{x}; \underline{y}) \wedge B_D(\underline{u}; \underline{v})\\
(A \vee B)_D(z, \underline{x}, \underline{u}; \underline{y}, \underline{v}) &:= (z = 0 \to A_D(\underline{x}; \underline{y})) \wedge (z \neq 0 \to B_D(\underline{u}; \underline{v}))\\
(A \to B)_D(\underline{U}, \underline{Y}; \underline{x}, \underline{v}) &:= A_D(\underline{x}; \underline{Y}\underline{x}\underline{v}) \to B_D(\underline{U}\underline{x}; \underline{v})\\
(\exists z^\tau A)_D(z, \underline{x}; \underline{y}) &:= A_D(\underline{x}; \underline{y})\\
(\forall z^\tau A)_D(\underline{X}; z, \underline{y}) &:= A_D(\underline{X}z; \underline{y})
\end{aligned}
$$

We define A^D to be the formula $\exists \underline{x} \forall \underline{y} A_D(\underline{x}; \underline{y})$.

As an immediate application, we get that if A is such that $t_D(A) = [\vec{\rho}, \vec{\theta}]$, and we have built $A_D(\underline{x}; \underline{y})$, then $t_D(\neg A) = [\vec{\rho} \to^2 \vec{\theta}, \vec{\rho}]$, $t_D(\neg\neg A) = [(\vec{\rho} \to^2 \vec{\theta}) \to^2 \vec{\rho}, \vec{\rho} \to^2 \vec{\theta}]$ and:

$$(\neg A)^D = \exists \underline{Y} \forall \underline{x} \neg A_D(\underline{x}; \underline{Y}\underline{x})$$

$$(\neg\neg A)^D = \exists \underline{X} \forall \underline{Y} \neg\neg A_D(\underline{X}\underline{Y}; \underline{Y}(\underline{X}\underline{Y}))$$

The following is the soundness theorem for this interpretation.

Theorem 2.2.3. *[27, 87] Let A be a formula of our language and $\mathcal{P}$ be a set of purely universal sentences. Suppose that:*

$$\textbf{WE-HA}^\omega + \textbf{AC} + \textbf{IP}^\omega_\forall + \mathcal{P} \vdash A.$$

Then there exists a term-tuple $\underline{t}$, its type being the first component $t_D(A)$ such that $Var(\underline{t}) \subseteq FV(A)$ and:

$$\textbf{WE-HA}^\omega + \mathcal{P} \vdash \forall \underline{y} A_D(\underline{t}; \underline{y}).$$

Moreover, this passage is computable, in the sense that there is an effective procedure which accepts as input a proof of A from the stated hypotheses and returns the corresponding proof of $\underline{t}$ mr A (and of course $\underline{t}$ as a byproduct).

Compared to modified realizability, functional interpretation accepts as additional lemmas only purely universal sentences (and restricts the non-constructive IP principle down to such formulas) and is forced to deal with quantifier-free extensionality. On the other hand, we see that it is able to interpret Markov's principle and hence we can combine it, as follows, in the following "main theorem on program extraction".

Theorem 2.2.4 ([46, Theorem 10.8])**.** *Let $A_0(x, y)$ be a quantifier-free formula containing only x and y as free variables (of type ρ and τ, respectively) and $\mathcal{P}$ be a set of purely universal sentences. Suppose that:*

$$\textbf{WE-PA}^\omega + \textbf{QF-AC} + \mathcal{P} \vdash \forall x \exists y A_0(x, y).$$

Then there is a closed term t of type $\tau\rho$ such that:

$$\textbf{WE-HA}^\omega + \mathcal{P} \vdash \forall x A_0(x, tx).$$

Proof. By the soundness of the Kuroda translation, we obtain (using throughout the observation before, of treating quantifier-free formulas as basic) that:

$$\textbf{WE-HA}^\omega + \textbf{QF-AC} + \textbf{M}^\omega + \mathcal{P} \vdash \neg\neg\forall x \neg\neg\exists y A_0(x,y).$$

Then, by pure intuitionistic logic, we get:

$$\textbf{WE-HA}^\omega + \textbf{QF-AC} + \textbf{M}^\omega + \mathcal{P} \vdash \forall x \neg\neg\exists y A_0(x,y).$$

Then we apply Markov's principle:

$$\textbf{WE-HA}^\omega + \textbf{QF-AC} + \textbf{M}^\omega + \mathcal{P} \vdash \forall x \exists y A_0(x,y).$$

We are now in position to apply the functional interpretation. We compute that

$$t_D(\forall x \exists y A_0(x,y)) = [(\rho \to \tau), ()]$$

and that

$$(\forall x \exists y A_0(x,y))_D(Y;x) = A_0(x, Yx),$$

where Y is a new variable of type $\rho \to \tau = \tau\rho$. Hence, by the soundness theorem, we have that there is a term t of type $\tau\rho$ with no free variables such that:

$$\textbf{WE-HA}^\omega + \mathcal{P} \vdash \forall x A_0(x, tx),$$

which is what we needed. □

The most powerful foundational system that has so far been studied from the viewpoint of proof mining is weakly extensional classical analysis in all finite types. It is obtained by adjoining to the system **WE-PA**$^\omega$ + **QF-AC** studied earlier, the axiom of dependent choice, formalized as:

$$\textbf{DC}^{\underline{\rho}}\colon\ \forall x^0 \forall \underline{y}^{\underline{\rho}} \exists \underline{z}^{\underline{\rho}} A(x, \underline{y}, \underline{z}) \to \exists \underline{f}^{\underline{\rho}(0)} \forall x^0 A(x, \underline{f}(x), \underline{f}(S(x))).$$

This combined system is usually referred to as $\mathcal{A}^\omega$. It is of a power comparable to the first-order, two-sorted theory usually denoted by $\mathbf{Z_2}$ or "full second-order arithmetic". The Gödelian functional interpretation admits an extension to this system devised by Spector [84], involving a construction known as the bar-recursive functionals. We do not present it in detail, as we will not use it. What we need to note is that the theory that results when adding the bar recursion constants and axioms to **WE-HA**$^\omega$ is unsound, i.e. the set-theoretic model defined above does not satisfy it. This is not a problem if we only care about the reason this construction was originally introduced, i.e. the provably total functions of analysis, since no discrepancy appears if we restrict ourselves to functionals of type 1.

Still, for more intricate work regarding this extended interpretation, we should, for lack of a standard model, appropriate a satisfactory one. This is the model of strongly majorizable functionals, which was proven by Bezem [11] to satisfy bar recursion and is well-behaved enough for actual applications, as we shall see in the following section where an expanded version of it will be presented.

Such was the state of proof interpretations circa 1990, when modern proof mining was starting to emerge under Ulrich Kohlenbach. His first major set of contributions consisted in developing the **monotone** variants of modified realizability and the functional interpretation. We illustrate the second, which modifies the sentence in the conclusion of the soundness theorem to one of the following form (in the case when A has no free variables):

$$\exists x(t^* \text{ maj } x \wedge \forall y A_D(x;y)).$$

The relation involved in the above is the majorization relation introduced by Howard [34], defined as follows:

$$x^* \text{ maj}_0 \; x := x^* \geq x$$

$$x^* \text{ maj}_{\tau\rho} \; x := \forall y^* \forall y (y^* \text{ maj}_\rho \; y \to x^* y^* \text{ maj}_\tau \; xy)$$

Therefore, one may crudely say that this interpretation serves to extract upper bounds instead of exact witnesses. As it is usual in this endeavour, this loss is part of a trade-off, and the gain is being able to add more sentences than just the universal ones as unexamined lemmas, i.e. we now permit sentences of the general form:

$$\forall \underline{a}^{\underline{\delta}} \exists \underline{b}^{\underline{\sigma}} \preceq_{\underline{\sigma}} \underline{ra} \forall \underline{c}^{\underline{\gamma}} B_0(\underline{a}, \underline{b}, \underline{c}),$$

an example of this being the weak König lemma (for a full proof, see [46, pp. 149–154]).

2.3 The metatheorems of proof mining

Proof mining in its 1990s incarnation was able to exploit the theorems in the previous section in order to obtain quantitative information in theorems of pure real analysis. It was also able to treat more complex structures like separable Hilbert spaces by suitably encoding them in the way it is usually done in computable analysis. The question was, how to treat completely abstract structures like axiomatic Hilbert spaces? Usually (for example in model theory), these were formalized as structures of some fixed signature in a logic similar to first-order logic, having a similar status with models of arithmetical theories. In studying Hilbert spaces model-theoretically, however, the arithmetical facts were relegated to the metatheory. This is unacceptable when one wants to extract quantitative information from a proof and hence all its steps need to be completely formalized.

This is where the second major breakthrough of Kohlenbach comes into the picture. The idea is to have the same system formalizing **both** the arithmetical aspects and the ones relevant to the abstract situation at hand. The way to go is to extend the higher-typed theories from the previous sections with additional types for the abstract structures and additional well-behaved constants and axioms to formalize the specific theory. We will now deal with the case of normed spaces.

The set of types for this system, $\mathbf{T}^X$, will be generated by two "primitive" types, the type 0 of natural numbers and a new abstract type X, representing elements from our space, forming, as before, a free algebra with a single binary operation $\to$, representing function types. For such a type ρ, we define the type $\hat{\rho}$ by replacing all occurences of X in ρ by 0.

Definition 2.3.1. *Such a type is* **small** *if it is of the form*

$$\rho \underbrace{(0)\ldots(0)}_{n \text{ times}},$$

where $\rho \in \{0, X\}$ *and* $n \geq 0$.

Definition 2.3.2. *Such a type is* **admissible** *if it is of the form*

$$\rho(\tau_n)\dots(\tau_1),$$

where $\rho \in \{0, X\}$, $n \geq 0$ *and* τ_1, ..., τ_n *are all small.*

Clearly, all small types are admissible.

Also, we add new constants for the various operations common to normed spaces, i.e. 0_X and 1_X of type X, $+_X$ of type $X(X)(X)$, $-_X$ of type $X(X)$, $\cdot_X$ of type $X(X)(1)$ (where $1 = 0(0)$ is the type of real numbers) and $\|\cdot\|_X$ of type $1(X)$. We allow infix notation and the "syntactic sugar" of writing $x -_X y$ for $x +_X (-_X y)$. Finally, we add the following axioms:

1. the equational, and hence purely universal, axioms for vector spaces;
2. $\forall x^X (\|x -_X x\|_X =_{\mathbb{R}} 0_{\mathbb{R}})$;
3. $\forall x^X y^X (\|x -_X y\|_X =_{\mathbb{R}} \|y -_X x\|_X)$;
4. $\forall x^X y^X z^X (\|x -_X z\|_X \leq_{\mathbb{R}} \|x -_X y\|_X +_{\mathbb{R}} \|y -_X z\|_X)$;
5. $\forall \alpha^1 x^X y^X (\|\alpha x -_X \alpha y\|_X =_{\mathbb{R}} \|\alpha\|_{\mathbb{R}} \cdot_{\mathbb{R}} \|x -_X y\|_X$;
6. $\forall \alpha^1 \beta^1 x^X (\|\alpha x -_X \beta x\|_X =_{\mathbb{R}} |\alpha -_{\mathbb{R}} \beta|_{\mathbb{R}} \cdot_{\mathbb{R}} \|x\|_X$;
7. $\forall x^X \forall y^X \forall u^X \forall v^X (\|(x +_X y) -_X (u +_X v)\|_X \leq_{\mathbb{R}} \|x -_X u\|_X +_{\mathbb{R}} \|y -_X v\|_X)$;
8. $\forall x^X y^X (\|(-_X x) -_X (-_X y)\|_X =_{\mathbb{R}} \|x -_X y\|_X)$;
9. $\forall x^X y^X (|\|x\|_X -_{\mathbb{R}} \|y\|_X|_{\mathbb{R}} \leq_{\mathbb{R}} \|x -_X y\|_X)$;
10. $\|1_X\|_X =_{\mathbb{R}} 1_{\mathbb{R}}$.

Note that the equality relation $x^X =_X y^X$ which is necessarily used in the expression of the vector space axioms is syntactically defined as $\|x -_X y\|_X =_{\mathbb{R}} 0_{\mathbb{R}}$. We define the equality for higher types as in the system $\mathcal{A}^\omega$, as extensional equality reducible to $=_0$ and $=_X$.

An issue when adding new constant symbols is their extensionality – roughly, as the base system admits only a quantifier-free rule of extensionality, it is not clear that for a new function symbol f that is added to the system (e.g. $+_X$ or $-_X$ from the above) one can prove in the new system a statement of the form

$$\forall x_1 ... \forall x_n \forall y_1 ... \forall y_n (\bigwedge_i x_i = y_i \to f(x_1, ..., x_n) = f(y_1, ..., y_n))$$

Some axioms above, like the eighth one, are written in this way purely to minimize the effort in writing such an extensionality proof; the rest of them yield it more readily in their classical forms. The result is that all new function symbols are provably extensional. The last axiom is added solely to ensure the non-triviality of the formalized space.

In order to formalize the fact that the space is Banach, i.e. its completeness, the following is done (see [46, pp. 432-434]). We first note that the following operation on X-valued sequences is term-definable in the system (cf. the earlier treatment of real numbers):

$$\widehat{x_n} := \begin{cases} x_n, & \text{if, for all } k < n, [d_X(\widehat{x_k, x_{k+1}})](k+1) <_{\mathbb{Q}} 6 \cdot 2^{-k-1} \\ x_k, & \text{where } k < n \text{ is the least such that } [d_X(\widehat{x_k, x_{k+1}})](k+1) \geq_{\mathbb{Q}} 6 \cdot 2^{-k-1} \end{cases}$$

where we have used explicitly the encoding of reals as functions. The operation above transforms a sequence into a Cauchy one of prescribed rate 2^{-n+3}. We now add a new constant C of type $X(X(0))$, used to assign the limit to such sequences. This is enforced by the following additional axiom:

$$\forall x^{X(0)} \forall k^0 (d_X(C(x), \widehat{x_k}) \leq_{\mathbb{R}} 2^{-k+3}).$$

We have therefore obtained the system $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}]$, formalizing Banach spaces.

We say that a formula in our language is a $\forall$-formula (resp. an $\exists$-formula) iff it is formed by adjoining a list of universal (resp. existential) quantifiers over variables of admissible types to a quantifier-free formula.

Now, if $(X, \|\|)$ is a Banach space, we define a canonical associated set-theoretic model (extending the previously defined, purely arithmetical one) $\mathcal{S}^{\omega,X} = \{\mathcal{S}_\rho\}_{\rho \in \mathbf{T}^X}$ in all finite types by putting $\mathcal{S}_0 := \mathbb{N}$, $\mathcal{S}_X := X$ and $\mathcal{S}_{\tau(\rho)} := \mathcal{S}_\tau^{\mathcal{S}_\rho}$ (i.e. the set-theoretic Hom-set), assigning to any language constant its standard value, except for 1_X, which can take any value of norm 1 – this is why we said "a" set-theoretic model. Also, we say that a sentence of our logical language is modeled by such a pair $(X, \|\|)$ iff it is satisfied in the usual Tarskian sense by all the possible models associated to it – i.e., regardless of the exact value 1_X, which, however, makes the tenth axiom to be satisfied in this sense and so the formalized space is guaranteed to be non-trivial. (Of course, there must be here some care taken w.r.t. the issues with the representation of real numbers.)

As we said near the close of the previous section, there is another relevant model associated to this kind of logical system. In order to introduce it, we define, for each $\rho \in \mathbf{T}^X$, the majorization relations $\gtrsim_\rho$ **syntactically** (i.e. **not** as $\gtrsim_\rho \subseteq \mathcal{S}_{\widehat{\rho}} \times \mathcal{S}_\rho$), in an inductive way, as follows:

$$\begin{aligned} x^* \gtrsim_0 x &:\Leftrightarrow x^* \geq x \\ x^* \gtrsim_X x &:\Leftrightarrow x^* \geq \|x\| \\ x^* \gtrsim_{\tau(\rho)} x &:\Leftrightarrow \forall y^*, y(y^* \gtrsim_\rho y \to x^* y^* \gtrsim_\tau xy) \\ &\qquad \wedge \forall y^*, y(y^* \gtrsim_{\widehat{\rho}} y \to x^* y^* \gtrsim_{\widehat{\tau}} x^* y). \end{aligned}$$

We can now define the model of **hereditarily strongly majorizable functionals**, $\mathcal{M}^{\omega,X} = \{\mathcal{M}_\rho\}_{\rho \in \mathbf{T}^X}$, by:

$$\begin{aligned} \mathcal{M}_0 &:= \mathbb{N} \\ \mathcal{M}_X &:= X \\ \mathcal{M}_{\tau(\rho)} &:= \{x \in \mathcal{M}_\tau^{\mathcal{M}_\rho} \mid \text{exists } x^* \in \mathcal{M}_{\widehat{\tau}}^{\mathcal{M}_{\widehat{\rho}}} \text{ such that } x^* \gtrsim_{\tau(\rho)} x\} \end{aligned}$$

(and interpreting the relevant $\gtrsim_\rho$'s inside the $\mathcal{M}_\rho$'s that have already been constructed, therefore defining them simultaneously).

One of the main uses of this majorizable model arises from the fact that, unlike the standard model, it is a model of bar recursion, which, as we have seen, is needed in the current state of the art to interpret the principle of dependent choice. Therefore, the proof of the general logical metatheorems involves some constant juggling between the two models (see [46, pp. 421-428]). As a consequence, the kind of sentences that one may freely add as axioms will be restricted here not only by the logical complexity, but also by the types involved. Here we see how the admissible types come into play – for such a type ρ, it is the fact (see [28, Lemma 5.7]) that $\mathcal{M}_\rho \subseteq \mathcal{S}_\rho$. This justifies the following definition.

Definition 2.3.3. *We say that a formula in our system is a* **Δ-sentence** *if it is of the following form:*

$$\forall \underline{a}^{\underline{\delta}} \exists \underline{b}^{\underline{\sigma}} \preceq_{\underline{\sigma}} \underline{ra} \forall \underline{c}^{\underline{\gamma}} B_0(\underline{a}, \underline{b}, \underline{c}),$$

where underlined letters represent tuples of variables or types, B_0 is quantifier-free and devoid of any additional variables, $\underline{r}$ is a **closed** *term tuple of the appropriate type, $\underline{\delta}$, $\underline{\sigma}$, $\underline{\gamma}$ are tuples of admissible types, and $\preceq$ is syntactic sugar for the following family of binary relations:*

$$x \preceq_0 y :\equiv x \leq y$$
$$x \preceq_X y :\equiv \|x\| \leq \|y\|$$
$$x \preceq_{\tau(\rho)} y :\equiv \forall z^\rho (x(z) \preceq_\tau y(z))$$

Definition 2.3.4. *The* **Skolem normal form** *of a Δ-sentence written as above is:*

$$\exists \underline{B}^{\underline{\sigma}(\underline{\delta})} \preceq_{\underline{\sigma}(\underline{\delta})} \underline{r} \forall \underline{a}^{\underline{\delta}} \forall \underline{c}^{\underline{\gamma}} B_0(\underline{a}, \underline{Ba}, \underline{c})$$

Notation 2.3.5. *If Δ is a set of Δ-sentences, we denote by $\widetilde{\Delta}$ the set of the Skolem normal forms of the sentences in the set Δ.*

Theorem 2.3.6 ([28, Lemma 5.11])**.** *Let $(X, \|\|)$ be a Banach space, $\mathcal{S}^{\omega,X}$ and $\mathcal{M}^{\omega,X}$ be models associated with it as above. Let Δ be a set of Δ-sentences. Suppose that $\mathcal{S}^{\omega,X} \models \Delta$. Then $\mathcal{M}^{\omega,X} \models \widetilde{\Delta}$.*

The following result is a slight modification of [28, Theorem 5.13 and Corollary 5.14]. The original metatheorems on which it is based were developed in [45, 24].

Theorem 2.3.7 (Logical metatheorem for Banach spaces)**.**
Let $\rho \in \mathbf{T}^X$ be an admissible type. Let $B_\forall(x, u)$ be a $\forall$-formula with at most x, u free and $C_\exists(x, v)$ an $\exists$-formula with at most x, v free. Let Δ be a set of Δ-sentences. Suppose that:

$$\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}] + \Delta \vdash \forall x^\rho (\forall u^0 B_\forall(x, u) \rightarrow \exists v^0 C_\exists(x, v)).$$

Then one can extract a partial functional $\Phi : S_{\widehat{\rho}} \rightharpoonup \mathbb{N}$, whose restriction to the strongly majorizable functionals of $S_{\widehat{\rho}}$ is a bar-recursively computable functional of $\mathcal{M}^\omega$, such that for all Banach spaces $(X, \|\|)$ having the property that any associated set-theoretic model of it satisfies Δ, we have that for all $x \in S_\rho$ and $x^* \in S_{\widehat{\rho}}$ such that $x^* \gtrsim_\rho x$, the following holds:

$$\forall u \leq \Phi(x^*) B_\forall(x, u) \rightarrow \exists v \leq \Phi(x^*) C_\exists(x, v).$$

In addition:

1. *If $\widehat{\rho}$ is equal to* 1, *then* Φ *is total.*
2. *All variables may occur as finite tuples satisfying the same restrictions.*
3. *If the proof in the system above proceeds without the use of the axiom of dependent choice, one can use solely the set-theoretical model $\mathcal{S}^{\omega,X}$, without any restriction to the majorizable functionals, and* Φ *is then a total computable functional which is higher-order (i.e. in the sense of Gödel) primitive recursive. Also, the additional restriction imposed on* ρ *is no longer necessary.*

So, given the above metatheorem, in order to check that a theorem is suitable for bound extraction, what we need to do is to check that its proof is formalizable in our system and that its statement has the required complexity with regard to the logical form and the types involved.

Similar metatheorems have been developed for other structures. We give some examples below. (The goal of the final chapter is to show that the class of L^p Banach spaces is well-behaved enough in order to admit such a metatheorem.)

The theory of Hilbert spaces together with a nonempty unbounded convex subset, denoted by $\mathcal{A}^\omega[X, \langle\cdot,\cdot\rangle, C]_{-b}$, is obtained from the theory above by adding the new constants c_X of type X and χ_C of type $0(X)$ and the following axioms:

1. $\forall x^X y^X (\|x +_X y\|_X^2 +_{\mathbb{R}} \|x -_X y\|_X^2 =_{\mathbb{R}} 2_{\mathbb{R}} \cdot_{\mathbb{R}} (\|x\|_X^2 +_{\mathbb{R}} \|y\|_X^2)$ (the parallelogram law, which is a necessary and sufficient condition for a normed structure to arise from an inner product structure, the last of which being necessarily unique);
2. $\forall x^X y^X \alpha^1 (\chi_C(x) =_0 0 \wedge \chi_C(y) =_0 0 \to \chi_C((1 -_{\mathbb{R}} \tilde{\alpha}) \cdot_X x +_X \tilde{\alpha} \cdot_X y) =_0 0)$;
3. $\chi_C(c_X) =_0 0$;
4. $\forall x^X (\chi_C(x) \leq_0 1)$.

In the axioms above, $\tilde{\lambda}$ represents the construction from [46, Definition 4.24] where a canonical representation of a real number in the interval $[0,1]$ is computably and stably associated to any representation of an arbitrary real number.

We may add a similar set axioms in order to obtain a suitable system for CAT(0) spaces. The system $\mathcal{A}^\omega[X, d, W, \mathrm{CAT}(0)]$ (a similar system is detailed in [46, pp. 388-389]), is constructed by adding the constants d_X of type $1(X)(X)$ (representing the metric) and W_X of type $X(1)(X)(X)$ (representing the function yielding convex combinations) and a standard axiomatization consisting of the universal closures of the following axioms (where $x =_X y$ is an abbreviation for $d_X(x,y) =_{\mathbb{R}} 0_{\mathbb{R}}$):

1. $d_X(x,x) =_{\mathbb{R}} 0_{\mathbb{R}}$;
2. $d_X(x,y) =_{\mathbb{R}} d_X(y,x)$;
3. $d_X(x,z) \leq_{\mathbb{R}} d_X(x,y) + d_X(y,z)$;
4. $d_X(z, W_X(x,y,\lambda)) \leq_{\mathbb{R}} (1_{\mathbb{R}} -_{\mathbb{R}} \tilde{\lambda}) \cdot_{\mathbb{R}} d_X(z,x) +_{\mathbb{R}} \tilde{\lambda} \cdot_{\mathbb{R}} d_X(z,y)$;
5. $d_X(W_X(x,y,\lambda_1), W_X(x,y,\lambda_2)) =_{\mathbb{R}} |\tilde{\lambda}_1 -_{\mathbb{R}} \tilde{\lambda}_2|_{\mathbb{R}} \cdot_{\mathbb{R}} d_X(x,y)$;

6. $W_X(x, y, \lambda) =_{\mathbb{R}} W_X(y, x, (1_{\mathbb{R}} -_{\mathbb{R}} \lambda))$;
7. $d_X(W_X(x, z, \lambda), W_X(y, w, \lambda)) \leq_{\mathbb{R}} (1_{\mathbb{R}} -_{\mathbb{R}} \tilde{\lambda}) \cdot_{\mathbb{R}} d_X(x, y) +_{\mathbb{R}} \tilde{\lambda} \cdot_{\mathbb{R}} d_X(z, w)$;
8. $d_X(x, W_X(y_1, y_2, 1/2))^2 \leq_{\mathbb{R}} (1/2)d_X(x, y_1)^2 + (1/2)d_X(x, y_2)^2 - (1/4)d_X(y_1, y_2)^2$.

Gerhardy and Kohlenbach [23] have also developed a series of metatheorems for so-called "semi-intuitionistic" systems – that is, extensions of intuitionistic ("Heyting") arithmetic by several principles that can be characterized as non-constructive. The analogous system to $\mathcal{A}^\omega[X, \langle\cdot,\cdot\rangle, C]_{-b}$ will be denoted by $\mathcal{A}_i^\omega[X, \langle\cdot,\cdot\rangle, C]_{-b}$ and will be based, instead, on fully extensional Heyting arithmetic in all finite types together with the full axiom of choice. The rest of the construction is analogous. Since full extensionality is included, one can surmise from the previous section that only modified realizability is suitable as an underlying interpretation. The interesting addition is that the **monotone** variant is used, which, as for the functional intepretation, is able to accept more powerful principles as axioms. Therefore, instead of the independence of premises principle used in modified realizability, we consider here the more powerful schema of comprehension for negated formulas:

$$CA_\neg : \quad (\exists \Phi \leq \lambda x.1)(\forall y)(\Phi(y) =_0 0 \leftrightarrow \neg A(y))$$

Chapter 3

Studying pseudocontractions

This chapter is dedicated to the class of pseudocontractions, introduced in the first chapter. The results involved can be found in the papers [81, 82, 65]. We shall also provide some context, in addition to that supplied by the first two chapters. All the results in this chapter are original, except otherwise noted.

The first section focuses on finding a relationship between two different constants involved in the study of convergence theorems in smooth Banach spaces. The second section surveys the history of such convergence theorems and of the efforts to extract quantitative information from them. The next two sections attempt to elucidate the generalization of strict pseudocontractions to 2-uniformly smooth Banach spaces. Another two sections focus on the parallel algorithm, giving both a detailed outline of the existing results and a quantitative treatment of the paper of López and Xu [71]. Finally, the last two sections present the recent work [50] on Fejér monotone sequences in totally bounded metric spaces and a new application of it to the Ishikawa iteration in its original context [35], i.e. Lipschitz pseudocontractions.

3.1 Relating two constants associated to uniform smoothness

But first, before tackling the pseudocontractions themselves, we shall pursue a related problem.

In a recent paper [90], that we will further examine below, H. Y. Zhou used a variant[1] of a function previously introduced by Cholamjiak and Suantai [16] – namely, for any Banach space E with a Fréchet differentiable norm (and therefore smooth, so j can be used), one can define the function $\beta^*_E : E \times (0,\infty) \to \mathbb{R}$ by:

$$\beta^*_E(x,t) := \sup\left\{ \frac{\|x+tv\|^2 - \|x\|^2}{t} - 2j(x)(v) \;\middle|\; v \in S(E) \right\},$$

[1]The original definition in [16, Lemma 3.2] was:

$$\beta^*_E(x,t) := \sup\left\{ \left| \frac{\|x+tv\|^2 - \|x\|^2}{t} - 2j(x)(v) \right| \;\middle|\; v \in S(E) \right\}.$$

which would make Zhou's condition (3.1) unnecessarily stronger.

for any $x \in E$ and $t \in (0, \infty)$. The use of this function was to put the condition that there is a $d \in [1, \infty)$ such that for any $x \in E$ and $t \in (0, \infty)$

$$\beta^*_E(x, t) \leq dt, \tag{3.1}$$

a restriction that would delimit a sufficiently well-behaved class of Banach spaces such that Zhou's convergence theorem could be proven.

Our first goal is to see that this condition is actually equivalent to 2-uniform smoothness.

By convention, we will set for any Banach space E and any $x \in E$, $\beta^*_E(x, 0) := 0$. We first note that for any Hilbert space H, any $x \in H$ and any $t \geq 0$, $\beta^*_H(x, t) = t$. Indeed, in that case, for any $x \in H$, $t > 0$ and $v \in S(H)$, we have that:

$$\frac{\|x + tv\|^2 - \|x\|^2}{t} - 2j(x)(v) = \frac{\|x + tv\|^2 - \|x\|^2 - 2\langle x, tv\rangle}{t} = \frac{\|x + tv - x\|^2}{t} = t.$$

The characterization lemma is the following.

Lemma 3.1.1. *Let E be a smooth Banach space. The following statements are equivalent:*

1. *E is 2-uniformly smooth, i.e. there is a $c > 0$ such that for all τ, $\rho_E(\tau) \leq c\tau^2$;*
2. *there is a $d > 0$ such that for all $x, y \in E$ we have that $\|x + y\|^2 \leq \|x\|^2 + 2j(x)(y) + d\|y\|^2$;*
3. *there is a $d > 0$ such that for all $x \in E$ and $t \geq 0$ we have that $\beta^*_E(x, t) \leq dt$ (i.e., Zhou's condition (3.1)).*

Moreover, the constants in (ii) and (iii) may be taken to be the same (and hence we have used the same designator).

Proof. The equivalence between (i) and (ii) is given by [89, Corollary 1'].

Suppose now that (ii) holds. Let $x \in E$ and $t \geq 0$. If $t = 0$, there is nothing to prove, so suppose $t > 0$. Let v be in $S(E)$. Then, by (ii), setting $y := tv$, we get that:

$$\|x + tv\|^2 \leq \|x\|^2 + 2tj(x)(v) + dt^2,$$

so

$$\frac{\|x + tv\|^2 - \|x\|^2}{t} - 2j(x)(v) \leq dt,$$

from which, by taking the supremum, the conclusion follows.

Suppose now that (iii) holds. Let $x, y \in E$. Again, if $y = 0$, there is nothing to prove. If $y \neq 0$, set $v := \frac{1}{\|y\|} \cdot y$ and $t := \|y\|$, so $tv = y$. Then we get that:

$$\frac{\|x + y\|^2 - \|x\|^2}{\|y\|} - \frac{2}{\|y\|} \cdot j(x)(y) \leq \beta^*_E(x, \|y\|) \leq d\|y\|,$$

and by multiplying by $\|y\|$ we obtain our desired result. □

We have therefore established the equivalence of Zhou's condition with 2-uniform smoothness, which was only mentioned as a special case in [90, p. 762]. We note that it is immediate that if d satisfies the two equivalent conditions, and $d \leq d'$, then d' also satisfies the condition. Hence we can always take $d \geq 1$.

We shall now sketch a way to compute the constant d – and that specific choice of d for a given c (whose value will be seen to be already greater than 1) will be denoted by d_c in the following sections.

Let E be a 2-uniformly smooth Banach space and let, therefore, $c > 0$ be such that $\rho_E(\tau) \leq c\tau^2$, for all τ. We shall use Lindenstrauss's classical formula from [68]:

$$\rho_E(t) = \sup_{\varepsilon \in [0,2]} \left(\frac{1}{2}\varepsilon t - \delta_{E^*}(\varepsilon) \right).$$

So for any suitable ε and t, we get that:

$$\delta_{E^*}(\varepsilon) \geq \frac{1}{2}\varepsilon t - ct^2.$$

The term on the right hand side is a quadratic function of t, which has as its maximum the value $\frac{\varepsilon^2}{16c}$. So for all ε, we have that:

$$\delta_{E^*}(\varepsilon) \geq \frac{1}{16c} \cdot \varepsilon^2.$$

Lemma 1 from [20] states that in this case, for any $x, y \in E^*$ with

$$\|x\|^2 + \|y\|^2 = 2$$

the following holds:

$$\frac{1}{2}\|x+y\| \leq 1 - k_1 \left(\frac{1}{2}\|x-y\| \right)^2,$$

where k_1 is given by the minimum of $\frac{1}{16c}$ and α, the constant α being such that

$$\sup_{t \in (0,1]} \frac{\sqrt{2-(1-t)^2} - 1 - t}{t^2} = -\alpha < 0.$$

The function to be maximized here can be seen to be equal to $\frac{-2}{t+1+\sqrt{2-(1-t)^2}}$, which increases along with its obviously increasing denominator. So the original function attains its maximum at $t := 1$, the maximum being $\sqrt{2} - 2$. The value of α is therefore $2 - \sqrt{2}$.

We denote by $L_2(E^*)$ the space of all functions $f : [0,1] \to E^*$ such that:

$$\int_0^1 \|f(t)\|^2 dt < \infty.$$

By Proposition 1 and "Added in proof" of [20], we get that for all $\varepsilon \in [0,2]$,

$$\delta_{L_2(E^*)}(\varepsilon) \geq k_2 \varepsilon^2,$$

where k_2 is computed by

$$k_2 := 2^{-2} \cdot 2^{-1} \cdot \min(k_1, 1) = \frac{\min\left(\frac{1}{16c}, 2 - \sqrt{2}\right)}{8}.$$

By the statement and proof of Lemma 2.1 from [77], we get that for all $x, y \in E^*$ and all $t \in (0,1)$,

$$\|tx + (1-t)y\|^2 \leq t\|x\|^2 + (1-t)\|y\|^2 - k_2 t(1-t)\|x-y\|^2,$$

which corresponds to equation (3.1) from [89].

Then, using the proof of the implication (i) $\Rightarrow$ (ii) from [89, Corollary 1], we get that for all $x, y \in E^*$ and all $f \in J(x)$,

$$\|x+y\|^2 \geq \|x\|^2 + 2f(y) + k_2\|y\|^2.$$

Finally, from the first lines of the proof of [89, Theorem 1'], we obtain that for all $x, y \in E$,

$$\|x+y\|^2 \leq \|x\|^2 + 2j(x)(y) + k_2^{-1}\|y\|^2.$$

We can therefore set

$$d_c := k_2^{-1} = \frac{8}{\min\left(\frac{1}{16c}, 2-\sqrt{2}\right)}. \tag{3.2}$$

We note that this bound is by no means an optimal one – we saw that for a Hilbert space one can simply take $d := 1$, whereas the formula would give $d_c := 64$ (using $c := \frac{1}{2}$, taken from the usual modulus of smoothness $\rho(\tau) := \sqrt{1+\tau^2} - 1 \leq \frac{\tau^2}{2}$). Still, the above argument shows that there is a simple method one can use to readily obtain a suitable $d \geq 1$ given the original smoothness constant c.

3.2 Past work in convergence theorems and quantitative information

We may now return to our problem of analysing proofs of convergence theorems. First, we speak a bit about the history and motivation of such results.

It is well-known that the classical method of Picard iterations, used to find the unique fixed point of a contraction, fails in the case of nonexpansive mappings. The first solution that has been found is the following scheme.

Definition 3.2.1. *Let E be a Banach space and $C \subseteq E$ a convex set. If $T : C \to C$, $x \in C$ and $t \in (0,1)$, then the* **Krasnoselski iteration corresponding to T, x and t** *is the unique sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_0 = x$ and for any $n \in \mathbb{N}$,*

$$x_{n+1} = (1-t)x_n + tTx_n.$$

The Krasnoselski iteration was first considered in 1955 by Krasnoselski [55] for the case $t = \frac{1}{2}$ and in 1957 by Schaefer [80] for a general t. It suits the purpose above, as shown e.g. by the following result.

Theorem 3.2.2 (Browder and Petryshyn (1967), [13, Theorem 8])**.** *Let H be a Hilbert space, $C \subseteq H$ a convex set and $T : C \to C$. Suppose that T is nonexpansive and C is closed and bounded. Let $x \in C$ and $t \in (0,1)$. Then the Krasnoselski iteration corresponding to T, x and t weakly converges to a fixed point of T.*

Later, the following more general iterative schema was found.

Definition 3.2.3. *Let E be a Banach space and $C \subseteq E$ a convex set. If $T : C \to C$, $x \in C$ and $(t_n)_{n\in\mathbb{N}} \subseteq (0,1)$, then the* **(Krasnoselski-)Mann iteration corresponding to T, x and $(t_n)_{n\in\mathbb{N}}$** *is the unique sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_0 = x$ and for any $n \in \mathbb{N}$, $x_{n+1} = T_{t_n}(x_n)$.*

Let t be in $(0,1)$. By letting $t_n = t$ for all $n \in \mathbb{N}$, we get the Krasnoselski iteration as a special case.

The following is the most useful known generalization of Theorem 3.2.2 for nonexpansive mappings.

Theorem 3.2.4 (Reich (1979), [78, Theorem 2])**.** *Let E be a uniformly convex Banach space with a Fréchet differentiable norm, $C \subseteq E$ a convex set and $T : C \to C$. Suppose that T is nonexpansive with $Fix(T) \neq \emptyset$ and that C is closed. Let $x \in C$ and $(t_n)_{n\in\mathbb{N}} \subseteq (0,1)$ such that*

$$\sum_{n=0}^{\infty} t_n(1-t_n) = \infty$$

Then the Mann iteration corresponding to T, x and $(t_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of T.

Efforts to extend this scheme to more general maps like pseudocontractions were not successful. Later, Chidume and Mutangadura [15] would exhibit an example of a Lipschitzian pseudocontractive map in $\mathbb{R}^2$ with a unique fixed point for which no Mann sequence converges.

Meanwhile, some alternate algorithms were proposed, the first of which being the one of Ishikawa [35], who deployed it successfully in the case of Lipschitzian pseudocontractions acting on a compact convex subset of a Hilbert space. It is defined as follows.

Definition 3.2.5. *Let E be a Banach space and $C \subseteq E$ a convex set. If $T : C \to C$, $x \in C$ and $(t_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}} \subseteq (0,1)$, then the* **Ishikawa iteration corresponding to T, x, $(t_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$** *is the unique sequence $(x_n)_{n\in\mathbb{N}}$ such that $x_0 = x$ and for any $n \in \mathbb{N}$,*

$$x_{n+1} = (1-t_n)x_n + t_n T(((1-s_n)x_n + s_n T x_n).$$

By letting $s_n = 0$ for all $n \in \mathbb{N}$, we get the Mann iteration as a special case.

What Ishikawa actually proved is the following.

Theorem 3.2.6 (Ishikawa (1974), [35])**.** *Let H be a Hilbert space and $C \subseteq H$ be a convex, compact subset. Let $T : C \to C$ be a Lipschitzian pseudocontraction. Let x be in C and $(t_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ be in $(0,1)$ such that*

1. *for all $n \in \mathbb{N}$, $t_n \leq s_n$;*
2. *$\sum_{n=0}^{\infty} t_n s_n = \infty$;*
3. *$\lim_{n\to\infty} s_n = 0$.*

Then the Ishikawa iteration corresponding to T, x, $(t_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ (strongly) converges to a fixed point of T.

As pointed out in [35], an example of a pair of sequences satisfying all three conditions is $t_n = s_n = \frac{1}{\sqrt{n+1}}$. Note also that Ishikawa, in the above result, does not assume *a priori* the existence of fixed points for T – this follows because of the compactness assumption of C, by an application of the theorem of Schauder.

In the first chapter, we saw that the class of k-strict pseudocontractions, for $k \in [0, 1)$, lies between nonexpansive and general pseudocontractive mappings (and, in particular, they are Lipschitzian functions). For these mappings, for reasons that will become obvious in the next section, the search for convergence theorem had more success. Even in the same paper where Theorem 3.2.2 above originated, the next result was proven.

Theorem 3.2.7 (Browder and Petryshyn (1967), [13, Theorem 12])**.** *Let H be a Hilbert space, $C \subseteq H$ a convex, closed, bounded set, k be in $(0, 1)$. Let $T : C \to C$ be a k-strict pseudocontraction. Let $x \in C$ and $t \in (0, 1-k)$.*

Then the Krasnoselski iteration corresponding to T, x and t weakly converges to a fixed point of T.

Still in the context of Hilbert spaces, the generalization of the above to the Mann iteration (using the same condition on the weights as Reich's) was proven only in 2007.

Theorem 3.2.8 (Marino and Xu (2007), [75, Theorem 3.1])**.** *Let H be a Hilbert space, $C \subseteq H$ a convex, closed set, k be in $(0, 1)$. Let $T : C \to C$ be a k-strict pseudocontraction with $Fix(T) \neq \emptyset$. Let $x \in C$ and $(t_n)_{n\in\mathbb{N}} \subseteq (0, 1-k)$ such that*

$$\sum_{n=0}^{\infty} t_n(1-k-t_n) = \infty$$

Then the Mann iteration corresponding to T, x, $(t_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of T.

So far, this section has been complementary to the first chapter, in the sense of specifying more specialized results of analysis, results of the kind that the program of proof mining has analyzed in the last twenty years. It is now the time to turn towards complementing the second chapter, i.e. see how the metatheorems that we have presented brought forward applications in this field.

The kind of information that proof mining is able to extract, consists, in order of prominence, of: witnesses and bounds for existentially quantified variables; weakening of premises of a certain kind; the extracted quantities' independence of certain parameters. The canonical example of an existentially quantified variable in ordinary mathematics comes from the definition of the limit of a sequence in a metric space, i.e.

$$\forall k \in \mathbb{N}\, \exists N \in \mathbb{N}\, \forall n \geq N \left(d(x_n, x) \leq \frac{1}{k+1} \right).$$

A witness for this existentially quantified N, also called *rate of convergence* for the sequence, as it will be defined in more detail further below, would consist of a formula giving it in terms of the k. Unfortunately, as it can be surmised from the general metatheorems, e.g. for Banach spaces, that we enuntiated in the last chapter, as the sentence above has three alternating quantifiers in a row

($\forall\exists\forall$), one cannot generally extract such a computable formula if the proof is non-constructive in the sense of using at least once the law of excluded middle (one can show that the existence of a general procedure for these cases would contradict the impossibility of the halting problem). Four avenues have generally been tried so far in proof mining, if the convergence of a sequence was under discussion. The first one is the extraction of the full rate of convergence in the rare case that the proof is fully or at least partially constructive. The second one is to settle for a weaker property, like the limit inferior, which may have a tractable $\forall\exists$ form (and if the sequence nonincreasingly tends to 0, the extracted modulus of liminf would also be a rate of convergence). The third one is to use some uniqueness properties of the limit in order to extract the rate of convergence from a distantly related property like the rate of asymptotic regularity. This technique was pioneered by Kohlenbach in his PhD thesis [41]. Finally, the fourth way consists of replacing the convergence of the sequence with an equivalent formulation (identifiable in logic as its Herbrand normal form), introduced by Tao [85, 86] under the name of *metastability*. The following sentence expresses the metastability of the sequence above:

$$\forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists N \in \mathbb{N}\, \forall i, j \in [N, N + g(N)] \quad \left(d(x_i, x_j) \leq \frac{1}{k+1} \right).$$

It is immediately seen that this sentence is of a reduced ($\forall\exists$) logical complexity. It is, however, a simple exercise, to check that it is classically (but not intuitionistically) equivalent to the assertion that the sequence under discussion is Cauchy. Therefore, one can now say that the fourth way is focused on obtaining a *rate of metastability* for the sequence, i.e. a formula giving a bound on the N in terms of the k and the g.

We have mentioned above the rate of asymptotic regularity, and this phrase begs for a definition. Given that the theorems above yield the (weak or strong) convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ to a fixed point of an operator T, one generally has that

$$\lim_{n\to\infty} d(x_n, Tx_n) = 0,$$

i.e. the sequence is T*-asymptotically regular*. This property is frequently one of the first steps done to prove the actual convergence of the sequence, and therefore the unwinding of its proof is inevitable. As it is in its turn a sentence that asserts the convergence of a sequence (this time a real-valued one), it falls under the same case-by-case discussion in the paragraph above. Still, one is bound to remark that this easier proof frequently falls into a case where the extraction of the full rate of convergence will be possible, and this rate will be called the *rate of (*T*-)asymptotic regularity* for the original sequence. We now give an example.

Kohlenbach has computed, using proof mining techniques, in [44, Theorem 3.4], a rate of T-asymptotic regularity for the Mann iteration of the conclusion of Theorem 3.2.4 of Reich. Before giving the rate, let us explain how it should look like as a formula, i.e. what it would depend on. First, we have a valid modulus η of uniform convexity for the space. Since we have supposed the existence of a fixed point, one of the parameters is a bound on the minimum distance between the initial point x and a fixed point of T. Finally, since the divergence of the series

$$\sum_{n=0}^{\infty} t_n(1 - t_n)$$

is presupposed, one might add in as a parameter a *rate of divergence* for it, i.e. a function $\theta : \mathbb{N} \to \mathbb{N}$

such that for any N,

$$\sum_{n=0}^{\theta(N)} t_n(1-t_n) \geq N.$$

(The exact reasons for adding these parameters and not others, e.g. like the sequence (t_n) itself, will become obvious in two sections, when we will discuss a complete extraction that we have carried out for a different sort of iteration.)

That being said, the rate of asymptotic regularity is the function $h^{(1)}_{b,\theta,\eta} : (0,\infty) \to \mathbb{N}$, defined for any $\varepsilon > 0$ by

$$h^{(1)}_{b,\theta,\eta}(\varepsilon) := \theta\left(\left\lceil \frac{3(b+1)}{2\varepsilon \cdot \eta(\frac{\varepsilon}{b+1})} \right\rceil\right),$$

i.e. we have that for any $\varepsilon > 0$ and any $n \geq h(\varepsilon)$,

$$d(x_n, Tx_n) \leq \varepsilon.$$

Moreover, using the remarks from the statement of the theorem in [44] it can be shown that in the case of Hilbert spaces, which have the well-known modulus of uniform convexity of $\eta(\varepsilon) := \frac{\varepsilon^2}{8}$, one can simplify the above rate to:

$$h^{(2)}_{b,\theta}(\varepsilon) := \theta\left(\left\lceil \frac{4(b+1)}{\varepsilon^2} \right\rceil\right). \tag{3.3}$$

Like the rate of divergence above, there are other possible pieces of quantitative information that one should take into account when analyzing a proof in nonlinear analysis. For example, if $(a_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative real numbers, a *modulus of liminf* of (a_n) is a mapping $\Delta : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ satisfying

$$\forall l \in \mathbb{N}\ \forall k \in \mathbb{N}\ \exists N \in [l, \Delta(l,k)] \quad \left(a_N \leq \frac{1}{k+1}\right).$$

One can easily see that $\liminf_{n\to\infty} a_n = 0$ if and only if (a_n) has a modulus of liminf.

In the situation where the nonnegative sequence is of the form $(d(x_n, Tx_n))$, we are often interested in a map $\Phi : \mathbb{N} \to \mathbb{N}$ such that:

$$\forall k \in \mathbb{N}\, \exists N \leq \Phi(k) \quad \left(d(x_n, Tx_n) \leq \frac{1}{k+1}\right).$$

It is clear that such a map may be obtained from a modulus of liminf of $(d(x_n, Tx_n))$ by setting $l := 0$. Since its existence indicates that the elements of the sequence (x_n) come arbitrarily close to being fixed points of the operator T, Φ is called an *approximate fixed point bound* for (x_n) with respect to T.

More recently, Ivan and Leuştean [36, Theorem 2.1] have computed a rate of asymptotic regularity for the Mann iteration in the conclusion of Theorem 3.2.8, of Marino and Xu, expressed by:

$$h^{(5)}_{b,\theta}(\varepsilon) := \theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right), \tag{3.4}$$

where θ is a rate of divergence, instead, for

$$\sum_{n=0}^{\infty} t_n(1-k-t_n).$$

3.3 Defining pseudocontractions in Banach spaces

Marino and Xu put forward in their paper as an open problem whether their result can be generalized to uniformly convex Banach space with a Fréchet differentiable norm, in the same vein as Reich's. Since then, various authors have tried to solve this problem to some degree.

The first problem is to define exactly what a k-strict pseudocontraction means in this context. The condition that we have given in the first chapter, i.e. that for any x, y in the space, we have that

$$j(x-y)((x-Tx)-(y-Ty)) \geq \frac{1-k}{2}\|(x-Tx)-(y-Ty)\|^2,$$

was the one used by Zhou [90] to prove the following theorem:

Theorem 3.3.1 (cf. [90, Theorem 3.1])**.** *Let E be a uniformly convex Banach space which is also uniformly smooth, $C \subseteq E$ a convex, closed set and $T : C \to C$. Let $d \in [1, \infty)$ be such that for any $x \in E$ and $t \in (0, \infty)$, condition (3.1) holds.*

Let k be in $(0, 1)$ and suppose that T is a k-strict pseudocontraction with $Fix(T) \neq \emptyset$. Let $x \in C$ and $(t_n)_{n\in\mathbb{N}} \subseteq (0, \frac{1-k}{2d})$ such that

$$\sum_{n=0}^{\infty} t_n = \infty.$$

Then the Mann iteration corresponding to T, x and $(t_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of T.

As we have seen, the additional condition (3.1) imposed on the Banach space is actually equivalent to 2-uniform smoothness. Now, another careful examination, this time of the definition of a k-strict pseudocontraction, will allow us to give simpler and immediate proofs of both Theorem 3.2.8 and Theorem 3.3.1.

For a given self-mapping of a convex set, $T : C \to C$, and a $t \in (0, 1)$, set $T_t := tT + (1-t)id_C$ – that is, for all $x \in C$, $T_t x = tTx + (1-t)x$. It is immediate that for all $t_1, t_2 \in (0, 1)$, $(T_{t_1})_{t_2} = T_{t_1 \cdot t_2}$. Also note that, for any t, T and T_t have the same fixed points. We now look at the following lemma, which was the one originally used to prove Theorem 3.2.7 above.

Lemma 3.3.2 (Browder and Petryshyn (1967), [13, Theorem 2])**.** *Let H be a Hilbert space and $C \subseteq H$ a convex set. If $T : C \to H$ is a mapping, T is a strict pseudocontraction iff there is a $t \in (0, 1)$ such that T_t is nonexpansive. More precisely, if T is a k-strict pseudocontraction, then for* **any** *$t \in (0, 1-k]$, T_t is nonexpansive.*

Our observation is that this result readily generalizes to 2-uniformly smooth Banach spaces and the above definition of a k-strict pseudocontraction.

Lemma 3.3.3. *Let E be a Banach space, $C \subseteq E$ a convex subset and $d \geq 1$ such that for any $x \in E$ and $t \geq 0$, $\beta^*_E(x, t) \leq dt$. Let $k \in (0, 1)$ and $T : C \to C$ a k-strict pseudocontraction.*

Let $t \in \left(0, \frac{1-k}{d}\right]$. Then T_t is nonexpansive. (In particular, $T_{\frac{1-k}{d}}$ is nonexpansive.)

Proof. Since $t \leq \frac{1-k}{d}$, we have that $dt - (1-k) \leq 0$.

Let $x, y \in E$. We have that:

$$\begin{aligned}
&\|T_t x - T_t y\|^2 \\
&= \|tTx + (1-t)x - tTy - (1-t)y\|^2 \\
&= \|(x-y) + (-t)((x-Tx) - (y-Ty))\|^2 \\
&\leq \|x-y\|^2 - 2tj(x-y)((x-Tx)-(y-Ty)) + dt^2\|(x-Tx)-(y-Ty)\|^2 \\
&\leq \|x-y\|^2 - t(1-k)\|(x-Tx)-(y-Ty)\|^2 + dt^2\|(x-Tx)-(y-Ty)\|^2 \\
&= \|x-y\|^2 + t(dt-(1-k))\|(x-Tx)-(y-Ty)\|^2 \\
&\leq \|x-y\|^2.
\end{aligned}$$

□

Using this, we may prove the following general convergence result.

Theorem 3.3.4. *Let E be a uniformly convex Banach space which is also 2-uniformly smooth, $C \subseteq E$ a convex, closed set and $T : C \to C$. Let, therefore, $d \geq 1$ be a constant satisfying conditions (ii) and (iii) from Lemma 3.1.1 (if, for example, $\rho_E(\tau) \leq c\tau^2$, for all τ, take $d := d_c$). Let k be in $(0,1)$ and suppose that T is a k-strict pseudocontraction with $Fix(T) \neq \emptyset$. Let $x \in C$ and $(t_n)_{n\in\mathbb{N}} \subseteq \left(0, \frac{1-k}{d}\right)$ such that*

$$\sum_{n=0}^{\infty} t_n \left(\frac{1-k}{d} - t_n\right) = \infty$$

Then the Mann iteration corresponding to T, x and $(t_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of T.

Proof. By Lemma 3.3.3, we have that $T_{\frac{1-k}{d}}$ is nonexpansive. For every $n \geq 0$, set $t'_n := t_n \cdot \frac{d}{1-k}$. Denote by $(x_n)_{n\in\mathbb{N}}$ the Mann iteration corresponding to T, x and $(t_n)_{n\in\mathbb{N}}$. Let $n \geq 0$. We have that:

$$\begin{aligned}
x_{n+1} &= t_n T x_n + (1-t_n)x_n \\
&= T_{t_n} x_n \\
&= T_{t'_n \cdot \frac{1-k}{d}} x_n \\
&= T_{t'_n}(T_{\frac{1-k}{d}} x_n) \\
&= t'_n T_{\frac{1-k}{d}} x_n + (1-t'_n)x_n.
\end{aligned}$$

We have then, that $(x_n)_{n\in\mathbb{N}}$ is the Mann iteration corresponding to $T_{\frac{1-k}{d}}$, x and $(t'_n)_{n\in\mathbb{N}}$. We seek to apply Theorem 3.2.4. For that we do the following verification:

$$\sum_{n=0}^{\infty} t'_n(1-t'_n) = \sum_{n=0}^{\infty} t_n \cdot \frac{d}{1-k}\left(1 - t_n \cdot \frac{d}{1-k}\right) = \left(\frac{d}{1-k}\right)^2 \sum_{n=0}^{\infty} t_n \left(\frac{1-k}{d} - t_n\right) = \infty.$$

We therefore get that $(x_n)_{n\in\mathbb{N}}$ weakly converges to a fixed point of $T_{\frac{1-k}{d}}$, which is also a fixed point of T. □

We may now achieve our stated goal of finding simpler proofs for the existing convergence theorems.

Proof of Theorem 3.3.1. Note that the hypothesis states that $(t_n)_{n\in\mathbb{N}} \subseteq (0, \frac{1-k}{2d})$. Then, for all n, $\frac{1-k}{d} - t_n \geq \frac{1-k}{2d}$, so:

$$\sum_{n=0}^{\infty} t_n \left(\frac{1-k}{d} - t_n\right) \geq \frac{1-k}{2d} \sum_{n=0}^{\infty} t_n = \infty.$$

We are therefore in the hypothesis of Theorem 3.3.4. □

Proof of Theorem 3.2.8. We have shown in the beginning of this section that one can take for a Hilbert space $d := 1$. The conclusion immediately follows. □

Note that for directly establishing the truth of Theorem 3.2.8, we would not have needed the full power of our lemmas, but only Lemma 3.3.2 itself. Also take into consideration, as a reviewer (of the paper in which these results appear) pointed out, that our results above can be considered special cases of the results of [17], specifically Lemma 3 and Theorem 12 (modulo different notations). Still, we feel that the above exposition, not being burdened by the extraneous details of the cyclic algorithm considered in [17], can shed light on the way in which earlier results on strict pseudocontractions, like the one in [75], can be deduced from older results like Reich's without rebuilding from the ground up the whole machinery on demiclosedness and monotonicity.

3.4 Quantitative information on pseudocontractions in Banach spaces

The proof of Theorem 3.3.4 shows that if $\theta : \mathbb{N} \to \mathbb{N}$ is a rate of divergence for the series

$$\sum_{n=0}^{\infty} t_n \left(\frac{1-k}{d} - t_n\right),$$

then, since $\left\lceil \left(\frac{1-k}{d}\right) \right\rceil = 1$ (remember that we always took $d \geq 1$), the same $\theta : \mathbb{N} \to \mathbb{N}$ is also a rate of divergence for the series

$$\sum_{n=0}^{\infty} t'_n(1 - t'_n).$$

In addition, an easy computation gives us that:

$$\|x_n - T_{\frac{1-k}{d}} x_n\| = \frac{1-k}{d} \|x_n - Tx_n\|.$$

We can, therefore, give the following result:

Theorem 3.4.1. *A rate of T-asymptotic regularity for the sequence in Theorem 3.3.4 is given by:*

$$h^{(3)}_{b,\theta,\eta}(\varepsilon) := \theta\left(\left\lceil \frac{3(b+1)d}{2\varepsilon(1-k)\cdot\eta\left(\frac{\varepsilon(1-k)}{(b+1)d}\right)} \right\rceil\right),$$

where b, η are as before and θ is a rate of divergence for the series

$$\sum_{n=0}^{\infty} t_n \left(\frac{1-k}{d} - t_n \right).$$

Proof. Applying Kohlenbach's result on the Mann iteration that we cited earlier and the remark on θ' from above, we get that for any $n \geq h^{(3)}_{b,\psi,\eta}(\varepsilon)$,

$$\|x_n - T_{\frac{1-k}{d}} x_n\| \leq \frac{\varepsilon(1-k)}{d}.$$

Now, by the computation before, we get that:

$$\|x_n - Tx_n\| \leq \frac{d}{1-k} \cdot \frac{\varepsilon(1-k)}{d} = \varepsilon.$$

□

Applying the same treatment to the Hilbert rate from (3.3), we get that a rate of T-asymptotic regularity in the case of Theorem 3.2.8 is:

$$h^{(4)}_{b,\theta}(\varepsilon) := \theta \left(\left\lceil \frac{4(b+1)}{(1-k)^2 \varepsilon^2} \right\rceil \right),$$

a quadratic rate not unlike the $h^{(5)}_{b,\theta}(\varepsilon) = \theta \left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil \right)$ one, obtained also with proof mining techniques in [36], for the same iteration.

3.5 General discussion on the parallel algorithm

An area of investigation closely related to the kind of iterative algorithms mentioned above has been the problem of finding a common fixed point of a (finite or infinite) family $(T_i)_i$ of self-mappings of a subset C like above. An iterative scheme that is useful in the case of a finite family $(T_i)_{1 \leq i \leq N}$ is the *parallel algorithm*, defined as follows. Let x be in C and $(t_n)_{n \in \mathbb{N}} \subseteq (0,1)$. For each $i \in \{1, \ldots, N\}$, let $(\lambda^{(n)}_i)_{n \in \mathbb{N}}$ be a sequence of positive real numbers such that, for any $n \in \mathbb{N}$:

$$\sum_{i=1}^{N} \lambda^{(n)}_i = 1.$$

Write, for all $n \geq 0$:

$$A_n := \sum_{i=1}^{N} \lambda^{(n)}_i T_i.$$

Let $(x_n)_{n \in \mathbb{N}}$ be the sequence defined by:

$$x_0 := x,$$

$$x_{n+1} := t_n x_n + (1-t_n) A_n x_n.$$

Then the sequence $(x_n)_{n \in \mathbb{N}}$ is the output of the parallel algorithm associated with the inputs T, x, (t_n) and $(\lambda^{(n)}_i)$.

Two remarks are in order here. Firstly, we see that the case $N = 1$ represents the already defined Mann iteration for finding a fixed point of a self-mapping and therefore, all the results pertaining to this algorithm immediately transfer to the case of a single mapping (i.e. the one treated in the last section and in [36]). Secondly, we now have to note that there exists another (equivalent) convention when working with Mann-like algorithms, pairing the t_n with the application of the appropriate mapping, i.e. the formula above would be:

$$x_{n+1} := t_n A_n x_n + (1 - t_n)x_n$$

We use the "$t_n x_n$" convention, in the description of the parallel algorithm and also further below, when formalizing the passage from nonexpansive to strictly pseudocontractive mappings, as it is the one used in [75, 71], in contrast to the convention used in the previous sections. One should be careful to check the convention used when comparing different hypotheses and convergence results.

When considering algorithms for finite families such as the one above, the intermediate result obtained during the proof of the convergence theorem will be still one of "asymptotic regularity", though one pertaining to the map(s) constructed as a byproduct of the algorithm (here, the A_n's), i.e. that:

$$\lim_{n\to\infty} \|x_n - A_n x_n\| = 0.$$

Given that A_n varies with n, such a result does not mean *a priori* that $(x_n)_n$ is an approximate fixed point sequence for any mapping – certainly not one given by the problem data. Therefore, what is actually relevant to the proof mining program is an asymptotic regularity related to the relevant mappings of the problem – that is, the T_i's. One might look for an associated rate of convergence for the statements:

$$\lim_{n\to\infty} \|x_n - T_i x_n\| = 0,$$

for each $i \in \{1, ..., N\}$. A concrete extraction of such a rate can be found, for example, in [40], for the Kuhfittig iteration.

López-Acedo and Xu, in 2007, have found sufficient conditions so that the parallel algorithm weakly converges to a fixed point of a finite family of strictly pseudocontractive mappings. Their result, using the notations introduced above, is expressed as follows.

Theorem 3.5.1 (López-Acedo and Xu (2007), [71, Theorem 3.3])**.** *Let H be a Hilbert space and $C \subseteq H$ a closed, convex set. Let $N \geq 1$, $(T_i : C \to C)_{1\leq i\leq N}$ a family of mappings and $(k_i)_{1\leq i\leq N} \subseteq (0, 1)$ such that each T_i is a k_i-strict pseudocontraction. Suppose that $\bigcap_{i=1}^N Fix(T_i) \neq \emptyset$. Set $k := \max_{1\leq i\leq N} k_i$. Let x be in C, $(t_n)_{n\in\mathbb{N}} \subseteq [k, 1]$ be such that*

$$\sum_{n=0}^{\infty} (t_n - k)(1 - t_n) = \infty.$$

Impose the conditions

$$\inf_{i,n} \lambda_i^{(n)} > 0$$

and

$$\sum_{j=0}^{\infty} \sqrt{\sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}|} < \infty$$

on $(\lambda_i^{(n)})$. Then the parallel algorithm associated with the inputs T, x, (t_n) and $(\lambda_i^{(n)})$ weakly converges to a common fixed point of the family $(T_i : C \to C)_{1\leq i\leq N}$.

Our goal will be to obtain rates of asymptotic regularity for this instance of the parallel algorithm – that is, a rate of convergence for each sequence $(\|x_n - T_i x_n\|)_{n\in\mathbb{N}}$, with the sequence $(x_n)_{n\in\mathbb{N}}$ being defined as before. As this is the first time that we carry out a complete extraction without making explicit use of similar past work, we will discuss in more detailed way the logical background of our results.

The following theorem is an adaptation of one by Gerhardy and Kohlenbach (see [24] and [46, Corollary 17.71]).

Theorem 3.5.2. *Let P be $\mathbb{N}$, $\mathbb{N}^{\mathbb{N}}$ or $\mathbb{N}^{\mathbb{N}\times\mathbb{N}}$, K an $\mathcal{A}^\omega$-definable compact metric space and ρ an admissible type, $B_\forall(u,y,z,f,n)$ a $\forall$-formula and $C_\exists(u,y,z,f,N)$ an $\exists$-formula (we assume that the free variables of the two formulas are at most the variables written as their arguments, such that the formula below shall be a sentence). Suppose that $\mathcal{A}^\omega[X,\langle\cdot,\cdot\rangle,C]_{-b}$ proves that:*

$$\forall u \in P \forall M \in \mathbb{N} \forall k \in \left[0, 1-\frac{1}{M+1}\right] \forall y \in K \forall z^C \forall f^{C\to C}$$

$$(f \text{ is } k\text{-strictly psc.} \wedge \forall n \in \mathbb{N} B_\forall \to \exists N \in \mathbb{N} C_\exists).$$

Then there exists a bar-recursively computable functional $\Phi : P\times\mathbb{N}\times\mathbb{N}\to\mathbb{N}$ such that in all models $\mathcal{S}^{\omega,X}$ such that for all $u\in P$, $b\in\mathbb{N}$, $M\in\mathbb{N}$, $y\in K$, $z\in C$ and $f: C\to C$ a $\left(1-\frac{1}{N+1}\right)$-strictly pseudocontractive mapping, one has that:

$$\|z\| \le b \wedge \|z - f(z)\| \le b \wedge \forall n \le \Phi(u,M,b) B_\forall \to \exists N \le \Phi(u,M,b) C_\exists.$$

If the proof in the system above proceeds without the use of the axiom of dependent choice, the Φ obtained is then a computable functional which is higher-order (i.e. in the sense of Gödel's System T) primitive recursive. In the case that the axiom of induction is used only for Σ^0_1 formulas, the functional is even primitive recursive in the ordinary sense of computability theory.

Proof. One must ensure that f is majorizable and that the additional premise of k-strict pseudocontractivity is universal. The latter is immediate, and the former follows because a k-strict pseudocontraction is, as we have seen, Lipschitzian of constant $\frac{1+k}{1-k}$. Another byproduct of our hypotheses is that, by the remarks in [46, p. 394], f is provably extensional, similarly to the functions that belong to the definition of an inner product space. □

Remark 3.5.3. *Instead of single premises with single universal quantifiers, one might also have finite conjunctions of such. Also, one might replace the universally quantified variables at the beginning with finite tuples of them. (By this argument, the k, being taken from a compact definable space, is assimilated with the y and therefore disappears as a separate dependency of the bound.)*

The above remark will be of use in our case, since we will be dealing with several self-mappings of the convex set C.

In a similar way, the metatheorem for purely intuitionistic systems is adapted to the case at hand.

Theorem 3.5.4 (cf. [23, Theorem 4.11]; there C is assumed to be bounded, but in our case, a majorant for f can still be achieved, as before)**.** *Let P be $\mathbb{N}$, $\mathbb{N}^{\mathbb{N}}$ or $\mathbb{N}^{\mathbb{N}\times\mathbb{N}}$, K an $\mathcal{A}^\omega_i$-definable compact metric space and $B(u,y,z,f)$ and $C(u,y,z,f,N)$ be arbitrary formulas (using the same convention* *about free variables as in the previous metatheorem). Suppose that* $\mathcal{A}^\omega_i[X, \langle\cdot,\cdot\rangle, C]_{-b} + CA_\neg$ *proves that:*

$$\forall u \in P \forall M \in \mathbb{N} \forall k \in \left[0, 1 - \frac{1}{M+1}\right] \forall y \in K \forall z^C \forall f^{C \to C} (f \text{ is } k\text{-strictly psc.} \wedge \neg B \to \exists N \in \mathbb{N} C).$$

Then there exists a functional $\Phi : P \times \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ *that is higher-order (i.e. in the sense of Gödel's System T) primitive recursive such that in all models* $\mathcal{S}^{\omega,X}$ *such that for all* $u \in P$*,* $b \in \mathbb{N}$*,* $M \in \mathbb{N}$*,* $y \in K$*,* $z \in C$ *and* $f : C \to C$ *a* $\left(1 - \frac{1}{M+1}\right)$*-strictly pseudocontractive mapping, one has that:*

$$\exists N \leq \Phi(u, M, b)(\neg B \to C).$$

Now, the first result of asymptotic regularity that we seek to quantify is the following:

Theorem 3.5.5. *Using the notations from the Introduction, assume that* $\sum_{n=0}^{\infty}(t_n - k)(1 - t_n) = \infty$ *and that* $\sum_{j=0}^{\infty}\sum_{i=1}^{N}|\lambda_i^{(j+1)} - \lambda_i^{(j)}| < \infty$*. Then* $\lim_{n\to\infty}\|x_n - A_n x_n\| = 0$*.*

The statement above is a convergence result, which again falls into the case-by-case discussion of three sections ago. If one examines the proof of the above result from [71], one can see that it makes some use of the law of the excluded middle, and hence the use of the second metatheorem is also not possible. Still, we shall find out that a rate of convergence can actually be extracted from the proof. This happens because the proof falls into the second half of the first case discussed there, i.e. it is "partially" constructive. To see why, we look at the way the proof is actually built. The first part of the proof, the one which uses the excluded middle, actually establishes the following result:

Lemma 3.5.6. *With the same notations, we have that* $\liminf_{n\to\infty}\|x_n - A_n x_n\| = 0$*.*

Because the sequence in question is composed of real positive numbers, the fact that 0 is its limit inferior is actually equivalent to 0 being a limit point of the sequence, which can be written as:

$$\forall \varepsilon \forall n \exists N \geq n \|x_N - A_N x_N\| < \varepsilon,$$

i.e. in a $\forall\exists$ form (note that we have used "$<$" instead of "$\leq$", following the previous discussion about the representation of real numbers).

One can, therefore, using the functional interpretation metatheorem, find a bound on the N in the sentence above, which will turn it into a sentence of the form that can be freely added as an axiom when extracting bounds from proofs using semi-intuitionistic systems. This is the case here, as the remainder of the proof uses no non-constructive principles, and hence can be analysed using modified realizability. We note that a similar problem has previously been encountered and solved by Leuştean [63] for the Ishikawa iteration as applied to nonexpansive mappings. (There is a hope that such cases may be more smoothly analysed using Oliva's hybrid interpretation [76], which allows one to use in the same proof, in a modular way, both modified realizability and functional interpretation combined with negative translation.)

All that remains now is to see that the metatheorems can be properly applied. By the remark in the previous section, one can replace the universal premise with a finite conjunction, i.e. add universal axioms to the framework. However, the premises that are at work here involve the convergence or divergence of series involving the families $(t_n)_n$ and $(\lambda_i^{(n)})_{i,n}$, which are clearly not universal. One solves this problem by adding rates of convergence and/or divergence to the statements and so turning them into universal ones. That is, we will take $\theta : \mathbb{N} \to \mathbb{N}$ be a rate of divergence for the series

$$\sum_{n=0}^{\infty} (t_n - k)(1 - t_n),$$

i.e., for all $N \in \mathbb{N}$,

$$\sum_{n=0}^{\theta(N)} (t_n - k)(1 - t_n) \geq N.$$

and $\gamma : (0, \infty) \to \mathbb{N}$ be a Cauchy modulus for the series

$$\sum_{j=0}^{\infty} \sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}|,$$

i.e., for all $n \in \mathbb{N}$ and $\varepsilon > 0$,

$$\sum_{j=\gamma(\varepsilon)+1}^{\gamma(\varepsilon)+n} \sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}| \leq \varepsilon.$$

One gets in this way two new potential parameters for the eventual rate of convergence. However, we shall not be flooded with such parameters, as $(t_n)_n$ and $(\lambda_i^{(n)})_{i,n}$ are drawn out from compact intervals ($[k, 1]$ and $[0, 1]$, respectively) and so they will not take part in the final expression. The only remnants of them will be the θ and the γ.

We now, as announced, use the obtained bound on the N (the "modulus of liminf") as an additional universal premise for the remainder of the proof, which is purely constructive, to get the desired rate of convergence for the asymptotic regularity sequence.

This discussion also, we hope, clarified the role of the parameters in Kohlenbach's result on the Mann iteration from Reich's theorem, which was presented earlier.

There is one final issue to discuss here: the extraction of a rate of convergence for the statement

$$\lim_{n \to \infty} \|x_n - T_i x_n\| = 0$$

from the one which regards the statement

$$\lim_{n \to \infty} \|x_n - A_n x_n\| = 0.$$

To do such an extraction, one might use the result which is implicit in the whole algorithm, namely that a fixed point of such a convex combination A of mappings is a fixed point of each mapping T_i. Fix an i. Then the statement just expressed can be written as:

$$\forall q \in C (\forall \delta > 0 \|q - Aq\| \leq \delta \to \forall \varepsilon > 0 \|q - T_i q\| < \varepsilon)$$

or as:

$$\forall q \in C \forall \varepsilon > 0 \exists \delta > 0 (\|q - Aq\| \leq \delta \to \|q - T_i q\| < \varepsilon)$$

We can apply, again, on this $\forall\exists$ statement, the general metatheorem of proof mining in order to get the δ as a function of ε. Then the rate of T_i-asymptotic regularity is immediately obtained.

3.6 The parallel algorithm – concrete results

We may now proceed to the results themselves.

3.6.1 The case of nonexpansive mappings

From now on, we will use the established notations and hypotheses from the definition of the parallel algorithm and from Theorem 3.5.1. Also, we will assume that there is a $b > 0$ and a $p \in \bigcap_{i=1}^{N} Fix(T_i)$ such that $\|x\| \leq b$ and $\|x - p\| \leq b$. Let $\theta : \mathbb{N} \to \mathbb{N}$ and $\gamma : (0, \infty) \to \mathbb{N}$ be defined as in the last section. Let $a > 0$ be such that

$$a \leq \inf_{i,n} \lambda_i^{(n)}.$$

We will do the proof for the case $k = 0$ at first (i.e. for nonexpansive mappings) and then, using again Lemma 3.3.2, derive the general result for strict pseudocontractions.

Lemma 3.6.1. *For all n, $t_n(1 - t_n)\|x_n - A_n x_n\|^2 \leq \|x_n - p\|^2 - \|x_{n+1} - p\|^2$.*

Proof. See [71, (3.9)], for $k := 0$. □

Lemma 3.6.2. *The sequence $(\|x_n - p\|)_{n\in\mathbb{N}}$ is nonincreasing and for all n, $\|x_n - p\| \leq b$.*

Proof. Immediate consequence of Lemma 3.6.1 and of the fact that $(t_n)_{n\in\mathbb{N}} \subseteq (0, 1)$. □

Proposition 3.6.3 ("modulus of liminf")**.** *Set, for all $\varepsilon > 0$ and $m \in \mathbb{N}$:*

$$\Delta_{b,\theta}(\varepsilon, m) := \theta(m + \left\lceil \frac{b^2}{\varepsilon^2} \right\rceil).$$

Then, for all $\varepsilon > 0$ and $m \in \mathbb{N}$, there is an $N \in [m, \Delta_{b,\theta}(\varepsilon, m)]$ such that $\|x_N - A_N x_N\| \leq \varepsilon$.

Proof. Set $\Psi := \sum_{n=m}^{\Delta_{b,\theta}(\varepsilon,m)} t_n(1 - t_n)\|x_n - A_n x_n\|^2$. We have, using Lemma 3.6.1, that:

$$\begin{aligned}
\Psi &\leq \sum_{n=0}^{\Delta_{b,\theta}(\varepsilon,m)} t_n(1 - t_n)\|x_n - A_n x_n\|^2 \\
&\leq \sum_{n=0}^{\Delta_{b,\theta}(\varepsilon,m)} (\|x_n - p\|^2 - \|x_{n+1} - p\|^2) \\
&= \|x_0 - p\|^2 - \|x_{\Delta_{b,\theta}(\varepsilon,m)+1} - p\|^2 \\
&\leq \|x - p\|^2 \\
&\leq b^2.
\end{aligned}$$

We argue by contradiction. Suppose, then, that for all $n \in [m, \Delta_{b,\theta}(\varepsilon, m)]$, $\|x_n - A_n x_n\| > \varepsilon$. Then:

$$\begin{aligned}
\Psi &> \varepsilon^2 \sum_{n=m}^{\Delta_{b,\theta}(\varepsilon,m)} t_n(1-t_n) \\
&= \varepsilon^2 \left(\sum_{n=0}^{\Delta_{b,\theta}(\varepsilon,m)} t_n(1-t_n) - \sum_{n=0}^{m-1} t_n(1-t_n) \right) \\
&\geq \varepsilon^2 \left(\sum_{n=0}^{\theta(m+\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil)} t_n(1-t_n) - m \right) \\
&\geq \varepsilon^2 (m + \left\lceil \frac{b^2}{\varepsilon^2} \right\rceil - m) \\
&= \varepsilon^2 \cdot \left\lceil \frac{b^2}{\varepsilon^2} \right\rceil \\
&\geq b^2.
\end{aligned}$$

Since we have by now proven that $\Psi \leq b^2$ and $\Psi > b^2$, we have obtained the desired contradiction. □

Lemma 3.6.4. *Each A_n is a nonexpansive mapping.*

Proof. We need to prove that a convex combination of nonexpansive mappings is nonexpansive. It is enough to prove this for two mappings, as the general case follows by induction. Let T_1 and T_2 be the two mappings and $T := (1-t)T_1 + tT_2$ be their convex combination. But if x and y are two points, then:

$$\begin{aligned}
\|Tx - Ty\| &= \|(1-t)(T_1 x - T_1 y) + t(T_2 x - T_2 y)\| \\
&\leq (1-t)\|T_1 x - T_1 y\| + t\|T_2 x - T_2 y\| \\
&\leq (1-t)\|x - y\| + t\|x - y\| \\
&= \|x - y\|.
\end{aligned}$$

□

Lemma 3.6.5. *We have that:*

1. $\|x_n - A_{n+1}x_{n+1}\| \leq (1-t_n)\|x_n - A_n x_n\| + \|x_{n+1} - A_{n+1}x_{n+1}\|$;
2. $\|A_n x_n - A_{n+1}x_{n+1}\| \leq (1-t_n)\|x_n - A_n x_n\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\|$;
3. $\|x_{n+1} - A_{n+1}x_{n+1}\| \leq (1-t_n)\|x_n - A_n x_n\| + t_n\|x_{n+1} - A_{n+1}x_{n+1}\| + (1-t_n)\|A_{n+1}x_{n+1} - A_n x_{n+1}\|$;
4. $\|x_{n+1} - A_{n+1}x_{n+1}\| \leq \|x_n - A_n x_n\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\|$;

Proof. 1. We have that:

$$\begin{aligned}
\|x_n - A_{n+1}x_{n+1}\| &\leq \|x_n - x_{n+1}\| + \|x_{n+1} - A_{n+1}x_{n+1}\| \\
&= (1-t_n)\|x_n - A_n x_n\| + \|x_{n+1} - A_{n+1}x_{n+1}\|.
\end{aligned}$$

2. We have that:

$$\begin{aligned}\|A_n x_n - A_{n+1}x_{n+1}\| &= \|A_n x_n - A_n x_{n+1} - (A_{n+1}x_{n+1} - A_n x_{n+1})\| \\ &\le \|A_n x_n - A_n x_{n+1}\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\| \\ &\le \|x_n - x_{n+1}\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\| \\ &= (1-t_n)\|x_n - A_n x_n\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\|.\end{aligned}$$

3. We have that:

$$\begin{aligned}\|x_{n+1} - A_{n+1}x_{n+1}\| &\le t_n\|x_n - A_{n+1}x_{n+1}\| + (1-t_n)\|A_n x_n - A_{n+1}x_{n+1}\| \\ &\le t_n(1-t_n)\|x_n - A_n x_n\| + t_n\|x_{n+1} - A_{n+1}x_{n+1}\| + \\ &\qquad + (1-t_n)^2\|x_n - A_n x_n\| + (1-t_n)\|A_{n+1}x_{n+1} - A_n x_{n+1}\| \\ &= (1-t_n)\|x_n - A_n x_n\| + t_n\|x_{n+1} - A_{n+1}x_{n+1}\| + \\ &\qquad + (1-t_n)\|A_{n+1}x_{n+1} - A_n x_{n+1}\|.\end{aligned}$$

4. Immediate from the last statement.

□

Theorem 3.6.6 ("rate of A_n-asymptotic regularity"). *Set, for all $\varepsilon > 0$:*

$$\begin{aligned}\Phi_{b,\theta,\gamma}(\varepsilon) &:= \Delta_{b,\theta}\left(\frac{\varepsilon}{2}, \gamma\left(\frac{\varepsilon}{6b}\right) + 1\right) \\ &= \theta\left(\gamma\left(\frac{\varepsilon}{6b}\right) + \left\lceil \frac{4b^2}{\varepsilon^2} \right\rceil + 1\right).\end{aligned}$$

Then for all $\varepsilon > 0$ we have that for all $n \ge \Phi_{b,\theta,\gamma}(\varepsilon)$, $\|x_n - A_n x_n\| \le \varepsilon$.

Proof. First we get, for any i and n, that:

$$\begin{aligned}\|T_i x_{n+1}\| &\le \|T_i x_{n+1} - p\| + \|p - x\| + \|x\| \\ &\le \|x_{n+1} - p\| + \|p - x\| + \|x\| \\ &\le \|x - p\| + \|p - x\| + \|x\| \\ &\le 3b.\end{aligned}$$

Using Lemma 3.6.5, we get that:

$$\begin{aligned}\|x_{n+1} - A_{n+1}x_{n+1}\| &\le \|x_n - A_n x_n\| + \|A_{n+1}x_{n+1} - A_n x_{n+1}\| \\ &\le \|x_n - A_n x_n\| + \left\|\sum_{i=1}^N (\lambda_i^{(n+1)} - \lambda_i^{(n)}) T_i x_{n+1}\right\| \\ &\le \|x_n - A_n x_n\| + 3b\sum_{i=1}^N |\lambda_i^{(n+1)} - \lambda_i^{(n)}|.\end{aligned}$$

and, by summing up, we have that for any $n, m \in \mathbb{N}$:

$$\|x_{n+m} - A_{n+m}x_{n+m}\| \le \|x_n - A_n x_n\| + 3b\sum_{j=n}^{n+m-1}\sum_{i=1}^N |\lambda_i^{(j+1)} - \lambda_i^{(j)}|. \tag{3.5}$$

We now apply Proposition 3.6.3 for $\frac{\varepsilon}{2}$ and $\gamma\left(\frac{\varepsilon}{6b}\right)+1$ to get that there is an N in the interval

$$\left[\gamma\left(\frac{\varepsilon}{6b}\right)+1,\ \Delta_{b,\theta}\left(\frac{\varepsilon}{2},\gamma\left(\frac{\varepsilon}{6b}\right)+1\right)=\Phi_{b,\theta,\gamma}(\varepsilon)\right]$$

such that:

$$\|x_N - A_N x_N\| \leq \frac{\varepsilon}{2}.$$

Take now an arbitrary $n \geq \Phi_{b,\theta,\gamma}(\varepsilon)$. Then, applying (3.5), we have that:

$$\begin{aligned}
\|x_n - A_n x_n\| &= \|x_{N+(n-N)} - A_{N+(n-N)} x_{N+(n-N)}\| \\
&\leq \|x_N - A_N x_N\| + 3b \sum_{j=N}^{n-1}\sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}| \\
&\leq \|x_N - A_N x_N\| + 3b \sum_{j=\gamma\left(\frac{\varepsilon}{6b}\right)+1}^{n-1}\sum_{i=1}^{N} |\lambda_i^{(j+1)} - \lambda_i^{(j)}| \\
&\leq \|x_N - A_N x_N\| + 3b \cdot \frac{\varepsilon}{6b} \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

where at the penultimate line we used the fact that γ is a Cauchy modulus. The statement is now proven. ☐

Note that using this alternate proof we have eliminated the square root from the original hypotheses of López-Acedo and Xu. This will be maintained when using the statement above for k-strict pseudocontractions, and also for the general convergence theorem, which does not use that specific hypothesis in its proof except in the preliminary asymptotic regularity result.

3.6.2 Individual rates of asymptotic regularity

As announced earlier, we shall derive the rates of asymptotic regularity for each of the T_i's.

Define the function $h : (0, \infty) \to (0, \infty)$, for any $\varepsilon > 0$, as follows:

$$h_{a,b}(\varepsilon) := \varepsilon + \sqrt{\frac{1-a}{a}}\sqrt{\varepsilon^2 + 2b\varepsilon}$$

Lemma 3.6.7. *Let $n \in \mathbb{N}$ and $z \in C$ such that $\|z - p\| \leq b$ and $\|z - A_n z\| \leq \varepsilon$. Then, for each i, $\|z - T_i z\| \leq h_{a,b}(\varepsilon)$.*

Proof. Remember that $A_n = \sum_{i=1}^{N} \lambda_i^{(n)} T_i$ and that p is a fixed point of all the T_i's and hence also of the A_n's. Then:

$$\|p - A_n z\| = \|A_n p - A_n z\| \leq \|p - z\| \leq b.$$

Let $j \in \{1, ..., N\}$. We have that:

$$\begin{aligned}
\|p-z\|^2 &\le \|p-A_nz\|^2 + \|z-A_nz\|^2 + 2\|p-A_nz\|\|z-A_nz\| \\
&\le \left\| p - \left(\left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left(\frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z \right) + \lambda_j^{(n)} T_j z \right) \right\|^2 + \|z-A_nz\|^2 + \\
&\quad + 2\|p-A_nz\|\|z-A_nz\| \\
&\le \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| p - \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z \right\|^2 + \lambda_j^{(n)} \|p-T_jz\|^2 - \\
&\quad - \lambda_j^{(n)} \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\|^2 + \varepsilon^2 + 2b\varepsilon \\
&\quad \text{by Proposition 1.1.3} \\
&\le \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \|p-z\|^2 + \lambda_j^{(n)} \|p-z\|^2 - \\
&\quad - \lambda_j^{(n)} \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\|^2 + \varepsilon^2 + 2b\varepsilon \\
&\le \|p-z\|^2 - \lambda_j^{(n)} \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\|^2 + \varepsilon^2 + 2b\varepsilon.
\end{aligned}$$

It follows that:

$$\lambda_j^{(n)} \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\|^2 \le \varepsilon^2 + 2b\varepsilon,$$

so

$$\left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\| \le \sqrt{\frac{1}{\lambda_j^{(n)} \left(\sum_{i\neq j} \lambda_i^{(n)} \right)}} \sqrt{\varepsilon^2 + 2b\varepsilon}.$$

We can now see that:

$$\begin{aligned}
\|z-T_jz\| &\le \|z-A_nz\| + \|A_nz-T_jz\| \\
&\le \varepsilon + \left(\sum_{i\neq j} \lambda_i^{(n)} \right) \left\| \frac{1}{\sum_{i\neq j}\lambda_i^{(n)}} \sum_{i\neq j} \lambda_i^{(n)} T_i z - T_j z \right\| \\
&\le \varepsilon + \sqrt{\frac{\sum_{i\neq j}\lambda_i^{(n)}}{\lambda_j^{(n)}}} \sqrt{\varepsilon^2 + 2b\varepsilon}.
\end{aligned}$$

All that remains is to show that

$$\frac{\sum_{i\neq j}\lambda_i^{(n)}}{\lambda_j^{(n)}} \le \frac{1-a}{a}.$$

But this is immediate, since $\lambda_j^{(n)} \ge a$, and hence $\frac{1}{\lambda_j^{(n)}} \le \frac{1}{a}$ and $\sum_{i\neq j} \lambda_i^{(n)} = 1 - \lambda_j^{(n)} \le 1 - a$. □

Theorem 3.6.8 ("rate of T_i-asymptotic regularity")**.** *Set, for all $\varepsilon > 0$:*

$$P_{a,b}(\varepsilon) := \min\left\{\frac{\varepsilon}{2}, \sqrt{\frac{a\varepsilon^2}{4(1-a)} + b^2} - b\right\}$$
$$\Phi'_{a,b,\theta,\gamma}(\varepsilon) := \Phi_{b,\theta,\gamma}(P_{a,b}(\varepsilon)).$$

Then for all $\varepsilon > 0$ we have that for all $n \geq \Phi'_{a,b,\theta,\gamma}(\varepsilon)$ and all i, $\|x_n - T_i x_n\| \leq \varepsilon$.

Proof. By Theorem 3.6.6, we have that $\|x_n - A_n x_n\| \leq P_{a,b}(\varepsilon)$. Since $\|x_n - p\| \leq \|x - p\| \leq b$, we can apply Lemma 3.6.7 to get that:

$$\|x_n - T_i x_n\| \leq h_{a,b}(P_{a,b}(\varepsilon)).$$

All we have to show, therefore, is that $h_{a,b}(P_{a,b}(\varepsilon)) \leq \varepsilon$, i.e. that:

$$P_{a,b}(\varepsilon) + \sqrt{\frac{1-a}{a}}\sqrt{P_{a,b}(\varepsilon)^2 + 2bP_{a,b}(\varepsilon)} \leq \varepsilon.$$

Now, it is immediate that the monic polynomial:

$$X^2 + 2bX - \frac{a\varepsilon^2}{4(1-a)}$$

has two roots, one negative and one equal to $\sqrt{\frac{a\varepsilon^2}{4(1-a)} + b^2} - b$. Since (by its definition) $P_{a,b}(\varepsilon) \geq 0$ and $P_{a,b}(\varepsilon) \leq \sqrt{\frac{a\varepsilon^2}{4(1-a)} + b^2} - b$, we have that:

$$P_{a,b}(\varepsilon)^2 + 2bP_{a,b}(\varepsilon) - \frac{a\varepsilon^2}{4(1-a)} \leq 0,$$

and hence that

$$P_{a,b}(\varepsilon)^2 + 2bP_{a,b}(\varepsilon) \leq \frac{a\varepsilon^2}{4(1-a)}.$$

We can now compute:

$$\begin{aligned} P_{a,b}(\varepsilon) + \sqrt{\frac{1-a}{a}}\sqrt{P_{a,b}(\varepsilon)^2 + 2bP_{a,b}(\varepsilon)} &\leq \frac{\varepsilon}{2} + \sqrt{\frac{1-a}{a}}\sqrt{\frac{a\varepsilon^2}{4(1-a)}} \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

□

3.6.3 The general case

We can at last consider the general case of a family of strict pseudocontractions – i.e. we will allow k to vary. Note that if $0 \leq k \leq k' < 1$ and T is a k-strict pseudocontraction, then it is also a k'-strict pseudocontraction. Hence, instead of considering each T_i to be a k_i-strict pseudocontraction, we can take k to be the maximum of the N k_i's and work, without loss of generality, with a finite family of k-strict pseudocontractions.

Now, if $T : C \to C$ is a k-strict pseudocontraction and $t \in (0, 1)$, denote by $T_t : C \to C$ the map defined by $T_t x := tx + (1-t)Tx$. (Again, we use the "opposite" convention from the one in the previous sections.) It is clear that for any such t, the maps T and T_t have the same fixed points. Note that $(T_{t_1})_{t_2} = T_{1-(1-t_1)(1-t_2)}$.

The main result of this section is the following.

Theorem 3.6.9 ("general rate of T_i-asymptotic regularity")**.** *Set, for all $\varepsilon > 0$:*

$$\Phi''_{a,b,k,\theta,\gamma}(\varepsilon) := \Phi'_{a,b,\theta,\gamma}((1-k)\varepsilon).$$

Then for all $\varepsilon > 0$ we have that for all $n \geq \Phi''_{a,b,k,\theta,\gamma}(\varepsilon)$ and all i, $\|x_n - T_i x_n\| \leq \varepsilon$.

Proof. Set, for all i, $T'_i := (T_i)_k$. Then, by Lemma 3.3.2, each such T'_i is nonexpansive. It is easy to see that if we set $A'_n := \sum_{i=1}^N \lambda_i^{(n)} T'_i$, then $A'_n = (A_n)_k$.

Set, for all n, $t'_n := 1 - \frac{1-t_n}{1-k}$. Since $t_n \in [k, 1]$, we have that $t'_n \in [0, 1]$.

Let $n \in \mathbb{N}$. Then we have that:

$$\begin{aligned} x_{n+1} &= t_n x_n + (1-t_n) A_n x_n \\ &= (A_n)_{t_n} x_n \\ &= (A_n)_{1-(1-k)(1-t'_n)} x_n \\ &= ((A_n)_k)_{t'_n} x_n \\ &= (A'_n)_{t'_n} x_n. \end{aligned}$$

So the sequence $(x_n)_n$ is also the output of the parallel algorithm for the mappings $(T'_i)_i$ and the weight sequence $(t'_n)_n$, with the same $\lambda_i^{(n)}$'s. Now, in order to apply Theorem 3.6.8, we must check that the θ is still valid. We have that $1 - t_n = (1-k)(1-t'_n)$ and that $t_n - k = (1-k)t'_n$, so:

$$\sum_{n=0}^{\theta(N)} t'_n(1-t'_n) = \frac{1}{(1-k)^2} \sum_{n=0}^{\theta(N)} (t_n - k)(1-t_n) \geq \frac{1}{(1-k)^2} \cdot N \geq N.$$

Hence, by Theorem 3.6.8 we have that for all $\varepsilon > 0$, for all $n \geq \Phi'_{a,b,\theta,\gamma}(\varepsilon)$ and for all i, $\|x_n - T'_i x_n\| \leq \varepsilon$. But $x_n - T'_i x_n = (1-k)(x_n - T_i x_n)$, so:

$$\|x_n - T_i x_n\| \leq \frac{\varepsilon}{1-k}.$$

So, for all $n \geq \Phi'_{a,b,\theta,\gamma}((1-k)\varepsilon)$, we have that:

$$\|x_n - T_i x_n\| \leq \varepsilon,$$

which was what we needed to show. ☐

As an aside, we note that one can remove the explicit appearance of k in the above bound by the following argument. By the definition of θ, we have that:

$$\sum_{n=0}^{\theta(1)} (t_n - k)(1-t_n) \geq 1.$$

Therefore, by the pigeonhole principle, we have that one of these first $\theta(1)+1$ terms of the series must be greater than $\frac{1}{\theta(1)+1}$. So there is an $n_0 \leq \theta(1)$ such that:

$$(t_{n_0}-k)(1-t_{n_0}) \geq \frac{1}{\theta(1)+1}.$$

But necessarily $(t_{n_0}-k)(1-t_{n_0}) \leq t_{n_0}-k \leq 1-k$, so $k \leq 1-\frac{1}{\theta(1)+1}$. We noted earlier that we can replace any k by a greater k', smaller than 1, and the strict pseudocontraction condition will be maintained. Therefore, in the above bound, we may freely replace k by $1-\frac{1}{\theta(1)+1}$ (taking note, in the process, that $\theta(1)$ plays the role of M from the statements of the two metatheorems).

3.7 Fejér monotonicity in the context of total boundedness

Frequently, the program of proof mining was advanced by identifying, within a concrete mathematical proof, an argument that could be said to form its core and to provide the greater part of the difficulty of extracting quantitative information. Once such a proof was completely analyzed, however, the argument could be unwinded separately and used verbatim in its quantitative form in the analysis of other similar proofs.

In a recent paper, U. Kohlenbach, L. Leuştean, and A. Nicolae [50] have studied a general line of argument used in convergence proofs in nonlinear analysis. Specifically, it is often the case that an iterative sequence is proven to be convergent to a point in a certain set F (e.g. the set of fixed points of an operator using which the sequence was constructed) if it sits inside a compact space, it is "Fejér monotone" with respect to F (that is, for all $q \in F$ and all $n \in \mathbb{N}$, $d(x_{n+1},q) \leq d(x_n,q)$) and it has "approximate fixed points", i.e. points which are, in a sense, near F. The main result in their paper is that all this can be made effective. For that to work, however, the three hypotheses must also be transformed into a quantified form. A "modulus of total boundedness" witnesses the space being compact. For the other two properties, one must formulate what exactly does it mean for a point to be "near" to F. This is done in terms of an approximation $F = \bigcap_{k\in\mathbb{N}} AF_k$, which helps formulate both the "modulus of uniform Fejér monotonicity" and the "approximate fixed point bound". The choice of an approximation to F, as well as the computation of these moduli, has been done for some classical sequences and it involves a certain degree of ingenuity. After all these parameters have been fixed, however, the rate of metastability surely follows.

We shall now present in more details their ideas, and in the next section we will apply them to the analysis of Theorem 3.2.6 above, of Ishikawa. We point out also that the Ishikawa iteration was already approached with proof mining methods in [62, 63] for nonexpansive mappings in uniformly convex geodesic spaces.

It is a basic fact in metric space theory that a metric space is compact iff it is complete and totally bounded. We should give now an explanation of what total boundedness means. However, our goal is to exploit this kind of notion quantitatively, and therefore we need a way to effectively measure it. Such a way will be a modulus of total boundedness, i.e. a space will be totally bounded iff it admits such a modulus. It is therefore sufficient, for the sake of exposition, to give a definition of the latter concept.

Definition 3.7.1 (cf. [50, Definition 2.2]). *A* **modulus of total boundedness** *for a space X is a function $\alpha : \mathbb{N} \to \mathbb{N}$ such that for any $k \in \mathbb{N}$ and any sequence $(x_n)_{n\in\mathbb{N}}$ in X there are $i < j$ in $[0, \alpha(k)]$ such that*

$$d(x_i, x_j) \leq \frac{1}{k+1}.$$

This quantitative version of total boundedness was used in [22] to obtain, also using proof mining, quantitative results in topological dynamics.

We will now proceed to define the rest of the needed moduli, in terms of the following definition. We say that a family $\{AF_k\}_{k\in\mathbb{N}}$ is an *approximation* to a set F iff

$$F = \bigcap_{k\in\mathbb{N}} AF_k$$

and for all k, $AF_{k+1} \subseteq AF_k$.

Definition 3.7.2 (cf. [50, Definition 3.4]). *A set F is called* **uniformly closed** *with respect to a approximation $\{AF_k\}_{k\in\mathbb{N}}$ with moduli $\delta_F, \omega_F : \mathbb{N} \to \mathbb{N}$ iff for all $k \in \mathbb{N}$ and $p, q \in X$ we have that if $q \in AF_{\delta_F(k)}$ and $d(p, q) \leq \frac{1}{\omega_F(k)+1}$ then $p \in AF_k$.*

Definition 3.7.3 (cf. [50, Definition 4.6]). *We say that a sequence $(x_n)_{n\in\mathbb{N}}$ is called* **uniformly Fejér monotone** *with respect to a approximation $\{AF_k\}_{k\in\mathbb{N}}$ of a set F with modulus χ if for all $n, m, r \in \mathbb{N}$, all $p \in AF_{\chi(n,m,r)}$ and all $l \leq m$ we have that*

$$d(x_{n+l}, p) \leq d(x_n, p) + \frac{1}{r+1}.$$

Definition 3.7.4 (cf. [50, Section 5]). *We say that a sequence $(x_n)_{n\in\mathbb{N}}$ has* **F-approximate fixed points** *with respect to a approximation $\{AF_k\}_{k\in\mathbb{N}}$ with modulus Φ (which is taken to be monotonely nondecreasing) if for all $k \in \mathbb{N}$ there is an $N \leq \Phi(k)$ such that $x_N \in AF_k$.*

In order to get monotonely nondecreasing moduli, one uses the following transformation. For any $f : \mathbb{N} \to \mathbb{N}$, one defines $f^M : \mathbb{N} \to \mathbb{N}$ by:

$$f^M(n) := \max_{i\leq n} f(i)$$

Then f^M is monotonely nondecreasing and for any n, we have that $f(n) \leq f^M(n)$.

Given two functions $G, H : \mathbb{R}_+ \to \mathbb{R}_+$, a sequence (u_n) in C is said to be (G, H)-*Fejér monotone* w.r.t. F if for all $n, m \in \mathbb{N}$ and all $p \in F$,

$$H(d(u_{n+m}, p)) \leq G(d(u_n, p)).$$

This is a natural generalization of Fejér monotonicity, which is obtained by putting $G = H = id_{\mathbb{R}^+}$. As in [50], we suppose that the mappings G, H satisfy the following properties: for all sequences (a_n) in $\mathbb{R}_+$,

$$(G) \quad \lim_{n\to\infty} a_n = 0 \text{ implies } \lim_{n\to\infty} G(a_n) = 0$$

and

$$(H) \quad \lim_{n\to\infty} H(a_n) = 0 \text{ implies } \lim_{n\to\infty} a_n = 0.$$

These properties allow us to obtain in the general setting some nice properties of Fejér monotone sequences, needed for proving strong convergence.

Equivalent quantitative versions of (G) and (H) assert the existence of moduli $\alpha_G : \mathbb{N} \to \mathbb{N}$ and $\beta_H : \mathbb{N} \to \mathbb{N}$ such that for all $k \in \mathbb{N}$ and all $a \in \mathbb{R}_+$,

$$a \leq \tfrac{1}{\alpha_G(k)+1} \text{ implies } G(a) \leq \tfrac{1}{k+1}$$
$$\text{and}$$
$$H(a) \leq \tfrac{1}{\beta_H(k)+1} \text{ implies } a \leq \tfrac{1}{k+1}.$$

We say that α_G is a G-modulus and β_H is an H-modulus.

The following uniform version of (G, H)-Fejér monotonicity was introduced in [50] and is another of the abovementioned notions needed to get our quantitative results.

Definition 3.7.5. *Let $G, H : \mathbb{R}_+ \to \mathbb{R}_+$. A sequence (u_n) in C is called* **uniformly** (G, H)**-Fejér monotone** *with respect to a approximation $\{AF_k\}_{k\in\mathbb{N}}$ of a set F with modulus $\chi : \mathbb{N}^3 \to \mathbb{N}$ if for all $n, m, r \in \mathbb{N}$, for all $p \in C$ with $p \in AF_{\chi(n,m,r)+1}$ and for all $l \leq m$ we have that*

$$H(d(u_{n+l}, p))) < G(d(u_n, p)) + \frac{1}{r+1}.$$

We are now in position to state the main result of [50].

Theorem 3.7.6 (cf. [50, Theorems 5.1 and 5.3])**.** *Assume that the space X has a modulus of total boundedness α. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X and $F \subseteq X$ having a decomposition*

$$F = \bigcap_{k\in\mathbb{N}} AF_k.$$

Assume further that:

(i) there are $G, H : \mathbb{R}_+ \to \mathbb{R}_+$ with α_G a G-modulus for G and β_H a H-modulus for H such that $(x_n)_{n\in\mathbb{N}}$ is uniformly (G, H)-Fejér monotone with respect to $\{AF_k\}_{k\in\mathbb{N}}$ with modulus χ;

(ii) $(x_n)_{n\in\mathbb{N}}$ has F-approximate fixed points with respect to $\{AF_k\}_{k\in\mathbb{N}}$ with modulus Φ.

We define the following functionals:

$$\chi_g(n, k) := \chi(n, g(n), k)$$
$$\chi_g^M(n, k) := \max_{i\leq n} \chi_g(i, k)$$
$$(\Psi_0)_{\Phi,\chi,\beta_H}(0, k, g) := 0$$
$$(\Psi_0)_{\Phi,\chi,\beta_H}(n + 1, k, g) := \Phi(\chi_g^M((\Psi_0)_{\Phi,\chi,\beta_H}(n, k, g), 2\beta_H(2k + 1) + 1))$$
$$\Psi_{\Phi,\chi,\alpha,\alpha_G,\beta_H}(k, g) := (\Psi_0)_{\Phi,\chi,\beta_H}(\alpha(\alpha_G(2\beta_H(2k + 1) + 1)), k, g)$$

Then $\Psi_{\Phi,\chi,\alpha,\alpha_G,\beta_H}$ is a rate of metastability for $(x_n)_{n\in\mathbb{N}}$, i.e. for all $k \in \mathbb{N}$, all $g : \mathbb{N} \to \mathbb{N}$ there is an $N \leq \Psi_{\Phi,\chi,\alpha,\alpha_G,\beta_H}(k, g)$ such that for all $i, j \in [N, N + g(N)]$,

$$d(x_i, x_j) \leq \frac{1}{k+1}.$$

Moreover, if we even further assume that F is uniformly closed w.r.t. $\{AF_k\}_{k\in\mathbb{N}}$ with moduli δ_F, ω_F, and if we further define:

$$l_{k,\omega_F} := \max\left\{k, \left\lceil\frac{\omega_F(k)-1}{2}\right\rceil\right\}$$

$$\lambda_{\chi,k,\delta_F}(n,m,r) := \max\{\delta_F(k), \chi(n,m,r)\}$$

$$\widetilde{\Psi}_{\Phi,\chi,\alpha,\alpha_G,\beta_H,\delta_F,\omega_F}(k,g) := \Psi_{\Phi,\lambda_{\chi,k,\delta_F},\alpha,\alpha_G,\beta_H}(l_{k,\omega_F}, g)$$

we have that for all $k \in \mathbb{N}$, all $g : \mathbb{N} \to \mathbb{N}$ there is an $N \leq \widetilde{\Psi}_{\Phi,\chi,\alpha,\alpha_G,\beta_H,\delta_F,\omega_F}(k,g)$ such that for all $i, j \in [N, N+g(N)]$,

$$d(x_i, x_j) \leq \frac{1}{k+1} \text{ and } x_i \in AF_k.$$

3.8 The Ishikawa iteration for Lipschitzian pseudocontractions

In order to study more closely the result of Theorem 3.2.6, we fix the following notations for the conditions that sequences (t_n), (s_n) in $[0,1]$ should satisfy:

$$\begin{array}{ll} \text{(A1)} & \lim_{n\to\infty} s_n = 0; \\ \text{(A2)} & \sum_{n=0}^{\infty} t_n s_n = \infty; \\ \text{(A3)} & t_n \leq s_n, \text{ for all } n \in \mathbb{N}. \end{array}$$

3.8.1 Some useful lemmas

Let H be a Hilbert space, $C \subseteq H$ a nonempty convex subset and $T : C \to C$ be a mapping. Furthermore, (t_n) and (s_n) are sequences of reals in $[0,1]$ and (x_n) is the Ishikawa iteration corresponding to T, x, $(t_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$.

In order for the computations to be less cumbersome, we shall also set for all $n \in \mathbb{N}$,

$$y_n := s_n T x_n + (1 - s_n) x_n,$$

so that we have, again for all $n \in \mathbb{N}$,

$$x_{n+1} = (1 - t_n) x_n + t_n T y_n.$$

Remark 3.8.1. *It is clear that $x_n - x_{n+1} = t_n(x_n - Ty_n)$, so $\|x_n - x_{n+1}\| \leq \|x_n - Ty_n\|$, and that $x_n - y_n = s_n(x_n - Tx_n)$, so $\|x_n - y_n\| \leq \|x_n - Tx_n\|$.*

Lemma 3.8.2. *Assume that T is L-Lipschitzian. Then $\|x_n - x_{n+1}\| \leq (1+L)\|x_n - Tx_n\|$.*

Proof. Using Remark 3.8.1, we have that:

$$\begin{aligned} \|x_n - x_{n+1}\| &\leq \|x_n - Ty_n\| \leq \|x_n - Tx_n\| + \|Tx_n - Ty_n\| \\ &\leq \|x_n - Tx_n\| + L\|x_n - y_n\| \\ &\leq (1+L)\|x_n - Tx_n\|. \end{aligned}$$

□

We recall the following well-known and useful equalities that hold in Hilbert spaces.

Lemma 3.8.3. *For any $x, y \in H$ and any $\lambda \in (0,1)$, the following identities hold:*

1. $\|\lambda x + (1-\lambda)y\|^2 = \lambda\|x\|^2 + (1-\lambda)\|y\|^2 - \lambda(1-\lambda)\|x-y\|^2$;
2. $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$ *and* $\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y\rangle$.

We shall denote, for any $y, w \in C$,

$$\sigma(y,w) := \|w - Tw\| + \|y - Tw\|.$$

Lemma 3.8.4. *Assume that T is a pseudocontraction. Then, for every $z, p \in C$,*

$$\|Tz - p\|^2 \le \|z-p\|^2 + \|z - Tz\|^2 + 2\|p - Tp\|\sigma(z,p). \tag{3.6}$$

Proof. Just follow the proof of [36, Lemma 3.2.(i)] (with $\kappa := 1$). □

The following equalities are immediate consequences of Lemma 3.8.3.(1).

Lemma 3.8.5. *For every $p \in C$, we have that:*

$$\begin{aligned}
\|x_{n+1} - p\|^2 &= t_n\|Ty_n - p\|^2 + (1-t_n)\|x_n - p\|^2 \\
&\quad - t_n(1-t_n)\|Ty_n - x_n\|^2 && (3.7)\\
\|y_n - p\|^2 &= s_n\|Tx_n - p\|^2 + (1-s_n)\|x_n - p\|^2 \\
&\quad - s_n(1-s_n)\|Tx_n - x_n\|^2 && (3.8)\\
\|y_n - Ty_n\|^2 &= s_n\|Tx_n - Ty_n\|^2 + (1-s_n)\|x_n - Ty_n\|^2 \\
&\quad - s_n(1-s_n)\|Tx_n - x_n\|^2 && (3.9)
\end{aligned}$$

Lemma 3.8.6. *Assume that T is a pseudocontraction and let $p \in C$.*

1. *We have that:*

$$\begin{aligned}
\|x_{n+1} - p\|^2 \le\ & \|x_n - p\|^2 + t_n s_n\|Tx_n - Ty_n\|^2 - t_n s_n(1-2s_n)\|Tx_n - x_n\|^2 \\
& - t_n(s_n - t_n)\|Ty_n - x_n\|^2 + 2\|p - Tp\|\big(\sigma(x_n,p) + \sigma(y_n,p)\big)
\end{aligned}$$

2. *Assume, furthermore, that T is L-Lipschitzian and that $(t_n), (s_n)$ satisfy (A3). Then we have:*

$$\begin{aligned}
\|x_{n+1} - p\|^2 \le\ & \|x_n - p\|^2 - t_n s_n(1 - 2s_n - L^2 s_n^2)\|x_n - Tx_n\|^2 \\
& + 2\|p - Tp\|\big(\sigma(x_n,p) + \sigma(y_n,p)\big).
\end{aligned} \tag{3.10}$$

Proof. The proof is a slightly modified version of the one from [35].

1. We get that:

$$\begin{aligned}
\|x_{n+1} - p\|^2 &= t_n\|Ty_n - p\|^2 + (1-t_n)\|x_n - p\|^2 - t_n(1-t_n)\|Ty_n - x_n\|^2 \\
&\quad \text{by (3.7)}
\end{aligned}$$

$$
\begin{aligned}
&\leq t_n(\|y_n - p\|^2 + \|y_n - Ty_n\|^2 + 2\|p - Tp\|\sigma(y_n, p)) \\
&\quad +(1-t_n)\|x_n - p\|^2 - t_n(1-t_n)\|Ty_n - x_n\|^2 \\
&\quad \text{by (3.6) with } z := y_n \\
&= t_n\|y_n - p\|^2 + (1-t_n)\|x_n - p\|^2 - t_n(1-t_n)\|Ty_n - x_n\|^2 \\
&\quad +t_ns_n\|Tx_n - Ty_n\|^2 + t_n(1-s_n)\|x_n - Ty_n\|^2 \\
&\quad -t_ns_n(1-s_n)\|Tx_n - x_n\|^2 + 2t_n\|p - Tp\|\sigma(y_n, p) \\
&\quad \text{by (3.9)} \\
&= t_ns_n\|Tx_n - Ty_n\|^2 + t_n(t_n - s_n)\|x_n - Ty_n\|^2 \\
&\quad +(1-t_n)\|x_n - p\|^2 + t_n\|y_n - p\|^2 - t_ns_n(1-s_n)\|Tx_n - x_n\|^2 \\
&\quad +2t_n\|p - Tp\|\sigma(y_n, p) \\
&= t_ns_n\|Tx_n - Ty_n\|^2 + t_n(t_n - s_n)\|x_n - Ty_n\|^2 \\
&\quad +(1-t_n)\|x_n - p\|^2 t_n(s_n\|Tx_n - p\|^2 + (1-s_n)\|x_n - p\|^2 \\
&\quad -s_n(1-s_n)\|Tx_n - x_n\|^2) - t_ns_n(1-s_n)\|Tx_n - x_n\|^2 \\
&\quad +2t_n\|p - Tp\|\sigma(y_n, p) \\
&\quad \text{by (3.8)} \\
&= t_ns_n\|Tx_n - Ty_n\|^2 + t_n(t_n - s_n)\|x_n - Ty_n\|^2 + \|x_n - p\|^2 \\
&\quad -2t_ns_n(1-s_n)\|Tx_n - x_n\|^2 + t_ns_n(\|Tx_n - p\|^2 - \|x_n - p\|^2) \\
&\quad +2t_n\|p - Tp\|\sigma(y_n, p) \\
&\leq t_ns_n\|Tx_n - Ty_n\|^2 + t_n(t_n - s_n)\|x_n - Ty_n\|^2 + \|x_n - p\|^2 \\
&\quad -2t_ns_n(1-s_n)\|Tx_n - x_n\|^2 + t_ns_n\|Tx_n - x_n\|^2 \\
&\quad +2t_ns_n\|p - Tp\|\sigma(x_n, p) + 2t_n\|p - Tp\|\sigma(y_n, p) \\
&\quad \text{by (3.6) with } z := x_n \\
&= \|x_n - p\|^2 + t_ns_n\|Tx_n - Ty_n\|^2 - t_ns_n(1-2s_n)\|Tx_n - x_n\|^2 \\
&\quad -t_n(s_n - t_n)\|Ty_n - x_n\|^2 + 2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big).
\end{aligned}
$$

2. If (A3) holds, then $t_n(s_n - t_n)\|Ty_n - x_n\|^2 \geq 0$. It follows that:

$$
\begin{aligned}
\|x_{n+1} - p\|^2 &\leq \|x_n - p\|^2 + t_ns_n\|Tx_n - Ty_n\|^2 - t_ns_n(1-2s_n)\|Tx_n - x_n\|^2 \\
&\quad +2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big) \\
&\leq \|x_n - p\|^2 + L^2t_ns_n\|x_n - y_n\|^2 - t_ns_n(1-2s_n)\|Tx_n - x_n\|^2 \\
&\quad +2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big) \\
&= \|x_n - p\|^2 + L^2t_ns_n^3\|x_n - Tx_n\|^2 - t_ns_n(1-2s_n)\|Tx_n - x_n\|^2 \\
&\quad +2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big) \\
&\quad \text{by Remark 3.8.1} \\
&= \|x_n - p\|^2 + t_ns_n(L^2s_n^2 - 1 + 2s_n)\|x_n - Tx_n\|^2 \\
&\quad +2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big) \\
&= \|x_n - p\|^2 - t_ns_n(1 - 2s_n - L^2s_n^2)\|x_n - Tx_n\|^2 \\
&\quad +2\|p - Tp\|\big(\sigma(x_n, p) + \sigma(y_n, p)\big).
\end{aligned}
$$

□

Lemma 3.8.7. *Assume that (s_n) satisfies (A1) and that β is a rate of convergence of (s_n). Set*

$$K := \beta\left(\left\lceil 1+\sqrt{2L^2+4}\right\rceil\right). \tag{3.11}$$

Then, for all $n \geq K$, $1-2s_n-L^2s_n^2 \geq \frac{1}{2}$.

Proof. Take $n \geq K$. Since β is a rate of convergence for the nonnegative sequence (s_n), whose limit is 0, we have that $s_n \leq \frac{1}{1+\lceil 1+\sqrt{2L^2+4}\rceil} \leq \frac{1}{2+\sqrt{2L^2+4}} = \frac{-2+\sqrt{2L^2+4}}{2L^2}$. It follows that $s_n + \frac{1}{L^2} \leq \frac{\sqrt{2L^2+4}}{2L^2}$, so $s_n^2 + \frac{2}{L^2}s_n + \frac{1}{L^4} \leq \frac{1}{2L^2} + \frac{1}{L^4}$ and $L^2s_n^2 + 2s_n \leq \frac{1}{2}$, hence the desired inequality. ☐

Let us, for all $n \in \mathbb{N}$, denote:

$$z_n := x_{n+K}. \tag{3.12}$$

In particular, we have that (z_n) is a subsequence of (x_n).

Lemma 3.8.8. *Assume that T is an L-Lipschitzian pseudocontraction, (t_n), (s_n) satisfy (A1) and (A3) and β is a rate of convergence of (s_n).*

1. *If C is bounded and b is an upper bound on the diameter of C, then for all $n \in \mathbb{N}$ and all $p \in C$,*

$$\|z_{n+1}-p\|^2 \leq \|z_n-p\|^2 - \frac{1}{2}t_ns_n\|z_n-Tz_n\|^2 + 8b\|p-Tp\|. \tag{3.13}$$

2. *If p is a fixed point of T, then for all $n \in \mathbb{N}$,*

$$\|z_{n+1}-p\|^2 \leq \|z_n-p\|^2 - \frac{1}{2}t_ns_n\|z_n-Tz_n\|^2. \tag{3.14}$$

Proof. Apply Lemma 3.8.7 and (3.10). For (i) use the fact that $2\|p-Tp\|(\sigma(x_n,p)+\sigma(y_n,p)) \leq 8b\|p-Tp\|$. ☐

3.8.2 An effective modulus of liminf

We shall use in this section the notations for the Ishikawa iteration introduced in the previous one. The following result is the first step in Ishikawa's proof of Theorem 3.2.6.

Proposition 3.8.9. *Assume that T has fixed points and that (t_n), (s_n) satisfy (A1)-(A3). Then $\liminf\limits_{n\to\infty}\|x_n-Tx_n\| = 0$ for all $x \in C$.*

The main result of this section is the following quantitative version of Proposition 3.8.9, giving us an effective and uniform modulus of liminf for $(\|x_n-Tx_n\|)$.

Theorem 3.8.10. *Assume that T has fixed points and that (t_n), (s_n) satisfy (A1)-(A3). Let β be a rate of convergence of (s_n) and θ be a rate of divergence of $\sum\limits_{n=0}^{\infty}t_ns_n$.*

Let us define $\Delta_{b,\theta}$, $\tilde{\Delta}_{b,L,\beta,\theta} : \mathbb{N}\times\mathbb{N}\to\mathbb{N}$ by

$$\Delta_{b,\theta}(l,k) := \theta(l+M), \quad \tilde{\Delta}_{b,L,\beta,\theta}(l,k) = K + \Delta_{b,\theta}(l,k),$$

with $K := \beta\left(\left\lceil 1+\sqrt{2L^2+4}\right\rceil\right)$, $M := 2(b^2+1)(k+1)^2$ *and* $b \in \mathbb{N}$ *is such that* $b \geq \|x_K - p\|$ *for some fixed point* p *of* T.

Then, for all $x \in C$,

1. $\liminf_{n\to\infty} \|z_n - Tz_n\| = 0$ *with modulus of liminf* $\Delta_{b,\theta}$;
2. $\liminf_{n\to\infty} \|x_n - Tx_n\| = 0$ *with modulus of liminf* $\tilde{\Delta}_{b,L,\beta,\theta}$.

Proof. Let $x \in C$, $p \in Fix(T)$ and b as in the hypothesis. We denote, for simplicity, $\Delta := \Delta_{b,\theta}(l,k)$.

1. We have to prove that

$$\forall l \in \mathbb{N}\ \forall k \in \mathbb{N}\ \exists N \in [l,\Delta]\ \left(\|z_N - Tz_N\| \leq \frac{1}{k+1}\right). \tag{3.15}$$

Remark first that, since θ is a rate of divergence for $\sum_{n=0}^{\infty} t_n s_n$ and t_n, s_n are sequences in $[0,1]$, we have that $\theta(n) \geq n-1$ for all $n \in \mathbb{N}$. Then $\Delta \geq l + M - 1 \geq l$, as $M \geq 1$.

By (3.14), we get that for all $n \in \mathbb{N}$,

$$\|z_{n+1} - p\|^2 \leq \|z_n - p\|^2 - \frac{1}{2} t_n s_n \|x_n - Tx_n\|^2. \tag{3.16}$$

As an immediate consequence, it follows that $\|z_{n+1} - p\| \leq \|z_n - p\|$ for all $n \in \mathbb{N}$. Thus, $b \geq \|x_K - p\| = \|z_0 - p\| \geq \|z_n - p\|$ for all $n \in \mathbb{N}$.

Assume by contradiction that (3.15) does not hold, hence $\|z_n - Tz_n\| > \frac{1}{k+1}$ for all $n \in [l,\Delta]$. Adding (3.16) for $n := l, \ldots, \Delta$, we get that

$$\|z_{\Delta+1} - p\|^2 \quad \leq \quad \|z_l - p\|^2 - \frac{1}{2}\sum_{n=l}^{\Delta} t_n s_n \|z_n - Tz_n\|^2 \leq b^2 - \frac{1}{2(k+1)^2}\sum_{n=l}^{\Delta} t_n s_n.$$

Remark now that

$$\sum_{n=l}^{\Delta} t_n s_n = \sum_{n=0}^{\theta(l+M)} t_n s_n - \sum_{n=0}^{l-1} t_n s_n \geq l + M - l = M.$$

It follows that

$$\|z_{\Delta+1} - p\|^2 \leq b^2 - \frac{1}{2(k+1)^2} M = -1.$$

We have obtained a contradiction.

2. By (1), there exists $N \in [l,\Delta]$ such that (3.15) holds. Let $\tilde{N} := K + N$. Then $l \leq N \leq \tilde{N} \leq K + \Delta = K + \Delta_{b,\theta}(l,k) = \tilde{\Delta}_{b,L,\beta,\theta}(l,k)$ and $x_{\tilde{N}} = z_N$, so

$$\|x_{\tilde{N}} - Tx_{\tilde{N}}\| = \|z_N - Tz_N\| \leq \frac{1}{k+1}.$$

□

Remark 3.8.11. *If C is bounded, then, obviously, the above theorem holds with $b \in \mathbb{N}$ being an upper bound on the diameter of C.*

We get some immediate consequences.

Corollary 3.8.12. *In the hypotheses of the above theorem, $\Delta'_{b,\theta} : \mathbb{N} \to \mathbb{N}$ is an approximate fixed point bound (with respect to T) for (z_n) and $\tilde{\Delta}'_{b,L,\beta,\theta} : \mathbb{N} \to \mathbb{N}$ is an approximate fixed point bound for (x_n), where*

$$\begin{aligned} \Delta'_{b,\theta}(k) &:= \Delta_{b,\theta}(0,k) = \theta(M), \text{ and} \\ \tilde{\Delta}'_{b,L,\beta,\theta}(k) &:= \tilde{\Delta}_{b,L,\beta,\theta}(0,k) = K + \theta(M). \end{aligned}$$

Proof. As indicated before, we may just let $l := 0$ in the above theorem. □

In the case when $t_n = s_n = \frac{1}{\sqrt{n+1}}$ we get a modulus of liminf of exponential growth.

Corollary 3.8.13. *In the hypotheses of the above theorem, assume further that $t_n = s_n = \frac{1}{\sqrt{n+1}}$. Then, for all $x \in C$, $\liminf_{n\to\infty} \|x_n - Tx_n\| = 0$ with modulus of liminf $\Gamma_{b,L}$, given by:*

$$\Gamma_{b,L}(l,k) := \left(\left\lceil 1 + \sqrt{2L^2+4}\right\rceil + 1\right)^2 + 4^{l+2(b^2+1)(k+1)^2}.$$

Proof. One can easily see that $\beta(k) := (k+1)^2$ is a rate of convergence for the sequence $\left(s_n = \frac{1}{\sqrt{n+1}}\right)$ and that $\theta(n) := 4^n$ is a rate of divergence for the sequence $\left(t_n s_n = \frac{1}{n+1}\right)$. □

Corollary 3.8.14. *In the hypotheses of the above theorem, we have that for all $x \in C$, $\liminf_{n\to\infty} \|x_n - x_{n+1}\| = 0$ with modulus of liminf $\hat{\Delta}_{b,L,\beta,\theta}$, given by:*

$$\hat{\Delta}_{b,L,\beta,\theta}(l,k) := \tilde{\Delta}_{b,L,\beta,\theta}(l,k'),$$

where $k' := \lceil (1+L)(1+k) \rceil$.

Proof. We know that there is an $N \in [l, \hat{\Delta}_{b,L,\beta,\theta}(l,k')]$ such that $\|x_N - Tx_N\| \leq \frac{1}{k'+1}$. Applying Lemma 3.8.2, we get that

$$\|x_N - x_{N+1}\| \leq (1+L)\|x_N - Tx_N\| \leq \frac{1+L}{k'+1} \leq \frac{1}{k+1},$$

which was what we needed to show. □

3.8.3 A rate of metastability

We may now begin to apply the results obtained in [50] and expounded upon in the last section.

As pointed out in [50, Lemma 7.1], if T is a uniformly continuous mapping, then $F := Fix(T)$ is uniformly closed with respect to the canonical decomposition $\{AF_k\}_{k\in\mathbb{N}}$, i.e. for each k,

$$AF_k := \{p \mid \|p - Tp\| \leq \frac{1}{k+1}\},$$

with moduli $\omega_F(k) = \max\{4k+3, \omega_T(4k+3)\}$ and $\delta_F(k) = 2k+1$, where ω_T is a modulus of uniform continuity of T – that is, a mapping $\omega_T : \mathbb{N} \to \mathbb{N}$ such that

$$\|p-q\| \leq \frac{1}{\omega_T(k)+1} \quad \text{implies} \quad \|Tp-Tq\| \leq \frac{1}{k+1}$$

for all $k \in \mathbb{N}$ and all $p, q \in C$.

Proposition 3.8.15. *Assume that T is an L-Lipschitzian pseudocontraction with $F \neq \emptyset$. Then F is a uniformly closed with respect to $\{AF_k\}_{k\in\mathbb{N}}$ with moduli*

$$\omega_F(k) = \lceil L \rceil (4k+4) \quad \textit{and} \quad \delta_F(k) = 2k+1.$$

Proof. Since T is L-Lipschitzian, it follows immediately that T is uniformly continuous with modulus $\omega_T(k) = \lceil L \rceil (k+1)$. □

Proposition 3.8.16. *Let $C \subseteq H$ be a bounded convex subset, $T : C \to C$ be an L-Lipschitzian pseudocontraction with $F \neq \emptyset$ and $b \in \mathbb{N}$ be an upper bound on the diameter of C. Assume that (t_n), (s_n) satisfy (A1) and (A3) and that β is a rate of convergence of (s_n). Then (z_n) is uniformly (G, H)-Fejér monotone with respect to $\{AF_k\}_{k\in\mathbb{N}}$ with modulus*

$$\chi_b(n, m, r) = 8bm(r+1),$$

where $G(a) = H(a) = a^2$. We note that $\alpha_G(k) = \left\lceil \sqrt{k} \right\rceil$ is a G-modulus for G and that $\beta_H(k) = (k+1)^2$ is a H-modulus for H.

Proof. Let $n, m, r \in \mathbb{N}, l \leq m$ and $p \in C$ be such that $\|p - Tp\| \leq \frac{1}{\chi(n,m,r)+1} = \frac{1}{8bm(r+1)+1}$. As a consequence of (3.13), we get that

$$\|z_{n+1} - p\|^2 \leq \|z_n - p\|^2 + 8b\|p - Tp\|. \tag{3.17}$$

It follows that

$$\begin{aligned} \|z_{n+l} - p\|^2 &\leq \|z_n - p\|^2 + 8bl\|p - Tp\| \quad \text{(by induction from (3.17))} \\ &\leq \|z_n - p\|^2 + 8bm\|p - Tp\| \leq \|z_n - p\|^2 + \frac{8bm}{8bm(r+1)+1} \\ &< \|z_n - p\|^2 + \frac{1}{r+1}. \end{aligned}$$

□

We may now proceed to give the main result of this section, namely a finitary, quantitative version of Theorem 3.2.6. Its proof may be found in the last subsubsection.

Theorem 3.8.17. *Let H be a Hilbert space, $C \subseteq H$ a nonempty totally bounded convex subset, $T : C \to C$ an L-Lipschitzian pseudocontraction with $F := Fix(T) \neq \emptyset$, (t_n), (s_n) sequences in $[0,1]$ satisfying (A1)-(A3) and (x_n) be the Ishikawa iteration starting with $x \in C$. Assume, furthermore, that γ is a modulus of total boundedness for C, $b \in \mathbb{N}$ is an upper bound on the diameter of C, β is a rate of convergence of (s_n) and θ is a rate of divergence of $\sum_{n=0}^{\infty} t_n s_n$.*

Let $\Sigma_{b,\theta,\gamma,\beta,L}$ and $\Omega_{b,\theta,\gamma,\beta,L} : \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$ be defined as in Table 3.1. Then

1. $\Sigma_{b,\theta,\gamma,\beta,L}$ *is a rate of metastability for* (x_n).
2. *There exists* $N \leq \Omega_{b,\theta,\gamma,\beta,L}(k,g)$ *such that*

$$\forall i,j \in [N, N+g(N)] \ \left(\|x_i - x_j\| \leq \frac{1}{k+1} \text{ and } \|x_i - Tx_i\| \leq \frac{1}{k+1} \right).$$

$$\begin{aligned}
&\Sigma_{b,\theta,\gamma,\beta,L}(k,g) := K + \tilde{\Sigma}_{b,\theta,\gamma}(k,h),\\
&\quad \tilde{\Sigma}_{b,\theta,\gamma} : \mathbb{N}\times\mathbb{N}^{\mathbb{N}} \to \mathbb{N}, \quad \tilde{\Sigma}_{b,\theta,\gamma}(k,g) := (\tilde{\Sigma}_0)_{b,\theta}(P,k,g),\\
&\quad (\tilde{\Sigma}_0)_{b,\theta} : \mathbb{N}\times\mathbb{N}\times\mathbb{N}^{\mathbb{N}} \to \mathbb{N}, \quad (\tilde{\Sigma}_0)_{b,\theta}(0,k,g) := 0,\\
&\quad (\tilde{\Sigma}_0)_{b,\theta}(n+1,k,g) := \theta^M\left(2(b^2+1)\big(Zg^M\left((\tilde{\Sigma}_0)_{b,\theta}(n,k,g)\right)+1\big)^2\right),\\
&\Omega_{b,\theta,\gamma,\beta,L}(k,g) := K + \tilde{\Omega}_{b,\theta,\gamma,L}(k,h),\\
&\quad \tilde{\Omega}_{b,\theta,\gamma,L} : \mathbb{N}\times\mathbb{N}^{\mathbb{N}} \to \mathbb{N}, \quad \tilde{\Omega}_{b,\theta,\gamma,L}(k,g) := (\tilde{\Omega}_0)_{b,\theta,L}(P_0,k,g),\\
&\quad (\tilde{\Omega}_0)_{b,\theta,L} : \mathbb{N}\times\mathbb{N}\times\mathbb{N}^{\mathbb{N}} \to \mathbb{N}, \quad (\tilde{\Omega}_0)_{b,\theta,L}(0,k,g) := 0,\\
&\quad (\tilde{\Omega}_0)_{b,\theta,L}(n+1,k,g) := \theta^M\Big(2(b^2+1)\big(\max\{2k+1, Z_0 g^M((\tilde{\Omega}_0)_{b,\theta,L}(n,k,g))\}+1\big)^2\Big),\\
&K := \beta\left(\lceil 1+\sqrt{2L^2+4}\rceil\right), \qquad h(n) := g(K+n), \qquad Z := 8b(8k^2+16k+10),\\
&P := \gamma\left(\left\lceil \sqrt{8k^2+16k+9}\right\rceil\right), \qquad k_0 := \left\lceil \frac{\lceil L\rceil(4k+4)-1}{2}\right\rceil,\\
&P_0 := \gamma\left(\left\lceil \sqrt{8k_0^2+16k_0+9}\right\rceil\right), \qquad Z_0 := 8b(8k_0^2+16k_0+10).
\end{aligned}$$

Table 3.1: Functionals and constants.

Theorem 3.8.17.(i) gives us a highly uniform rate of metastability $\Sigma_{b,\theta,\gamma,\beta,L}$, which depends only on the Lipschitz constant L, an upper bound b on the diameter of C and a modulus of total boundedness γ for C, and the rates β, θ associated to the sequences $(t_n), (s_n)$. As an immediate consequence, we get the Cauchyness of (x_n) for totally bounded convex C. Using [50, Remark 5.5], we may see that Theorem 3.8.17.(ii) is indeed the true finitization of Ishikawa's original statement, i.e. it implies back not only the convergence of the iterative sequence, but also the fact that its limit point is a fixed point of T.

Corollary 3.8.18. *In the hypotheses of the above theorem, assume further that* $t_n = s_n = \frac{1}{\sqrt{n+1}}$. *Then there exists* $N \leq \Omega'_{b,\gamma,L}(k,g)$ *such that*

$$\forall i,j \in [N, N+g(N)] \ \left(\|x_i - x_j\| \leq \frac{1}{k+1} \text{ and } \|x_i - Tx_i\| \leq \frac{1}{k+1} \right),$$

where $\Omega'_{b,\gamma,L}(k,g) := K_0 + (\Omega'_0)_{b,L}(P_0,k,h)$, with $K_0 := \left(\left\lceil 1+\sqrt{2L^2+4}\right\rceil + 1\right)^2$,

$$\begin{aligned}
(\Omega'_0)_{b,L}(0,k,g) &:= 0,\\
(\Omega'_0)_{b,L}(n+1,k,g) &:= 4^{2(b^2+1)\left(\max\{2k+1, 8b(8k_0^2+16k_0+10)g^M((\Omega'_0)_{b,L}(n,k,g))\}+1\right)^2}.
\end{aligned}$$

and h, P_0, k_0 as in Table 3.1.

Proof. Use the moduli from Corollary 3.8.13. □

3.8.3.1 Proof of Theorem 3.8.17

1. **Claim:** $\tilde{\Sigma}_{b,\theta,\gamma}$ is a rate of metastability for (z_n).

   Proof of claim: By Proposition 3.8.16, (z_n) is uniformly (G, H)-Fejér monotone w.r.t. F with modulus
$$\chi_b(n, m, r) = 8bm(r + 1),$$
where $G(a) = H(a) = a^2$ with moduli
$$\alpha_G(k) = \left\lceil \sqrt{k} \right\rceil \quad \text{and} \quad \beta_H(k) = (k + 1)^2.$$
Define $\Phi : \mathbb{N} \to \mathbb{N}$ by
$$\Phi(k) := \theta^M(2(b^2 + 1)(k + 1)^2) \tag{3.18}$$
Then Φ is nondecreasing and Φ is an an approximate fixed point bound for (z_n) by Corollary 3.8.12 and the fact that $\Phi(k) \geq \theta(2(b^2 + 1)(k + 1)^2)$ for all k.

   We may now apply Theorem 3.7.6 for F and (z_n). Using the notations from [50, Theorem 5.1], we get in our setting that
$$\chi_g(n, k) = 8(k + 1)bg(n), \quad \chi_g^M(n, k) = 8(k + 1)bg^M(n),$$
$$P = \gamma\left(\left\lceil \sqrt{8k^2 + 16k + 9} \right\rceil\right)$$
and
$$\Psi_0(0, k, g, \Phi, \chi, \beta_H) = 0$$
$$\Psi_0(n+1, k, g, \Phi, \chi, \beta_H) = \theta^M\Big(2(b^2 + 1)(Zg^M(\Psi_0(n, k, g, \Phi, \chi, \beta_H)) + 1)^2\Big).$$
By definition, we have that $\Psi_0(n, k, g, \Phi, \chi, \beta_H) = (\tilde{\Sigma}_0)_{b,\theta}(n, k, g)$. It follows that
$$\Psi(k, g, \Phi, \chi, \alpha_G, \beta_H, \gamma) = \Psi_0(P, k, g, \Phi, \chi, \beta_H) = (\tilde{\Sigma}_0)_{b,\theta}(P, k, g) = \tilde{\Sigma}_{b,\theta,\gamma}(k, g).$$
Thus, the claim is proven. ■

Let $k \in \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$ be arbitrary. Applying the claim, we get $N \leq \tilde{\Sigma}_{b,\theta,\gamma}(k, h)$ such that for all $i, j \in [N, N + h(N)] = [N, N + g(K + N)]$,
$$\|z_i - z_j\| \leq \frac{1}{k + 1}.$$
Define $\tilde{N} := K + N$. Then $\tilde{N} \leq K + \tilde{\Sigma}_{b,\theta,\gamma}(k, h) = \Sigma_{b,\theta,\gamma,\beta,L}(k, g)$ and $x_{\tilde{N}} = z_N$. Let $i, j \in [\tilde{N}, \tilde{N} + g(\tilde{N})] = [K + N, K + N + g(K + N)]$ and take $i_0 := i - K, j_0 := j - K$. Then $i_0, j_0 \in [N, N + g(K + N)]$ and so:
$$\|x_i - x_j\| = \|x_{i_0+K} - x_{j_0+K}\| = \|z_{i_0} - z_{j_0}\| \leq \frac{1}{k + 1}.$$

2. We apply now the second part of Theorem 3.7.6, i.e. [50, Theorem 5.3], for F and (z_n). Using the notations from [50, Theorem 5.3] and using Proposition 3.8.15 we get in our setting that

$$k_0 = \left\lceil \frac{\lceil L \rceil (4k+4) - 1}{2} \right\rceil, \quad \chi_{k,\delta_F}(n,m,r) = \max\{2k+1, \chi(n,m,r)\}$$

It follows that $\Psi_0(n, k_0, g, \Phi, \chi_{k,\delta_F}, \beta_H) = (\tilde{\Omega}_0)_{b,\theta,L}(n,k,g)$ and that the quantity

$$\Psi(k_0, g, \Phi, \chi_{k,\delta_F}, \alpha_G, \beta_H, \gamma)$$

is equal to:

$$\Psi_0(P_0, k_0, g, \Phi, \chi_{k,\delta_F}, \beta_H) = (\tilde{\Omega}_0)_{b,\theta,L}(P_0, k, g) = \tilde{\Omega}_{b,\theta,\gamma,L}(k,g).$$

Thus, we have obtained that for all $k \in \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$, there exists $N \leq \tilde{\Omega}_{b,\theta,\gamma,L}(k,g)$ such that

$$\forall i,j \in [N, N+g(N)] \ \left(\|z_i - z_j\| \leq \frac{1}{k+1} \text{ and } \|z_i - Tz_i\| \leq \frac{1}{k+1} \right).$$

As in (i), one gets immediately that (ii) holds.

Chapter 4

The proximal point algorithm

The proximal point algorithm is a fundamental tool of convex optimization, going back to Martinet [74], Rockafellar [79] and Brézis and Lions [12]. Since its inception, the schema turned out to be highly versatile, covering in its various developments, *inter alia*, the problems of finding zeros of monotone operators, minima of convex functions and fixed points of nonexpansive mappings. For a general introduction to the field in the context of Hilbert spaces, see the book of Bauschke and Combettes [8].

We pursue two goals in this chapter, corresponding to its two sections. The first is to derive general and abstract versions of the proximal point algorithm that subsume all the known instances that bear the name and to try to identify the most natural formulations among them. The second is to identify abstractly a subcase that appears repeatedly, that we dub the "uniform case", and to apply a variant of Kohlenbach's technique (mentioned in the last chapter as the "third way" to get around the main logical obstacle of the convergence property) to derive full and surprisingly uniform rates of convergence. Alongside these, we will also derive some minor pieces of quantitative information related to the algorithm.

These results can be found in [64, 66].

4.1 Abstract forms of the algorithm

We shall fix a complete CAT(0) space X in this and all the subsequent sections.

4.1.1 The general conditions

We first state the most general, albeit unnatural, conditions which imply convergence for the algorithm.

Theorem 4.1.1 (General Proximal Point Algorithm)**.** *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of self-mappings such that $F := \bigcap_{n\in\mathbb{N}} Fix(T_n) \neq \emptyset$. Let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in X$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$. Assume that:*

(i) the T_n's all satisfy the (P_2) property (in particular, they may all be firmly nonexpansive);

(ii) for all $n, m \in \mathbb{N}$ and $w \in X$, $d(T_n w, T_m w) \leq \frac{|\gamma_n - \gamma_m|}{\gamma_n} d(w, T_n w)$;

(iii) the sequence $\left(\frac{d(x_n, x_{n+1})}{\gamma_n}\right)_{n\in\mathbb{N}}$ is nonincreasing.

Then the sequence $(x_n)_{n\in\mathbb{N}}$ Δ-converges to an element of F.

We shall now begin a series of propositions that will culminate in a proof of the above theorem. Fix, then, a family $(T_n : X \to X)_{n\in\mathbb{N}}$ and a sequence $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$. Denote by F the set of common fixed points of all the T_n's and suppose that it is nonempty. Let $x \in X$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$. Assume that the three conditions from the theorem above hold.

Lemma 4.1.2. *All the T_n's have the same set of fixed points (which is equal to F).*

Proof. Follows immediately from the second condition: if w is a fixed point of a T_n and $m \in \mathbb{N}$, then

$$d(w, T_m w) =\leq \frac{|\gamma_n - \gamma_m|}{\gamma_n} d(w, w) = 0.$$

□

Lemma 4.1.3. *For all $z \in F$ and all $n \in \mathbb{N}$, we have that:*

$$d^2(x_{n+1}, z) \leq d^2(x_n, z) - d^2(x_n, x_{n+1}).$$

Proof. Let $z \in F$ and $n \in \mathbb{N}$. Since T_n satisfies the (P_2) property and $z \in Fix(T_n)$, by (1.4), we have that:

$$\begin{aligned} d^2(x_{n+1}, z) &= d^2(T_n x_n, z) \\ &\leq d^2(x_n, z) - d^2(x_n, T_n x_n) \\ &= d^2(x_n, z) - d^2(x_n, x_{n+1}). \end{aligned}$$

□

Corollary 4.1.4. *The sequence $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t. F.*

Choose now a $z \in F$. Let $b \in \mathbb{N}$ be such that $d(x, z) \leq b$.

Corollary 4.1.5. *For all $n \in \mathbb{N}$, we have that $d(x_n, z) \leq b$.*

Corollary 4.1.6. *We have that $\sum_{n=0}^{\infty} d^2(x_n, x_{n+1}) \leq b^2$.*

Proof. Let $N \in \mathbb{N}$. Then:

$$\sum_{n=0}^{N} d^2(x_n, x_{n+1}) \leq \sum_{n=0}^{N} (d^2(x_n, z) - d^2(x_{n+1}, z)) = d^2(x_0, z) - d^2(x_{N+1}, z) \leq d^2(x, z) \leq b^2.$$

□

Corollary 4.1.7. *We have that* $\lim_{n\to\infty} d^2(x_n, x_{n+1}) = 0$*, and hence that* $\lim_{n\to\infty} d(x_n, x_{n+1}) = 0$*.*

Proof. By the fact that $\sum_{n=0}^{\infty} d^2(x_n, x_{n+1})$ is convergent. □

Suppose, from now on, that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$ with rate of divergence $\theta : \mathbb{N} \to \mathbb{N}$, i.e. for all $K \in \mathbb{N}$ we have that

$$\sum_{n=0}^{\theta(K)} \gamma_n^2 \geq K.$$

Proposition 4.1.8. *We have that* $\lim_{n\to\infty} \frac{d(x_n,x_{n+1})}{\gamma_n} = 0$*. More precisely, for all* $\varepsilon > 0$ *and all* $n \geq \theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)$*, we have that* $\frac{d(x_n,x_{n+1})}{\gamma_n} \leq \varepsilon$*.*

Proof. Let $\varepsilon > 0$. Assume, by way of contradiction, that for all $i \in \left\{0, ..., \theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)\right\}$, one has that $\frac{d(x_i,x_{i+1})}{\gamma_i} > \varepsilon$. Then:

$$\sum_{i=0}^{\theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)} d^2(x_i, x_{i+1}) = \sum_{i=0}^{\theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)} \gamma_i^2 \cdot \frac{d^2(x_i, x_{i+1})}{\gamma_i^2} > \sum_{i=0}^{\theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)} \gamma_i^2 \varepsilon^2 \geq \varepsilon^2 \left\lceil \frac{b^2}{\varepsilon^2} \right\rceil \geq b^2,$$

contradicting Corollary 4.1.6. Hence we get that there is an $i \in \left\{0, ..., \theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)\right\}$ such that $\frac{d(x_i,x_{i+1})}{\gamma_i} \leq \varepsilon$. Since $\left(\frac{d(x_n,x_{n+1})}{\gamma_n}\right)_{n\in\mathbb{N}}$ is nonincreasing, for all $n \geq \theta\left(\left\lceil \frac{b^2}{\varepsilon^2} \right\rceil\right)$ one has, since $n \geq i$, that $\frac{d(x_n,x_{n+1})}{\gamma_n} \leq \frac{d(x_i,x_{i+1})}{\gamma_i} \leq \varepsilon$. □

Lemma 4.1.9. *For all* $m \in \mathbb{N}$*, we have that* $\lim_{n\to\infty} d(x_n, T_m x_n) = 0$*.*

Proof. Let $m \in \mathbb{N}$. We have that for all $n \in \mathbb{N}$:

$$\begin{aligned} d(x_n, T_m x_n) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, T_m x_n) \\ &= d(x_n, x_{n+1}) + d(T_n x_n, T_m x_n) \\ &\leq d(x_n, x_{n+1}) + \frac{|\gamma_n - \gamma_m|}{\gamma_n} d(x_n, T_n x_n) \\ &\leq 2d(x_n, x_{n+1}) + \gamma_m \cdot \frac{d(x_n, x_{n+1})}{\gamma_n}. \end{aligned}$$

Our conclusion follows by applying Corollary 4.1.7 and Proposition 4.1.8. □

Lemma 4.1.10. *For all* $m \in \mathbb{N}$ *and* $w \in X$*, we have that:*

$$\limsup_{n\to\infty} (d(x_n, T_m w) - d(x_n, w)) \leq 0.$$

Proof. Let $m \in \mathbb{N}$ and $w \in X$. Since T_m satisfies the (P_2) property, T_m is nonexpansive, so we have that:

$$\begin{aligned} d(x_n, T_m w) - d(x_n, w) &\leq d(x_n, T_m x_n) + d(T_m x_n, T_m w) - d(x_n, w) \\ &\leq d(x_n, T_m x_n). \end{aligned}$$

Applying Lemma 4.1.9, our conclusion follows. □

Proposition 4.1.11. *Let $w \in X$ be such that there is a subsequence $(u_n)_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$, having w as its Δ-limit. Then $w \in F$.*

Proof. Let $m \in \mathbb{N}$. Take $(v_n)_{n\in\mathbb{N}}$ to be a subsequence of $(u_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} d(v_n, T_m w) = \liminf_{n\to\infty} d(u_n, T_m w).$$

Then, by Lemma 4.1.10, we have that:

$$\limsup_{n\to\infty}(d(v_n, T_m w) - d(v_n, w)) \leq 0.$$

and so that:

$$\liminf_{n\to\infty}(d(v_n, T_m w) - d(v_n, w)) \leq 0.$$

Then we get that:

$$\begin{aligned}
0 &\geq \liminf_{n\to\infty}(d(v_n, T_m w) - d(v_n, w)) \\
&\geq \liminf_{n\to\infty} d(v_n, T_m w) - \limsup_{n\to\infty} d(v_n, w) \\
&= \lim_{n\to\infty} d(v_n, T_m w) - \limsup_{n\to\infty} d(v_n, w) \\
&= \limsup_{n\to\infty} d(v_n, T_m w) - \limsup_{n\to\infty} d(v_n, w).
\end{aligned}$$

Hence, we have that:

$$\limsup_{n\to\infty} d(v_n, T_m w) \leq \limsup_{n\to\infty} d(v_n, w).$$

But since $\mathcal{A}((v_n)_{n\in\mathbb{N}}) = \{w\}$, we have that $T_m w = w$.

Since m was chosen arbitrarily, we have that $w \in F$. □

We are now in a position to prove our main result.

Proof of Theorem 4.1.1. Having proven Proposition 4.1.11, we simply note that applying Proposition 1.2.6 and Corollary 4.1.4 yields our conclusion. □

4.1.2 Jointly firmly nonexpansive families

The following set of definitions represent, in our opinion, the most natural general conditions under which we can talk meaningfully about the proximal point algorithm. In addition, all known variants of the algorithm (that we shall see in the next subsection) fall under the strongest of the definitions here (the "jointly firmly nonexpansive" one), while the weakest of them (the "jointly (P_2)" one) implies the even weaker conditions considered above.

Definition 4.1.12. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. We say that the family $(T_n)_{n\in\mathbb{N}}$ is* **jointly firmly nonexpansive with respect to** *the sequence $(\gamma_n)_{n\in\mathbb{N}}$ if for all $n, m \in \mathbb{N}$, $x, y \in X$ and all $\alpha, \beta \in [0,1]$ such that $(1-\alpha)\gamma_n = (1-\beta)\gamma_m$ we have that:*

$$d(T_n x, T_m y) \leq d((1-\alpha)x + \alpha T_n x, (1-\beta)y + \beta T_m y).$$

Proposition 4.1.13. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then each T_n is firmly nonexpansive.*

Proof. We apply the condition for $m = n$ and remark that we may then take for any $t \in (0,1)$, $\alpha = \beta = t$. The condition then becomes what we need to show. □

The following property is the analogue of the (P_2) condition, as shown by the two propositions that follow.

Definition 4.1.14. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. We say that the family $(T_n)_{n\in\mathbb{N}}$ is* **jointly** (P_2) **with respect to** *the sequence $(\gamma_n)_{n\in\mathbb{N}}$ if for all $n,m \in \mathbb{N}$ and all $x,y \in X$ we have that:*

$$\frac{1}{\gamma_n}(d^2(x,T_my) - d^2(x,T_nx) - d^2(T_nx,T_my)) \geq \frac{1}{\gamma_m}(d^2(T_nx,T_my) + d^2(y,T_my) - d^2(y,T_nx)).$$

Proposition 4.1.15. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then each T_n satisfies the (P_2) property.*

Proof. Again, we apply the condition for $m = n$. □

Proposition 4.1.16. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$.*

Proof. Let $m,n \in \mathbb{N}$ and $x,y \in X$. We need to show that:

$$\frac{1}{\gamma_n}(d^2(x,T_my) - d^2(x,T_nx) - d^2(T_nx,T_my)) \geq \frac{1}{\gamma_m}(d^2(T_nx,T_my) + d^2(y,T_my) - d^2(y,T_nx)).$$

Let $\alpha \in \left(1 - \min\left\{\frac{\gamma_m}{\gamma_n}, 1\right\}, 1\right)$ be arbitrarily chosen. Set $\beta_\alpha := 1 - (1-\alpha)\frac{\gamma_n}{\gamma_m}$.

Then we have that indeed $\beta_\alpha \in (0,1)$, $(1-\alpha)\gamma_n = (1-\beta_\alpha)\gamma_m$ and (writing β instead of β_α):

$$d^2(T_nx,T_my) \leq d^2((1-\alpha)x + \alpha T_nx, (1-\beta)y + \beta T_my).$$

Applying the inequality in the definition of a CAT(0) space twice (once for each argument), we eventually obtain that:

$$\begin{aligned} d^2(T_nx,T_my) \leq &(1-\alpha)(1-\beta)d^2(x,y) + (1-\beta)\alpha d^2(T_nx,y) + (1-\alpha)\beta d^2(x,T_my) + \\ &+ \alpha\beta d^2(T_nx,T_my) - \alpha(1-\alpha)d^2(x,T_nx) - \beta(1-\beta)d^2(y,T_my). \end{aligned}$$

If we divide the above by $1-\alpha \neq 0$, we get that:

$$\begin{aligned} \frac{1-\alpha\beta}{1-\alpha}d^2(T_nx,T_my) \leq &(1-\beta)d^2(x,y) + \frac{(1-\beta)\alpha}{1-\alpha}d^2(T_nx,y) + \beta d^2(x,T_my) - \\ &- \alpha d^2(x,T_nx) - \frac{\beta(1-\beta)}{1-\alpha}d^2(y,T_my). \end{aligned}$$

We may easily compute that:

$$\frac{1-\alpha\beta}{1-\alpha} = 1 + \alpha\frac{\gamma_n}{\gamma_m}$$

$$\frac{(1-\beta)\alpha}{1-\alpha} = \alpha\frac{\gamma_n}{\gamma_m}$$

$$\frac{\beta(1-\beta)}{1-\alpha} = \left(1-(1-\alpha)\frac{\gamma_n}{\gamma_m}\right)\frac{\gamma_n}{\gamma_m}$$

Therefore, by letting $\alpha \to 1$ and taking note that $\beta = 1-(1-\alpha)\frac{\gamma_n}{\gamma_m}$, we get that:

$$\left(1+\frac{\gamma_n}{\gamma_m}\right)d^2(T_nx, T_my) \leq \frac{\gamma_n}{\gamma_m}d^2(T_nx, y) + d^2(x, T_my) - d^2(x, T_nx) - \frac{\gamma_n}{\gamma_m}d^2(y, T_my).$$

Dividing by γ_n, we obtain our required inequality. □

Using the mapping of Berg and Nikolaev introduced in the first chapter, we may express the jointly (P_2) condition as:

$$\frac{1}{\gamma_n}\langle\overrightarrow{T_nxT_my}, \overrightarrow{xT_nx}\rangle \geq \frac{1}{\gamma_m}\langle\overrightarrow{T_nxT_my}, \overrightarrow{yT_my}\rangle. \tag{4.1}$$

Proposition 4.1.17. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then any two mappings of the family have the same fixed points.*

Proof. It suffices to show that for all $m, n \in \mathbb{N}$ and $z \in Fix(T_n)$ we have that $z \in Fix(T_m)$. Let m, n, z be as such. By (4.1) and the fixed point property, we get that:

$$\frac{1}{\gamma_n}\langle\overrightarrow{zT_mz}, \overrightarrow{zz}\rangle \geq \frac{1}{\gamma_m}\langle\overrightarrow{zT_mz}, \overrightarrow{zT_mz}\rangle.$$

The left hand side is equal to 0 (by simple expansion or by applying properties (ii)-(iv) of Proposition 1.1.5) and the right hand side is equal to $d^2(z, T_mz)$ by the first axiom of Proposition 1.1.5. It follows that $T_mz = z$. □

Proposition 4.1.18. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then for all $m, n \in \mathbb{N}$ and all $w \in X$, we have that:*

$$d(T_nw, T_mw) \leq \frac{|\gamma_n - \gamma_m|}{\gamma_n}d(w, T_nw).$$

Proof. Let $m, n \in \mathbb{N}$. We shall denote, for simplicity, $T := T_n$, $U := T_m$, $\lambda := \gamma_n$, $\mu := \gamma_m$.

We want to show that for all w,

$$d(Tw, Uw) \leq \frac{|\lambda - \mu|}{\lambda}d(w, Tw).$$

If $Tw = Uw$, the statement is trivially true. Let, then, $w \in X$ be such that $Tw \neq Uw$.

We have that:

$$\frac{1}{\lambda}\langle\overrightarrow{TwUw},\overrightarrow{wTw}\rangle \geq \frac{1}{\mu}\langle\overrightarrow{TwUw},\overrightarrow{wUw}\rangle,$$

and by multiplying by $(-\lambda)$, we get that:

$$\langle\overrightarrow{TwUw},\overrightarrow{Tww}\rangle \leq \frac{\lambda}{\mu}\langle\overrightarrow{TwUw},\overrightarrow{Uww}\rangle. \tag{4.2}$$

By a simple expansion of $\langle\cdot,\cdot\rangle$, we can prove that:

$$d^2(Tw,Uw) = d^2(w,Uw) - d^2(w,Tw) + 2\langle\overrightarrow{TwUw},\overrightarrow{Tww}\rangle. \tag{4.3}$$

By exchanging the roles of T and U in the above equation, we obtain that:

$$d^2(Uw,Tw) = d^2(w,Tw) - d^2(w,Uw) + 2\langle\overrightarrow{UwTw},\overrightarrow{Uww}\rangle. \tag{4.4}$$

By (4.2) and (4.3), we get that:

$$d^2(Tw,Uw) \leq d^2(w,Uw) - d^2(w,Tw) + \frac{2\lambda}{\mu}\langle\overrightarrow{TwUw},\overrightarrow{Uww}\rangle.$$

Multiplying (4.4) by $\frac{\lambda}{\mu}$, we get that:

$$\frac{\lambda}{\mu}d^2(Uw,Tw) \leq \frac{\lambda}{\mu}d^2(w,Tw) - \frac{\lambda}{\mu}d^2(w,Uw) + \frac{2\lambda}{\mu}\langle\overrightarrow{UwTw},\overrightarrow{Uww}\rangle.$$

Summing the last two inequalities, we obtain that:

$$\left(1+\frac{\lambda}{\mu}\right)d^2(Tw,Uw) \leq \left(\frac{\lambda}{\mu}-1\right)(d^2(w,Tw)-d^2(w,Uw)),$$

and by multiplying by μ, that:

$$(\lambda+\mu)d^2(Tw,Uw) \leq (\lambda-\mu)(d^2(w,Tw)-d^2(w,Uw)).$$

We distinguish two cases, according to the sign of $\lambda-\mu$.

When $\lambda-\mu$ is negative, and therefore $|\lambda-\mu| = \mu-\lambda$, we have that:

$$\begin{aligned}(\lambda+\mu)d^2(Tw,Uw) &\leq (\lambda-\mu)(d^2(w,Tw)-d^2(w,Uw))\\ &= (\mu-\lambda)(d^2(w,Uw)-d^2(w,Tw))\\ &\leq (\mu-\lambda)((d(w,Tw)+d(Tw,Uw))^2-d^2(w,Tw))\\ &= (\mu-\lambda)(d^2(w,Tw)+2d(w,Tw)d(Tw,Uw)+d^2(Tw,Uw)-d^2(w,Tw))\\ &= (\mu-\lambda)d(Tw,Uw)(d(Tw,Uw)+2d(w,Tw)).\end{aligned}$$

Dividing by $d(Tw,Uw)\neq 0$, we obtain that:

$$(\lambda+\mu)d(Tw,Uw) \leq 2(\mu-\lambda)d(w,Tw) + (\mu-\lambda)d(Tw,Uw),$$

so that:

$$2\lambda d(Tw,Uw) \leq 2(\mu-\lambda)d(w,Tw)$$

and

$$d(Tw, Uw) \leq \frac{\mu - \lambda}{\lambda} d(w, Tw) = \frac{|\lambda - \mu|}{\lambda} d(w, Tw).$$

Now, when $\lambda - \mu$ is positive, and hence $|\lambda - \mu| = \lambda - \mu$, we proceed as follows. Since, by the reverse triangle inequality for metric spaces,

$$d(w, Uw) \geq \big|d(Tw, Uw) - d(w, Tw)\big|,$$

we have that:

$$d^2(w, Uw) \geq d^2(Tw, Uw) - 2d(w, Tw)d(Tw, Uw) + d^2(w, Tw).$$

We obtain:

$$\begin{aligned}(\lambda + \mu)d^2(Tw, Uw) &\leq (\lambda - \mu)(d^2(w, Tw) - d^2(w, Uw)) \\ &\leq (\lambda - \mu)(d^2(w, Tw) - d^2(Tw, Uw) + 2d(w, Tw)d(Tw, Uw) - d^2(w, Tw)) \\ &= (\lambda - \mu)d(Tw, Uw)(2d(w, Tw) - d(Tw, Uw)).\end{aligned}$$

Again, dividing by $d(Tw, Uw) \neq 0$, we get that:

$$(\lambda + \mu)d(Tw, Uw) \leq 2(\lambda - \mu)d(w, Tw) - (\lambda - \mu)d(Tw, Uw),$$

so that:

$$2\lambda d(Tw, Uw) \leq 2(\lambda - \mu)d(w, Tw)$$

and

$$d(Tw, Uw) \leq \frac{\lambda - \mu}{\lambda} d(w, Tw) = \frac{|\lambda - \mu|}{\lambda} d(w, Tw).$$

□

Proposition 4.1.19. *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Let $x \in X$ and set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$.*

Then the sequence $\left(\frac{d(x_n, x_{n+1})}{\gamma_n}\right)_{n\in\mathbb{N}}$ is nonincreasing.

Proof. Let $n \in \mathbb{N}$. By the joint (P_2) property, we get that:

$$\frac{1}{\gamma_n}\langle \overrightarrow{T_n x_n T_{n+1} x_{n+1}}, \overrightarrow{x_n T_n x_n}\rangle \geq \frac{1}{\gamma_{n+1}}\langle \overrightarrow{T_n x_n T_{n+1} x_{n+1}}, \overrightarrow{x_{n+1} T_{n+1} x_{n+1}}\rangle,$$

so:

$$\frac{1}{\gamma_n}\langle \overrightarrow{x_{n+1} x_{n+2}}, \overrightarrow{x_n x_{n+1}}\rangle \geq \frac{1}{\gamma_{n+1}}\langle \overrightarrow{x_{n+1} x_{n+2}}, \overrightarrow{x_{n+1} x_{n+2}}\rangle.$$

We obtain that (using at the second inequality the Cauchy-Schwarz property):

$$\begin{aligned}0 &\leq \frac{1}{\gamma_n}\langle \overrightarrow{x_{n+1} x_{n+2}}, \overrightarrow{x_n x_{n+1}}\rangle - \frac{d^2(x_{n+1}, x_{n+2})}{\gamma_{n+1}} \\ &= \gamma_{n+1}\left(\frac{1}{\gamma_n \gamma_{n+1}}\langle \overrightarrow{x_{n+1} x_{n+2}}, \overrightarrow{x_n x_{n+1}}\rangle - \frac{d^2(x_{n+1}, x_{n+2})}{\gamma_{n+1}^2}\right) \\ &\leq \gamma_{n+1}\left(\frac{d(x_n, x_{n+1})}{\gamma_n} \cdot \frac{d(x_{n+1}, x_{n+2})}{\gamma_{n+1}} - \frac{d^2(x_{n+1}, x_{n+2})}{\gamma_{n+1}^2}\right) \\ &= \gamma_{n+1} \cdot \frac{d(x_{n+1}, x_{n+2})}{\gamma_{n+1}}\left(\frac{d(x_n, x_{n+1})}{\gamma_n} - \frac{d(x_{n+1}, x_{n+2})}{\gamma_{n+1}}\right).\end{aligned}$$

If $\frac{d(x_{n+1},x_{n+2})}{\gamma_{n+1}} \neq 0$, $\frac{d(x_n,x_{n+1})}{\gamma_n} - \frac{d(x_{n+1},x_{n+2})}{\gamma_{n+1}} \geq 0$, so $\frac{d(x_{n+1},x_{n+2})}{\gamma_{n+1}} \leq \frac{d(x_n,x_{n+1})}{\gamma_n}$. Otherwise, if $\frac{d(x_{n+1},x_{n+2})}{\gamma_{n+1}} = 0$, clearly $\frac{d(x_{n+1},x_{n+2})}{\gamma_{n+1}} \leq \frac{d(x_n,x_{n+1})}{\gamma_n}$. □

Theorem 4.1.20 (Abstract Proximal Point Algorithm)**.** *Let $(T_n : X \to X)_{n\in\mathbb{N}}$ and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$ (in particular, they may be jointly firmly nonexpansive). Suppose that the common fixed point set of the T_n's is nonempty and that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in X$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ Δ-converges to an element of F.*

Proof. We seek to apply Theorem 4.1.1. We need to check that the three conditions are satisfied. But that is guaranteed by Proposition 4.1.15, Proposition 4.1.18 and Proposition 4.1.19. □

4.1.3 Concrete instances

A recent breakthrough in the study of proximal point iterations was achieved by Bačák [4], who proved the weak convergence in CAT(0) spaces (that is, Δ-convergence) of the variant of the algorithm used to find minima of convex, lower semicontinuous (lsc) functions. For reasons of generality, we shall study his recent result before moving on to older ones. Let us detail the statement of his theorem.

For any convex, lsc function $f : X \to (-\infty, +\infty]$, define its *Moreau-Yosida resolvent* or its *proximal point mapping*, $J_f : X \to X$, for any $x \in X$, as:

$$J_f(x) := \operatorname{argmin}_{y\in X} \left[f(y) + \frac{1}{2} d^2(x,y) \right].$$

We usually consider the resolvent of f of order $\gamma > 0$, which is simply the resolvent of γf, namely:

$$J_{\gamma f}(x) = \operatorname{argmin}_{y\in X} \left[\gamma f(y) + \frac{1}{2} d^2(x,y) \right] = \operatorname{argmin}_{y\in X} \left[f(y) + \frac{1}{2\gamma} d^2(x,y) \right].$$

This definition (albeit without that factor of 2) first appeared in [38]. The argmin is unique for convex lsc functions, and so the operator is indeed well-defined, as witnessed by [38, Lemma 2].

We establish the following basic property.

Proposition 4.1.21. *Let $f : X \to (-\infty, +\infty]$ be convex lsc and $x \in X$. Then f attains its minimum in x iff x is a fixed point of J_f.*

Proof. Suppose first that x is a minimizer of f, i.e. for any $y \in X$, $f(x) \leq f(y)$. But that means that for all y,

$$f(x) + \frac{1}{2} d^2(x,x) \leq f(y) + \frac{1}{2} d^2(x,y),$$

therefore x is also the argmin of the right hand side w.r.t. y – that is, $J_f(x)$.

Suppose now that $J_f(x) = x$. Then for all $y \in X$, as before,

$$f(x) \leq f(y) + \frac{1}{2} d^2(x,y).$$

Let $w \in X$. The for any $t \in (0,1)$ we have the following (using the fact that f is convex):

$$\begin{aligned} f(x) &\leq f((1-t)x+tw) + \frac{1}{2}d^2(x,(1-t)x+tw) \\ &\leq (1-t)f(x) + tf(w) + \frac{1}{2}t^2d^2(x,w). \end{aligned}$$

Subtracting $(1-t)f(x)$ and dividing by t, we obtain that:

$$f(x) \leq f(w) + \frac{1}{2}td^2(x,w),$$

and by letting $t \to 0$, we obtain $f(x) \leq f(w)$. Since w was chosen arbitrarily, we get that x is a minimizer of f. □

Now, if $\gamma > 0$, since f has the same minimizers as γf, it follows that $Fix(J_{\gamma f})$ is also the set of minimizers of f.

The following two propositions were proven in [38] (and they do not depend on the lack of that extra factor of 2).

Proposition 4.1.22 ([38, Lemma 4])**.** *Let $f : X \to (-\infty, +\infty]$ be convex lsc. Then for any $\gamma > 0$, $J_{\gamma f}$ is nonexpansive.*

Proposition 4.1.23 ([38, Corollary 1])**.** *Let $f : X \to (-\infty, +\infty]$ be convex lsc. Then for any $\gamma > 0$, any $x \in X$ and any $t \in [0,1]$, we have that:*

$$J_{(1-t)\gamma f}((1-t)x + tJ_{\gamma f}(x)) = J_{\gamma f}(x).$$

The following proposition establishes the link with the previous sections.

Proposition 4.1.24. *Let $f : X \to (-\infty, +\infty]$ be convex lsc and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Then the family $(J_{\gamma_n f})_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n\in\mathbb{N}}$.*

Proof. Let $m, n \in \mathbb{N}$, $x, y \in X$ and $\alpha, \beta \in [0,1]$ be such that

$$(1-\alpha)\gamma_n = (1-\beta)\gamma_m =: \delta.$$

Then, using Proposition 4.1.22 and Proposition 4.1.23 and the fact that :

$$\begin{aligned} d(J_{\gamma_n f}x, J_{\gamma_m f}y) &= d(J_{(1-\alpha)\gamma_n f}((1-\alpha)x + \alpha J_{\gamma_n f}x), J_{(1-\beta)\gamma_m f}((1-\beta)y + \beta J_{\gamma_m f}y)) \\ &= d(J_{\delta f}((1-\alpha)x + \alpha J_{\gamma_n f}x), J_{\delta f}((1-\beta)y + \beta J_{\gamma_m f}y)) \\ &\leq d((1-\alpha)x + \alpha J_{\gamma_n f}x, (1-\beta)y + \beta J_{\gamma_m f}y). \end{aligned}$$

□

We may now derive the convergence theorem for a concrete proximal point algorithm.

Theorem 4.1.25 (Proximal Point Algorithm for Convex Lsc Functions)**.** *Let $f : X \to (-\infty, +\infty]$ be a convex lsc function that has at least one minimizer and let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in X$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := J_{\gamma_n f}x_n$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ Δ-converges to a minimizer of f.*

Proof. For all n, put $T_n := J_{\gamma_n f}$. Since, on one hand, by the remarks above, all T_n's have as fixed point set the set of the minimizers of f and on the other hand, by Proposition 4.1.24, the family $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n\in\mathbb{N}}$, we may apply Theorem 4.1.20 to derive our conclusion. □

The above is a slightly weaker variant, with a different proof, of Bačák's result [4, Theorem 1.4], since we used the stronger assumption $\sum_{n=0}^{\infty}\gamma_n^2 = \infty$ instead of $\sum_{n=0}^{\infty}\gamma_n = \infty$.

We give another application. Let $T : X \to X$ be a nonexpansive mapping. For all $x \in X$ and $\gamma > 0$ we define the auxiliary mapping $G_{T,x,\gamma} : X \to X$, for any $y \in X$, by:

$$G_{T,x,\gamma}(y) := \frac{1}{1+\gamma}x + \frac{\gamma}{1+\gamma}y.$$

This mapping can easily be seen to be a Lipschitzian contraction of constant $\frac{\gamma}{1+\gamma} \in (0,1)$. Therefore it admits a unique fixed point, which we shall denote by $R_{T,\gamma}x$. We have therefore defined a mapping $R_{T,\gamma} : X \to X$, called the *resolvent of order* γ of T, which satisfies, for any $x \in X$:

$$R_{T,\gamma}x = \frac{1}{1+\gamma}x + \frac{\gamma}{1+\gamma}TR_{T,\gamma}x.$$

Proposition 4.1.26. *Let $T : X \to X$ be a nonexpansive mapping and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Then the family $(R_{T,\gamma_n})_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n\in\mathbb{N}}$.*

Proof. Let $m, n \in \mathbb{N}$, $x, y \in X$ and $\alpha, \beta \in [0,1]$ be such that

$$(1-\alpha)\gamma_n = (1-\beta)\gamma_m =: \delta.$$

We need to show that:

$$d(R_{T,\gamma_n}x, R_{T,\gamma_m}y) \leq d((1-\alpha)x + \alpha R_{T,\gamma_n}x, (1-\beta)y + \beta R_{T,\gamma_m}y).$$

Put

$$u := (1-\alpha)x + \alpha R_{T,\gamma_n}x$$

and

$$v := (1-\beta)y + \beta R_{T,\gamma_m}y.$$

Therefore we must show that:

$$d(R_{T,\gamma_n}x, R_{T,\gamma_m}y) \leq d(u,v).$$

Now, since

$$u = (1-\alpha)x + \alpha R_{T,\gamma_n}x$$

and

$$R_{T,\gamma_n}x = \frac{1}{1+\gamma_n}x + \frac{\gamma_n}{1+\gamma_n}TR_{T,\gamma_n}x,$$

we might apply Lemma 1.1.4 to obtain that:

$$R_{T,\gamma_n}x = (1-\nu)u + \nu TR_{T,\gamma_n}x,$$

where

$$\nu = \frac{(1-\alpha)\frac{\gamma_n}{1+\gamma_n}}{1-\alpha\cdot\frac{\gamma_n}{1+\gamma_n}} = \frac{\delta}{1+\delta}.$$

We remark that $\nu \neq 1$. Similarly, we might show that:

$$R_{T,\gamma_m} y = (1-\nu)v + \nu T R_{T,\gamma_m} y.$$

We now have, applying (1.1) and the nonexpansiveness of T, that:

$$\begin{aligned} d(R_{T,\gamma_n} x, R_{T,\gamma_m} y) &= d((1-\nu)u + \nu T R_{T,\gamma_n} x, (1-\nu)v + \nu T R_{T,\gamma_m} y) \\ &\leq (1-\nu)d(u,v) + \nu d(T R_{T,\gamma_n} x, T R_{T,\gamma_m} y)) \\ &\leq (1-\nu)d(u,v) + \nu d(R_{T,\gamma_n} x, R_{T,\gamma_m} y)). \end{aligned}$$

Therefore we have obtained that:

$$(1-\nu)d(R_{T,\gamma_n} x, R_{T,\gamma_m} y) \leq (1-\nu)d(u,v),$$

and by dividing by $1-\nu \neq 0$, we obtain what we wanted to show. □

Theorem 4.1.27 (Proximal Point Algorithm for Nonexpansive Mappings)**.** *Let $T : X \to X$ be a nonexpansive mapping that has at least one fixed point and let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in X$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := R_{T,\gamma_n} x_n$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ Δ-converges to a fixed point of T.*

Proof. For all n, put $T_n := R_{T,\gamma_n}$. Since, on one hand, is immediate that for any $\gamma > 0$, $Fix(R_{T,\gamma}) = Fix(T)$ and on the other hand, by Proposition 4.1.26, the family $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n\in\mathbb{N}}$, we may apply Theorem 4.1.20 to derive our conclusion. □

We have therefore obtained a new proof of a result of Bačák and Reich [5, Proposition 1.5].

4.1.4 The case of Hilbert spaces

We will now focus on peculiarities of our results specific to Hilbert spaces. Fix now, for this section, a Hilbert space H. Recall, first, the following basic identities (cf. Lemma 3.8.3).

Lemma 4.1.28. *Let H be a Hilbert space and $x, y, z \in H$. Then:*

(i) $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$;

(ii) $\langle z-y, x-z\rangle + \frac{1}{2}\|z-x\|^2 = \frac{1}{2}(\|y-x\|^2 - \|y-z\|^2)$.

Proof. (i) Immediate.

(ii) We have that:

$$\begin{aligned} \|y-x\|^2 &= \|x-y\|^2 \\ &= \|(z-y)+(x-z)\|^2 \\ &= \|z-y\|^2 + \|x-z\|^2 + 2\langle z-y, x-z\rangle \\ &= \|y-z\|^2 + \|x-z\|^2 + 2\langle z-y, x-z\rangle \end{aligned}$$

Rearranging the terms and dividing by two, we obtain the desired equality.

□

As per the first chapter, we know that a Hilbert space is a particular case of a CAT(0) space, that here firm nonexpansiveness coincides with the (P_2) property and that Δ-convergence reduces to weak convergence.

We now recall the following result.

Proposition 4.1.29 (e.g. [8, Theorem 5.10])**.** *Let $(x_n)_{n\in\mathbb{N}} \subseteq H$ and $C \subseteq H$ such that $int(C) \neq \emptyset$ and $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to C. Then $(x_n)_{n\in\mathbb{N}}$ is a strongly convergent sequence.*

Hence we might give the following instance of our main theorem.

Theorem 4.1.30 (General Proximal Point Algorithm for Hilbert spaces)**.** *Let $(T_n : H \to H)_{n\in\mathbb{N}}$ be a family of self-mappings such that $F := \bigcap_{n\in\mathbb{N}} Fix(T_n) \neq \emptyset$. Let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in H$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$. Assume that:*

(i) the T_n's are all firmly nonexpansive;

(ii) for all $n, m \in \mathbb{N}$ and $w \in X$, $\|T_n w - T_m w\| \leq \frac{|\gamma_n - \gamma_m|}{\gamma_n} \|w - T_n w\|$;

(iii) the sequence $\left(\frac{\|x_n - x_{n+1}\|}{\gamma_n}\right)_{n\in\mathbb{N}}$ is nonincreasing.

Then:

(a) the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to an element of F;

(b) if $int(F) \neq \emptyset$, the convergence is strong.

Proof. Everything above follows from Theorem 4.1.1 and Proposition 4.1.29. In addition, we show that independently of the statement of (a) we have that in the case (b) necessarily the limit is an element of F. Let x^* be the strong limit of $(x_n)_{n\in\mathbb{N}}$. Let $m \in \mathbb{N}$ and $\varepsilon > 0$. Let, then, $n \in \mathbb{N}$ be such that $d(x_n, x^*) \leq \varepsilon$ and, by Lemma 4.1.9, such that $d(x_n, T_m x_n) \leq \varepsilon$. Then:

$$\begin{aligned} d(x^*, T_m x^*) &\leq d(x^*, x_n) + d(x_n, T_m x_n) + d(T_m x_n, T_m x^*) \\ &\leq d(x^*, x_n) + d(x_n, T_m x_n) + d(x_n, x^*) \\ &\leq 3\varepsilon. \end{aligned}$$

Since ε was chosen arbitrarily, we have that $x^* \in Fix(T_m)$. Since m was chosen arbitrarily, we have that $x^* \in F$. □

Another reminder from the first chapter is the following relation: that for any $a, b, c, d \in H$ we have that:

$$\langle \overrightarrow{ab}, \overrightarrow{cd} \rangle = \langle a - b, c - d \rangle = \langle b - a, d - c \rangle.$$

It may also be proven, as follows, that joint firm nonexpansiveness coincides with the joint (P_2) property. This is the more general version of the proof that we omitted in the first chapter (where we gave reference to [8]) of the statement that a (P_2) mapping in a Hilbert space is firmly nonexpansive.

Proposition 4.1.31. *Let $(T_n : H \to H)_{n\in\mathbb{N}}$ be a family of mappings and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$. Suppose that $(T_n)_{n\in\mathbb{N}}$ is jointly (P_2) with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Then $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$.*

Proof. Let $m, n \in \mathbb{N}$, $x, y \in X$ and $\alpha, \beta \in [0,1]$ be such that

$$(1-\alpha)\gamma_n = (1-\beta)\gamma_m =: \delta.$$

We know that:

$$\frac{1}{\gamma_n}\langle T_n x - T_m y, x - T_n x\rangle \geq \frac{1}{\gamma_m}\langle T_n x - T_m y, y - T_m y\rangle. \tag{4.5}$$

We want to show that:

$$\|T_n x - T_m y\| \leq \|((1-\alpha)x + \alpha T_n x) - ((1-\beta)y + \beta T_m y)\|.$$

But since, by a simple computation, we have the following two identities:

$$(1-\alpha)x + \alpha T_n x = T_n x + \delta\frac{x - T_n x}{\gamma_n},$$

$$(1-\beta)y + \beta T_m y = T_m y + \delta\frac{y - T_m y}{\gamma_m},$$

we may write that:

$$\begin{aligned}\|((1-\alpha)x + \alpha T_n x) - ((1-\beta)y + \beta T_m y)\|^2 &= \left\|(T_n x - T_m y) + \delta\left(\frac{x - T_n x}{\gamma_n} - \frac{y - T_m y}{\gamma_m}\right)\right\|^2 \\ &= \|T_n x - T_m y\|^2 + \delta^2\left\|\frac{x - T_n x}{\gamma_n} - \frac{y - T_m y}{\gamma_m}\right\|^2 \\ &\quad + 2\delta\left\langle T_n x - T_m y, \frac{x - T_n x}{\gamma_n} - \frac{y - T_m y}{\gamma_m}\right\rangle.\end{aligned}$$

In order to show that the right hand side is greater than $\|T_n x - T_m y\|^2$, which is what we are aiming to prove here, it is sufficient to show that

$$\left\langle T_n x - T_m y, \frac{x - T_n x}{\gamma_n} - \frac{y - T_m y}{\gamma_m}\right\rangle \geq 0.$$

But that is true by (4.5). □

We may now state the following instance of Theorem 4.1.20.

Theorem 4.1.32 (Abstract Proximal Point Algorithm for Hilbert spaces)**.** *Let $(T_n : H \to H)_{n\in\mathbb{N}}$ and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to the sequence $(\gamma_n)_{n\in\mathbb{N}}$. Suppose that the common fixed point set of the T_n's is nonempty and that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in H$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := T_n x_n$. Then:*

(a) the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to an element of F;

(b) if $int(F) \neq \emptyset$, the convergence is strong.

It is clear that Theorems 4.1.25 and 4.1.27 adapt as well. We now present an application that is specific to Hilbert spaces. This may be considered the classical case of the algorithm: finding zeros of maximally monotone operators (a problem that we alluded to in the first chapter when discussing pseudocontractions).

A *multi-valued operator* on H is a function $A : H \to 2^H$. A *zero* of it will be an element $x \in H$ such that $0 \in A(x)$. We call such an operator $A : H \to 2^H$ *monotone* if for all $x, y, u, v \in H$ such that $u \in A(x)$ and $v \in A(y)$ we have that:

$$\langle x - y, u - v \rangle \geq 0.$$

In addition, we call it *maximally monotone* when it is maximal among the set of all monotone operators represented as binary relations $A \subseteq H \times H$ and ordered by set-theoretic inclusion. If we define the *resolvent* of A as the relation

$$J_A := (id + A)^{-1}$$

where the operations are those involving binary relations, we know from the general theory of monotone operators that when A is maximally monotone, J_A is a single-valued function with the whole of H as its domain. Also, it is easy to check that the zeros of A coincide with the fixed points of J_A. We shall usually consider the resolvent of A of order $\gamma > 0$, which is simply the resolvent of γA. Such an operation preserves both the zeros, on the side of the monotone operator, and the fixed points, on the side of the resolvent.

Lemma 4.1.33. *Let $A : H \to 2^H$ be a maximally monotone operator and $\gamma > 0$. Then, for all $x \in H$,*

$$\frac{x - J_{\gamma A} x}{\gamma} \in A(J_{\gamma A} x).$$

Proof. Let $x \in H$. Since $J_{\gamma A} x = (id + \gamma A)^{-1} x$, $x \in (id + \gamma A)(J_{\gamma A} x)$. We obtain, successively, that $x \in J_{\gamma A} x + \gamma A(J_{\gamma A} x)$, that $x - J_{\gamma A} x \in \gamma A(J_{\gamma A} x)$ and that $\frac{x - J_{\gamma A} x}{\gamma} \in A(J_{\gamma A} x)$. □

Proposition 4.1.34. *Let $A : H \to 2^H$ be a maximally monotone operator and $(\gamma_n)_{n \in \mathbb{N}} \subseteq (0, \infty)$. Then the family $(J_{\gamma_n A})_{n \in \mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n \in \mathbb{N}}$.*

Proof. Let $n, m \in \mathbb{N}$ and $x, y \in H$. We have, by Lemma 4.1.33 that:

$$\frac{x - J_{\gamma_n A} x}{\gamma_n} \in A(J_{\gamma_n A} x)$$

and

$$\frac{y - J_{\gamma_m A} y}{\gamma_m} \in A(J_{\gamma_m A} y).$$

By the monotonicity of A we obtain that:

$$\left\langle J_{\gamma_n A} x - J_{\gamma_m A} y, \frac{x - J_{\gamma_n A} x}{\gamma_n} - \frac{y - J_{\gamma_m A} y}{\gamma_m} \right\rangle \geq 0,$$

therefore:

$$\frac{1}{\gamma_n} \langle J_{\gamma_n A} x - J_{\gamma_m A} y, x - J_{\gamma_n A} x \rangle \geq \frac{1}{\gamma_m} \langle J_{\gamma_n A} x - J_{\gamma_m A} y, y - J_{\gamma_m A} y \rangle.$$

□

We may now derive the convergence theorem for the corresponding proximal point algorithm.

Theorem 4.1.35 (Proximal Point Algorithm for Maximally Monotone Operators)**.** *Let $A : H \to 2^H$ be a maximally monotone operator that has at least one zero and let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ be such that $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Let $x \in H$. Set $x_0 := x$ and for all $n \in \mathbb{N}$, $x_{n+1} := J_{\gamma_n A} x_n$. Then:*

(a) the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a zero of A;

(b) if $int(F) \neq \emptyset$, the convergence is strong.

Proof. For all n, put $T_n := J_{\gamma_n A}$. Since, on one hand, by the introductory remarks above, all T_n's have as fixed point set the set of the zeros of A and on the other hand, by Proposition 4.1.34, the family $(T_n)_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to $(\gamma_n)_{n\in\mathbb{N}}$, we may apply Theorem 4.1.32 to derive our conclusion. □

We have derived in the process the classical theorem which is presented in [8, Theorem 23.41.(i)].

4.2 Quantitative results

We shall now gather some general quantitative information that we may use later. Fix the given data of Theorem 4.1.1 as we did in the section proving it. We reformulate the quantitative result that we have already obtained in Proposition 4.1.8.

Proposition 4.2.1. *Let $\Sigma_{b,\theta} : \mathbb{N} \to \mathbb{N}$ be defined, for all k, by $\Sigma_{b,\theta}(k) := \theta(b^2(k+1)^2)$. For all $k \in \mathbb{N}$ and all $n \geq \Sigma_{b,\theta}(k)$, we have that*

$$\frac{d(x_n, x_{n+1})}{\gamma_n} \leq \frac{1}{k+1}.$$

The following result shall be useful in a later subsection.

Proposition 4.2.2. *Let $\Delta_b : \mathbb{N} \to \mathbb{N}$ be defined, for all k, by $\Delta_b(k) := b^2(k+1)^2$. Then for all $k \in \mathbb{N}$ there is an $n \leq \Delta_b(k)$ such that*

$$d(x_n, x_{n+1}) \leq \frac{1}{k+1}.$$

Proof. Suppose, on the contrary, that there is a k such that for all $n \leq \Delta_b(k)$ we have that $d(x_n, x_{n+1}) \geq \frac{1}{k+1}$. Then, using Corollary 4.1.6, we have that:

$$b^2 \geq \sum_{n=0}^{\Delta_b(k)} d^2(x_n, x_{n+1}) \geq \sum_{n=0}^{\Delta_b(k)} \frac{1}{(k+1)^2} = \frac{\Delta_b(k)+1}{(k+1)^2} > b^2,$$

which is a contradiction. □

We now derive a quantitative version of Proposition 4.1.29. This will give us a rate of metastability for our sequence $(x_n)_{n\in\mathbb{N}}$ in the case of Theorem 4.1.30.(b).

Proposition 4.2.3. *Let H be a Hilbert space, $(x_n)_{n\in\mathbb{N}} \subseteq H$ and $C \subseteq H$. Assume that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone w.r.t. C and that there is a $p^* \in C$ and $r > 0$ such that the closed ball of center p^* and radius r is contained in C. Also, we shall presume that this p^* is the point z chosen earlier, so that $\|x - p^*\| \leq b$. For any $g : \mathbb{N} \to \mathbb{N}$, define $\chi_g : \mathbb{N} \to \mathbb{N}$ recursively, as follows:*

$$\chi_g(0) := 0$$

$$\chi_g(n+1) := \chi_g(n) + g(\chi_g(n))$$

Set now, for any $k \in \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$,

$$\Phi_{b,r}(k,g) := \chi_g\left(\left\lceil \frac{b^2(k+1)}{2r} \right\rceil\right).$$

Then $\Phi_{b,r}$ is a rate of metastability for $(x_n)_{n\in\mathbb{N}}$.

Proof. Suppose, on the contrary, that there are $k \in \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$ such that for all $n \leq \Phi_{b,r}(k,g)$ there are i_n, j_n with $n \leq i_n < j_n \leq n + g(n)$ and $\|x_{i_n} - x_{j_n}\| > \frac{1}{k+1}$.

We claim first that for any $m \in \mathbb{N}$ we have that:

$$\|x_m - x_{m+1}\| \leq \frac{1}{2r}(\|p^* - x_m\|^2 - \|p^* - x_{m+1}\|^2)$$

To prove that, first set $h := \frac{1}{\|x_m - x_{m+1}\|}(x_m - x_{m+1})$. Note that $\|h\| = 1$, so $\|p^* + rh - p^*\| = r$. We have, then, that $p^* + rh \in C$. By Lemma 4.1.28.(ii), we have that:

$$\langle x_{m+1} - (p^* + rh), x_m - x_{m+1}\rangle + \frac{1}{2}\|x_{m+1} - x_m\|^2 = \frac{1}{2}(\|(p^* + rh) - x_m\|^2 - \|(p^* + rh) - x_{m+1}\|^2) \geq 0$$

at the last inequality applying the Fejér monotonicity. But then:

$$r\langle h, x_m - x_{m+1}\rangle \leq \langle x_{m+1} - p^*, x_m - x_{m+1}\rangle + \frac{1}{2}\|x_{m+1} - x_m\|^2 = \frac{1}{2}(\|p^* - x_m\|^2 - \|p^* - x_{m+1}\|^2)$$

having applied again Lemma 4.1.28.(ii) at the last equality.

But then:

$$\langle h, x_m - x_{m+1}\rangle \leq \frac{1}{2r}(\|p^* - x_m\|^2 - \|p^* - x_{m+1}\|^2),$$

which was what we sought to show, since $\langle h, x_m - x_{m+1}\rangle = \|x_m - x_{m+1}\|$. The claim being proven, we have that:

$$\begin{aligned}
\frac{1}{k+1} &< \|x_{i_n} - x_{j_n}\| \\
&= \left\| \sum_{m=i_n}^{j_n-1} (x_m - x_{m+1}) \right\| \\
&\leq \sum_{m=i_n}^{j_n-1} \|x_m - x_{m+1}\| \\
&\leq \frac{1}{2r} \sum_{m=i_n}^{j_n-1} (\|p^* - x_m\|^2 - \|p^* - x_{m+1}\|^2) \\
&= \frac{1}{2r}(\|p^* - x_{i_n}\|^2 - \|p^* - x_{j_n}\|^2).
\end{aligned}$$

It follows that:

$$\|p^* - x_{i_n}\|^2 - \|p^* - x_{j_n}\|^2 > \frac{2r}{k+1}.$$

We have, then, by the nonincreasingness of $(\|p^* - x_k\|^2)_{k\in\mathbb{N}}$, that, for all $n \leq \Phi_{b,r}(k,g)$:

$$\|p^* - x_n\|^2 - \|p^* - x_{n+g(n)}\|^2 > \frac{2r}{k+1}.$$

In particular, since we have that for any n, $\chi_g(n+1) = \chi_g(n) + g(\chi_g(n))$, we can write:

$$\|p^* - x_{\chi_g(n)}\|^2 - \|p^* - x_{\chi_g(n+1)}\|^2 > \frac{2r}{k+1},$$

for all n such that $\chi_g(n) \leq \Phi_{b,r}(k,g) = \chi_g\left(\left\lceil \frac{b^2(k+1)}{2r} \right\rceil\right)$ – that is, for all $n \leq \left\lceil \frac{b^2(k+1)}{2r} \right\rceil$, since the function χ_g is nondecreasing. Summing these relations for $n \in \{0, \ldots, \left\lceil \frac{b^2(k+1)}{2r} \right\rceil - 1\}$, we obtain that:

$$\sum_{n=0}^{\left\lceil \frac{b^2(k+1)}{2r} \right\rceil - 1} (\|p^* - x_{\chi_g(n)}\|^2 - \|p^* - x_{\chi_g(n+1)}\|^2) > \frac{2r}{k+1} \left\lceil \frac{b^2(k+1)}{2r} \right\rceil$$

But then, noting that the sum is telescoping and that $x_{\chi_g(0)} = x_0 = x$ and $x_{\chi_g\left(\left\lceil \frac{b^2(k+1)}{2r} \right\rceil\right)} = x_{\Phi_{b,r}(k,g)}$, we have that:

$$\|p^* - x\|^2 - \|p^* - x_{\Phi_{b,r}(k,g)}\|^2 > \frac{2r}{k+1} \left\lceil \frac{b^2(k+1)}{2r} \right\rceil$$

so:

$$\begin{aligned}
\|p^* - x_{\Phi_{b,r}(k,g)}\|^2 &< \|p^* - x\|^2 - \frac{2r}{k+1} \left\lceil \frac{b^2(k+1)}{2r} \right\rceil \\
&\leq \|p^* - x\|^2 - \frac{2r}{k+1} \cdot \frac{b^2(k+1)}{2r} \\
&= \|p^* - x\|^2 - b^2 \\
&\leq b^2 - b^2 \\
&= 0,
\end{aligned}$$

which is a contradiction. □

4.2.1 The uniform case

Definition 4.2.4. *Let $T : X \to X$ and $C \subseteq X$ such that $T(C) \subseteq C$. Let $\phi : [0,\infty) \to [0,\infty)$ be an increasing function which vanishes only at 0. We say that T is* **uniformly firmly nonexpansive** *on C with modulus ϕ if for any $x, y \in C$ and all $t \in [0,1]$, we have that:*

$$d^2(Tx,Ty) \leq d^2((1-t)x + tTx, (1-t)y + tTy) - 2(1-t)\phi(d(Tx,Ty)).$$

Definition 4.2.5. *Let $T : X \to X$ and $C \subseteq X$ such that $T(C) \subseteq C$. Let $\phi : [0,\infty) \to [0,\infty)$ be an increasing function which vanishes only at 0. We say that T is* **uniformly** *(P_2) on C with modulus ϕ if for any $x, y \in C$, we have that:*

$$2d^2(Tx,Ty) \leq d^2(x,Ty) + d^2(y,Tx) - d^2(x,Tx) - d^2(y,Ty) - 2\phi(d(Tx,Ty)).$$

The uniformly (P_2) condition may be expressed using the Berg-Nikolaev quasi-linearization function as follows:

$$\langle \overrightarrow{TxTy}, \overrightarrow{xTx} \rangle \geq \langle \overrightarrow{TxTy}, \overrightarrow{yTy} \rangle + \phi(d(Tx, Ty)), \tag{4.6}$$

and in light of that, it is seen to be equivalent in the context of Hilbert spaces with the definition given in [7, Section 3.4].

Proposition 4.2.6. *Let $T : X \to X$ and $C \subseteq X$ such that $T(C) \subseteq C$. Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Suppose that T is uniformly firmly nonexpansive on C with modulus ϕ. Then T is uniformly (P_2) on C with the same modulus ϕ.*

Proof. Let $x, y \in C$. Applying twice, like in the proof of Proposition 4.1.16, the inequality in the definition of a CAT(0) space on the uniformly firm nonexpansiveness condition for a $t \in (0, 1)$, we obtain that:

$$\begin{aligned} d^2(Tx, Ty) \leq (1-t)^2 d^2(x, y) + t(1-t)d^2(Tx, y) + t(1-t)d^2(x, Ty) + t^2 d^2(Tx, Ty) \\ - t(1-t)d^2(x, Tx) - t(1-t)d^2(y, Ty) - 2(1-t)\phi(d(Tx, Ty)). \end{aligned}$$

Dividing by $1 - t \neq 0$, we get that:

$$\begin{aligned} (1+t)d^2(Tx, Ty) \leq (1-t)d^2(x, y) + td^2(Tx, y) + td^2(x, Ty) \\ - td^2(x, Tx) - td^2(y, Ty) - 2\phi(d(Tx, Ty)), \end{aligned}$$

and by taking $t \to 1$ we obtain what we needed. □

Proposition 4.2.7. *Let $T : X \to X$ and $C \subseteq X$ such that $T(C) \subseteq C$. Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Suppose that T is uniformly (P_2) on C with modulus ϕ. Let $x \in C$ and $z \in C \cap Fix(T)$. Then we have that:*

$$\phi(d(Tx, z)) \leq d(x, Tx)d(Tx, z).$$

Proof. Using the definition above, take $y := z$. So we get that

$$d^2(Tx, z) \leq d^2(x, z) - d^2(x, Tx) - 2\phi(d(Tx, z)).$$

Since $d^2(x, z) \leq (d(x, Tx) + d(Tx, z))^2 \leq d^2(x, Tx) + d^2(Tx, z) + 2d(x, Tx)d(Tx, z)$, we obtain the desired inequality. □

Corollary 4.2.8. *Let $T : X \to X$ and $C \subseteq X$ such that $T(C) \subseteq C$. Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Suppose that T is uniformly (P_2) on C with modulus ϕ. Then the set $C \cap Fix(T)$ is at most a singleton.*

Proof. In Proposition 4.2.7, by taking $x \in Fix(T)$, we obtain that $\phi(d(x, z)) = 0$. Since ϕ vanishes only at 0, we have that $d(x, z) = 0$. □

We now return to the framework established in Theorem 4.1.1 that we have fixed above. Remember that we have taken $b \in \mathbb{N}$ such that $d(x, z) \leq b$. Set C to be the closed ball of center z and radius b. By the results obtained in the previous sections, we have that for all n, $T_n(C) \subseteq C$.

Whereas C may have been just fixed, it is understood that the definitions in the next subsection, like the ones above, apply to a general C.

We shall impose in the sequel the condition that there is a $\phi : [0,\infty) \to [0,\infty)$, an increasing function which vanishes only at 0, such that for all n, T_n is uniformly (P_2) on C with modulus $\gamma_n\phi$. It is clear than in this case z is the unique fixed point of the T_n's in C and that it is the point to which the sequence $(x_n)_{n\in\mathbb{N}}$ weakly converges. In addition, we have the following result.

Theorem 4.2.9. *In these circumstances, the convergence is strong. Moreover, set, for any $k \in \mathbb{N}$,*

$$\Psi_{b,\theta,\phi}(k) := \Sigma_{b,\theta}\left(\left\lceil \frac{2b}{\phi\left(\frac{1}{k+1}\right)} \right\rceil\right) + 1,$$

where $\Sigma_{b,\theta}$ is the one from Proposition 4.2.1.

Then $\Psi_{b,\theta,\phi}$ is a rate of convergence for $(x_n)_{n\in\mathbb{N}}$.

Proof. Let $k \in \mathbb{N}$ and $n \geq \Psi_{b,\theta,\phi}(k)$. We must show that

$$d(x_n, z) \leq \frac{1}{k+1}.$$

Set $n' := n - 1$. Then

$$n' \geq \Sigma_{b,\theta}\left(\left\lceil \frac{2b}{\phi\left(\frac{1}{k+1}\right)} \right\rceil\right).$$

By Proposition 4.2.1, we have that:

$$\frac{d(x_{n'}, x_{n'+1})}{\gamma_{n'}} \leq \frac{1}{\left\lceil \frac{2b}{\phi\left(\frac{1}{k+1}\right)} \right\rceil + 1} \leq \frac{1}{\frac{2b}{\phi\left(\frac{1}{k+1}\right)}} = \frac{1}{2b} \cdot \phi\left(\frac{1}{k+1}\right).$$

Applying Proposition 4.2.7 for $x := x_{n'}$, $T := T_{n'}$ (and hence ϕ becomes $\gamma_{n'}\phi$), we get that:

$$\gamma_{n'}\phi(d(T_{n'}x_{n'}, z)) \leq d(x_{n'}, T_{n'}x_{n'})d(T_{n'}x_{n'}, z).$$

Since $x_{n'+1} = T_{n'}x_{n'}$, we have that:

$$\phi(d(x_{n'+1}, z)) \leq \frac{d(x_{n'}, x_{n'+1})}{\gamma_{n'}} \cdot d(x_{n'+1}, z) \leq \frac{1}{2b} \cdot \phi\left(\frac{1}{k+1}\right) \cdot b = \frac{1}{2}\phi\left(\frac{1}{k+1}\right).$$

Since ϕ is an increasing function which vanishes only at 0, we get that

$$d(x_{n'+1}, z) \leq \frac{1}{k+1},$$

which is what we wanted to show, since $n = n' + 1$. □

The result above is very surprising – not because it exhibits a full rate of convergence, since the technique is not fundamentally new – but because of the way the sequence of weights $(\gamma_n)_{n\in\mathbb{N}}$ disappears in the middle of the proof, even though it is not taken from a compact interval like the weight sequences from the previous chapter. It would be interesting to find out if there is a logical explanation behind this.

4.2.1.1 Applications

We shall now check that the conditions introduces above are indeed satisfied by non-trivial particular cases in the concrete instances that we have presented.

Firstly, we shall tackle Theorem 4.1.25.

Definition 4.2.10. *Let $\psi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. A function $f : X \to (-\infty, \infty]$ is called* **uniformly convex** *on C with modulus ψ if for all $x, y \in C$ and all $t \in [0, 1]$ we have that:*

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - t(1-t)\psi(d(x, y)).$$

Proposition 4.2.11. *Let $\psi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Let $f : X \to (-\infty, \infty]$ be a lsc function which is uniformly convex on C with modulus ψ. Let $\gamma > 0$ and $x, z \in C$. We assume that $J_{\gamma f}(C) \subseteq C$. Then we have that:*

$$d^2(J_{\gamma f}x, z) \leq d^2(x, z) - d^2(x, J_{\gamma f}x) - 2\gamma(f(J_{\gamma f}x) - f(z)) - 2\gamma\psi(d(z, J_{\gamma f}x)).$$

Proof. By the definition of $J_{\gamma f}$, we have that for all $a \in C$,

$$f(J_{\gamma f}x) + \frac{1}{2\gamma}d^2(x, J_{\gamma f}x) \leq f(a) + \frac{1}{2\gamma}d^2(x, a).$$

Let $t \in (0, 1)$ be arbitrary. Note that, by our definition of CAT(0) spaces:

$$d^2(x, (1-t)z + tJ_{\gamma f}x) \leq (1-t)d^2(x, z) + td^2(x, J_{\gamma f}x) - t(1-t)d^2(z, J_{\gamma f}x).$$

Applying the first inequality (multiplied by γ) on $a := (1-t)z + tJ_{\gamma f}x$ and using the above and the uniform convexity, we get that:

$$\begin{aligned}\gamma f(J_{\gamma f}x) + \frac{1}{2}d^2(x, J_{\gamma f}x) \leq &\gamma((1-t)f(z) + tf(J_{\gamma f}x) - t(1-t)\psi(d(z, J_{\gamma f}x))) \\ &+ \frac{1}{2}((1-t)d^2(x, z) + td^2(x, J_{\gamma f}x) - t(1-t)d^2(z, J_{\gamma f}x)).\end{aligned}$$

Dividing by $1 - t \neq 0$, we get that:

$$\gamma(f(J_{\gamma f}x) - f(z)) \leq \frac{1}{2}(d^2(x, z) - d^2(x, J_{\gamma f}x) - td^2(z, J_{\gamma f}x)) - \gamma t\psi(d(z, J_{\gamma f}x)).$$

By letting $t \to 1$, we obtain our conclusion. □

Corollary 4.2.12. *Let $\psi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Let $f : X \to (-\infty, \infty]$ be a lsc function which is uniformly convex on C with modulus ψ. We assume that $J_{\gamma f}(C) \subseteq C$. Let $\gamma > 0$ and $x, y \in C$. Then we have that:*

$$d^2(J_{\gamma f}x, J_{\gamma f}y) \leq d^2(x, y) - 4\gamma\psi(d(J_{\gamma f}x, J_{\gamma f}y)).$$

Proof. By taking in the above proposition $z := J_{\gamma f}y$, we obtain that:

$$d^2(J_{\gamma f}x, J_{\gamma f}y) \leq d^2(x, J_{\gamma f}y) - d^2(x, J_{\gamma f}x) - 2\gamma(f(J_{\gamma f}x) - f(J_{\gamma f}y)) - 2\gamma\psi(d(J_{\gamma f}x, J_{\gamma f}y)).$$

By interchanging the roles of x and y in the above, we get:

$$d^2(J_{\gamma f}y, J_{\gamma f}x) \leq d^2(y, J_{\gamma f}x) - d^2(y, J_{\gamma f}y) - 2\gamma(f(J_{\gamma f}y) - f(J_{\gamma f}x)) - 2\gamma\psi(d(J_{\gamma f}y, J_{\gamma f}x)).$$

Summing up, we obtain:

$$2d^2(J_{\gamma f}x, J_{\gamma f}y) + d^2(x, J_{\gamma f}x) + d^2(y, J_{\gamma f}y) \leq d^2(x, J_{\gamma f}y) + d^2(y, J_{\gamma f}x) - 4\gamma\psi(d(J_{\gamma f}y, J_{\gamma f}x)).$$

By [10, Theorem 6], we have that:

$$d^2(x, J_{\gamma f}y) + d^2(y, J_{\gamma f}x) \leq d^2(x, y) + d^2(J_{\gamma f}x, J_{\gamma f}y) + d^2(x, J_{\gamma f}x) + d^2(y, J_{\gamma f}y),$$

from which we get our conclusion. □

Theorem 4.2.13. *Let $\psi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Let $f : X \to (-\infty, \infty]$ be a lsc function which is uniformly convex on C with modulus ψ. Let $\gamma > 0$. We assume that $J_{\lambda f}(C) \subseteq C$ for all $\lambda > 0$. Then $J_{\gamma f}$ is uniformly firmly nonexpansive on C with modulus $2\gamma\psi$.*

Proof. Let $x, y \in C$ and $t \in [0, 1]$. By the corollary above and Proposition 4.1.23, we get:

$$\begin{aligned}
d^2(J_{\gamma f}x, J_{\gamma f}y) &= d^2(J_{(1-t)\gamma f}((1-t)x + tJ_{\gamma f}x), J_{(1-t)\gamma f}((1-t)y + tJ_{\gamma f}y)) \\
&\leq d^2((1-t)x + tJ_{\gamma f}x, (1-t)y + tJ_{\gamma f}y) \\
&\quad - 4(1-t)\gamma\psi(d(J_{(1-t)\gamma f}((1-t)x + tJ_{\gamma f}x), J_{(1-t)\gamma f}((1-t)y + tJ_{\gamma f}y))) \\
&= d^2((1-t)x + tJ_{\gamma f}x, (1-t)y + tJ_{\gamma f}y) \\
&\quad - 4(1-t)\gamma\psi(d(J_{\gamma f}x, J_{\gamma f}y)).
\end{aligned}$$

□

Therefore, for a lsc function which is uniformly convex on C with modulus ψ, we might supplement the result of Theorem 4.1.25 with the one of Theorem 4.2.9 above, taking $\phi := 2\psi$.

Fix now a Hilbert space H. We have the following analogue of the classical result.

Proposition 4.2.14. *Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Let $T : H \to H$ be uniformly (P_2) on C with modulus ϕ. Suppose that $T(C) \subseteq C$. Then T is uniformly firmly nonexpansive on C with the same modulus ϕ.*

Proof. Let $x, y \in C$ and $t \in [0, 1]$. We know that:

$$\langle Tx - Ty, x - Tx\rangle \geq \langle Tx - Ty, y - Ty\rangle + \phi(\|Tx - Ty\|).$$

We want to show that:

$$\|Tx - Ty\|^2 \leq \|((1-t)x + tTx) - ((1-t)y + tTy)\|^2 - 2(1-t)\phi(\|Tx - Ty\|).$$

We have that:

$$\begin{aligned}
\|((1-t)x + tTx) - ((1-t)y + tTy)\|^2 &= \|(Tx - Ty) + (1-t)((x - Tx) - (y - Ty))\|^2 \\
&= \|Tx - Ty\|^2 + (1-t)^2\|(x - Tx) - (y - Ty)\|^2 \\
&\quad + 2(1-t)\langle Tx - Ty, (x - Tx) - (y - Ty)\rangle \\
&\geq \|Tx - Ty\|^2 + 0 + 2(1-t)\phi(\|Tx - Ty\|),
\end{aligned}$$

which is what we had needed. □

Now we may tackle Theorem 4.1.35.

Definition 4.2.15 (e.g. [8, p. 323])**.** *Let $A : H \to 2^H$ be a multi-valued operator. Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. We call A* **uniformly monotone** *on C with modulus ϕ if for all $x, y \in C$ and $u, v \in H$ with $u \in A(x)$ and $v \in A(y)$ we have that:*

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

Theorem 4.2.16. *Let $\phi : [0, \infty) \to [0, \infty)$ be an increasing function which vanishes only at 0. Let $A : H \to 2^H$ be a maximally monotone operator which is uniformly monotone on C with modulus ϕ. Let $\gamma > 0$. Then $J_{\gamma A}$ is uniformly firmly nonexpansive on C with modulus $\gamma\phi$.*

Proof. Let $x, y \in C$. We have by Lemma 4.1.33 that:

$$\frac{x - J_{\gamma A}x}{\gamma} \in A(J_{\gamma A}x)$$

and

$$\frac{y - J_{\gamma A}y}{\gamma} \in A(J_{\gamma A}y).$$

By the uniform monotonicity of A we obtain that:

$$\left\langle J_{\gamma A}x - J_{\gamma A}y, \frac{x - J_{\gamma A}x}{\gamma} - \frac{y - J_{\gamma A}y}{\gamma}\right\rangle \geq \phi(\|J_{\gamma A}x - J_{\gamma A}y\|)$$

and therefore that:

$$\langle J_{\gamma A}x - J_{\gamma A}y, x - J_{\gamma A}x\rangle \geq \langle J_{\gamma A}x - J_{\gamma A}y, y - J_{\gamma A}y\rangle + \gamma\phi(\|J_{\gamma A}x - J_{\gamma A}y\|),$$

using (4.6) and Proposition 4.2.14. □

Therefore, for a maximally monotone operator which is uniformly monotone on C with modulus ϕ, we might supplement the result of Theorem 4.1.35 with the one of Theorem 4.2.9 above, with the same ϕ. We have established a quantitative version of [8, Theorem 23.41.(ii)].

We may finally remark that all the "uniform" variants of the various functions that we considered above have as a special case the "strong" variants, i.e. when the moduli ϕ or ψ that we have defined are functions of the form $t \mapsto \beta t^2$, where $\beta > 0$.

4.2.2 Convergence in finitely many steps

Still in the Hilbert space setting, we present a quantitative criterion that guarantees that the sequence reaches its limit in finitely many steps.

Theorem 4.2.17. *Suppose that there is an $\eta > 0$ such that for all $n \in \mathbb{N}$ and for all $w \in C$ with $\|w - z\| \leq \eta$ we have that $T_n w = z$. Then for all $n \geq \Delta_b\left(\left\lceil\frac{2}{\eta}\right\rceil\right) + 1$, we have that $x_n = z$, where Δ_b is the one from Proposition 4.2.2.*

Proof. From Proposition 4.2.2, we have that there is an $m \leq \Delta_b\left(\left\lceil \frac{2}{\eta} \right\rceil\right)$ such that

$$\|x_m - x_{m+1}\| \leq \frac{1}{\left\lceil \frac{2}{\eta} \right\rceil + 1} \leq \frac{\eta}{2}.$$

Now, set:

$$w := z + x_m - x_{m+1} + \frac{\eta}{2} \cdot \frac{x_{m+1} - z}{\|x_{m+1} - z\|}.$$

Therefore,

$$\|w - z\| \leq \|x_m - x_{m+1}\| + \frac{\eta}{2} \leq \eta,$$

and so $T_m w = z$. Using the definition of firmly nonexpansive mappings in Hilbert spaces, we have that:

$$\begin{aligned}
\|x_{m+1} - z\|^2 &= \|T_m x_m - T_m w\|^2 \\
&\leq \|x_m - w\|^2 - \|(id - T_m)x_m - (id - T_m)w\|^2 \\
&= \|x_m - w\|^2 - \|x_m - T_m x_m - w + z\|^2 \\
&= \|x_m - w\|^2 - \|T_m x_m - x_m + w - z\|^2 \\
&= \left\| x_{m+1} - z - \frac{\eta}{2} \cdot \frac{x_{m+1} - z}{\|x_{m+1} - z\|} \right\|^2 - \left\| \frac{\eta}{2} \cdot \frac{x_{m+1} - z}{\|x_{m+1} - z\|} \right\|^2 \\
&= \left(1 - \frac{\eta}{2\|x_{m+1} - z\|}\right)^2 \|x_{m+1} - z\|^2 - \frac{\eta^2}{4} \\
&= \|x_{m+1} - z\|^2 - \eta\|x_{m+1} - z\| + \frac{\eta^2}{4} - \frac{\eta^2}{4} \\
&= \|x_{m+1} - z\|^2 - \eta\|x_{m+1} - z\|,
\end{aligned}$$

from which is follows that $x_{m+1} = z$. Therefore $x_{m+2} = T_{m+1}x_{m+1} = T_{m+1}z = z$ and so on, inductively, we get that for all $n \geq m + 1$ we have that $x_n = z$. Since $m \leq \Delta_b\left(\left\lceil \frac{2}{\eta} \right\rceil\right)$, the conclusion follows. □

4.2.3 The totally bounded case

The techniques for deriving rates of metastability in the context of Fejér monotone sequences in totally bounded spaces were also applied to the proximal point algorithm for maximally monotone operators in Hilbert spaces in the last section of [50]. What we shall do here is to further explore the logical foundations of this result in order to give a treatment of Bačák's convergence theorem. We shall actually use the condition originally imposed for this case,

$$\sum_{i=0}^{\infty} \gamma_i = \infty,$$

instead of the "squared" one from the previous sections. Bačák's proof is what we are going to build upon, mainly, in our quantitative analysis from the viewpoint of proof mining.

The logical system for CAT(0) spaces is enriched by this assumption of total boundedness for the space in the following way. We add a constant α of type 0(0) to represent the modulus and the universal closure of the following formula as its defining sentence:

$$\exists I, J \leq_{\mathbb{N}} \alpha(k) \left(I <_{\mathbb{N}} J \wedge d_X(x_I, x_J) \leq_{\mathbb{R}} \frac{1}{k+1} \right).$$

Since the existential quantifiers are bounded, the sentence is considered to be purely universal and hence it does not affect in any way the extraction of the bound. The only difference will now be that the bound will also depend on the constant α.

There are some further modifications to the system which are required by the specific problem of the proximal point algorithm. Since what we are going to do is, roughly, to choose a point $x \in X$ and then, given a sequence of weights $(\gamma_n)_{n\in\mathbb{N}}$, iterate the algorithm in the form $x_0 := x$, $x_{n+1} := J_{\gamma_n f} x_n$, for any n, the sole modifications necessary to the system are to add a constant γ of type $1(0)$ and one J of type $X(X)(0)$, together with a defining axiom, the universal closure of:

$$\gamma(n) f(J(n)(x)) +_{\mathbb{R}} \frac{1}{2} d_X(x, J(n)(x))^2 \leq_{\mathbb{R}} \gamma(n) f(y) +_{\mathbb{R}} \frac{1}{2} d(x, y)^2.$$

Since the resolvents are all nonexpansive, the operator J is "bounded" in a certain technical way, and hence the extracted quantities will depend additionally only on the sequence $(\gamma_n)_{n\in\mathbb{N}}$.

We may now proceed to study the algorithm in itself. We fix a convex lsc function $f : X \to (-\infty, +\infty]$, an $x \in X$ and $\gamma = \{\gamma_n\}_{n\in\mathbb{N}} \in (0, \infty)$ such that $\sum_{n=0}^{\infty} \gamma_n = \infty$. We then iterate the proximal point algorithm from this data, i.e. we set $x_0 := x$ and for all n, $x_{n+1} := J_{\gamma_n f} x_n$. We aim to find the minimizers of f and so we set $F := Argmin(f)$. We remark that this set, and also the proximal mappings of any order, do not vary with f up to the addition of a constant, and therefore, as in [4], we may consider w.l.o.g. $\min(f) = 0$.

For all n, we have that $Fix(J_{\gamma_n f}) = F$, and so we have the following approximation:

$$F = \bigcap_{k\in\mathbb{N}} AF_k = \bigcap_{k\in\mathbb{N}} \left\{ x \in X \mid \text{for all } i \leq k,\ d(x, J_{\gamma_i f} x) \leq \frac{1}{k+1} \right\}.$$

This approximation will turn out to be convenient for the results we are aiming for.

Proposition 4.2.18. *With respect to the above approximation, F is uniformly closed with moduli $\delta_F(k) := 2k + 1$, $\omega_F(k) := 4k + 3$.*

Proof. Let $q \in AF_{2k+1}$. Then for all $i \leq 2k+1$, $d(q, J_{\gamma_i f} q) \leq \frac{1}{2k+2}$.

Let p be such that $d(p, q) \leq \frac{1}{4k+4}$. We need to show that for all $i \leq k$, $d(p, J_{\gamma_i f} p) \leq \frac{1}{k+1}$.

Let $i \leq k$. Then:

$$\begin{aligned} d(p, J_{\gamma_i f} p) &\leq d(p, q) + d(q, J_{\gamma_i f} q) + d(J_{\gamma_i f} q, J_{\gamma_i f} p) \\ &\leq d(p, q) + d(q, J_{\gamma_i f} q) + d(q, p) \\ &\leq \frac{1}{4k+4} + \frac{1}{2k+2} + \frac{1}{4k+4} \\ &= \frac{1}{k+1}, \end{aligned}$$

where we have used at the second inequality the fact that $J_{\gamma_i f}$ is nonexpansive. ∎

Lemma 4.2.19. *(i) For all $n \in \mathbb{N}$ and $p \in X$,*

$$d(x_{n+1}, p) \leq d(x_n, p) + d(p, J_{\gamma_n f} p).$$

(ii) For all $n, m \in \mathbb{N}$ and $p \in X$,

$$d(x_{n+m}, p) \leq d(x_n, p) + \sum_{i=n}^{i=n+m-1} d(p, J_{\gamma_i f} p).$$

Proof. (i) We have that:

$$d(x_{n+1}, p) = d(J_{\gamma_n f} x_n, p) \leq d(J_{\gamma_n f} x_n, J_{\gamma_n f} p) + d(J_{\gamma_n f} p, p) \leq d(x_n, p) + d(p, J_{\gamma_n f} p).$$

(ii) Easy induction on (i).

□

Lemma 4.2.20. *The sequence $(x_n)_{n \in \mathbb{N}}$ is uniformly Fejér monotone w.r.t. the above approximation with modulus $\chi(n, m, r) := \max\{n + m - 1, m(r + 1)\}$.*

Proof. Let $n, m, r \in \mathbb{N}$, $p \in AF_{\chi(n,m,r)}$ and $l \leq m$. From the previous lemma, we have that

$$d(x_{n+l}, p) \leq d(x_n, p) + \sum_{i=n}^{i=n+l-1} d(p, J_{\gamma_i f} p).$$

We then see that:

$$\sum_{i=n}^{i=n+l-1} d(p, J_{\gamma_i f} p) \leq \sum_{i=n}^{i=n+m-1} d(p, J_{\gamma_i f} p) \leq \frac{m}{\chi(n, m, r) + 1} \leq \frac{1}{r + 1},$$

where at the second-to-last inequality we used that $\chi(n, m, r) \geq n + m - 1$ and at the last one, that $\chi(n, m, r) \geq m(r + 1)$. □

Lemma 4.2.21. *Let $p \in F$. Then:*

(i) For all n, $d^2(x_{n+1}, p) \leq d^2(x_n, p) - d^2(x_n, x_{n+1})$.

(ii) The sequence $(f(x_n))_{n \in \mathbb{N}}$ is nonincreasing.

Proof. (i) This is an immediate consequence of the last equation in the proof of (7) from [4].

(ii) This is used without proof in [4], and hence we shall justify it. Let $n \in \mathbb{N}$. We must show that $f(x_{n+1}) \leq f(x_n)$. By the definition of $J_{\gamma n f}$ and considering that $x_{n+1} = J_{\gamma n f} x_n$, we have that:

$$\gamma_n f(x_{n+1}) + \frac{1}{2} d^2(x_n, x_{n+1}) \leq \gamma_n f(x_n) + \frac{1}{2} d^2(x_n, x_n) = \gamma_n f(x_n),$$

and so, since $d^2(x_n, J_{\gamma_n f} x_n) \geq 0$,

$$\gamma_n f(x_{n+1}) \leq \gamma_n f(x_n).$$

□

We shall, from now on, suppose that there is a $p \in F$ and we take $b \in \mathbb{R}$ with $d(x,p) \leq b$. Since the result is crucially based on the divergence of the series $\sum_{n=0}^{\infty} \gamma_n$, one must therefore add it as an axiom to the system. However, the problem is that the divergence statement is not purely universal. In order to rectify that, we replace the existentially quantified variable with a new constant, which will be a rate of divergence θ for the series. The compromise will be that the extracted bound will additionally depend on this θ.

The following lemma highlights two different intermediate results that are used in the convergence proofs. These results refer in their turn to some real-valued sequences and must be then added as premises in order to do the final extraction. Two remarks are in order here: the first is that the existentially quantified variables must be again replaced by rates (defined in the previous section) as for the series above, and the second is that the results exhibit some other tactic mentioned in the Introduction for dealing with intractable statements about sequences – they settle themselves for the weaker property of the limit inferior that has $\forall\exists$ complexity, which in the second one is actually equivalent to convergence, since the sequence consists of non-negative real numbers and was proven in the previous lemma to be nonincreasing. Still, these rates are sufficient for the purpose mentioned in the first remark.

Lemma 4.2.22. *We have that:*

(i) $\liminf_{n\to\infty} d(x_n, x_{n+1}) = 0$ *with modulus of liminf* $\Delta_b(k,L) := \lceil b^2(k+1)^2 \rceil + L - 1$.

(ii) $\lim_{n\to\infty} f(x_n) = 0$, *with (monotonely nondecreasing) rate of convergence*

$$\beta_{b,\theta}(k) := \theta^M(\lceil b^2(k+1) \rceil) + 1.$$

Proof. (i) From Lemma 4.2.21.(i), we get that for all j, k,

$$\sum_{n=j}^{k} d^2(x_n, x_{n+1}) \leq \sum_{n=0}^{k} (d^2(x_n,p) - d^2(x_{n+1},p)) \leq d^2(x_0,p) - d^2(x_{k+1},p) \leq d^2(x,p) \leq b^2.$$

Suppose, by way of contradiction, that for all $N \in [L, \Delta_b(k,L)]$, we have $d(x_n, x_{n+1}) > \frac{1}{k+1}$. Then we have that:

$$(\Delta_b(k,L) - L + 1)\frac{1}{(k+1)^2} < \sum_{n=L}^{\Delta_b(k,L)} d^2(x_n, x_{n+1}) \leq b^2,$$

from which we get $\Delta_b(k,L) < b^2(k+1)^2 + L - 1$, a contradiction.

(ii) Since, by Lemma 4.2.21.(ii), $(f(x_n))_{n\in\mathbb{N}}$ is nonincreasing, all we have to show is that

$$f(x_{\beta_{b,\theta}(k)}) \leq \frac{1}{k+1}.$$

Suppose that $f(x_{\beta_{b,\theta}(k)}) > \frac{1}{k+1}$. Then, using [4, (8)], we have (taking note that λ_k coresponds to our γ_{k-1}) that

$$\frac{1}{k+1} < f(x_{\beta_{b,\theta}(k)}) \leq \frac{b^2}{\sum_{i=0}^{\theta^M(\lceil b^2(k+1) \rceil)} \lambda_i} \leq \frac{b^2}{\sum_{i=0}^{\theta(\lceil b^2(k+1) \rceil)} \lambda_i} \leq \frac{b^2}{\lceil b^2(k+1) \rceil} \leq \frac{1}{k+1},$$

a contradiction.

∎

Theorem 4.2.23. *The sequence $(x_n)_{n\in\mathbb{N}}$ has F-approximate fixed points (still w.r.t. the above approximation) with (monotonely nondecreasing) modulus*

$$\Phi_{b,\theta,\gamma}(k) := \lceil b^2(k+1)^2\rceil + \beta_{b,\theta}(\lceil 2m_k(k+1)^2\rceil),$$

where $m_k := \max_{0\le i\le k}\gamma_i$.

Proof. We have to show that for all $k\in\mathbb{N}$ there is an $N\le\Phi_{b,\theta,\gamma}(k)$ such that for all $i\le k$,

$$d(x_N, J_{\gamma_i f}x_N)\le\frac{1}{k+1}.$$

Let $k\in\mathbb{N}$. Set $c:=\beta_{b,\theta}(\lceil 2m_k(k+1)^2\rceil)$. Applying Lemma 4.2.22.(i) we obtain that there is an $N\in[c,\Delta_b(k,c)]$ such that

$$d(x_N,x_{N+1})\le\frac{1}{k+1}.$$

Now, we have that:

$$N\le\Delta_b(k,c)=\lceil b^2(k+1)^2\rceil+c-1\le\lceil b^2(k+1)^2\rceil+\beta_{b,\theta}(\lceil 2m_k(k+1)^2\rceil)=\Phi_{b,\theta,\gamma}(k).$$

Since $c\le N$ we have, using Lemma 4.2.22.(ii), that

$$f(x_N)\le f(x_c)=f(x_{\beta_{b,\theta}(\lceil 2m_k(k+1)^2\rceil)})\le\frac{1}{\lceil 2m_k(k+1)^2\rceil}.$$

On the other hand, for all $i\le k$, we have, by the definition of $J_{\gamma_i f}$, that:

$$\gamma_i f(J_{\gamma_i f}x_N)+\frac{1}{2}d^2(x_N,J_{\gamma_i f}x_N)\le\gamma_i f(x_N)+\frac{1}{2}d^2(x_N,x_N)=\gamma_i f(x_N).$$

Since $f(J_{\gamma_i f}x_N)\ge\min(f)=0$, we have that:

$$d^2(x_N,J_{\gamma_i f}x_N)\le 2\gamma_i f(x_N)\le 2m_k\frac{1}{\lceil 2m_k(k+1)^2\rceil}\le\frac{1}{(k+1)^2}.$$

This N then satisfies our requirements. □

We may now apply the theorem on the metastability of Fejér monotone sequences and state the main result of this subsection.

Theorem 4.2.24. *Let (X,d) be a CAT(0) space which has modulus of total boundedness α. Let $f: X\to(-\infty,]infty]$ be convex lsc. Let $x\in X$ and $(\gamma_n)_{n\in\mathbb{N}}\subseteq(0,\infty)$. We assume that f has a minimizer p and take $b\ge d(x,p)$. Take θ to be a rate of divergence for the series $\sum_{n=0}^{\infty}\gamma_n$. Set $x_0:=x$ and for all n, $x_{n+1}:=x_n$. We define the following functionals (where χ and $\Phi_{b,\theta,\gamma}$ are the one defined in the previous results):*

$$\chi_g(n,k):=\chi(n,g(n),k)$$

$$\chi_g^M(n,k):=\max_{i\le n}\chi_g(i,k)$$

$$(\Psi_0)_{b,\theta,\gamma}(0,k,g):=0$$

$$(\Psi_0)_{b,\theta,\gamma}(n+1,k,g) := \Phi_{b,\theta,\gamma}(\chi_g^M((\Psi_0)_{b,\theta,\gamma}(n,k,g),4k+3))$$
$$\Psi_{b,\theta,\gamma,\alpha}(k,g) := (\Psi_0)_{b,\theta,\gamma}(\alpha(4k+3),k,g)$$

Then $\Psi_{b,\theta,\gamma,\alpha}$ *is a rate of metastability for* $(x_n)_{n\in\mathbb{N}}$*, i.e. for all* $k \in \mathbb{N}$*, all* $g : \mathbb{N} \to \mathbb{N}$ *there is an* $N \leq \Psi_{b,\theta,\gamma,\alpha}(k,g)$ *such that for all* $i, j \in [N, N+g(N)]$*,*

$$d(x_i, x_j) \leq \frac{1}{k+1}.$$

Proof. Apply Theorem 3.7.6. □

Chapter 5

Proof mining in L^p spaces

As we hinted in the second chapter, the formalization of analysis used in proof mining is not the only way, and not even the only useful way of interfacing logic with analysis. One of the first methods to represent real numbers in a logic in a built-in matter was attempted in the 1960s – see, e.g., the book on continuous model theory by Chang and Keisler [14]. Later, Ben Yaacov and others realized that the lack of fruitful lines of research out of that logic was due to an unfortunate choice of parameters – specifically, the truth values could vary wildly along an arbitrary compact Hausdorff space (instead of just the interval $[0,1]$), while equality itself was tightly restricted to binary values. Their efforts led to what has been called "continuous first-order logic", a system in which many celebrated and relatively advanced results of 20th century model theory could be reasonably translated – see [9] for an introduction. Another strand of developments came from Henson's positive-bounded logic, introduced in [29] and later shown to be largely equivalent to continuous first-order logic. Despite this fact, due to its later exhaustive treatment by Henson and Iovino focusing on the model-theoretic ultraproduct construction [30], this logic was subject to an investigation from which it resulted that, in combination with the aforementioned ultraproducts, it could be used to prove uniformity results in nonlinear analysis and ergodic theory – see the recent paper of Avigad and Iovino [3].

We can now ask the question of whether the methods of proof mining are sufficiently powerful to provide us with all uniformity results given to us by the model-theoretic properties of positive-bounded formulas. (Proof theory already can be considered to have some upper hand in the matter of being able to deal with weak forms of extensionality.) The answer, as presented in the 2016 paper of Günzel and Kohlenbach [28], is in the affirmative. To give a rough sketch, the positive-bounded formulas are there translated into a special class of higher-order formulas denoted by $\mathcal{PBL}$, which are then turned into Δ-formulas, a class of formulas which can be freely added as additional axioms, with no negative consequences to the bound extraction procedure, as per the classical metatheorems of proof mining. A new metatheorem is then obtained for the classes of spaces which could be axiomatized by positive-bounded formulas. In addition, the treatment of a "uniform boundedness principle" tries to clarify just what exactly is the role played by the ultraproduct construction. Examples are given of such classes of spaces, and the translations for each set of axioms into the higher-order language are given explicitly, together with their metatheorems. Notable among these are the L^p and BL^pL^q Banach lattices, which are usually defined by a construction, but for which axiomatic characterizations into positive-bounded logic have been found, for the last one by Henson and Raynaud [31]. The continual addition of new classes of spaces to the list of targets of logical metatheorems has been long-pursued within proof mining – see, e.g. Leuştean's metatheorem on $\mathbb{R}$-trees [60] or Kohlenbach and Nicolae's on CAT(κ) spaces [53].

The contribution of this chapter is to find an appropriate treatment of the class of L^p Banach spaces in themselves, as defined in the first chapter. It turns out – see [59, 70] for detailed expositions – that these spaces can be given an implicit characterization, which resembles a bit the axiomatization of BL^pL^q lattices which was analysed by Günzel and Kohlenbach. Notably, and in contrast to that, this characterization does not use at all the natural lattice structure. What we shall do is to show how it may be modified in order to build from it a logical system that (i) accurately represents the $L^p(\mu)$ spaces relatively to their standard models (Theorem 5.1.7); (ii) allows for a bound extraction metatheorem (Theorem 5.1.9); and (iii) admits an internal proof that the standard modulus of uniform convexity is valid for this class of spaces (Theorem 5.2.5). In addition, we shall draw, in Theorem 5.1.8, a similar conclusion to that of [31]: the class of L^p spaces is axiomatizable in the language of positive-bounded logic within the class of pure Banach spaces, therefore obtaining a more concrete proof of this classical result which is shown in [29] using ultraproducts. These results can all be found in [83].

5.1 The Δ-axiomatization of $L^p(\mu)$ Banach spaces

The goal of this section is to describe an extension of the theory in the previous section, one that can formalize the concept of an $L^p(\mu)$ Banach space. Since such spaces are usually defined explicitly, as equivalence classes of p-integrable real-valued functions on a measurable space, it is clear that an implicit characterization is needed. Such a characterization in terms of the natural lattice structure of $L^p(\mu)$ spaces was used in [28] in order to provide a logical metatheorem for this class of spaces. For our use, however, it is more helpful to use the following characterization, for which references are [70, 59] and which uses solely the Banach space structure. In the sequel, we shall denote by $\mathbb{R}^n_p$ the Euclidean vector space $\mathbb{R}^n$ endowed with the standard p-norm.

Definition 5.1.1. *Let X and Y be linearly isomorphic Banach spaces. The* **Banach-Mazur distance** *between X and Y is*

$$d(X, Y) := \inf\{\|L\|\|L^{-1}\| \mid L \text{ is a linear isomorphism between } X \text{ and } Y\}.$$

Definition 5.1.2. *Let p, $\lambda > 1$. We say that a Banach space X is an $\mathcal{L}_{p,\lambda}$* **space** *if for each finite dimensional subspace Y of X there exists a finite dimensional subspace Z of X such that $Y \subseteq Z$ and $d(Z, \mathbb{R}_p^{\dim_{\mathbb{R}} Z}) \leq \lambda$.*

Theorem 5.1.3 ([69, 88])**.** *Let $p > 1$. A Banach space X is isometric to some $L_p(\mu)$ space iff for all $\varepsilon > 0$, X is an $\mathcal{L}_{p,1+\varepsilon}$ space.*

The first step in converting the above characterization into a logical axiomatization consists of the following quantitative bounding lemmas. In proving them, we shall use an argument adapted from [31, Proposition 3.7].

Lemma 5.1.4. *Let X be the L^p space on a measure space $(\Omega, \mathcal{F}, \mu)$. Then, for all x_1,..., x_n in X of norm less than 1, and for all $N \in \mathbb{N}_{\geq 1}$, there is a subspace $C \subseteq X$ and y_1,..., y_n in C such that C is of dimension at most $(2nN + 1)^n$, it is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y_i\| \leq \frac{1}{N}$.*

Proof. For any $f : \Omega \to \mathbb{R}$, we denote by $|f| : \Omega \to \mathbb{R}$ the function defined, for all $\omega \in \Omega$, by $|f|(\omega) := |f(\omega)|$.

We fix from the beginning some representatives for x_1,..., x_n, denoting them by the same designators, and we note that all constructions below will be well-defined w.r.t. the a.e.-equality equivalence relation. We set $\varphi := \sum_{j=1}^n |x_j|$ and, for each $i \in \{1, \ldots, n\}$ and $k \in \{0, \ldots, nN-1\}$:

$$A_{i,k} := \{\omega \in \Omega \mid \frac{k}{nN}\varphi(\omega) < |x_i(\omega)| \leq \frac{k+1}{nN}\varphi(\omega)\},$$
$$A_{i,k,+} := \{\omega \in A_{i,k} \mid x_i(\omega) > 0\}, \quad A_{i,k,-} := \{\omega \in A_{i,k} \mid x_i(\omega) < 0\},$$
$$A_{i,\otimes} := \{\omega \in \Omega \mid x_i(\omega) = 0\}.$$

Clearly, for all i, we have that $\Omega = \bigcup_{k=0}^{nN-1}(A_{i,k,+} \cup A_{i,k,-}) \cup A_{i,\otimes}$ and this is a disjoint union in all of its components.

For each i, put $y_i := \sum_{k=0}^{nN-1} \frac{k}{nN}(\mathbb{1}_{A_{i,k,+}} - \mathbb{1}_{A_{i,k,-}}) \cdot \varphi$. Let $i \in \{1, \ldots, n\}$ and $\omega \in \Omega$ be such that $x_i(\omega) > 0$. Then, by the above, there is a unique k_0 such that $\omega \in A_{i,k_0,+}$ and there is not any k such that $\omega \in A_{i,k,-}$. Therefore, $y_i(\omega) = \frac{k_0}{nN} \cdot \varphi(\omega)$. As $\omega \in A_{i,k_0,+}$, $x_i(\omega) \leq \frac{k_0+1}{nN}\varphi(\omega)$, so $x_i(\omega) - y_i(\omega) \leq \frac{\varphi(\omega)}{nN}$. Since we also have that $x_i(\omega) > \frac{k_0}{nN}\varphi(\omega) = y_i(\omega)$ (so $x_i(\omega) - y_i(\omega) > 0$), we get that $|x_i(\omega) - y_i(\omega)| \leq \frac{\varphi(\omega)}{nN}$. Analogously, we might prove this result for $x_i(\omega) = 0$ and $x_i(\omega) < 0$. We have therefore established that for all i, $|x_i - y_i| \leq \frac{1}{nN} \cdot \varphi$. From that we get that for all i,

$$\|x_i - y_i\| \leq \frac{1}{nN} \cdot \|\varphi\| \leq \frac{1}{nN}\sum_{j=1}^n \|x_j\| \leq \frac{1}{N}.$$

Returning to the disjoint union from before, we remark that, for different i's, those sets might overlap. Therefore, for each $l : \{1, \ldots, n\} \to ((\{0, \ldots, nN-1\} \times \{+, -\}) \cup \{\otimes\})$, set:

$$B_l := \bigcap_{i=1}^n A_{i,l(i)}$$

so

$$\Omega = \bigcup_l B_l$$

is a disjoint union. For each such l, of which there are $(2nN+1)^n$, set now:

$$z_l := \mathbb{1}_{B_l} \cdot \varphi.$$

We have, then, for each i, that:

$$\begin{aligned} y_i &= \sum_{k=0}^{nN-1} \frac{k}{nN}(\mathbb{1}_{A_{i,k,+}} - \mathbb{1}_{A_{i,k,-}}) \cdot \varphi \\ &= \sum_{k=0}^{nN-1} \frac{k}{nN}\left(\sum_{l(i)=(k,+)} \mathbb{1}_{B_l} - \sum_{l(i)=(k,+)} \mathbb{1}_{B_l}\right) \cdot \varphi \\ &= \sum_{k=0}^{nN-1} \frac{k}{nN}\left(\sum_{l(i)=(k,+)} z_l - \sum_{l(i)=(k,+)} z_l\right), \end{aligned}$$

i.e. a linear combination of z_l's.

Let D be the set of all l's such that $z_l \neq 0$. We take C to be the space spanned by all the z_l's with $l \in D$. It clearly contains, by the above, all the y_i's and is of dimension at most (actually, equal, as we shall see) the cardinality of D, which is in turn at most $(2nN+1)^n$. It remains to show that it is isometric to $\mathbb{R}_p^D$. If $l \in D$, then:

$$0 \neq \|z_l\| = \left(\int_\Omega |z_l|^p \mathrm{d}\mu\right)^{\frac{1}{p}} = \left(\int_{B_l} |\varphi|^p \mathrm{d}\mu\right)^{\frac{1}{p}}.$$

We can now show that the linear map $f : \mathbb{R}_p^D \to C$, defined on the standard basis vectors by $f(e_l) := \frac{1}{\|z_l\|} \cdot z_l$ is an isometry. Let $v \in \mathbb{R}_p^D$, so there exist $(\lambda_l)_{l\in D}$ such that $v = \sum_{l\in D} \lambda_l e_l$. Then we have that:

$$\begin{aligned}
\|f(v)\| &= \left\|\sum_{l\in D} \frac{\lambda_l}{\|z_l\|} \cdot z_l\right\| \\
&= \left(\int_\Omega \left|\sum_{l\in D} \frac{\lambda_l}{\|z_l\|} \cdot \mathbb{1}_{B_l} \cdot \varphi\right|^p \mathrm{d}\mu\right)^{\frac{1}{p}} \\
&= \left(\sum_{l\in D} \int_{B_l} \left|\frac{\lambda_l}{\|z_l\|}\right|^p \cdot |\varphi|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} && \text{(as the } B_l\text{'s are disjoint)} \\
&= \left(\sum_{l\in D} \left|\frac{\lambda_l}{\|z_l\|}\right|^p \int_{B_l} |\varphi|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} \\
&= \left(\sum_{l\in D} |\lambda_l|^p\right)^{\frac{1}{p}} \\
&= \left\|\sum_{l\in D} \lambda_l e_l\right\| \\
&= \|v\|,
\end{aligned}$$

and we are done. □

Lemma 5.1.5. *The statement of Lemma 5.1.4 is still valid if we require that all y_i's are of norm less than 1 and we allow for C to be of dimension at most $(4nN+1)^n$.*

Proof. We apply Lemma 5.1.4 for our x_i's, but with N replaced by $2N$. We therefore obtain a subspace $C \subseteq X$ and y'_1,..., y'_n in C such that C is of dimension at most $(4nN+1)^n$, it is isometric to $\mathbb{R}_p^{\dim_\mathbb{R} C}$ and for all i, $\|x_i - y'_i\| \leq \frac{1}{2N}$. For each i, if $\|y'_i\| \geq 1$, set $y_i := \frac{y'_i}{\|y'_i\|}$, else put $y_i := y'_i$. For the "unmodified" y_i's, clearly $\|x_i - y_i\| \leq \frac{1}{N}$. The others are certainly still in C, so we must only show for them that $\|x_i - y_i\| \leq \frac{1}{N}$.

Set $\alpha_i := \frac{1}{\|y'_i\|}$. Since $\|y'_i\| \leq \|x_i\| + \|y'_i - x_i\| \leq 1 + \frac{1}{2N}$, we get that $\frac{1-\alpha_i}{\alpha_i} \leq \frac{1}{2N}$, so:

$$\|x_i - y_i\| = \|x_i - \alpha_i y'_i\| \leq \|x_i - y'_i\| + \|y'_i - \alpha_i y'_i\| \leq \frac{1}{2N} + (1-\alpha_i)\|y'_i\| = \frac{1}{2N} + \frac{1-\alpha_i}{\alpha_i} \leq \frac{1}{N},$$

and we are done. □

Lemma 5.1.6. *Let X be a Banach space that satisfies the conclusion of Lemma 5.1.5. Then, for all x_1,..., x_n in X of norm exactly 1, and for all $N \in \mathbb{N}_{\geq 1}$, there is a subspace $C \subseteq X$ and y_1,..., y_n of norm exactly 1 in C such that C is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y_i\| \leq \frac{1}{N}$.*

Proof. Let x_1,..., x_n in X of norm exactly 1, and $N \in \mathbb{N}_{\geq 1}$. We apply our hypothesis (i.e. the conclusion of Lemma 5.1.5) for these x_i's and we set N to be $2N$. We therefore obtain a subspace $C \subseteq X$ and y'_1,..., y'_n in C of norm at most 1 such that C is of dimension at most $(8nN+1)^n$ (note that we no longer care about this), it is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y'_i\| \leq \frac{1}{2N}$. For each i, we have that $1 = \|x_i\| \leq \|y'_i\| + \|x_i - y'_i\| \leq \|y'_i\| + \frac{1}{2N}$, from which we get that $\|y'_i\| \geq 1 - \frac{1}{2N} > 0$. We may therefore set $\alpha_i := \frac{1}{\|y'_i\|}$ and $y_i := \alpha_i y'_i$. Those vectors are of norm 1 and still in C, so what remains to be shown is that for each i, $\|x_i - y_i\| \leq \frac{1}{N}$.

For each i, $\|y'_i\| \leq 1$, so $\alpha_i - 1 \geq 0$. Then from the relation $\|y'_i\| \geq 1 - \frac{1}{2N}$ obtained above, we get that $\frac{\alpha_i - 1}{\alpha_i} \leq \frac{1}{2N}$, so:

$$\|x_i - y_i\| = \|x_i - \alpha_i y'_i\| \leq \|x_i - y'_i\| + \|y'_i - \alpha_i y'_i\| \leq \frac{1}{2N} + (\alpha_i - 1)\|y'_i\| = \frac{1}{2N} + \frac{\alpha_i - 1}{\alpha_i} \leq \frac{1}{N}.$$

□

Plugging in the above result into the "commutativity of approximation" argument of [70, p. 198], one obtains that an $L^p(\mu)$ space is actually a $\mathcal{L}_{p,1+\varepsilon}$ space for all $\varepsilon > 0$, thereby proving the "only if" direction of Theorem 5.1.3. (Actually, the whole reason for the presence of Lemma 5.1.6 here was to clarify why we can force the norm 1 constraint on the y_i's in that argument.) What is in fact relevant here is the extra information this detour gives us through Lemma 5.1.5, namely the equivalence of the two conditions of the theorem with a third one, expressed as follows:

> *for all x_1,..., x_n in X of norm less than 1 and for all $N \in \mathbb{N}_{\geq 1}$, there is a subspace $C \subseteq X$ and y_1,..., y_n in C of norm less than 1 such that C is of dimension at most $(4nN+1)^n$, it is isometric to $\mathbb{R}_p^{\dim_{\mathbb{R}} C}$ and for all i, $\|x_i - y_i\| \leq \frac{1}{N}$.*

The advantage of the condition above is that it is both intrinsic and quantitative, therefore amenable to a logical axiomatization.

Table 5.1 shows one such axiomatization (into a crude first-order-like language), i.e. the characterization of the space is expressed by the simultaneous validity of all $A_{n,N}$ sentences. With that in mind, by closely examining the formulas, one can easily see that they represent a straightforward translation of the condition from before.

Table 5.2, where we have used some of the notations from [28, Definitions 7.9 and 7.10], shows how one may translate the infinite family of axioms $A_{n,N}$ into the one axiom B which is, like the one in [28], representable as a Δ-sentence. Let us see some details of the translation. Firstly, we remark that the operation $\widetilde{v} := \frac{v}{\max\{\|v\|,1\}}$ that we used excused us from writing the antecedent from $A_{n,N}$. Then we see that by substituting into $\psi_m(\underline{z})$ all λ_i's with 0, except for one which we set to 1,

$$\psi_m(\underline{z}) := \forall \underline{\lambda} \left(\|\sum_{i=1}^m \lambda_i z_i\| = (\sum_{i=1}^m |\lambda_i|^p)^{\frac{1}{p}} \right)$$
$$\psi'_{m,n}(\underline{y}, \underline{z}) := \bigwedge_{k=1}^n (\exists \underline{\lambda} (y_k = \sum_{i=1}^m \lambda_i z_i))$$
$$\psi''_{n,N}(\underline{x}, \underline{y}) := \bigwedge_{k=1}^n \left(\|x_k - y_k\| \leq \frac{1}{N+1} \wedge \|y_k\| \leq 1 \right)$$
$$\varphi_{n,m,N}(\underline{x}) := \exists \underline{y} \exists \underline{z} \left(\psi_m(\underline{z}) \wedge \psi'_{m,n}(\underline{y}, \underline{z}) \wedge \psi''_{n,N}(\underline{x}, \underline{y}) \right)$$
$$\phi_{n,N}(\underline{x}) := \bigvee_{0 \leq m \leq (4nN+1)^n} \varphi_{n,m,N}(\underline{x})$$
$$A_{n,N} := \forall \underline{x} \left((\bigwedge_{k=1}^n \|x_k\| \leq 1) \to \phi_{n,N}(\underline{x}) \right)$$

Table 5.1: A first axiomatization.

$$\psi(m, z) := \forall \lambda^{1(0)(0)} \left(\|\sum_{i=1}^m |\lambda(i)|_{\mathbb{R}} \cdot_X z(i)\| =_{\mathbb{R}} (\sum_{i=1}^m |\lambda(i)|_{\mathbb{R}}^p)^{1/p} \right)$$
$$\psi'(m, n, y, z, \lambda) := \forall k \preceq_0 (n-1) \left(y(k+1) =_X \sum_{i=1}^m \lambda(i) \cdot_C z(i) \right)$$
$$\psi''(n, N, x, y) := \forall k \preceq_0 (n-1) \left(\left\| \widetilde{x(k+1)} - y(k+1) \right\| \leq_{\mathbb{R}} \frac{1}{N} \wedge \|y(k+1)\| \leq_{\mathbb{R}} 1 \right)$$
$$\varphi(n, m, N, x, y, z, \lambda) := \psi(m, z) \wedge \psi'(m, n, y, z, \lambda) \wedge \psi''(n, N, x, y)$$
$$B := \forall n^0, N^0 \geq 1 \forall x^{X(0)} \exists y, z \preceq_{X(0)(0)} 1_{X(0)(0)} \exists \lambda^{1(0)(0)(0)} \in [-2, 2] \exists m \preceq_0 (4nN+1)^n$$
$$\varphi(n, m, N, x, y, z, \lambda)$$

Table 5.2: The Δ-axiomatization.

we obtain the fact that all z_i's are of norm one. We have also postulated that all y_k's are of norm less than 1. Thus, if we have, as in $\psi'_{m,n}(\underline{y}, \underline{z})$, that for a given k:

$$y_k = \sum_{i=1}^m \lambda_i z_i,$$

the formula $\psi_m(\underline{z})$ tells us further that:

$$1 \geq \|y_k\| \geq \left\| \sum_{i=1}^m \lambda_i z_i \right\| = \left(\sum_{i=1}^m |\lambda_i|^p \right)^{\frac{1}{p}},$$

from which we get that each such λ_i is in the interval $[-1, 1]$. These results allow us to correspondingly bound the y, the z and the λ (which are now properly functionals) in the axiom B. Another such bounding comes from the $(4nN+1)^n$ established before (i.e. here it matters that the characterization is quantitative), which helped us eliminate the potentially infinite disjunction in Table 5.1 (where such constraints were not yet relevant) and the unbounded existential quantifier in Table 5.2 (which would have hindered us in presenting the axiom B as a Δ-sentence). As a curiosity, we note that choosing to present B as a single axiom and not as an infinite schema like in Table 5.1, i.e. taking advantage of the arithmetic already present in the framework, adds a bit of strength to the system, given the fact that we do not work here with any sort of ω-rule.

We denote by $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}, L^p]$ the extension of the system $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}]$ by the constant c_p of type 1, together with the axiom $1_\mathbb{R} \leq_\mathbb{R} c_p$ and the axiom B from above. From the above discussion, the following soundness theorem holds.

Theorem 5.1.7 (cf. [28, Propositions 3.5 and 7.12])**.** *Let X be a Banach space and $p \geq 1$. Denote by $\mathcal{S}^{\omega,X}$ its associated set-theoretic model and let the constant c_p in our extended signature take as a value the canonical representation of the real number p. Then $\mathcal{S}^{\omega,X}$ is a model of $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}, L^p]$ iff X is isomorphic to some $L^p(\Omega, \mathcal{F}, \mu)$ space.*

In a parallel way to the one suggested in [31], by some similar arguments to the ones used above to construct the required higher-order system, one could perform reasonable transformations to the formulas in Table 5.1, obtaining a new, concrete proof of the following classical result of Henson.

Theorem 5.1.8. *The subclass of Banach spaces which are isomorphic to spaces of the form $L^p(\mu)$ is axiomatizable in positive-bounded logic.*

Analogously to the treatment done in [28] for the classes of Banach lattices, we may now state the corresponding metatheorem for the system devised above.

Theorem 5.1.9 (Logical metatheorem for $L^p(\mu)$ Banach spaces, cf. [28, Theorems 5.13 and 7.13])**.** *Let $\rho \in \mathbf{T}^X$ be an admissible type. Let $B_\forall(x,u)$ be a $\forall$-formula with at most x, u free and $C_\exists(x,v)$ an $\exists$-formula with at most x, v free. Let Δ be a set of Δ-sentences. Suppose that:*

$$\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}, L^p] + \Delta \vdash \forall x^\rho(\forall u^0 B_\forall(x,u) \to \exists v^0 C_\exists(x,v)).$$

Then one can extract a partial functional $\Phi : S_{\widehat{\rho}} \rightharpoonup \mathbb{N}$, whose restriction to the strongly majorizable functionals of $S_{\widehat{\rho}}$ is a bar-recursively computable functional of $\mathcal{M}^\omega$, such that for all $L^p(\mu)$ Banach spaces $(X, \|\|)$ having the property that any associated set-theoretic model of it satisfies Δ, we have that for all $x \in S_\rho$ and $x^* \in S_{\widehat{\rho}}$ such that $x^* \gtrsim_\rho x$, the following holds:

$$\forall u \leq \Phi(x^*) B_\forall(x,u) \to \exists v \leq \Phi(x^*) C_\exists(x,v).$$

All the additional considerations from Theorem 2.3.7 also apply here.

Proof. This theorem extends Theorem 2.3.7. The two additional axioms are Δ-axioms, and the constant c_p is majorized (as in [46, Lemma 17.8]) by $M(b) := \lambda n.j(b2^{n+2}, 2^{n+1}-1)$, where j is the Cantor pairing function and $b \in \mathbb{N}$ such that $b \geq p$ (e.g., $b := \lceil (c_p(0))_\mathbb{Q} \rceil + 1$). We note that the Φ depends on p only via this upper bound b. □

5.2 The derivation of the modulus of uniform convexity

The axiomatization that we have just obtained has, essentially, the form of a comparison principle with respect to the p-normed Euclidean spaces. This suggests that it may be particularly application-friendly. Let us see why this is the case. Suppose that we have an existing mathematical theorem regarding L^p spaces. The particularization of the proof to the Euclidean case is likely to be easily derivable in our higher systems of arithmetic (with the possible addition of universal lemmas), since statements about integrals are reduced to statements about sums and powers of real numbers. The second step would be to translate the result along the ε-close approximation of our characterization, a translation involving a sequence of boundings which is likely to leave the original statement intact if it is well-behaved enough. We shall now illustrate this general strategy on a classical result on L^p spaces.

Uniform convexity is a fundamental notion in the theory of Banach spaces. As per [28, Section 6.4], the property can be formalized as:

$$\forall k^0 \exists n^0 \forall x_1, x_2 \preceq_X 1_X \left(\left\| \frac{1}{2}(x_1 + x_2) \right\| \leq 1 - 2^{-n} \to \|x_1 - x_2\| < 2^{-k} \right).$$

and it is suitable for bound extraction. We note that, in the above statement, like in the definition of the convergence of a sequence, a bound (for n, in this case) is also a witness. Also, with the logical issues now resolved, we note that, for the ease of understanding, we shall work with ε-style characterizations. Therefore, following [49, Section 2.1], we define a **modulus of uniform convexity** for a Banach space to be a function $\eta : (0, 2] \to (0, \infty)$ such that for any $\varepsilon > 0$ and any x_1 and x_2 with $\|x_1\| \leq 1$, $\|x_2\| \leq 1$ and $\|x - y\| \geq \varepsilon$, we have that

$$\left\| \frac{1}{2}(x_1 + x_2) \right\| \leq 1 - \eta(\varepsilon).$$

We make the observation that what is usually called "the" modulus of uniform convexity of a space is the "optimal" such modulus, i.e. for each $\varepsilon > 0$ we take as $\eta(\varepsilon)$ the greatest value of δ that works for all suitable x_1, x_2, i.e. the minimum of the expression $1 - \|\frac{1}{2}(x_1 + x_2)\|$. The goal of this section is to derive a modulus of uniform convexity for $L^p(\mu)$ spaces using only the axiomatization established in the previous section. We will consider, for simplicity, $p \geq 2$, i.e. we add the additional admissible axiom $2 \leq_{\mathbb{R}} c_p$ to our system. We begin with some results of real analysis. The following lemma and corollary are standard in the literature.

Lemma 5.2.1. *For all $x_1, x_2 \geq 0$, $x_1^p + x_2^p \leq (x_1^2 + x_2^2)^{p/2}$.*

Proof. The case $x_2 = 0$ is clear. If $x_2 \neq 0$, we can divide by x_2^p and we notice that we only have to prove that for all $t \geq 0$, $t^p + 1 \leq (t^2 + 1)^{p/2}$. Consider the function $f : \mathbb{R} \to \mathbb{R}$, defined, for all t, by $f(t) := (t^2 + 1)^{p/2} - t^p - 1$. Since $f'(t) = \frac{p}{2}(t^2 + 1)^{(p/2)-1} \cdot 2t - pt^{-1} \geq pt^{p-2} \cdot t - pt^{p-1} = 0$ and $f(0) = 0$, we obtain that for all t, $f(t) \geq 0$, and hence the conclusion. □

Corollary 5.2.2. *For all $a, b \in \mathbb{R}$, $\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p \leq \frac{1}{2}(|a|^p + |b|^p)$.*

Proof. We substitute into the above lemma $x_1 := \left|\frac{a+b}{2}\right|$ and $x_2 := \left|\frac{a-b}{2}\right|$. Since $\left|\frac{a+b}{2}\right|^2 + \left|\frac{a-b}{2}\right|^2 = \frac{1}{2}(a^2 + b^2)$, we obtain that:

$$\begin{aligned}
\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p &\leq \left(\frac{1}{2}(a^2 + b^2)\right)^{p/2} \\
&\leq \frac{1}{2}((a^2)^{p/2} + (b^2)^{p/2}) \\
&= \frac{1}{2}(|a|^p + |b|^p),
\end{aligned}$$

where the last inequality follows from the convexity of the function $t \mapsto t^p$ on $(0, \infty)$, for any $p \geq 2$. □

Set, now, for all $a, d \in (0,1)$, $\sigma(a,d) := a - (1 - ((1-a^p)^{1/p} + d)^p)^{1/p}$.

Lemma 5.2.3. *For all $a, d \in (0,1)$, $\sigma(a,d) > 0$.*

Proof. Since $d > 0$, we have that $(1-a^p)^{1/p} < (1-a^p)^{1/p} + d$, so

$$1 - a^p < ((1-a^p)^{1/p} + d)^p.$$

From that we successively obtain:

$$a^p > 1 - ((1-a^p)^{1/p} + d)^p,$$

$$a > (1 - ((1-a^p)^{1/p} + d)^p)^{1/p},$$

$$a - (1 - ((1-a^p)^{1/p} + d)^p)^{1/p} > 0.$$

□

Lemma 5.2.4. *For all $a, d \in (0,1)$ and all $\delta \in (0, \sigma(a,d))$, we have that:*

$$(1 - (a-\delta)^p)^{1/p} \leq (1-a^p)^{1/p} + d.$$

Proof. Clearly $\sigma(a,d) < a$, so $(a-\delta)^p$ is well-defined. Now, since

$$\delta \leq a - (1 - ((1-a^p)^{1/p} + d)^p)^{1/p},$$

we obtain, successively, that:

$$a - \delta \geq (1 - ((1-a^p)^{1/p} + d)^p)^{1/p},$$

$$(a-\delta)^p \geq 1 - ((1-a^p)^{1/p} + d)^p,$$

$$1 - (a-\delta)^p \leq ((1-a^p)^{1/p} + d)^p,$$

$$(1 - (a-\delta)^p)^{1/p} \leq (1-a^p)^{1/p} + d.$$

□

Note that the statements of Corollary 5.2.2 and Lemma 5.2.4 are universal and therefore it is admissible to add them as supplementary axioms – denote them by C_1 and C_2. We are now in a position to state the main theorem of this section.

Theorem 5.2.5. *Provably in the system $\mathcal{A}^\omega[X, \|\cdot\|, \mathcal{C}, L^p] + \{2 \leq_{\mathbb{R}} c_p; C_1; C_2\}$, the function $\eta : (0,2] \to (0,\infty)$, defined, for any $\varepsilon > 0$, by $\eta(\varepsilon) := 1 - (1 - (\frac{\varepsilon}{2})^p)^{1/p}$, is a modulus of uniform convexity.*

Proof. Let $\varepsilon > 0$. Take $x_1, x_2 \in X$ with $\|x_1\|, \|x_2\| \leq 1$ and $\|x_1 - x_2\| \geq \varepsilon$. Let $c \in (0,1)$. Set $\delta := \min\{\frac{c}{2}, \frac{\sigma(\frac{\varepsilon}{2}, \frac{c}{2})}{2}\}$. Take y_1, y_2, $z_1, \ldots, z_m$ like in our axiomatization (e.g., from Table 5.1) such that for all $k \in \{1,2\}$,

$$\|x_k - y_k\| \leq \delta, \quad \|y_k\| \leq 1.$$

Write now:

$$y_1 = \sum_{i=1}^m \lambda_i z_i, \quad y_2 = \sum_{i=1}^m \mu_i z_i.$$

We have that:

$$\begin{aligned}\left\|\frac{y_1+y_2}{2}\right\|^p + \left\|\frac{y_1-y_2}{2}\right\|^p &= \left\|\sum_{i=1}^m \frac{\lambda_i+\mu_i}{2} z_i\right\|^p + \left\|\sum_{i=1}^m \frac{\lambda_i-\mu_i}{2} z_i\right\|^p \\ &= \sum_{i=1}^m \left(\left|\frac{\lambda_i+\mu_i}{2}\right|^p + \left|\frac{\lambda_i-\mu_i}{2}\right|^p\right) \\ &\leq \frac{1}{2}\sum_{i=1}^m (|\lambda_i|^p + |\mu_i|^p) \\ &= \frac{1}{2}(\|y_1\|^p + \|y_2\|^p) \\ &\leq 1.\end{aligned}$$

Assume that $\|y_1 - y_2\| \geq \rho$. Then we get that

$$\left\|\frac{y_1+y_2}{2}\right\| \leq \left(1 - \left(\frac{\rho}{2}\right)^p\right)^{1/p}.$$

Incidentally, what we have shown above is the validity of η as a modulus of uniform convexity for the $\mathbb{R}^m_p$ spaces (with $p \geq 2$).

Note that:

$$\varepsilon \leq \|x_1 - x_2\| \leq \|x_1 - y_1\| + \|y_1 - y_2\| + \|y_2 - x_2\| \leq \|y_1 - y_2\| + 2\delta$$

and hence we may take $\rho := \varepsilon - 2\delta > 0$ (since $\delta < \sigma(\frac{\varepsilon}{2}, \frac{c}{2}) < \frac{\varepsilon}{2}$). We have obtained that:

$$\left\|\frac{y_1+y_2}{2}\right\| \leq \left(1 - \left(\frac{\varepsilon}{2} - \delta\right)^p\right)^{1/p}.$$

On the other hand,

$$\|x_1+x_2\| \leq \|y_1+y_2\| + \|(x_1+x_2)-(y_1+y_2)\| \leq \|y_1+y_2\| + \|x_1-y_1\| + \|x_2-y_2\| \leq \|y_1+y_2\| + 2\delta,$$

so

$$\left\|\frac{x_1+x_2}{2}\right\| \leq \left\|\frac{y_1+y_2}{2}\right\| + \delta \leq \left(1 - \left(\frac{\varepsilon}{2} - \delta\right)^p\right)^{1/p} + \delta.$$

Since $0 < \delta < \sigma(\frac{\varepsilon}{2}, \frac{c}{2})$, we have that:

$$\left(1 - \left(\frac{\varepsilon}{2} - \delta\right)^p\right)^{1/p} \leq \left(1 - \left(\frac{\varepsilon}{2}\right)^p\right)^{1/p} + \frac{c}{2}.$$

Also, we know that $\delta \leq \frac{c}{2}$, so we finally obtain that:

$$\left\|\frac{x_1+x_2}{2}\right\| \leq \left(1 - \left(\frac{\varepsilon}{2}\right)^p\right)^{1/p} + c.$$

Now, since $c \in (0,1)$ was arbitrarily chosen, we can apply the very definition of the relation $\leq_\mathbb{R}$ in our system ([46, p. 80]) in order to get that:

$$\left\|\frac{x_1+x_2}{2}\right\| \leq \left(1 - \left(\frac{\varepsilon}{2}\right)^p\right)^{1/p},$$

showing, indeed, that η is a modulus of uniform convexity. □

Bibliography

[1] D. Ariza-Ruiz, L. Leuştean, G. López-Acedo, Firmly nonexpansive mappings in classes of geodesic spaces. *Transactions of the American Mathematical Society* 366, 4299–4322, 2014.

[2] D. Ariza-Ruiz, G. López-Acedo, A. Nicolae, The asymptotic behavior of the composition of firmly nonexpansive mappings. *J. Optim. Theory Appl.* 167, 409–429, 2015.

[3] J. Avigad, J. Iovino, Ultraproducts and metastability. *New York J. of Math.* 19, 713–727, 2013.

[4] M. Bačák, The proximal point algorithm in metric spaces. *Israel J. Math.* 194, no. 2, 689–701, 2013.

[5] M. Bačák, S. Reich, The asymptotic behavior of a class of nonlinear semigroups in Hadamard spaces. *J. Fixed Point Theory Appl.* 16, 189–202, 2014.

[6] M. Bačák, I. Searston, B. Sims, Alternating projections in CAT(0) spaces. *J. Math. Anal. Appl.* 385, no. 2, 599–607, 2012.

[7] S. Bartz, H. Bauschke, S. M. Moffat, X. Wang, The Resolvent Average of Monotone Operators: Dominant and Recessive Properties. *SIAM J. Optim.*, 26(1), 602–634, 2016.

[8] H. Bauschke, P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, 2010.

[9] I. Ben Yaacov, A. Berenstein, C. W. Henson, A. Usvyatsov, *Model theory for metric structures.* In: *Model theory with applications to algebra and analysis.* Vol. 2, 315–427, London Math. Soc. Lecture Note Ser., vol. 350, Cambridge Univ. Press, Cambridge, 2008.

[10] I. D. Berg, I. G. Nikolaev, Quasilinearization and curvature of Alexandrov spaces. *Geom. Dedicata* 133, 195–218, 2008.

[11] M. A. Bezem, Strongly majorizable functionals of finite type: a model for bar recursion containing discontinuous functionals. *J. Symb. Logic* 50, 652–660, 1985.

[12] H. Brézis, P. Lions, Produits infinis de résolvantes. *Israel J. Math.* 29, 329–345, 1978.

[13] F. E. Browder, W. V. Petryshyn, Construction of fixed points of nonlinear mappings in Hilbert spaces. *J. Math. Anal. Appl.* 20:197–228, 1967.

[14] C. C. Chang, H. J. Keisler, *Continuous model theory.* Princeton University Press, 1966.

[15] C. E. Chidume, S. A. Mutangadura, An example on the Mann iteration method for Lipschitz pseudo-contractions. *Proc. Amer. Math. Soc.* 129, 2359–2363, 2001.

[16] P. Cholamjiak, S. Suantai, Weak and strong convergence theorems for a countable family of strict pseudocontractions in Banach spaces. *Optimization* 62, no. 2, 255–270, 2013.

[17] V. Colao, G. Marino, Common fixed points of strict pseudocontractions by iterative algorithms. *J. Math. Anal. Appl.* 382(2):631–644, 2011.

[18] S. Dhompongsa, W. Kirk, B. Sims, Fixed points of uniformly Lipschitzian mappings. *Nonlinear Anal.*, 65, pp. 762–772, 2006.

[19] R. Espínola, A. Fernández-León, CAT(k)-spaces, weak convergence and fixed points. *J. Math. Anal. Appl.*, 353, 410–427, 2009.

[20] T. Figiel, An example of infinite dimensional reflexive Banach space non-isomorphic to its Cartesian square. *Studia Mathematica* 42.3:295–306, 1972.

[21] E. L. Fuster, Moduli and constants... what a show! http://www.uv.es/llorens/Documento.pdf, 2006.

[22] P. Gerhardy, Proof mining in topological dynamics. *Notre Dame J. Form. Log.* **49**, 431–446, 2008.

[23] P. Gerhardy, U. Kohlenbach, Strongly uniform bounds from semi-constructive proofs. *Ann. Pure Applied Logic* vol. 141, 89–107, 2006.

[24] P. Gerhardy, U. Kohlenbach, General logical metatheorems for functional analysis. *Trans. Amer. Math. Soc.* 360, 2615–2660, 2008.

[25] J.-Y. Girard, Une extension de l'interpretation de Gödel à l'analyse, et son application à l'elimination des coupures dans l'analyse et dans le théorie des types. In: J. E. Fenstad (ed.), *Proc. of the Second Scandinavian Logic Symposium*, 63–92. North-Holland, Amsterdam, 1971.

[26] K. Gödel, Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums* v.4, 34–38, 1933.

[27] K. Gödel, Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica* 12, 280–287, 1958.

[28] D. Günzel, U. Kohlenbach, Logical metatheorems for abstract spaces axiomatized in positive bounded logic. *Advances in Mathematics* vol. 290, 503–551, 2016.

[29] C. W. Henson, Nonstandard hulls of Banach spaces. *Israel Journal of Mathematics* 25, 108–144, 1976.

[30] C. W. Henson, J. Iovino, *Ultraproducts in analysis*. In: *Analysis and Logic*. London Mathematical Society Lecture Notes Series, vol. 262, 1–113, 2002.

[31] C. W. Henson, Y. Raynaud, On the theory of $L_p(L_q)$-Banach lattices. *Positivity* 11, no. 2, 201–230, 2007.

[32] D. Hilbert, Über das Unendliche. *Mathematische Annalen* 95, 161–190, 1926.

[33] D. Hilbert, P. Bernays, *Grundlagen der Mathematik. I.* (German) Zweite Auflage. Die Grundlehren der mathematischen Wissenschaften, Band 40 Springer-Verlag, Berlin-New York, 1968.

[34] W. A. Howard, Hereditarily majorizable functionals of finite type. In: A. Troelstra, *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics 344, pp. 454–461, Springer, Berlin, 1973.

[35] S. Ishikawa, Fixed points by a new iteration method. *Proc. Amer. Math. Soc.* **44**, 147–150, 1974.

[36] D. Ivan, L. Leuştean, A rate of asymptotic regularity for the Mann iteration of k-strict pseudo-contractions. *Numer. Funct. Anal. Optimiz.* 36:792–798, 2015.

[37] J. Jost, Equilibrium maps between metric spaces. *Calc. Var. Partial Differential Equations* 2, 173–204, 1994.

[38] J. Jost, Convex functionals and generalized harmonic maps into spaces of non positive curvature. *Comment. Math. Helvetici* 70, no. 4, 659–673, 1995.

[39] T. Kato, Nonlinear semigroups and evolution equations. *J. Math. Soc. Japan* 19:508–520, 1967.

[40] M. A. A. Khan, U. Kohlenbach, Bounds on Kuhfittig's iteration schema in uniformly convex hyperbolic spaces. *J. Math. Anal. Appl.* 403:633–642, 2013.

[41] U. Kohlenbach, Theorie der majorisierbaren und stetigen Funktionale und ihre Anwendung bei der Extraktion von Schranken aus inkonstruktiven Beweisen: Effektive Eindeutigkeitsmodule bei besten Approximationen aus ineffektiven Beweisen. PhD Thesis, Frankfurt am Main, 1990.

[42] U. Kohlenbach, Analysing proofs in analysis. In: W. Hodges, M. Hyland, C. Steinhorn, J. Truss (eds.), *Logic: from Foundations to Applications*. European Logic Colloquium (Keele, 1993), 225–260, Oxford University Press, 1996.

[43] U. Kohlenbach, Relative constructivity. *J. Symbolic Logic* 63:1218–1238, 1998.

[44] U. Kohlenbach. Uniform asymptotic regularity for Mann iterates. *J. Math. Anal. Appl.*, Vol. 279, 531–544, 2003.

[45] U. Kohlenbach, Some logical metatheorems with applications in functional analysis. *Trans. Amer. Math. Soc.* vol. 357, no. 1, 89–128, 2005.

[46] U. Kohlenbach, *Applied proof theory: Proof interpretations and their use in mathematics*. Springer Monographs in Mathematics, Springer-Verlag, Berlin-Heidelberg, 2008.

[47] U. Kohlenbach, Recent progress in proof mining in nonlinear analysis. *IFCoLog Journal of Logics and their Applications* 10, 3357–3406, 2017.

[48] U. Kohlenbach, L. Leuştean, Mann iterates of directionally nonexpansive mappings in hyperbolic spaces, *Abstract and Applied Analysis*, No. 8, 449–477, 2003.

[49] U. Kohlenbach, L. Leuştean, On the computational content of convergence proofs via Banach limits. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* Vol. 370, No. 1971, 3449–3463, 2012.

[50] U. Kohlenbach, L. Leuştean, A. Nicolae, Quantitative results on Fejér monotone sequences. *Communications in Contemporary Mathematics* 20, 1750015 [42 pp.], 2018.

[51] U. Kohlenbach, G. López-Acedo, A. Nicolae, Quantitative asymptotic regularity for the composition of two mappings. *Optimization* vol. 66, pp. 1291–1299, 2017.

[52] U. Kohlenbach, P. Oliva, Proof mining: a systematic way of analysing proofs in mathematics. *Proc. Steklov Inst. Math.*, vol. 242, 136–164, 2003.

[53] U. Kohlenbach, A. Nicolae, A proof-theoretic bound extraction theorem for CAT(κ)-spaces. *Studia Logica* vol. 105, pp. 611–624, 2017.

[54] D. Körnlein, *Quantitative Analysis of Iterative Algorithms in Fixed Point Theory and Convex Optimization*. PhD thesis, Darmstadt, 2016.

[55] M. A. Krasnoselski, Two remarks on the method of successive approximation. *Usp. Math. Nauk (N.S.)* 10:123–127, 1955.

[56] G. Kreisel, Mathematical significance of consistency proofs. *J. Symbolic Logic*, 23, 155–182, 1958.

[57] G. Kreisel, Interpretation of analysis by means of constructive functionals of finite types. In: A. Heyting (Ed.), *Constructivity in Mathematics*, 101–128, North-Holland, Amsterdam, 1959.

[58] G. Kreisel et al., *Reports of Seminar on the Foundations of Analysis ("Stanford report")*. Technical report, Stanford University, Summer 1963.

[59] H. E. Lacey, *The isometric theory of classical Banach spaces*. Die Grundlehren der mathematischen Wissenschaften, Band 208. Springer-Verlag, New York-Heidelberg, 1974.

[60] L. Leuştean, Proof mining in $\mathbb{R}$-trees and hyperbolic spaces. *Electronic Notes in Theoretical Computer Science*, 165, 95–106. In G. Mints, R. de Queiroz (eds.), *Proceedings of the 13th Workshop on Logic, Language, Information and Computation (WoLLIC 2006)*, Stanford University, CA, USA, 18-21 July 2006.

[61] L. Leuştean, A quadratic rate of asymptotic regularity for CAT(0)-spaces. *J. Math. Anal. Appl.*, Vol. 325, No. 1, 386–399, 2007.

[62] L. Leuştean, Nonexpansive iterations in uniformly convex W-hyperbolic spaces. In: A. Leizarowitz, B. S. Mordukhovich, I. Shafrir, A. Zaslavski (eds.), *Nonlinear Analysis and Optimization I: Nonlinear Analysis*. Cont. Math. 513, 193–209, Amer. Math. Soc., Providence, RI, 2010.

[63] L. Leuştean, An application of proof mining to nonlinear iterations. *Ann. Pure Appl. Logic* **165**, 1484–1500, 2014.

[64] L. Leuştean, A. Nicolae, A. Sipoş, An abstract proximal point algorithm. *Journal of Global Optimization*, Volume 72, Issue 3, 553–577, 2018.

[65] L. Leuştean, V. Radu, A. Sipoş, Quantitative results on the Ishikawa iteration of Lipschitz pseudo-contractions. *Journal of Nonlinear and Convex Analysis*, Volume 17, Number 11, 2277–2292, 2016.

[66] L. Leuştean, A. Sipoş, An application of proof mining to the proximal point algorithm in CAT(0) spaces. In: A. Bellow, C. Calude, T. Zamfirescu (eds.), *Mathematics Almost Everywhere. In Memory of Solomon Marcus* (pp. 153–168), World Sci. Publ., Hackensack, NJ, 2018.

[67] T. C. Lim, Remarks on some fixed point theorems. *Proc. Amer. Math. Soc.* 60, 179–182, 1976.

[68] J. Lindenstrauss, On the modulus of smoothness and divergent series in Banach spaces. *Michigan Math. J.* 10:241–252, 1963.

[69] J. Lindenstrauss, A. Pelczynski, Absolutely summing operators in $\mathcal{L}_p$ spaces and their applications. *Studia Math.*, 29:275–326, 1968.

[70] J. Lindenstrauss, L. Tzafriri, *Classical Banach spaces.* Lecture Notes in Mathematics, Vol. 338. Springer-Verlag, Berlin-New York, 1973.

[71] G. López-Acedo, H.-K. Xu, Iterative methods for strict pseudo-contractions in Hilbert spaces. *Nonlinear Anal.* 67, no. 7, 2258–2271, 2007.

[72] H. Luckhardt, *Extensional Gödel Functional Interpretation.* Springer Lecture Notes in Mathematics 306, 1973.

[73] H. Luckhardt, Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. *J. Symb Log.* 54, 234–263, 1989.

[74] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle* 4, 154–158, 1970.

[75] G. Marino, H.-K. Xu, Weak and strong convergence theorems for strict pseudo-contractions in Hilbert spaces. *J. Math. Anal. Appl.* 329, no. 1, 336–346, 2007.

[76] P. Oliva, Hybrid functional interpretations of linear and intuitionistic logic. *J. Logic Comput.* 22:305–328, 2012.

[77] B. Prus, R. Smarzewski, Strongly unique best approximations and centers in uniformly convex spaces. *J. Math. Anal. Appl.* Volume 121, Issue 1, 10–21, 1987.

[78] S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.* 67, no. 2, 274–276, 1979.

[79] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optimization* 14, 877–898, 1976.

[80] H. Schaefer, Über die Methode sukzessiver Approximationen. *Jber. Deutsch. Math. Verein.* 59, Abt. 1, 131–140, 1957.

[81] A. Sipoş, A note on the Mann iteration for k-strict pseudocontractions in Banach spaces. *Numerical Functional Analysis and Optimization*, Volume 38, Issue 1, 80–90, 2017.

[82] A. Sipoş, Effective results on a fixed point algorithm for families of nonlinear mappings. *Annals of Pure and Applied Logic*, Volume 168, Issue 1, 112–128, 2017.

[83] A. Sipoş, Proof mining in L^p spaces. *Journal of Symbolic Logic*, Volume 84, Issue 4, 1612–1629, 2019.

[84] C. Spector, Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics. In: J. C. E. Dekker (Ed.), *Proc. Sympos. Pure Math.* 5, pp. 1–27, Amer. Math. Soc., Providence, RI, 1962.

[85] T. Tao, Soft analysis, hard analysis, and the finite convergence principle. Essay posted May 23, 2007, appeared in: T. Tao, *Structure and Randomness: Pages from Year One of a Mathematical Blog*. Amer. Math. Soc., Providence, RI, 2008.

[86] T. Tao, Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* 28, 657–688, 2008.

[87] A. Troelstra, *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics 344, Springer, Berlin, 1973.

[88] L. Tzafriri, Remarks on contractive projections in L_p-spaces. *Israel J. Math.*, 7:9–15, 1969.

[89] H.-K. Xu, Inequalities in Banach spaces with applications. *Nonlinear Anal.*, 16:1127–1138, 1991.

[90] H. Y. Zhou, Weak convergence theorems for strict pseudo-contractions in Banach spaces. *Acta Math. Sin. (Engl. Ser.)* 30, no. 5, 755–766, 2014.