Quantitative Analysis of Iterative Algorithms in Fixed Point Theory and Convex Optimization

Vom Fachbereich Mathematik
der Technischen Universität Darmstadt

zur Erlangung des Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigte Dissertation

von

M.Sc. Daniel Körnlein
aus Frankfurt a. M.

Referent:	Prof. Dr. Ulrich Kohlenbach
1. Korreferent:	Prof. Dr. Genaro López Acedo
2. Korreferent:	Prof. Dr. Thomas Streicher
Tag der Einreichung:	3. Februar 2016
Tag der mündlichen Prüfung:	12. April 2016

Darmstadt 2016
– D17 –

Acknowledgments

I am very grateful to my advisor Prof. Dr. Ulrich Kohlenbach, from whom I have learned virtually everything I know about Proof Theory and who always found the time and patience to answer any questions I had, and to discuss the challenging problems I encountered during my time as a PhD student. I am also very grateful to Prof. Dr. Genaro López Acedo, who hosted me in Seville in September 2014, where the foundations for the results of Chapter 8 were laid out. I also thank the Department of Mathematics at Technische Universität Darmstadt and the Deutsche Forschungsgemeinschaft (Project 1737/5-2) for providing the financial support for this thesis.

Darmstadt, May 25, 2016

Deutsche Zusammenfassung

Die vorliegende Dissertation befasst sich mit der Anwendung logischer Methoden auf konkrete mathematische Konvergenzresultate der Fixpunkttheorie mit dem Ziel, uniforme Konvergenzschranken aus ihren jeweiligen Beweisen zu extrahieren. Dies geschieht mithilfe sogenannter Metatheoreme und den zugrundeliegenden Ideen der mathematischen Logik, welche unter allgemeinen Bedingungen an den Beweis die Existenz und Konstruierbarkeit solcher Schranken garantieren. Diese Bedingungen an den Beweis umfassen mitunter auch den Gebrauch von klassischer Logik und gewisser idealer Prinzipien. Tatsächlich kann es sogar wünschenswert sein, dass ein Beweis möglichst starken Gebrauch solcher prima facie inkonstruktiver Argumente macht, da eine beweistheoretische Analyse in solchen Fällen besonders interessante logische Phänomene aufweist.

Bevor wir in Kapitel 5 in die Analyse konkreter Beweise einsteigen, motivieren wir zunächst das „proof mining"-Programm aus mathematischer und logischer Sicht und gehen dabei besonders auf die allgemeine Form der zu erwartenden Resultate ein. Anschließend führen wir auf minimalistische Art die zentralen logischen Grundlagen ein. Danach ordnen wir die wichtigsten in dieser Arbeit vorkommenden analytischen Konzepte im Lichte der Logik und Metatheoreme ein.

Die erste logische Analyse befasst sich mit sogenannten Halpern-Iterationen für nichtexpansive Abbildungen. Zunächst extrahieren wir eine Konvergenrate für die asymptotische Regularität unter gewissen Bedingungen von Xu in beliebigen normierten Räumen. Zudem extrahieren wir eine Rate der Metastabilität in gleichmäßig glatten Banachräumen relativ zu einer Metastabilitätsrate für die Resolvente der betrachteten nichtexpansiven Abbildung. Da eine solche Metastabilitätsrate für die Resolvente in Hilberträumen bekannt ist, erhalten wir eine unabhängige Metastabilitätsrate in diesen Fall. Da die Halpern Iteration für lineare Abbildungen genau der Folge der Cesàro-Mittel entspricht, erhalten wir als Korollar hierzu insbesondere eine Metastabilitätsrate für den Birkhoffschen Ergodensatz, einer Verallgemeinerung des bekannten von Neumannschen Ergodensatzes.

Danach wenden wir uns Konvergenzresultaten für Bruck-Iterationen gegen Fixpunkte pseudokontraktiver Abbildungen zu. Fixpunkte pseudokontraktiver Abbildungen entsprechen genau den Nullstellen sogenannter akkretiver Operatoren. Akkretive Operatoren spielen im Gebiet der partiellen Differentialgleichungen eine gewichtige Rolle, da sie nichtlineare Evolutionsgleichungen beschreiben. Dementsprechend spiegeln die Nullstellen dieser Operatoren die Gleichgewichtspunkte der entsprechenden Evolutionsgleichung wieder. Einerseits beweisen wir eine Metastabilitätsrate der Bruck-Iteration für Lipschitzstetige Pseudokontraktionen in Hilberträumen, deren Komplexität sich nur unwesentlich von der Rate aus Kapitel 5 für die Teilklasse der nichtexpansiven Abbildungen unterscheidet. Andererseits extrahieren wir eine Metastabilitätsrate für den Fall, dass der Operator nur demistetig ist. Demistetigkeit ist eine wesentlich liberalere Bedingung als Lipschitzstetigkeit. Tatsächlich sind solche Operatoren im Allgemeinen sogar unstetig bezüglich der starken Topologie auf Bild und Definitionsbereich. Dies äußert sich auch in den Konvergenzschranken, welche wesentlich komplexer sind.

Als nächstes wenden wir uns dem Variationsungleichungsproblem der konvexen Optimierung zu. Dieses kann mittels der „Hybrid Steepest Descent“-Methode gelöst werden, deren Konvergenz beweistheoretisch analysiert wird. Neben einer quantitativen Version der Konvergenz für den Fall eines einzelnen Operators, welche unter anderem eine Metastabilitätsrate umfasst, behandeln wir auch die Konvergenz der „Hybrid Steepest Descent“-Methode für endliche Familien von Operatoren.

Im letzten Kapitel arbeiten wir auf eine Verallgemeinerung der Resultate aus Kapitel 5 vom Hilbertraum zu gleichmäßig glatten und gleichmäßig konvexen Banachräumen hin. Dafür geben wir einen alternativen Beweis an ohne Gebrauch des Zornschen Lemmas, in welchem der Gebrauch des Satzes von Tychonoff, welcher ebenfalls äquivalent zum Auswahlaxiom ist, auf die schwache Kompaktheit beschränkter, abgeschlossener und konvexer Teilmengen solcher Banachräume. Wir argumentieren, dass letzteres von besagten Metatheoremen möglicherweise abgedeckt ist. Eine endgültige Antwort allerdings muss Gegenstand künftiger Forschung bleiben.

Contents

1 Introduction

The ongoing program of 'proof mining' aims to extract new, quantitative information in the form of bounds and rates from prima facie noneffective proofs in mathematics. In doing so, proof mining and the thesis at hand draws a bridge between mathematics and logic, prompting a lively interaction between mathematical practice and logical theory. This thesis applies proof mining paradigms to several convergence results in fixed point theory and nonlinear optimization, resulting in new complexity information in the form of rates of convergence or rates of metastability.

A first – yet from the proof mining perspective trivial – example is Banach's famous fixed point theorem; The theorem itself asserts that any contraction mapping on a complete metric space possesses a unique fixed point, and that this fixed point may be approximated by means of the Picard iteration. The proof, on the other hand, exhibits the well known rate of convergence, which can be read off immediately from it. Moreover, the rate is uniform in the mapping, the underlying metric space and the starting point in that it only depends on a Lipschitz constant for the mapping in question and an upper bound on its initial displacement. In other words, for given contraction constant q and positive real number b, the rate of convergence is valid for the class of all metric spaces, all mappings on this metric space with contraction constant q and all starting points that are displaced by the operator by a distance of at most b.

The reason that Banach's fixed point theorem admits such a uniform rate of convergence and that its proof divulges the rate so easily has two reasons: It is constructive and convergence to the limit point is monotone. In general, however, neither is the case. It is precisely in these cases that proof mining, or rather the general logical theorems and tools behind the name, come into play.

Under vastly general conditions on the proof that admit non-constructive reasoning in the form of ideal principles and classical logic, so-called metatheorems guarantee the existence of uniform complexity information. For instance, a large part of classical analysis is covered by those metatheorems. Furthermore, the complexity information is not only guaranteed to exist, but can be extracted from the proof at hand.

The original proof is moreover transformed into a new proof of the new statement which exhibits the additional complexity information. The new proof then exhibits no trace of its proof-theoretic manipulation and is carried out without reference to any mathematical logic. This has the further advantage that the complexity bound is not only valid, but its proof is easily accessible to the experts of the respective mathematical field and can be published in the corresponding journals.

1.1 Plan of this Thesis

In Chapter 2, we first discuss the historic background of proof mining, highlighting in particular classic ideas and theorems that prompted its investigation. We then go into some detail of the nature of results one can expect in practice, as well as the difficulties and limitations one will encounter. The chapter is rounded off by motivating the general form of frequently obtained results.

Chapter 3 provides a minimalistic introduction to the proof-theoretic background. It introduces the most important formal proof system, focusing on practical aspects rather than the vast logical background. The main metatheorem on proof mining is then stated in a form tailored to the context of the thesis.

The most important analytical preliminaries are covered in Chapter 4, with a strong emphasis on their compatibility with the logical framework and the metatheory covered in Chapter 3. We also provide and solve a toy example of a task in proof mining that could realistically present itself in practice.

Beginning with Chapter 5, we present the main results of the thesis. We start off with a full rate of convergence for the asymptotic regularity under conditions due to Xu in arbitrary normed spaces of the Halpern iteration, which approximates fixed points of nonexpansive maps. If the underlying space is uniformly smooth and complete in its norm, we furthermore give a rate of metastability for the Halpern iteration relative to a rate of metastability for its resolvent. For Hilbert spaces, the latter is known, so we combine the two results to obtain an unconditional rate of metastability for the Halpern iteration. Since the Halpern iteration reduces to the Cesàro mean for linear mappings, these results apply in particular to von Neumann's Mean Ergodic Theorem. The main results of Chapter 5 have been published in [69]. In Chapter 9, we work towards generalizing the rate of metastability for the resolvent from Hilbert spaces to the much broader class of uniformly convex and uniformly smooth Banach spaces, providing a partial result and outlining a strategy for further investigation. Moreover, we answer a question of purely analytical nature left open in [58].

In Chapter 6, we generalize the results of Chapter 5 from nonexpansive mappings to the broader class of pseudocontractions, which appear naturally in, for instance, nonlinear evolution equations. This broader class calls for an adapted iteration scheme due to Bruck, while the complexity of the rates is essentially maintained over those obtained in Chapter 5. These results have appeared in [71]. Chapter 7 further generalizes the results of Chapter 6 from Lipschitzian to demicontinuous pseudocontractions, i.e. pseudocontractive maps which are continuous from the strong to the weak topology, albeit at the expense of significantly more involved rates. We then discuss the interesting logical phenomena that stem from admitting norm-discontinuous mappings. The results of Chapter 7 have been submitted and are currently under peer review [68].

Finally, we give in Chapter 8 metastability results in nonlinear optimization for the Hybrid Steepest Descent Method, an explicit algorithm of convex optimization with numerous practical applications. The rates found in this chapter highlight several interesting properties of the employed proof-theoretic methods. First of all, one witnesses the complete modularity of the Dialectica interpretation and the advantages derived therefrom. Moreover, the theorem provides a prime example of the potential complexity of the finitary combinatorial core hidden behind ideal principles: The proof of the central lemma increases from a few lines in its original form to more than ten pages in its quantitative form. Finally, as predicted by metatheory, the types of the involved operators blows up throughout the proof, only to collapse again to yield the usual rate of metastability.

2 Proof Mining

2.1 What is Proof Mining?

The program of proof mining has its roots in the 1950's, when Kreisel asked the central question [27, 78]

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

To make sense of this paradigmatic question, it is necessary to give meaning to 'restricted means', and the additional information we wish to extract from proofs using only these restricted means. The former is usually interpreted to mean that a given proof of a theorem A can be formalized in some logical system. The proof is then transformed using proof interpretations into a new proof of a transformed theorem A^I, where the statement A^I now exhibits additional information.

Roughly speaking, the new information exhibited in A^I will be in the form of witnesses, finite lists of candidates or bounds for the existential quantifiers present or hidden in the original theorem A. For instance, if $A \equiv \forall x \, \exists y \, B(x, y)$, a witness is a term t such that $\forall x \, B(x, tx)$ is true; A list of candidates is a tuple $t_1, \ldots, t_n$ such that $\forall x \, (B(x, t_1 x) \vee \ldots \vee B(x, t_n x))$ is true; A bound is a term t such that $\forall x \, \exists y \leq tx \, B(x, y)$ is true. For the latter, the relation $\leq$ can have different meanings, depending on the range of the quantifiers. For example, if y ranges over the natural numbers, then '$\leq$' is the usual order on the naturals.

On the other hand, if $A = \forall x \, A_{qf}(x)$, where $A_{qf}(x)$ is quantifier-free, then the statement is sometimes called *complete* in the sense of Bishop [9] as it does not ask for witnessing data. In such cases, A^I will not exhibit any new information. However, it is interesting to observe that an implication between two complete theorems can be viewed as incomplete. To illustrate this point, suppose we have two functions $f, g : [0, 1] \to [0, 1]$ and a theorem A asserting that any root of f is also a root of g, i.e. $A \equiv \forall x \in [0, 1] \, (f(x) = 0 \to g(x) = 0)$. The additional, quantitative content one can hope to extract from a proof of A becomes apparent if one chooses a suitable prenexation, which is given by

$$\forall x \in [0, 1] \, \forall n \in \mathbb{N} \, \exists k \in \mathbb{N} \left(|f(x)| \leq 2^{-k} \to |g(x)| \leq 2^{-n} \right). \tag{2.1}$$

As before, the statement A asks for witnessing data for the existential quantifier, which has a very intuitive meaning in this case; Namely, given a natural number n, how close to zero must $f(x)$ be for $g(x)$ to be 2^{-n}-close to zero. It is also noteworthy that in this special case, any bound on k is already a witness. This is due to the fact that the hypothesis of the implication is monotone in k, i.e. whenever the hypothesis is satisfied by some natural number k and $k \leq k'$, then it is also satisfied by k'. Monotonicity of formulas can be very helpful in practice as it often allows for the elimination of a quantifier.

We now introduce the Brouwer-Heyting-Kolmogorov ('BHK') interpretation, which is considered as the foundation for intuitionistic, or constructive, logic. As such, it is closely related to our cause. In fact, although it would be a gross misuse in view of its original purpose, it could be interpreted as a first, informal blueprint for a proof interpretation capable of exhibiting new, constructive content inherent in constructive proofs.

The *BHK interpretation* consists of the following clauses:

(i) There is no proof of falsity $\perp$.

(ii) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(iii) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n such that q is a proof of A if $n = 0$ and q is a proof of B if $n \neq 0$.

(iv) A proof of $A \to B$ is a construction q which transforms any hypothetical proof s of A into a proof $q(s)$ of B.

(v) A proof of $\forall x\, A(x)$ is a construction q which produces for every construction c_d of an element d of the domain a proof $q(c_d)$ of $A(d)$.

(vi) A proof of $\exists x\, A(x)$ is a pair (c_d, q), where c_d is a construction of an element d of the domain and q is a proof of $A(d)$.

One could argue that 'explanation' is more accurate than 'interpretation', since the BHK interpretation is formally not a proof interpretation, see e.g. [92]. For intuitionistic number theory ('Heyting Arithmetic'), which is classical number theory ('Peano Arithmetic') without the law of the excluded middle $A \vee \neg A$, the BHK interpretation has a very natural formal implementation; Kreisel's modified realizability [64, 65]. In this specific case, 'restricted means' in Kreisel's original question then refers to provability within intuitionistic number theory. Modified realizability can, in the light of the BHK interpretation, be viewed as a form of 'bookkeeping' of witnesses throughout the proof by means of partial recursive functions, while the absence of the law of the excluded middle ensures that no content can be conjured 'out of thin air' to begin with (more on this later, when we discuss limitations to the extractibility of witnesses and bounds). The reader is referred to Troelstra [97] and Troelstra, van Dalen [98] for further information on intuitionistic and classical number theory as well as recursive realizability.

The BHK interpretation and recursive realizability have two main drawbacks from our point of view. Let us first recall the earlier example of the theorem

$$A \equiv \forall x \in [0, 1]\, (f(x) = 0 \to g(x) = 0)$$

for self-mappings f and g of the unit interval. Under the BHK interpretation, this statement is computationally empty because A is built up using only $\to$ and $\forall$, while the BHK interpretation only asks for constructions in the $\vee$ and $\exists$ clauses.

Moreover, proof interpretations implementing the BHK are, by design, only applicable to constructive proofs. It is worth noting that, on the one hand, this immediately blocks a workaround to the issue outlined in the previous paragraph; In intuitionistic systems, a theorem need not be provably equivalent to any of its prenexations. In fact, given a proof q of some theorem A in the sense of the BHK interpretation, it is impossible, in general, to extend q to a proof q' of any prenexation of A. On the other hand, our aim will be to apply proof mining methods to concrete theorems in mathematical research, so the restrictions on the proof should be as broad as possible. As Herbrand's theorem shows, which we state now, it is actually not necessary to restrict oneself to constructive theories.

Herbrand's Theorem. *Suppose $A \equiv \forall x\, \exists y\, A_{qf}(x, y)$, where A is quantifier-free. Then A is provable in predicate logic without equality if and only if there exist finitely many terms $t_1 x, \ldots, t_n x$ built up using x, and the constant symbols, free variables and function symbols of A, such that*

$$A(x, t_1 x) \vee \ldots \vee A(x, t_n x)$$

is a tautology.

Herbrand's theorem as stated above is directly applicable to open theories, i.e. theories axiomatized by purely universal formulas, by using the axioms used in a given proof as implicative assumptions. The Herbrand disjunction is then not necessarily a tautology, but provable in the theory. Although the restriction of being open is necessary for this strategy, it is possible to extend it to Peano Arithmetic and formal systems accommodating abstract or represented spaces using the Dialectica interpretation, as we will see later on.

Still, it is already interesting to note that, in the presence of classical logic, one has to restrict the logical complexity of the underlying theorem. In fact, Herbrand's theorem is sharp in the sense that there exists a logically valid sentence $A \equiv \exists x\, \forall y\, A_{qf}(x, y)$ in the language of Peano Arithmetic such that, for no finite list $t_1, \ldots, t_n$, Peano arithmetic proves $\forall y\, P(t_1, y) \vee \ldots \vee \forall y\, P(t_n, y)$, see Kohlenbach [49].

Therefore, when given an arbitrary formula in prenex normal form, it is necessary to pass first to the so-called Herbrand normal form: Any formula in prenex normal form $A :\equiv \exists x_1\, \forall y_1 \ldots \exists x_n\, \forall y_n\, A_{qf}(x_1, y_1, \ldots, x_n, y_n)$ is valid if and only if its Herbrand normal form A^H is valid, which is given by

$$A^H :\equiv \exists x_1, \ldots, x_n\, A_{qf}\big(x_1, f_1(x_1), x_2, f_2(x_1, x_2), \ldots, x_n, f_n(x_1, \ldots, x_n)\big),$$

where $f_1, \ldots, f_n$ are new function symbols. Then, Herbrand's theorem is applicable, and we get a list of candidates for the new theorem A^H. The variables $x_1, \ldots, x_n$ that satisfy A^H, viewed as functionals $\Phi_1, \ldots, \Phi_n$ in their function arguments $\underline{f} := f_1, \ldots, f_n$, are then said to satisfy Kreisel's no-counterexample interpretation [62, 63] of A, i.e.

$$A_{qf}\left(\Phi_1(\underline{f}), f_1(\Phi_1(\underline{f})), \ldots, \Phi_n(\underline{f}), f_n(\Phi_1(\underline{f}), \ldots, \Phi_n(\underline{f}))\right).$$

In this case, we write $\Phi_1, \ldots, \Phi_n$ n.c.i. A. The term 'no-counterexample interpretation' stems from the observation that the negation $\neg A$ is equivalent to the prenex formula $\forall x_1 \exists y_1 \ldots \forall x_n \exists y_n \neg A_{qf}(x_1, y_1, \ldots, x_n, y_n)$, so the functions $f_1, \ldots, f_n$ attempt to produce counterexamples to choices of $x_1, \ldots, x_n$. The functionals $\Phi_1, \ldots, \Phi_n$, in turn, refute any such strategy of producing counterexamples for A.

To clarify these points, let us examine the no-counterexample interpretation of a hypothetical theorem that some real sequence $(a_n) \subset \mathbb{R}$ is Cauchy, i.e. a theorem $A = \forall \varepsilon > 0 \, \exists n \in \mathbb{N} \, \forall m \in \mathbb{N} \, (|a_{n+m} - a_n| \leq \varepsilon)$. The Herbrand normal form of A then reads $A^H :\equiv \forall \varepsilon > 0 \, \exists n \in \mathbb{N} \, (|a_{n+g(n)} - a_n| \leq \varepsilon)$, and a hypothetical list of candidates for '$\exists n \in \mathbb{N}$' can then be transformed into a bound on n by taking their maximum:

$$\forall \varepsilon > 0 \, \exists n \leq \Phi(\varepsilon, g) \, \big(|a_{n+g(n)} - a_n| \leq \varepsilon\big) \, .$$

A bound Φ n.c.i. A can then be transformed into a bound Φ' that satisfies the slightly different, but equivalent form (where we use the notation $[n; n + g(n)] := \{n, \ldots, n + g(n)\}$)

$$\forall \varepsilon > 0 \, \exists n \leq \Phi(\varepsilon, g) \, \forall i, j \in [n; n + g(n)] \, (|a_i - a_j| \leq \varepsilon) \, .$$

This form has become famous under the name of 'metastability', a term introduced by Tao [95,96], who independently rediscovered the concept in 2008. A thorough motivation of metastability and its variants will be given in Section 2.3 and in subsequent chapters respectively.

A decisive defect of the no-counterexample interpretation is that it fails to be sound with respect to the modus ponens, which is illustrated by the following example. Suppose theorem $B \equiv \forall u \, \exists v \, B_{qf}(u, v)$ is shown by means of a lemma of the form $A \equiv \forall x_1 \, \exists y_1 \ldots A_{qf}(x_1, y_1, \ldots)$, and that we have functionals Φ n.c.i. A and Ψ n.c.i. $(A \to B)^{pr}$, where we choose the prenexation

$$(A \to B)^{pr} \equiv \exists x_1 \, \forall y_1 \ldots \forall u \, \exists v \, (A_{qf}(x_1, y_1, \ldots) \to B(u, v))$$

Then, there is in general no term in Gödel's system $\mathbf{T}$ that transforms uniformly in A the functionals Φ and Ψ into a functional that satisfies the no-counterexample interpretation of B. In fact, one needs Spector's bar recursion [90] of lowest type or Feferman's μ-operator [26] for the construction of the required functional, see Kohlenbach [47,49].

This problem arises from the fact that the no-counterexample interpretation remains at type-level two; that is, it only involves counterfunctions which are functions of the ground type into itself, and functionals mapping such counterfunctions back into the ground type. As a result, irrespective of the prenex normal form one chooses for $A \to B$, one needs increasingly stronger instances of the axiom of choice to prove $((A \to B)^{pr})^H \to (A \to B)^{pr}$ as A or B become increasingly complex.

If, on the other hand, one allows for arbitrary (finite) types, one can find to each formula A a formula $(A')^D$ in prenex normal form such that the equivalence $A \leftrightarrow (A')^D$ is provable using only the axiom of choice for *quantifier-free* formulas. This is achieved by means of Kuroda's negative translation [66] followed by Gödel's famous Dialectica interpretation [34]. We now outline how this fact can be used to interpret any proof in Peano Arithmetic in all finite types in a fashion that allows for the extraction of witnesses.

(i) All formulas A in the proof are transformed into $(A')^D \equiv \forall \underline{x}\, \exists \underline{y}\, A_{qf}(\underline{x}, \underline{y})$, which is in prenex normal form, by means of Kuroda's negative translation A' and the Dialectica translation A^D.

(ii) Since the equivalence of A and $(A')^D$ was provable in Peano Arithmetic (its weakly extensional version in all finite types, which we tacitly assume from now on) with quantifier-free choice, its negative translation $\neg\neg\forall \underline{x}\, \neg\neg\exists \underline{y}\, A_{qf}(\underline{x}, \underline{y})$ is provable in Heyting arithmetic with quantifier-free choice.

(iii) This is intuitionistically equivalent to $\forall \underline{x}\, \neg\neg\exists \underline{y}\, A_{qf}(\underline{x}, \underline{y})$ (see e.g. [21]). Using Markov's principle

$$\neg\neg\exists \underline{x}\, A_{qf}(\underline{x}) \to \exists \underline{x}\, A_{qf}(\underline{x}),$$

one can therefore prove in Heyting Arithmetic with quantifier-free choice and Markov's principle the original formula $(A')^D$.

(iv) By the soundness of the Dialectica interpretation with respect to the Axiom of Choice and Markov's principle, one can extract terms $\underline{t}$ from the proof such that $(A')^D(\underline{x}, \underline{tx})$.

In essence, one avoids the use of the full Axiom of Choice, which is not admissible in the presence of classical logic, by instead increasing the types of the quantified variables. Moreover, this method manages to push all traces of classical logic through Markov's principle, which in turn has a trivial interpretation under the Dialectica, resulting essentially in instances of $A \to A$.

This is best illustrated by means of the following example: For a formula $A :\equiv \forall u\, \exists v\, \forall x\, \exists y\, A_{qf}(u, v, x, y)$, the Kuroda negative translation is given by

$$A' \equiv \neg\neg\forall u\, \neg\neg\exists v\, \forall x\, \neg\neg\exists y\, A_{qf}(u, v, x, y)$$

Using Markov's principle, one can discard the innermost double negation, while the elimination of the outermost double negation is even intuitionistically valid:

$$A' \leftrightarrow \forall u\, \neg\neg\exists v\, \forall x\, \exists y\, A_{qf}(u, v, x, y)$$

Then, using only quantifier-free choice

$$\begin{aligned} A' &\leftrightarrow \forall u\, \neg\neg\exists v\, \exists Y\, \forall x\, A_{qf}(u, v, x, Yx) \\ &\leftrightarrow \forall u\, \neg\neg\forall X\, \exists v\, \exists Y\, A_{qf}(u, v, XY, Y(XY)). \end{aligned}$$

Intuitionistically, $\neg\neg\forall x\, B(x) \to \forall x\, \neg\neg B(x)$ for any formula B, so we can use Markov's principle once more to obtain

$$\forall u\, \forall X\, \exists v\, \exists Y\, A_{qf}(u, v, XY, Y(XY)) \equiv (A')^D.$$

Terms Φ and V then solve $(A')^D$ if

$$\forall u \, \forall X \, A_{qf}(u, VuX, X(\Phi uY), \Phi(X(\Phi uY))).$$

While this might seem artificial at first, it turns out that (provably in Heyting Arithmetic, see Kohlenbach [49]), $(A')^D \to A^H$, so a solution of $(A')^D$ can be easily transformed into a solution of the no-counterexample interpretation of A using typed λ-terms and elementary number-theoretic operations, see [49] . Moreover, for theorems of the form $\forall\exists\forall$, the no-counterexample interpretation and the *ND* interpretation (negative translation followed by the Dialectica) actually coincide. Moreover, in Chapters 8 and 9, particularly Section 9.4, we will actually see concrete examples of convergence theorems which use lemmas whose *ND* interpretations are substantially different from their no-counterexample interpretations.

While only applicable to Peano Arithmetic and its finite-type extension in its original form, the method of using negative translation and a monotone variant of the Dialectica interpretation due to Kohlenbach [45] have been extended to more general theories by Kohlenbach and others, see for example [31, 37, 48, 49]. The general idea is that one can enrich the base system by adding new structure along with defining axioms as long as the new axioms fulfill certain a-priori criteria, which we discuss in the Chapter 3. In this way, one can add abstract metric, normed or hyperbolic spaces, Polish spaces or duality mappings, to name but a few.

While 'proof mining' in the sense of Kreisel's original question has been applied to number theory [77, 78], combinatorics [5, 33], algebra [23] and computer science [6], the systematic approach outlined in the previous paragraph has mainly led to numerous proof-mining applications in nonlinear functional analysis and numerical analysis, in particular to approximation theory, ergodic theory and fixed point theory, including the results presented in this thesis. For a survey of proof theory in these areas up until 2008, see [50]. Results in proof mining since 2008 are among [2, 29, 30, 52, 55, 56, 60, 75].

2.2 Proof Mining in Practice

So far, we have taken a first glimpse at the general idea behind proof mining. Apart from the fact that vast amounts of the mathematical structure used in metric fixed point theory are axiomatized by formulas which are admissible for Dialectica-based interpretations, fixed point theory exhibits a plethora of iteration schemes for various classes of operators. From a proof mining perspective, the convergence statements found in the literature for these iterations not only provide highly interesting opportunities to put the logical machinery to work, but also allows to observe new phenomena.

For instance, we give in Chapter 9 a partial solution to the Dialectica interpretation of the existence of the so-called sunny nonexpansive retraction onto the fixed point set of a nonexpansive mapping in uniformly convex and uniformly smooth Banach spaces. Bruck's original existence proof is highly noneffective and well beyond the reach of existing metatheorems; it uses Tychonoff's theorem to show that the Cartesian product indexed over a closed, bounded and convex subset of the underlying space is weakly compact, and subsequently proceeds using Zorn's lemma. After a detailed proof-theoretic analysis guided by the metatheorems presented in Chapter 3, we transform the original proof into a 'less inconstructive' proof. We argue that the new proof is within the scope of existing metatheorems, although the actual extraction of the bounds will remain the subject of future research.

This is in stark contrast to the Hilbert space case, where the sunny nonexpansive retraction is simply the metric projection, a concept that is well-understood and whose Dialectica interpretation (or rather, its Shoenfield variant [89], which was shown by Streicher and Kohlenbach [91] to coincide with original Dialectica interpretation followed by the Krivine negative translation) has been treated extensively by Kohlenbach in [51]. In fact, to prove for any given $\varepsilon > 0$ the existence of an ε-good projection, the induction axiom over purely existential sentences suffices, while the Axiom of Countable Choice is required to produce a sequence of ε-good approximations, which in turn implies the existence of the metric projection.

Another interesting logical phenomenon can be witnessed in Chapter 7, where we examine Bruck's iteration scheme for demicontinuous pseudocontractions T. Since the metatheorems formulated in the next chapter only allow the rule of extensionality for quantifier-free formulas instead of full extensionality, it is almost always necessary to impose uniform norm-continuity on the operator T, which is much stronger than demicontinuity. The proof-theoretic analysis, however, shows not only that uniform continuity is not necessary for the extraction of bounds, but also that the demicontinuity leaves no trace in the bound. It is only a-posteriori that this behavior fits nicely within the existing theory, since demicontinuity is only used in the original proof to show the existence of the resolvent (z_t) defined implicitly as the unique path satisfying $z_t = tTz_t + (1-t)z_0$ for all $t \in [0, 1)$ and an arbitrary but fixed anchor point z_0. Therefore, one could actually drop the demicontinuity assumption in the original theorem and instead add the sequence and its defining axiom directly to the formal system, making the existence of the path (z_t) an axiom instead of a lemma. As opposed to demicontinuity, the new axiom is complete in the sense introduced earlier, so it does not contribute to the bound. Since the implication "demicontinuity $\Rightarrow$ existence of (z_t)" is valid, the bound is then also valid under the original assumption.

Generally, while the metatheorems do provide a-priori criteria for the extractibility of bounds, proofs generally require a fair amount of preprocessing prior to the actual extraction, either in the form of finding suitable formulations for the hypotheses, or the elimination of unnecessary uses of inconstructive ideal principles. For both of these tasks, one relies heavily on proof-theoretic tools. They not only provide guidance how to preprocess the proof and its hypotheses, but also ensure that these steps are carried out in a systematic way.

Therefore, we stress that this thesis and proof mining in general is not only concerned with applying existing logical machinery to proofs covered by the known metatheorems. In fact, finding new applications and extending the known metatheory is of equal importance. This usually involves a close interaction between logical theory and mathematical practice, where a proof is examined with respect to the ideal principles used therein. This is usually not a straightforward task since proofs frequently use highly noneffective methods that can be eliminated in favor of principles to which the metatheorems can be extended, thereby pushing the boundary of the mathematical theory that is covered by metatheory.

2.3 Metastability

Suppose we have a theorem that some sequence (x_n) in a Banach space converges to an element x, i.e.

$$\forall \varepsilon > 0 \, \exists n \in \mathbb{N} \, \forall m \geq n \, (\|x_m - x\| \leq \varepsilon).$$

A natural follow-up is to ask how fast (x_n) converges. Ideally, one would like to have a rate of convergence, that is, a function $r : (0, \infty) \to \mathbb{N}$ such that

$$\forall \varepsilon > 0 \, \exists n \leq r(\varepsilon) \, \forall m \geq n \, (\|x_m - x\| \leq \varepsilon),$$

However, as we discussed in the remarks following Herbrand's Theorem, $\forall\exists\forall$ statements shown by means of classical logic in general do not allow for any term or bound extraction. A rate of metastability, i.e. a function $\Phi : (0, \infty) \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$ such that

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Phi(\varepsilon, g) \, \forall i, j \in [n; n + g(n)] \; (\|x_i - x_j\| \leq \varepsilon) \, . \tag{2.2}$$

on the other hand, is guaranteed to exist under vastly general circumstances. Although it presents slightly weaker complexity information than a full rate of convergence, we will see in the respective chapters that a rate of convergence is even ruled out in all cases where we provide a rate of metastability. Rates of metastability are therefore – in some sense – optimal in these cases.

Moreover, the new statement (2.2) is now finitary since, for given $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, it only involves a finite initial segment of the sequence (x_n). Moreover, in the process of finding a rate of metastability, ideal principles used in the original proof are frequently eliminated. In fact, statement (2.2) trivially implies the original convergence result; A single application of the Axiom of Choice over quantifier-free statements implies that the sequence is Cauchy, and completeness implies its convergence.

In this thesis, this is the case for the main results in Chapters 7 and 8. Therefore, proof mining can also be seen as following the modified Hilbert program, which consists in reducing mathematics to constructive reasoning. In fact, many theorems are a-priori noneffective through their use of ideal principles, but a close inspection guided by the logical machinery found in Chapter 3 allows one to eliminate them. This observation, in turn, can be used as a blueprint for similar situations and turned into a new metatheorem. In this way, a highly interesting interplay between mathematical practice and proof-theoretic theory arises.

Another aspect of rates of metastability and other quantitative results obtained using these metatheorems is their striking uniformity. Suppose for a moment that the sequence (x_n) was designed to approximate a fixed point of some mapping T on a Banach space. If the conditions of a metatheorem are fulfilled, we are not only guaranteed the existence of a rate of metastability, but also uniformity in the concrete space and the concrete operator. The bound will only depend on certain bounds derived from the conditions on X and T. For instance, if T was assumed to be Lipschitz continuous, the rate of metastability would depend on an upper bound of the Lipschitz constant. If the existence of a fixed point of T was additionally assumed, then we would also require an upper bound on the norm of any fixed point of T.

3 The Logical Framework

In this chapter, we give a brief introduction to the logical preliminaries mentioned in the introduction. We aim to provide a minimal summary of the machinery to make this thesis self-contained. To this end, we first define the formal system $\mathcal{A}^\omega$, before turning to majorization and Kohlenbach's metatheorems. For a comprehensive and extensive book treatment, we refer the reader to [49]. To complete the chapter, we sketch how the formal system and the metatheorems accommodate for concrete theorems in fixed point theory and nonlinear optimization.

3.1 The System $\mathcal{A}^\omega$

In the first chapter, we stressed that it is essential to allow for arbitrary finite types. Therefore, we work in a many-sorted logical framework that accommodates each finite type. The set of all types, denoted by $\mathbf{T}$, is defined inductively by the clauses

(T1) $0 \in \mathbf{T}$ and

(T2) $\rho \in \mathbf{T}$ and $\tau \in \mathbf{T}$ imply $\rho \to \tau \in \mathbf{T}$.

The base type 0 corresponds to the type of the natural numbers, while the intended meaning of $\tau \to \rho$ is the type of functions mapping objects of type τ to objects of type ρ. For instance $(0 \to 0) \to 0$ is the type of functionals mapping number-theoretic functions, i.e. functions from the naturals to the naturals, to natural numbers. An example of such a function is the 'evaluation at zero' operator, which asks for a number-theoretic function and evaluates it at zero.

For the base type, we have equality $=_0$ with the usual equality axioms

(I1) $\forall x^0 \, (x =_0 x)$ (symmetry),

(I2) $\forall x^0, y^0 \, (x =_0 y \to y =_0 x)$ (reflexivity) and

(I3) $\forall x^0, y^0, z^0 \, (x =_0 y \wedge y =_0 z \to x =_0 z)$ (transitivity).

Higher-type equality is a defined notion, reduced to the ground type. To this end, observe first that any type $\tau \in \mathbf{T}$ can be uniquely written in the form $\tau = \tau_1 \to (\tau_2 \to \ldots \to (\tau_n \to 0))$. Equality $s =_\tau t$ between two terms s and t of type $\tau = \tau_1 \to (\tau_2 \to \ldots \to (\tau_n \to 0))$ is then an abbreviation for

$$\forall x_1^{\tau_1}, \ldots, \forall x_n^{\tau_n} \, (s x_1 \ldots x_n =_0 t x_1 \ldots x_n).$$

To populate the base type, we add a constant symbol 0, which is not to be confused with the base type, and a function symbol S of type $0 \to 0$ for the successor function, along with the axioms

(S1) $\forall x^0\,(S(x) \neq 0)$ and

(S2) $\forall x^0\,(S(x) = S(y) \to x = y)$.

The types are populated by means of the recursor R, the projector Π and the combinator Σ, which allow for primitive recursion, higher-type recursion and λ-abstraction and are therefore essential for the soundness of the proof interpretations. Their defining axioms are highly technical and therefore left out in this thesis. The inclined reader is again referred to [49] for details.

Furthermore, we have the schema of induction

$$A(0) \wedge \forall x^0\,(A(x) \to A(S(x))) \to \forall x^0 A(x)$$

and a rule of extensionality for quantifier-free formulas

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

where A_0 is quantifier free and s^ρ, t^ρ and r^τ are terms of arbitrary type. Observe that this is much weaker than having full extensionality, i.e. the schema

$$\forall z^\rho\,\forall x_1^{\rho_1}, y_1^{\rho_1}, \ldots, x_n^{\rho_n}\,\forall y_n^{\rho_n}\left(\bigwedge_{i=1}^{n}(x_i =_{\rho_i} y_i) \to z x_1 \ldots x_n =_0 z y_1 \ldots y_n\right)$$

for all types $\rho = \rho_1 \to (\rho_2 \to \ldots(\rho_n \to 0))$. Full extensionality, however, is never admissible in the presence of classical logic if one wishes to extract bounds or witnesses from proofs, see Kohlenbach [49].

In practice, when extensionality is used in combination with classical logic, one then has to ensure that the operator in question is provably extensional. Consider for example some operator $T : C \to C$, where C is a subset of a Banach space X. For the metatheorems stated later in this section, proving extensionality of X is possible if and only if the operator satisfies a uniform continuity requirement.

The system $\mathcal{A}^\omega$ consists of the axioms and rules over classical logic and the schema of quantifier-free choice

$$\forall x^\rho\,\exists y^\tau\,A_{qf}(x, y) \to \exists Y^{\rho\to\tau}\,\forall x^\rho\,A_{qf}(x, Y(x)),$$

where $\rho := \rho_1 \to (\rho_2 \to \ldots(\rho_{n-1} \to \rho_n))$, and the schema of dependent choice

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

3.1.1 Extension to normed spaces

The system $\mathcal{A}^\omega$ introduced in the previous section forms the basis of the metatheorems employed in proof mining. While Polish spaces, i.e. complete separable metric spaces, can be represented therein, it is necessary to extend the system if one wishes to talk about abstract structures like arbitrary normed spaces. Since the fixed point theory of real Banach spaces is the central topic of this thesis, we now outline how some classes of real Banach spaces can be incorporated in $\mathcal{A}^\omega$.

We start out by extending $\mathcal{A}^\omega$ to $\mathcal{A}^\omega[X, \|\cdot\|]$, which is the base system together with an abstract vector space. First, our type system T is extended to T^X by adding a second base sort X:

(T1′) $0 \in \mathbf{T}^X$, $X \in \mathbf{T}^X$ and

(T2′) $\rho \in \mathbf{T}^X$ and $\tau \in \mathbf{T}^X$ imply $\rho \to \tau \in \mathbf{T}^X$.

Next, we add function symbols for vector addition $+_X$, vector subtraction $-_X$, scalar multiplication $\cdot_X$ and the norm $\|\cdot\|_X$. Vector addition and subtraction are clearly of type $X \to (X \to X)$. For the latter two, observe that real numbers are encoded by objects of type $0 \to 0$ in the base system $\mathcal{A}^\omega$ (see [49] for details). Therefore, scalar multiplication is of type $(0 \to 0) \to (X \to X)$, and $\|\cdot\|_X$ is of type $X \to (0 \to 0)$.

When adding the defining axioms for the function symbols we just introduced, some caution is required. The details of this step (to be found in [49]) are of purely technical nature, so we do not include them here. It is, however, important to know that these axioms ensure that the abstract structure behaves in the expected way, and that all new functions symbols are provably extensional.

Completeness of the space can formalized by adding a completeness operator C, which maps Cauchy sequences in X, i.e. objects of type $0 \to X$, to their limit element of X. Thus, C is of type $(0 \to X) \to X$. The general idea on how to form the completion is identical to that of forming the completion of the rational numbers, i.e. the real numbers, in the base system $\mathcal{A}^\omega$. The details can again be found in [49] and are omitted here. In proof mining practice, however, the completeness of the space is rarely used since end-products are usually rates of metastability or asymptotic regularity. The latter is completely independent of completeness properties as it only claims $\|Tx_n - x_n\| \to 0$ for some operator $T : X \to X$ and some sequence $(x_n) \subset X$. While the former is, in fact, a quantitative version of a Cauchy statement, it is of the form

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Phi(g, \varepsilon) \, \forall i, j \in [n; n + g(n)] \; (\|x_i - x_j\| \leq \varepsilon) \, .$$

Therefore, it no longer claims the existence of the limit point itself, as it only implies Cauchyness in the absence of completeness. In fact, rates of metastability in practice almost always hold without completeness, and imply the original theorem only if one adds completeness.

Therefore, we do not officially include the completeness operator in the extension $\mathcal{A}^\omega[X, \|\cdot\|]$ of $\mathcal{A}^\omega$ to normed spaces.

3.1.2 Adding a convex subset C

Subsets of X are represented through characteristic functions. One also needs to add a new constant symbol that witnesses the nonemptyness of the subset. Additional properties of the set need to be axiomatized by purely universal axioms. For convex subsets C, for instance, one needs to add a convexity axiom for the characteristic function, which is purely universal, to obtain the system $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$. It should be noted that the newly introduced characteristic function is necessarily discontinuous, so it cannot be provably extensional in $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$. Hence, we only have the extensionality rule available for it.

3.2 The Main Theorem on Proof Mining in Normed Spaces

We are now in a position to state the main metatheorem on proof mining in normed spaces:

Theorem 3.2.1 (Main Theorem on Proof Mining in Normed Spaces, [49]). *Let P be an $\mathcal{A}^\omega$-definable Polish space and $B_\forall$ and $C_\exists$ be purely universal and purely existential formulas in the language of $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$ respectively. Suppose that $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$ proves a sentence*

$$\forall x \in P\, \forall z^C\, \forall f^{C\to C}\big(f \textit{ nonexpansive} \wedge \forall u^0\, B_\forall(x, y, z, f, u) \to \exists v^0 C_\exists(x, y, z, f, v)\big),$$

then one can extract from the proof a computable functional $\Phi : \mathbb{N}^\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that, for all representatives $r_x : \mathbb{N} \to \mathbb{N}$ of $x \in P$ and all bounds $b \in \mathbb{N}$,

$$\begin{aligned}\forall x \in P\, \forall z^C\, \forall f^{C\to C}\big(&f \textit{ nonexpansive} \wedge \|z\|_X \leq_\mathbb{R} n \wedge \|f(z) - z\|_x \leq_\mathbb{R} n \\ &\wedge \forall u^0 \leq \Phi(r_x, n)\, B_\forall(x, y, z, f, u) \to \exists v^0 \leq \Phi(r_x, b)\, C_\exists(x, y, z, f, v)\big)\end{aligned}$$

holds in all nonempty real normed linear spaces X and nonempty convex subsets C.

Remark 3.2.2. The subrecursive complexity of the bound extracted from the proof reflects the level of noneffectiveness of the proof. The bounds extracted in this thesis are all primitive recursive of lowest types.

As stated above, the metatheorem only applies to the case of a single nonexpansive mapping $f : C \to C$ and a single vector parameter $z \in C$ in a real normed space X with nonempty convex subset C. However, one can easily extend it to other scenarios, as we will see now.

For instance, one can enrich the base system $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$ even further by adding a function symbol J of type $X \to (X \to 1)$ along with the purely universal axioms expressing that J is the normalized duality mapping on X, cf. Definition 4.2.3. The metatheorem then extends to this scenario.

In fact, one can impose *any* further restrictions on $\mathcal{A}^\omega[X, \|\cdot\|, C]_{-b}$ as long as

(i) The additional axioms are purely universal, and

(ii) Any new function symbols used in the axiomatization are majorizable in the sense of '$\gtrsim$'.

The notion $\gtrsim$ is an extension of Bezem's strong majorization relation to the abstract type X due to Kohlenbach (see, for instance, [49]). The details of the relation are once more beyond the scope of our modest overview. For the types X, $X \to X$ or $0 \to X$ (or C, $C \to C$ or $0 \to C$) however, it reduces to the following simple clauses:

Definition 3.2.3. (i) An object n^0 of ground type 0, i.e. a natural number, majorizes a vector $x \in X$ (or $x \in C$) if it is a norm bound: $\|x\| \leq_{\mathbb{R}} n$.

(ii) An object $M^{0\to 0}$ of type $0 \to 0$, i.e. a number-theoretic function, majorizes a function $f : X \to X$ (or $C \to C$) if it is nondecreasing and it transforms a norm bound on the input into a norm bound on the output: $\|x\| \leq_{\mathbb{R}} n \to \|f(x)\| \leq_{\mathbb{R}} M(n)$.

(iii) An object $M^{0\to 0}$ of type $0 \to 0$, i.e. a number-theoretic function, majorizes a function $f : 0 \to X$ (or $0 \to C$) if it is nondecreasing and $\|f(n)\| \leq_{\mathbb{R}} M(n)$ for all n^0. In other words, a nondecreasing $M^{0\to 0}$ majorizes a sequence in X or C if it produces for each n^0 a norm bound $M(n)$ on the n-th element of the sequence.

For the majority of this thesis, these three clauses suffice. It is only in Chapter 8 that majorization of higher types becomes relevant. We then extend the above clauses to the cases required in that chapter. Apart from real inner product spaces, one can for instance treat uniformly smooth and uniformly convex spaces, notions which are discussed in Chapter 4. The extracted rates and bounds will then additionally depend on the respective moduli.

4 Analytical Preliminaries

4.1 Geometry of Banach Spaces

By virtue of its inner product $\langle \cdot, \cdot \rangle$, any Hilbert space exhibits rich geometrical structure. For example, to name but a few, the parallelogram and polarization identities hold; there is a canonical notion of orthogonality; projections of points onto linear subspaces exist and are unique; the norm induced by the inner product is Fréchet differentiable; and its unit sphere does not contain any line segments. It is interesting to note that, except for the parallelogram and polarization identities, the aforementioned properties do not characterize Hilbert space. In fact, though they fail to be Hilbert spaces, many of the classical Banach spaces like the L_p and l_p spaces, $1 < p < \infty$, $p \neq 2$, share some of those rich geometric properties. Therefore, considerable effort has gone into studying Banach spaces that satisfy analogues of some of the geometric properties that Hilbert spaces possess.

For fixed-point theory, in particular, the extensive study of the geometry of certain classes of Banach spaces has proven to be especially fruitful. Numerous techniques hinging on geometrical considerations employed in the Hilbert space setting generalize to large classes of Banach spaces. Perhaps the most prominent one is the class of *uniformly convex* Banach spaces introduced by Clarkson in 1936, which we examine in the next subsection.

4.1.1 Uniformly Convex Banach Spaces

Definition 4.1.1 (Clarkson [20])**.** A Banach space X is called *uniformly convex* if for all $\varepsilon \in (0, 2]$, there exists a $\delta \in (0, 1]$ such that for all $x, y \in X$

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

The geometric interpretation of uniform convexity is rather simple: A Banach space is uniformly convex if the midpoint of a variable chord of the unit ball of the space cannot approach the surface of the ball unless the length of the cord approaches zero (cf. Clarkson [20]). Clarkson also introduced the slightly weaker notion of *strict convexity* of Banach spaces, which states that the unit sphere of the space does not contain a line segment:

Definition 4.1.2 (Clarkson [20])**.** A Banach space X is called *strictly convex* if for all $x, y \in X$,

$$\|x\| \leq 1,\ \|y\| \leq 1 \text{ and } x \neq y \text{ imply } \left\| \frac{x + y}{2} \right\| < 1.$$

From the proof-mining perspective, the former notion is very well-behaved. Uniform convexity may be formalized as

$$\forall \varepsilon \in (0,2]\, \exists \delta \in (0,1]\, \forall x,y \in X \left(\|x\| \le 1 \wedge \|y\| \le 1 \wedge \|x-y\| > \varepsilon \longrightarrow \left\| \frac{x+y}{2} \right\| \le 1-\delta \right).$$

If we introduce a Skolem function $\eta : (0,2] \to (0,1]$ for "$\exists\delta$" above, then the strict convexity of X can be added as a purely universal axiom to the formal system $\mathcal{A}^\omega[X, \|\cdot\|, \eta]$ arising from $\mathcal{A}^\omega[X, \|\cdot\|]$ by adding a new function symbol η:

$$\forall \varepsilon \in (0,2]\, \forall x,y \in B \left(\|x\| \le 1 \wedge \|y\| \le 1 \wedge \|x-y\| > \varepsilon \longrightarrow \left\| \frac{x+y}{2} \right\| \le 1-\eta(\varepsilon) \right). \tag{4.1}$$

The Skolem function $\eta : (0,2] \to (0,1]$ will be called a *modulus of convexity* for X. Moreover, the main metatheorem, i.e. Theorem 3.2.1, extends to the system $\mathcal{A}^\omega[X, \|\cdot\|, \eta]$.

Proposition 4.1.3. *(i) Any Hilbert space is uniformly convex with modulus*

$$\eta_H(\varepsilon) := 1 - \sqrt{1 - \frac{\varepsilon^2}{4}}$$

(ii) The L^P-spaces, $1 < p < \infty$ are uniformly convex with modulus

$$\eta_{L^p}(\varepsilon) := \begin{cases} \frac{p-1}{8} \cdot \varepsilon^2 & \text{for } 1 < p < 2, \\ 1 - \left(1 - \left(\frac{\varepsilon}{2}\right)^p\right)^{\frac{1}{p}}, & \text{for } 2 \le p < \infty. \end{cases}$$

(iii) L_1 and L_∞ fail to be uniformly convex.

It is interesting to note that in the literature, there already exists an alternate notion of "modulus of convexity". Although it essentially captures the same concept, it is somewhat unnatural; Given *any* Banach space X, the alternate modulus is the function $\delta_X : [0,2] \to [0,1]$ defined by

$$\delta_X(\varepsilon) := \inf \left\{ 1 - \left\| \frac{x+y}{2} \right\| \,:\, \|x\| \le 1,\ \|y\| \le 1,\ \|x-y\| \ge \varepsilon \right\}. \tag{4.2}$$

One should remark that any Banach space has a modulus of convexity in the sense of (4.2), even if it is not strictly convex. A Banach space X is then uniformly convex if and only if $\delta_X(\varepsilon) > 0$ for any $\varepsilon > 0$, and in this case δ_X is *the best* modulus of convexity in our sense, i.e. (4.1). However, for certain concrete cases it might not be trivial to compute explicitly in ε the best modulus δ_X, whereas finding any modulus of convexity η could be easier. For example, in the case of the L_p-spaces for $2 \le p < \infty$, it is known that the modulus η_{L^p} stated in 4.1.3 is optimal, so $\eta_{L^p} = \delta_{L^P}$, while this is not the case for $1 < p < 2$. In fact, for $1 < p < 2$, one only has the implicit form (see [39])

$$\left|1 - \delta_{L^p}(\varepsilon) + \frac{\varepsilon}{2}\right|^p + \left|1 - \delta_{L^p}(\varepsilon) - \frac{\varepsilon}{2}\right|^p = 2$$

for δ_{L^p}.

It is also interesting to examine how the notion of strict convexity behaves with respect to monotone functional interpretation. To this end, observe that strict convexity is trivially equivalent to the formula

$$\forall x, y \in B\, \forall \varepsilon \in (0,2]\, \exists \delta \in (0,1] \left(\|x-y\| > \varepsilon \longrightarrow \left\| \frac{x+y}{2} \right\| \le 1 - \delta \right). \tag{4.3}$$

Now, suppose that we add an axiom Δ to the theory $\mathcal{A}^\omega[X, \|\cdot\|]$ such that we could prove strict convexity of X. Then, monotone functional interpretation would allow us to extract from the proof a positive lower bound on "$\exists\delta$" which is uniform in x and y, and hence, by monotonicity, a uniform realizer $\eta : (0,2] \to (0,1]$. But any realizer of "$\exists\delta$" in (4.3) that is uniform in x and y is already a modulus of convexity for X. Consequently, X is then even uniformly convex.

As we have just seen, it is impossible to axiomatize strict convexity of Banach spaces in our theory $\mathcal{A}^\omega[X, \|\cdot\|]$. Therefore, when analyzing from a proof mining perspective theorems on strictly convex Banach spaces, it is necessary to view the theorem in the slightly less general setting of uniformly convex Banach spaces, and then carry out the analysis.

To illustrate the preceding considerations, consider the following well-known fact.

Proposition 4.1.4. *Let C be a convex subset of a Banach space X and let $T : C \to C$ be nonexpansive. If the norm of X is strictly convex, the fixed point set is convex.*

Suppose that we encountered this statement as a lemma in some proof that is the subject of a proof-theoretic analysis. What is the quantitative content we need to extract? To answer this question, let us write the statement as a logical formula. For simplicity, we restrict ourselves to the midpoint of two fixed points rather than any convex combination.

$$\forall x, y \in C \left(\forall \delta > 0 \left(\|Tx - x\| \le \delta \wedge \|Ty - y\| \le \delta \right) \longrightarrow \forall \varepsilon > 0 \left\| T\Big(\frac{x+y}{2}\Big) - \frac{x+y}{2} \right\| \le \varepsilon \right).$$

The functional interpretation of this statement (which is equivalent to the monotone functional interpretation due to the monotonicity of the premise) is given by

$$\exists \delta : (0,\infty) \to (0,\infty)\, \forall x, y \in C\, \forall \varepsilon > 0$$
$$\|Tx - x\| \le \delta(\varepsilon) \wedge \|Ty - y\| \le \delta(\varepsilon) \longrightarrow \left\| T\Big(\frac{x+y}{2}\Big) - \frac{x+y}{2} \right\| \le \varepsilon.$$

From the standard proof of Proposition 4.1.4, one can extract the function $\delta : (0,\infty) \to (0,\infty)$ above. In fact, as we now show, one can take $\delta(\varepsilon) := \varepsilon \cdot \eta(\varepsilon)/2$.

Proposition 4.1.5. *Let X be a uniformly convex Banach space with modulus of convexity η. Suppose $T : C \to C$ is nonexpansive, where $C \subseteq X$ is convex with* $\operatorname{diam} C \le 1$*. Then, for all $\varepsilon > 0$ and all $x, y \in C$,*

$$\|Tx - x\| \le \frac{\varepsilon \cdot \eta(\varepsilon)}{2} \quad \textit{and} \quad \|Ty - y\| \le \frac{\varepsilon \cdot \eta(\varepsilon)}{2} \quad \textit{imply} \quad \left\| T\Big(\frac{x+y}{2}\Big) - \frac{x+y}{2} \right\| < \varepsilon.$$

Proof. Let z be the midpoint of x and y. Then, for $\delta := \max\{\|Tx - x\|, \|Ty - y\|\}$,

$$\begin{aligned}\|x - Tz\| + \|z - y\| &\leq \|x - Tx\| + \|Tx - Tz\| + \|z - y\| \\ &\leq \|x - z\| + \|z - y\| + \delta \\ &= \|x - y\| + \delta.\end{aligned}$$

and so $\|x - y\| \geq \|x - Tz\| + \|z - y\| - \delta$. Moreover,

$$\begin{aligned}\|x - Tz\| + \|Tz - y\| &\leq \|x - Tx\| + \|Tx - Tz\| + \|y - Ty\| + \|Ty - Tz\| \\ &\leq \|x - z\| + \|z - y\| + 2\delta \\ &= \|x - y\| + 2\delta,\end{aligned}$$

and so $\|x - y\| \geq \|x - Tz\| + \|Tz - y\| - 2\delta$. Similarly,

$$\begin{aligned}\|x - z\| + \|Tz - y\| &\leq \|x - z\| + \|y - Ty\| + \|Ty - Tz\| \\ &\leq \|x - z\| + \|z - y\| + \delta \\ &= \|x - y\| + \delta,\end{aligned}$$

and so $\|x - y\| \geq \|x - z\| + \|Tz - y\| - \delta$. To summarize, we have the following relations:

$$\begin{cases} \|x - y\| = \|x - z\| + \|z - y\|, \\ \|x - y\| \geq \|x - Tz\| + \|z - y\| - \delta, \\ \|x - y\| \geq \|x - Tz\| + \|Tz - y\| - 2\delta, \\ \|x - y\| \geq \|x - z\| + \|Tz - y\| - \delta. \end{cases} \tag{4.4}$$

Now suppose $\|Tz - z\| > \varepsilon$. Then, by the uniform convexity of X,

$$\begin{aligned}\left\|\frac{x - Tz}{2} + \frac{x - z}{2}\right\| &< \max\{\|x - Tz\|, \|x - z\|\} - \frac{\varepsilon}{2} \cdot \eta(\varepsilon), \text{ and} \\ \left\|\frac{y - Tz}{2} + \frac{y - z}{2}\right\| &< \max\{\|y - Tz\|, \|y - z\|\} - \frac{\varepsilon}{2} \cdot \eta(\varepsilon).\end{aligned}$$

But then

$$\begin{aligned}&\max\{\|x - Tz\|, \|x - z\|\} + \max\{\|y - Tz\|, \|y - z\|\} - \varepsilon \cdot \eta(\varepsilon) \\ &\qquad > \left\|\frac{x - Tz}{2} + \frac{x - z}{2}\right\| + \left\|\frac{y - Tz}{2} + \frac{y - z}{2}\right\| \\ &\qquad = \left\|x - \frac{Tz - z}{2}\right\| + \left\|y - \frac{Tz - z}{2}\right\| \\ &\qquad \geq \|x - y\| \\ &\qquad \overset{(4.4)}{\geq} \max\{\|x - Tz\|, \|x - z\|\} + \max\{\|y - Tz\|, \|y - z\|\} - 2\delta,\end{aligned}$$

which implies $\varepsilon \cdot \eta(\varepsilon) < 2\delta = 2 \cdot \max\{\|Tx - x\|, \|Ty - y\|\}$, contradicting the hypothesis. □

4.1.2 Uniformly Smooth Banach Spaces

Another fundamental geometric property in the context of Banach spaces is *smoothness*, and, closely related to it, differentiability of the norm.

Definition 4.1.6 (see e.g. [44, 94])**.** Let X be a normed space and B be the unit ball. If the limit

$$\lim_{t\to 0} \frac{\|x+ty\| - \|x\|}{t}$$

(i) exists for each x, y on the unit sphere, then the norm is *Gâteaux differentiable* and X is called *smooth*;

(ii) is attained uniformly in x, then the norm is called *uniformly Gâteax differentiable*;

(iii) is attained uniformly in y, then the norm is called *Fréchet differentiable*;

(iv) is attained uniformly in x and y, then the norm is called *uniformly Fréchet differentiable* and X is called *uniformly smooth.*

Observe that the norm of a linear space X is Gâteaux differentiable if and only if there exists for each $x \in B$ a unique supporting hyperplane to the unit ball at x. The notion of uniform smoothness is frequently defined equivalently as follows:

Theorem 4.1.7. *A space is uniformly smooth if and only if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in X$ with $\|x\| = 1$ and $\|y\| \leq \delta$,*

$$\|x+y\| + \|x-y\| \leq 2\|x\| + \varepsilon\|y\|. \tag{4.5}$$

As in the case of the uniformly convex spaces, given a uniformly smooth space X, we say that a function $\tau_X : (0,\infty) \to (0,\infty)$ providing '$\exists\delta$' above is called a *modulus of smoothness* for X. This, again, should not be confused with the notion of 'modulus of smoothness' $\rho_X : [0,\infty) \to [0,\infty)$ sometimes found in the literature, which is defined for any normed space X as the function

$$\rho_X(t) := \sup\left\{\frac{\|x+y\| + \|x-y\|}{2} - 1 : \|x\| = 1,\ \|y\| = t\right\} \tag{4.6}$$

As before, this alternate modulus is defined for all normed spaces X, and the space is uniformly smooth if and only if $\lim_{t\to 0^+} \rho_X(t)/t \to 0$.

Many of the classical Banach spaces share smoothness properties, including the L^p spaces for $1 < p < \infty$. Moreover, (uniform) smoothness and (uniform) convexity are dual notions, as we discuss in the subsequent section.

As with uniform convexity, the metatheorem 3.2.1 can be extended to uniformly smooth Banach spaces by adding to the system $\mathcal{A}^\omega[X, \|\cdot\|]$ a new function symbol τ to the language and formulating an axiom expressing that τ is a modulus of uniform smoothness. We should point out that in this case, some caution is required with the formulation of that axiom, see Kohlenbach and Leuştean [58].

4.2 Duality

Given a normed space X, we denote its dual by X^*. Apart from the strong topology on X, the weak topology plays a central role in fixed point theory, which we recall here for convenience.

Definition 4.2.1. For each $\varepsilon > 0$, we define the set

$$U(x_0; x^*, \varepsilon) = \{x \in X : |\langle x - x_0, x^* \rangle| < \varepsilon\}.$$

The *weak topology* on X is the topology generated by the class of all sets which are expressible in the form $U(x_0, x^*, \varepsilon)$.

The weak topology is the coarsest topology under which all $x^* \in X^*$ are continuous. Since a significant portion of fixed point theory is concerned with iteration algorithms, it is noteworthy to observe that a sequence (x_n) in X converges (strongly) to $x \in X$ if $\|x_n - x\| \to 0$. In this case, we write $x_n \to x$. Likewise, (x_n) converges weakly to $x \in X$ if

$$\lim_{n\to\infty} \langle x_n - x, x^* \rangle = 0, \quad \text{for all } x^* \in X,$$

and we write $x_n \rightharpoonup x$ or, alternatively, $\text{w-lim}\, x_n = x$. Moreover, strong convergence implies weak convergence since

$$\langle x_n - x, x^* \rangle \leq \|x_n - x\| \cdot \|x^*\|.$$

Another basic observation is that the norm is weakly lower semicontinuous:

Lemma 4.2.2. *The norm on a Banach space is weak-lower semicontinuous, i.e. $x_n \rightharpoonup x$ implies $\|x\| \leq \liminf_{n\to\infty} \|x_n\|$ for all sequences (x_n) in X.*

A central tool in fixed point theory is the Banach-Alaoglu Theorem, which states that the unit ball of any normed space is compact with respect to the weak topology:

Banach-Alaoglu Theorem. *The unit ball of a reflexive Banach space is weakly compact.*

The compactness of the unit ball in the weak topology is frequently used to prove first the weak convergence of an iteration (x_n), and the demiclosedness principle then implies that (x_n) is even strongly convergent given it is asymptotically regular, i.e. given that $\|Tx_n - x_n\|$ is a null sequence.

4.2.1 The Duality Mapping

We now introduce another important tool called the normalized duality mapping, which we usually refer to as simply 'the duality mapping'.

Definition 4.2.3. Let X be a real Banach space.

(i) The normalized duality mapping $J : X \to \mathcal{P}(X)$ is a multivalued mapping defined by

$$Jx := \left\{x^* \in X^* : \langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2\right\}$$

(ii) Any mapping $j : X \to X$ such that $jx \in Jx$ for all $x \in X$ is called a duality selection mapping of J.

It is a well-known consequence of the Hahn-Banach theorem that Jx is nonempty for all $x \in X$. We now gather some properties of the normalized duality mapping.

Lemma 4.2.4. *Let X be a real Banach space and let X^* be its dual. Then*

(i) The mapping J is monotone, that is

$$\langle x - y, jx - jy \rangle \geq 0, \quad \textit{for all } x, y \in X,\ jx \in Jx \textit{ and } jy \in Jy.$$

(ii) J is homogeneous, i.e. $J(\lambda x) = \lambda Jx$ for all $x \in X$ and $\lambda \in \mathbb{R}$.

(iii) For all $x, y \in X$ and $j(x + y) \in J(x + y)$,

$$\|x + y\|^2 \leq \|x\|^2 + 2 \langle y, j(x + y) \rangle .$$

Lemma 4.2.5 (Kato [43], Giles [32])**.** *Let $x, y \in X$. The following are equivalent:*

(i) $\|x\| \leq \|x + \lambda y\|$ for all $\lambda \geq 0$;

(ii) there exists $jx \in Jx$ with $\langle y, jx \rangle \geq 0$.

Observe that, by definition, any duality selection mapping is, by definition linear in the first argument and homogeneous in the second. Moreover, it satisfies a polarization *in*equality instead of a polarization equality; In Hilbert spaces, the duality mapping is single-valued, with $J(x) = \{x\}$ for all x, so Lemma 4.2.4(iii) holds in Hilbert spaces with equality and is simply the well-known polarization identity. This inequality is a classical result of Petryshyn [83].

Therefore, given a duality selection mapping j, the mapping $\langle \cdot\, , j(\cdot) \rangle$ plays the same role in Banach spaces as the inner product does in Hilbert spaces. This is further illustrated by Lemma 4.2.5, since condition (i) states that x is acute to y, and condition (ii) expresses this fact in terms of the duality mapping and hence, in Hilbert space, the inner product.

4.2.2 Duality of Smoothness and Convexity

We round off the glimpse at the geometry of Banach spaces by connecting the concepts of smoothness, convexity and the duality mapping.

Theorem 4.2.6. *Let X be a Banach space.*

(i) If X^* is uniformly smooth, then X is uniformly convex.

(ii) If X^* is uniformly convex, then X is uniformly smooth.

It is a well-known fact that, if X is uniformly smooth or uniformly convex, then X is reflexive, so we obtain the following important corollary.

Corollary 4.2.7. *Let X be a Banach space. Then X is uniformly smooth if and only if X^* is uniformly convex.*

As a result of Kohlenbach and Leuştean [58], a modulus of smoothness for a uniformly smooth Banach space can be transformed into a modulus of convexity for its dual X^*; If X is uniformly smooth with modulus τ, then X^* is uniformly convex with modulus $\eta(\varepsilon) := \frac{\varepsilon}{4} \cdot \tau\left(\frac{\varepsilon}{2}\right)$.

This identity highlights a further advantage of the notion of smoothness and convexity moduli used in this thesis. For the alternative versions δ and ρ defined in (4.2) and (4.6) respectively, the Lindenstrauss duality formula only gives the implicit relationship

$$\rho_X(\delta) := \sup\left\{\frac{\delta\varepsilon}{2} - \delta_{X^*}(\varepsilon) : 0 \leq \varepsilon \leq 2\right\}.$$

It is also well-known that the duality mapping on a Banach space X is single-valued whenever X is smooth, and hence also in uniformly convex or uniformly smooth spaces. In the latter case, J is not only single valued, but its unique duality selection mapping (which we also denote by J) is norm-to-norm uniformly continuous on bounded subsets of X, which is formalized by

$$\forall d > 0\, \forall \varepsilon > 0\, \exists \delta > 0\, \forall x, y \in B_d(0)\ (\|x - y\| \leq \delta \to \|Jx - Jy\| \leq \varepsilon)\,.$$

A *modulus of continuity* $\omega : (0,\infty) \times (0,\infty) \to (0,\infty)$ is a function witnessing '$\exists\delta$' uniformly in x and y. As a first instance of our metatheorems, we are guaranteed[1] the existence such a modulus ω in the form of a term in the formal system $\mathcal{A}^\omega[X, \|\cdot\|, \tau]$. The extraction of a modulus of continuity from a modulus of smoothness has been carried out by Kohlenbach and Leusţean in [58].

Theorem 4.2.8 (Kohlenbach and Leuştean [58])**.** *(i) If X is uniformly smooth with modulus τ, then X^* is uniformly convex with modulus $\eta(\varepsilon) := \frac{\varepsilon}{4} \cdot \tau\left(\frac{\varepsilon}{2}\right)$.*

(ii) If X^* is uniformly convex with modulus η, then J is uniformly continuous with modulus

$$\omega(d, \varepsilon) := \begin{cases} \dfrac{\varepsilon}{3} \cdot \eta\left(\dfrac{\varepsilon}{d}\right) & \text{if } \varepsilon \leq 2 \text{ and } d \geq 1, \\[2ex] \dfrac{\varepsilon}{3} \cdot \eta(\varepsilon) & \text{if } \varepsilon \leq 2 \text{ and } d < 1, \\[2ex] \dfrac{2}{3} \cdot \eta\left(\dfrac{2}{d}\right) & \text{if } \varepsilon > 2 \text{ and } d \geq 1, \\[2ex] \dfrac{2}{3} \cdot \eta(2) & \text{if } \varepsilon > 2 \text{ and } d < 1. \end{cases}$$

[1] after verifying that the proof of "uniformly smooth" implies "J is uniformly continuous on bounded subsets" is formalizable in $\mathcal{A}^\omega[X, \|\cdot\|, \tau]$

In practice, uniform smoothness is often required of the underlying space X with the sole purpose to guarantee the uniform smoothness of the duality mapping. Therefore, as remarked by Kohlenbach and Leuştean [58], one can alternatively assume that X is a *space with a uniformly continuous duality selection mapping*, which is a pair (X, J) of a real Banach space X together with a mapping $J : X \to X^*$ satisfying

(i) $\langle x, Jx\rangle = \|x\|^2 = \|Jx\|^2$ for all x in X, and

(ii) J is norm-to-norm uniformly continuous on bounded subsets of X.

It is clear that any uniformly smooth real Banach space X can be endowed with a mapping J that turns (X, J) into a space with a uniformly continuous duality selection mapping. The converse direction, however, is left open in [58]. We will answer the question in the affirmative later on.

In view of this, one can alternatively axiomatize uniformly smooth Banach spaces as Banach spaces with a uniformly continuous duality selection mapping. Instead of the modulus of smoothness and its corresponding axiom, one can therefore add two functions $J : C \times C \to \mathbb{R}$ and $\omega : (0,\infty) \times (0,\infty) \to (0,\infty)$ to the language. One then needs two new axioms, one to formalize that J is a duality selection mapping and another one to express that ω is a modulus of uniform continuity:

Definition 4.2.9 (Kohlenbach and Leuştean [58])**.** Let X be a Banach space with a uniformly continuous duality selection mapping J. A map $\omega : (0,\infty) \times (0,\infty) \to (0,\infty)$ is called a modulus of uniform continuity for J if for all $M, \varepsilon > 0$,

$$\|x\|, \|y\| \leq M \text{ and } \|x - y\| < \omega(M, \varepsilon) \text{ implies } \|J(x) - J(y)\| < \varepsilon.$$

Moreover, James [41] proved the following representation theorem for smooth and uniformly convex Banach spaces using the duality mapping. Giles [32] later proved the same result for continuous semi-inner-product spaces which are uniformly convex and complete in their norm, a notion that turns out to be equivalent. However, the latter proof is more accessible from a proof-mining perspective.

Theorem 4.2.10 (James [41], Giles [32])**.** *Let X be a smooth and uniformly convex Banach space. Then, to every x^* in X^* there exists a y in X such that*

$$\langle x, x^*\rangle = \langle x, J(y)\rangle, \quad \textit{for all } x \in X.$$

5 Halpern's Iteration for Nonexpansive Mappings

5.1 Introduction

In order to motivate the Halpern iteration, we recall von Neumann's famous Mean Ergodic Theorem.

Theorem 5.1.1 (von Neumann [99], Birkhoff [8]). *Let X be a uniformly convex Banach space and $S : X \to X$ be a linear operator that satisfies $\|Sx\| \leq \|x\|$ for all $x \in X$. Then, the Césaro means*

$$x_{n+1} := \frac{1}{n+1} \cdot \sum_{k=0}^{n} S^{(k)}(x_0).$$

converge strongly for any starting point x_0.

For nonlinear mappings S, the Césaro means only converge weakly, while strong convergence is in general false. Therefore, several variants thereof that coincide for linear mappings S have been considered over the years, perhaps most notably the Halpern iteration:

Definition 5.1.2 (Halpern [38]). Let X be a normed linear space, $C \subseteq X$ a closed and convex subset and $S : C \to C$ be a self-mapping. For starting point $x_0 \in C$, anchor $u \in C$ and control sequence $(\alpha_n) \subset [0,1]$, the Halpern iteration (x_n) is defined by

$$x_{n+1} := \alpha_n u + (1 - \alpha_n) S x_n. \tag{5.1}$$

Observe that, if S is linear, $u = x_0$ and $\alpha_n = 1/(n+2)$, this iteration is identical to the sequence of Césaro means.

For nonexpansive mappings $S : C \to C$ for bounded subsets $C \subset H$ of a Hilbert space H, Halpern [38] gave a set of necessary and a different set of sufficient conditions for (α_n) under which the scheme (5.1) converges strongly to a fixed point of S. However, these conditions allowed no conclusion whether his iteration converges for the choice $\alpha_n = 1/(n+2)$ that turns (5.1) into the Césaro mean for linear S.

Wittmann [100] answered this question in the affirmative in 1992: If S is assumed to have a fixed point (or, alternatively, C is bounded) and the sequence (α_n) satisfies

The results of this Chapter have been published in [69].

(i) $\lim_{n\to\infty} \alpha_n = 0$, (ii) $\sum_{n=0}^{\infty} \alpha_n = \infty$, (iii) $\sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < \infty$,

then the Halpern iteration converges strongly to the fixed point closest to the starting point x_0. This result presents a vast generalization of the Mean Ergodic Theorem for Hilbert spaces.

Using the proof-theoretic metatheorems of Kohlenbach [49, 58] that we briefly introduced in Chapter 3, Leuştean [73] extracted from Wittmann's proof a rate of asymptotic regularity for general normed and even hyperbolic spaces, i.e a rate of convergence for $\|Sx_n - x_n\| \to 0$ under the assumption that the Halpern iteration remains bounded, which is always the case if S has a fixed point. The rate is highly uniform in the sense that it does not depend on the set C, the operator S or the specific choice of the sequence (α_n), but only on witnesses for the existential quantifiers in conditions (i) to (iii) above and a bound on, in essence, the sequence (x_n).

Strong convergence is then established by Wittmann using the metric projection of x_0 onto the fixed point set and weak sequential compactness applied to the iteration sequence. As shown by Avigad, Gerhary and Towsner in [2], there cannot be a computable bound on the rate of convergence even for the special case where $\alpha_n = 1/(n+2)$ and S is linear. In this case, the Halpern iteration coincides with the ergodic average, and so Wittmann's theorem implies von Neumann's mean ergodic theorem.

More recently, Neumann [82] even showed that there are (computable) nonexpansive mappings f on the Hilbert cube, i.e. the space of sequences $(x_n) \in \ell_2$ with $|x_n| \leq 1$ for all n, that have no computable fixed points, and so no computable sequence approximating any fixed point of f can have a computable rate of convergence. Therefore, a uniform rate of convergence is ruled out not only for the Halpern iteration, but for *any* computable iteration approximating fixed points of *any* class of mappings that include nonexpansions on Hilbert space.

On the other hand, a uniform rate of metastability as introduced in Section 2.3 is guaranteed to exist by the metatheorems and was extracted by Kohlenbach in [52]:

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Phi(\varepsilon, g) \forall i, j \in [n; n + g(n)] \left(\|x_i - x_j\| < \varepsilon \right), \tag{5.2}$$

where $[n; n + g(n)] := \{n, n+1, \ldots, n+g(n)\}$. The bound is highly uniform in the sense that it does not depend on the operator S, the starting point x_0, the anchor u or the specific Hilbert space. Apart from rates of convergence and divergence for the conditions (i) to (iii), it only depends on an upper bound on the distance of the starting point from some fixed point of S. In the case $\alpha_n = 1/(n+2)$, Kohlenbach [52] also improved the exponential rate of asymptotic regularity to a quadratic one. Moreover, these results were also generalized to CAT(0) spaces [57] and CAT(κ) spaces [74].

Closely related to Wittmann's result is the following Theorem.

Theorem 5.1.3 (Browder [11])**.** *Let H be a Hilbert space, S a nonexpansive mapping of H into H. Suppose that there exists a bounded closed convex subset C of H mapped by S into itself. Let u be an arbitrary point of C, and for each t with $0 < t < 1$, let $S_t x = tu + (1-t)Sx$.*

Then S_t is a strict contraction of H with ratio t, S_t has a unique fixed point z_t in C, and (z_t) converges as $t \to 0$ strongly in H to a fixed point v of S in C. The fixed point v is uniquely specified as the fixed point of S closest to u.

The proof is structured similarly to the proof of Wittmann's theorem in that its ineffective part consists of a projection onto the fixed point set and weak sequential compactness, this time applied to (z_t). In fact, the proof theoretic analysis of Browder's theorem, also carried out in [52], exhibits interesting parallels to the aforementioned one.

There is also an elementary proof due to Halpern [38] for the special case where C is the closed unit ball of H, which can easily be generalized to arbitrary bounded closed convex subsets. The non-effectivity of Halpern's proof stems from the monotone convergence principle, i.e. that every monotone sequence in the real unit interval converges. A metastable version of this can be found on page 30 of [49]. Using this, a simpler rate of metastability was extracted in [52].

Reich generalized Browder's Theorem to uniformly smooth Banach spaces.

Theorem 5.1.4 (Reich [85])**.** *Let C be a closed convex subset of a uniformly smooth Banach space X and let $S : C \to C$ be a nonexpansive mapping with a fixed point. For $u \in C$ and $t \in (0, 1)$, let (z_t) be defined by the equation $z_t := tu + (1 - t)\, Sz_t$. Then (z_t) converges strongly to a fixed point of S as $t \to 0$.*

Recall from the analytical preliminaries of Chapter 4 that a Banach space X is said to be smooth if the limit

$$\lim_{t\to 0} \frac{\|x + ty\| - \|x\|}{t}$$

exists for all x, y in the unit sphere. If the limit is attained uniformly in x, then X is said to have a uniformly Gâteaux differentiable norm. If the limit is attained uniformly in x, y, then X is called uniformly smooth. In this case, the normalized duality map $J : X \to X^*$, defined as

$$J(x) = \left\{x^* \in X^* : \langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2\right\}.$$

is single-valued and norm-to-norm uniformly continuous on bounded subsets of X. Using Reich's theorem, Shioji and Takahashi [88] generalized Wittmann's result to uniformly smooth Banach spaces. The proof is highly noneffective due to the use of Banach limits, whose existence thus far has only been established making substantial reference to the Axiom of Choice. This difficulty was overcome by Kohlenbach and Leuştean [58] by the observation that the Banach limits in the proof could be replaced by Cesàro means, which, in turn, are covered by the aforementioned methods.

Xu proved the following variant, which uses neither weak compactness (as in Wittmann's proof) nor Banach limits (as in Shioji and Takahashi's proof).

Theorem 5.1.5 (Xu [102])**.** *Let X be a uniformly smooth Banach space, C be a closed convex subset of X, and $S : C \to C$ be a nonexpansive mapping with a fixed point. Let $u, x_0 \in C$ be given, Assume that $(\alpha_n) \subset [0, 1]$ satisfies the control conditions (i), (ii) and*

(iii)' $\lim_{n\to\infty}(\alpha_n - \alpha_{n-1})/\alpha_n = 0$

Then the sequence (x_n) *defined by*

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) S x_n$$

converges strongly to a fixed point of S.

Observe that Xu [102] showed that Xu's condition (iii)' does not imply (iii), while Remark 2.3.3 of Schade [86] shows that the converse is also not true in general. However, they both cover the most important case $1/(n+2)$.

We extract two quantitative versions of Theorem 5.1.5, namely an explicit and highly uniform rate of convergence for $\|x_n - Sx_n\| \to 0$ for general Banach spaces X, and a rate of metastability Φ relative to a rate of metastability of the resolvent (z_t) in the uniformly smooth case (cf. Theorem 5.1.4).

For Hilbert spaces, a rate of metastability for the resolvent is known, so we get in this case a concrete rate of metastability for the Halpern iteration, which is then of the form

$$(L(\underline{a}) \circ \tilde{g})^{(B(\underline{a}))}(0),$$

where $\tilde{g}(n) := \max\{g(i) : i \leq n\} + n$, and g is the counterfunction in equation (5.2). The functions B and L depend only on the tuple $\underline{a}$ parameterizing the sequences, the points u, x_0 and the sequence (α_n), but *not* on the counterfunction g. This was guaranteed a priori by a general metatheorem of Kohlenbach and Safarik [61], and stems from the fact that the proof only uses a limited amount of the law-of-excluded-middle.

It is still an open problem to extract a rate of metastability for (z_t) in uniformly smooth and uniformly convex Banach spaces, which would include the L_p-spaces (with $1 < p < \infty$, $p \neq 2$); until now, this has only been done in Hilbert space [52] so far. We tackle this problem in Chapter 9, but can so far only give a partial solution. Moreover, in Chapter 6 we provide a rate of metastability for the resolvent (z_t) for the much broader class of pseudocontractions, however still only in Hilbert space.

5.2 Technical Lemmas

Lemma 5.2.1 (Xu [102]). *Let* (s_n) *be a sequence of nonnegative real numbers satisfying for all nonnegative integers* n

$$s_{n+1} \leq (1 - \alpha_n) s_n + \alpha_n \beta_n + \gamma_n,$$

where $(\alpha_n) \subset [0,1]$ *is divergent in sum,* $(\beta_n) \subset \mathbb{R}$ *satisfies* $\limsup_{n\to\infty} \beta_n \leq 0$ *and* $(\gamma_n) \subset [0, \infty)$ *is convergent in sum. Then* (s_n) *is a null sequence.*

The following two lemmas can both be seen as quantitative versions of this lemma.

Lemma 5.2.2. *Let* (s_n) *be a sequence of real numbers satisfying*

$$s_{n+1} \leq (1 - \alpha_n) s_n + \alpha_n \beta_n + \gamma_n,$$

where $(\alpha_n) \subset [0,1]$, $(\beta_n) \subset \mathbb{R}$ and $(\gamma_n) \subset [0,\infty)$ are real sequences and $C > 0$ is such that $\beta_n \leq C$ for all n from some m on. Then

$$s_{n+1} \leq \left(\prod_{k=m}^{n} (1-\alpha_k)\right) s_m + \left(1 - \prod_{k=m}^{n} (1-\alpha_k)\right) C + \sum_{k=m}^{n} \gamma_k,$$

for all $n \geq m$.

Proof. The induction start $n = m$ is clear. Now let $n > m$. Then

$$\begin{aligned}
s_{n+1} &\leq (1-\alpha_n) s_n + \alpha_n \beta_n + \gamma_n \\
&\leq (1-\alpha_n) \left(\prod_{k=m}^{n-1} (1-\alpha_k) s_m + \left(1 - \prod_{k=m}^{n-1} (1-\alpha_k)\right) C + \sum_{k=m}^{n-1} \gamma_k\right) + \alpha_n \beta_n + \gamma_n \\
&\leq \left(\prod_{k=m}^{n} (1-\alpha_k)\right) s_m + \left((1-\alpha_n) - \prod_{k=m}^{n} (1-\alpha_k)\right) C + \alpha_n \cdot C + \sum_{k=m}^{n} \gamma_k \\
&= \left(\prod_{k=m}^{n} (1-\alpha_k)\right) s_m + \left(1 - \prod_{k=m}^{n} (1-\alpha_k)\right) C + \sum_{k=m}^{n} \gamma_k.
\end{aligned}$$

□

Lemma 5.2.3. *Let (s_n) be a sequence of non-negative real numbers bounded by some constant $C \in \mathbb{N}$. Furthermore, let (s_n) satisfy*

$$s_{n+1} \leq (1-\alpha_n) s_n + \alpha_n \beta_n + \gamma_n, \quad \textit{for all } n \geq 0,$$

where $(\alpha_n) \subset [0,1]$, $(\beta_n) \subset \mathbb{R}$ and $(\gamma_n) \subset [0,\infty)$ are real sequences. Moreover, suppose that $S_1, S_2, S_3 : (0,\infty) \to \mathbb{N}$ such that

(i) $\forall \varepsilon > 0 \prod_{k=0}^{S_1(\varepsilon)} (1-\alpha_k) \leq \varepsilon$

(ii) $\forall \varepsilon > 0 \forall n \geq S_2(\varepsilon)\, \beta_n \leq \varepsilon$,

(iii) $\forall \varepsilon > 0 \forall i \geq j \geq S_3(\varepsilon) \sum_{n=i}^{j} \gamma_n \leq \varepsilon$.

Finally, assume that $D : (0,\infty) \to (0,\infty)$ satisfies

$$0 < D(\varepsilon) \leq \prod_{n=0}^{\max\{S_2(\varepsilon/3), S_3(\varepsilon/3)\}} (1-\alpha_k).$$

Then

$$\forall \varepsilon > 0 \forall n \geq \Phi(\varepsilon, C, S_1, S_2, S_3, D)\, (s_{n+1} \leq \varepsilon),$$

where

$$\Phi(\varepsilon, C, S_1, S_2, S_3, D) = \max\left\{S_1\left(\frac{\varepsilon \cdot D(\varepsilon)}{3C}\right), S_2\left(\frac{\varepsilon}{3}\right), S_3\left(\frac{\varepsilon}{3}\right)\right\},$$

Proof. Let $\varepsilon > 0$ be given. Set $m := \max\{S_2(\varepsilon/3), S_3(\varepsilon/3)\}$. Then, for all $n \geq \Phi \geq m$, Lemma 5.2.2 implies

$$\begin{aligned} s_{n+1} &\leq s_m \prod_{k=m}^{n} (1-\alpha_k) + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= s_m \cdot \frac{\prod_{k=0}^{n} (1-\alpha_k)}{\prod_{k=0}^{m-1} (1-\alpha_k)} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \leq C \cdot \frac{\varepsilon \cdot D(\varepsilon)}{3C} \cdot \frac{1}{D(\varepsilon)} + \frac{2\varepsilon}{3} = \varepsilon. \end{aligned}$$

□

5.3 Main Theorems

The following is essentially Proposition 6.1 of [87]. We include a proof for completeness.

Proposition 5.3.1. *Let X be a normed space, C be a closed convex subset of X and $S : C \to C$ be a nonexpansive mapping with a fixed point. Suppose that $M > 0$ and $x_0, u \in C$ are such that $2\max\{\|p - x_0\|, \|p - u\|\} \leq M$ for some fixed point p of S. Assume that $(\alpha_n) \subset (0,1)$ and $R_1, R_2, R_3 : (0,\infty) \to \mathbb{N}$ satisfy*

(i) $\forall \varepsilon > 0 \forall n \geq R_1(\varepsilon) (\alpha_n \leq \varepsilon)$,

(ii) $\forall \varepsilon > 0 \prod_{k=1}^{R_2(\varepsilon)} (1-\alpha_k) \leq \varepsilon$,

(iii) $\forall \varepsilon > 0 \forall n \geq R_3(\varepsilon) (|\alpha_n - \alpha_{n-1}| \leq \varepsilon\alpha_n)$.

Suppose moreover that $D : (0,\infty) \to (0,\infty)$ satisfies

$$0 < D(\varepsilon) \leq \prod_{k=0}^{R_3(\varepsilon/3M)} (1-\alpha_k).$$

Then the sequence (x_n) generated by

$$x_{n+1} := \alpha_n u + (1-\alpha_n) S x_n$$

is asymptotically regular with rate

$$\psi(\varepsilon) := \max\left\{R_1\left(\frac{\varepsilon}{2M}\right), R_2\left(\frac{\varepsilon \cdot D(\varepsilon/2)}{6C}\right), R_3\left(\frac{\varepsilon}{6M}\right)\right\},$$

i.e.,

$$\forall \varepsilon > 0 \, \forall n \geq \psi(\varepsilon) \, (\|x_n - Sx_n\| \leq \varepsilon).$$

Proof. Observe that

$$\begin{aligned}\|x_{n+1} - p\| &= \|\alpha_n(u-p) + (1-\alpha_n)(Sx_n - p)\| \\ &\leq \alpha_n \|u-p\| + (1-\alpha_n)\|x_n - p\|.\end{aligned}$$

Thus, by induction, $\|x_n - p\| \leq \max\{\|p - x_0\|, \|p-u\|\} \leq M/2$. Thus, for all integers $n \geq 1$, $\|x_{n+1} - x_n\| \leq M$ and $\|u - Sx_{n-1}\| \leq \|u-p\| + \|Sp - Sx_{n-1}\| \leq M$. Hence

$$\begin{aligned}\|x_{n+1} - x_n\| &= \|(\alpha_n - \alpha_{n-1})(u - Sx_{n-1}) + (1-\alpha_n)(Sx_n - Sx_{n-1})\| \\ &\leq (1-\alpha_n)\|x_n - x_{n-1}\| + |\alpha_n - \alpha_{n-1}|\,\|u - Sx_{n-1}\| \\ &\leq (1-\alpha_n)\|x_n - x_{n-1}\| + M|\alpha_n - \alpha_{n-1}| \\ &= (1-\alpha_n)\|x_n - x_{n-1}\| + \alpha_n\beta_n,\end{aligned}$$

where

$$\beta_n := M\frac{|\alpha_n - \alpha_{n-1}|}{\alpha_n}.$$

Lemma 5.2.3 with $\gamma_n = 0$ then implies that for all $n \geq \Phi\left(\frac{\varepsilon}{2}, M, R_2, R_3\left(\frac{\cdot}{M}\right), \mathbf{0}, D\right)$, where $\mathbf{0}$ denotes the function that is constant and equal to 0,

$$\|x_{n+1} - x_n\| \leq \frac{\varepsilon}{2}.$$

Moreover, $\|x_{n+1} - Sx_n\| = \alpha_n\|u - Sx_n\| \leq M\alpha_n$. Thus, in total, we get

$$\|x_n - Sx_n\| \leq \|x_n - x_{n+1}\| + \|x_{n+1} - Sx_n\| \leq \varepsilon$$

for all $n \geq \psi(\varepsilon) = \max\left\{R_1\left(\frac{\varepsilon}{2M}\right), \Phi\left(\frac{\varepsilon}{2}, M, R_2, R_3\left(\frac{\cdot}{M}\right), \mathbf{0}, D\right)\right\}$. In other words, ψ is a rate of asymptotic regularity for (x_n). □

Corollary 5.3.2. *If, in the situation of Proposition 5.3.1, $\alpha_n = 1/(n+2)$, $M \geq 1$ and $\varepsilon \leq 3/2$, then*

$$\psi(\varepsilon) = \left\lfloor \frac{12M\left\lfloor\frac{3M}{\varepsilon}\right\rfloor}{\varepsilon}\right\rfloor \leq \left\lfloor\frac{36M^2}{\varepsilon^2}\right\rfloor. \tag{5.3}$$

Remark 5.3.3. Kohlenbach and Leuştean [57] also extracted a quadratic rate of asymptotic regularity of Halpern iterates in W-hyperbolic spaces (which are more general than Banach spaces) under slightly different requirements on (α_n). Both sets of conditions on (α_n) include the natural choice $1/(n+2)$; this corollary states an alternative rate of asymptotic regularity for this special case. In fact, the two rates are almost identical and have the same complexity.

Theorem 5.3.4. *Let X be a uniformly smooth Banach space, whose normalized duality mapping J has modulus of uniform continuity ω, C be a closed convex subset of X and $S : C \to C$ be a nonexpansive mapping with a fixed point. Suppose that $M > 0$ and $x_0, u \in C$ are such that $2\max\{\|p - x_0\|, \|p-u\|\} \leq M$ for some fixed point p*

of S. Suppose that the sequence $\left(z_{1/m}\right)_{m\geq 1}$ of Theorem 5.1.4 is Cauchy with rate of metastability K, i.e.,

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq K(\varepsilon, g) \forall k, l \in [n, n+g(n)] \left\| z_{1/k} - z_{1/l} \right\| \leq \varepsilon.$$

Assume that $(\alpha_n) \subset (0,1)$, $R_1, R_2, R_3 : (0,\infty) \to \mathbb{N}$ satisfy conditions (i),(ii) and (iii) of the previous proposition. Moreover, suppose that $E : \mathbb{N} \to \mathbb{N}$ satisfies

$$0 < E(k) \leq \prod_{n=0}^{k} (1-\alpha_n).$$

Then the sequence (x_n) generated by

$$x_{n+1} := \alpha_n u + (1-\alpha_n) S x_n$$

converges strongly to a fixed point of S. Moreover,

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Sigma(\varepsilon, g, M, K, E, R_1, R_2, R_3, \omega) \forall k, l \in [n, n+g(n)] \\ \left(\|x_k - x_l\| \leq \varepsilon\right), \quad (5.4)$$

where $\Sigma := \max\left\{\Gamma, \max\left\{R_2\left(\frac{E(k)\cdot\varepsilon^2}{12M^2}\right), \tilde{\Gamma} \leq k \leq \Gamma\right\}\right\}$ and

$$\begin{aligned}
&\varepsilon_0 := \min\{\delta, \omega(M,\delta)\}, \quad \delta := \frac{\varepsilon^2}{144M}, \quad D(\varepsilon) := E(R_3(\varepsilon/3M)), \\
f^*(k) &:= f(k+m_0) + m_0, \quad m_0 := \left\lceil \frac{72M^2}{\varepsilon^2} \right\rceil, \quad \tilde{E}(k) := E(\varphi(k)), \\
f(k) &:= \left\lceil \max\left\{ 24M^2 \cdot \frac{\max\left\{g^*\left(R_2\left(\frac{\tilde{E}(k)\cdot\varepsilon^2}{12M^2}\right)\right), \varphi(k)\right\} - \varphi(k)}{\varepsilon^2} - k, 0 \right\} \right\rceil, \\
g^*(k) &:= k + g(k), \quad \varphi(k) := \psi\left(\frac{\varepsilon^2}{72Mk}\right) \\
&\Gamma := \max\{\varphi(k),\, m_0 \leq k \leq K(\varepsilon_0, f^*) + m_0\}, \\
&\tilde{\Gamma} := \min\{\varphi(k),\, m_0 \leq k \leq K(\varepsilon_0, f^*) + m_0\}
\end{aligned}$$

and ψ is as defined in Proposition 5.3.1, i.e., a rate of asymptotic regularity for (x_n).

Proof. Set $z_m := z_{1/m}$. First observe that, as in the proof of Proposition 5.3.1, $\|x_n - p\| \leq M/2$ Moreover,

$$\|z_m - p\| = \|\frac{1}{m}(u-p) + (1-\frac{1}{m})(Sz_m - Sp)\| \leq \frac{1}{m}\|u-p\| + (1-\frac{1}{m})\|z_m - p\|.$$

Therefore, we have the estimates

$$\|x_m - p\|, \|z_m - p\| \leq M/2 \text{ for all integers } m \geq 0, \quad (5.5)$$

and hence also

$$\|Sx_m - p\|, \|Sz_m - p\| \leq M/2 \text{ for all integers } m \geq 0,$$

which will be crucial in the rest of the proof.

Now let $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$ be given. Then there is some $K_1 \leq K(\varepsilon_0, f^*)$ such that $\|z_k - z_l\| \leq \varepsilon_0$ for all $k, l \in [K_1, K_1 + f^*(K_1)]$. Set $K_0 := m_0 + K_1 \leq m_0 + K(\varepsilon_0, f^*)$. Then, the interval I of positive integers defined by $I := [K_0, K_0 + f(K_0)] = [K_1 + m_0, K_1 + m_0 + f(K_1 + m_0)] \subseteq [K_1, K_1 + f^*(K_1)]$ and so we have $\|z_m - z_{K_0}\| \leq \varepsilon_0$ for all $m \in I$, and $K_0 \geq m_0 \geq 72M^2/\varepsilon^2$. Consequently, if we let

$$\beta_n^m := 2\langle u - z_m, J(x_n - z_m)\rangle - \frac{M^2}{m},$$

then, since $\|z_{K_0} - z_m\| \leq \varepsilon_0 \leq \omega(M, \delta)$, we obtain $2\langle u - z_m, J(x_n - z_m) - J(x_n - z_{K_0})\rangle \leq 2\|u - z_m\| \cdot \|J(z_{K_0} - z_m)\| \leq 2\delta(\|u - p\| + \|p - z_m\|) \leq 2M\delta$ for all $m \in I$,

$$\begin{aligned}
\beta_n^m - \beta_n^{K_0} &= 2\langle u - z_m, J(x_n - z_m)\rangle - 2\langle u - z_{K_0}, J(x_n - z_{K_0})\rangle + \left(\frac{1}{K_0} - \frac{1}{m}\right) M^2 \\
&\leq 2\langle u - z_m, J(x_n - z_m) - J(x_n - z_{K_0})\rangle + 2\langle z_{K_0} - z_m, J(x_n - z_{K_0})\rangle + \frac{M^2}{K_0} \\
&\leq \frac{\varepsilon^2}{72M} \cdot M + \frac{\varepsilon^2}{72M} \cdot M + \frac{\varepsilon^2}{72M^2} \cdot M^2 = \frac{\varepsilon^2}{24}
\end{aligned}$$

for all $m \in I$. Moreover, by the subdifferential inequality (see Lemma 4.2.4(iii)), we obtain analogously to [102],

$$\begin{aligned}
\|z_m - x_n\|^2 &\leq (1 - \frac{1}{m})^2 \|Sz_m - x_n\|^2 + \frac{2}{m}\langle u - x_n, J(z_m - x_n)\rangle \\
&\leq (1 - \frac{1}{m})^2 (\|Sz_m - Sx_n\| + \|Sx_n - x_n\|)^2 \\
&\quad + \frac{2}{m}\left(\|z_m - x_n\|^2 + \langle u - z_m, J(z_m - x_n)\rangle\right) \\
&\leq \left(1 + \frac{1}{m^2}\right)\|z_m - x_n\|^2 + \|Sx_n - x_n\|(2\|z_m - x_n\| + \|Sx_n - x_n\|) \\
&\quad + \frac{2}{m}\langle u - z_m, J(z_m - x_n)\rangle .
\end{aligned}$$

Therefore,

$$\begin{aligned}
2\langle u - z_m, J(x_n - z_m)\rangle &\leq \frac{1}{m}\|z_m - x_n\|^2 + m\|Sx_n - x_n\|(2\|z_m - x_n\| + \|Sx_n - x_n\|) \\
&\leq \frac{M^2}{m} + 3Mm\|x_n - Sx_n\|,
\end{aligned}$$

so $\beta_n^m \leq 3mM\|Sx_n - x_n\|$. Since ψ is a rate of asymptotic regularity for x_n, we know that $\beta_n^{K_0} \leq \varepsilon^2/24$ for all $n \geq n_0 := \psi\left(\frac{\varepsilon^2}{72MK_0}\right) = \varphi(K_0)$ and so $\beta_n^m \leq \frac{\varepsilon^2}{24} + \beta_n^{K_0} \leq \varepsilon^2/12$

for all $n \geq n_0$ and $m \in I$. Consequently, applying the subdifferential inequality (Lemma 4.2.4(iii)) yields

$$\begin{aligned}\|x_{n+1} - z_m\|^2 &= \|(1-\alpha_n)(Sx_n - z_m) + \alpha_n (u - z_m)\|^2 \\ &\leq (1-\alpha_n)^2 \|Sx_n - z_m\|^2 + 2\alpha_n \langle u - z_m, J(x_{n+1} - z_m)\rangle \\ &= (1-\alpha_n)^2 \|Sx_n - Sz_m + Sz_m - z_m\|^2 + \alpha_n \beta_{n+1}^m + \frac{\alpha_n M^2}{m}.\end{aligned}$$

Applying again the subdifferential inequality and observing that $Sz_m - z_m = (Sz_m - u)/m$,

$$\begin{aligned}\|x_{n+1} - z_m\|^2 &\leq (1-\alpha_n)^2 \left(\|Sx_n - Sz_m\|^2 + 2\langle Sz_m - z_m, J(Sx_n - z_m)\rangle\right) \\ &\quad + \alpha_n \beta_{n+1}^m + \frac{\alpha_n M^2}{m} \\ &\leq (1-\alpha_n)^2 \left(\|x_n - z_m\|^2 + \frac{2}{m}\langle Sz_m - u, J(Sx_n - z_m)\rangle\right) \\ &\quad + \alpha_n \beta_{n+1}^m + \frac{\alpha_n M^2}{m} \\ &\leq (1-\alpha_n) \left(\|x_n - z_m\|^2 + \frac{2}{m}\langle Sz_m - u, J(Sx_n - z_m)\rangle\right) \\ &\quad + \alpha_n \beta_{n+1}^m + \frac{\alpha_n M^2}{m} \\ &\leq (1-\alpha_n)\left(\|x_n - z_m\|^2 + \frac{2M^2}{m}\right) + \alpha_n \beta_{n+1}^m + \frac{\alpha_n M^2}{m} \\ &\leq (1-\alpha_n)\|x_n - z_m\|^2 + \alpha_n \beta_{n+1}^m + \frac{2M^2}{m}.\end{aligned}$$

Thus, if we apply Lemma 5.2.2 with $\gamma_n = \frac{2M^2}{m}$, we obtain for $n > n_0$

$$\begin{aligned}\|x_n - z_m\|^2 &\leq \left(\prod_{k=n_0}^{n-1}(1-\alpha_k)\right)\|x_{n_0} - z_m\|^2 + \left(1 - \prod_{k=n_0}^{n-1}(1-\alpha_k)\right)\frac{\varepsilon^2}{12} + \frac{2M^2}{m}(n - n_0) \\ &\leq \left(\prod_{k=n_0}^{n-1}(1-\alpha_k)\right) M^2 + \frac{\varepsilon^2}{12} + \frac{2M^2}{m}(n - n_0). \qquad (5.6)\end{aligned}$$

Therefore, for all $n \geq n_1 := \max\left\{R_2\left(\frac{E(n_0)\cdot\varepsilon^2}{12M^2}\right), n_0\right\}$,

$$\|x_n - z_m\|^2 \leq \frac{\varepsilon^2}{6} + \frac{2M^2}{m}(n - n_0). \qquad (5.7)$$

Now observe that

$$\begin{aligned}K_0 + f(K_0) &= \left\lceil \max\left\{ 24M^2 \cdot \frac{g^*\left(\max\left\{R_2\left(\frac{\tilde{E}(K_0)\cdot\varepsilon^2}{12M^2}\right), n_0\right\}\right) - \varphi(K_0)}{\varepsilon^2}, K_0 \right\} \right\rceil \\ &\geq 24M^2 \cdot \frac{g^*\left(\max\left\{R_2\left(\frac{E(n_0)\cdot\varepsilon^2}{12M^2}\right), n_0\right\}\right) - \varphi(K_0)}{\varepsilon^2} \\ &= 24M^2 \cdot \frac{g^*(n_1) - n_0}{\varepsilon^2}.\end{aligned}$$

Thus, if we set $P := K_0 + f(K_0) \in I$, we have

$$P \geq \frac{24M^2 \left(g(n_1) + n_1 - n_0\right)}{\varepsilon^2}.$$

Then, for all $n \in [n_1, n_1 + g(n_1)]$

$$\frac{2M^2}{P}(n - n_0) \leq \frac{2M^2 \left(n_1 + g(n_1) - n_0\right)}{2 \cdot 12M^2 \left(g(n_1) + n_1 - n_0\right)} \varepsilon^2 = \frac{\varepsilon^2}{12}.$$

Then (5.7) yields

$$\|x_n - z_P\| \leq \frac{\varepsilon}{2}.$$

Thus, for all $k, l \in [n_1, n_1 + g(n_1)]$,

$$\|x_k - x_l\| \leq \|x_k - z_P\| + \|x_l - z_P\| \leq \varepsilon.$$

The claim follows from the observation that, since $m_0 \leq K_0 \leq m_0 + K(\varepsilon_0, f^*)$, $n_0 \in [\tilde{\Gamma}, \Gamma]$ and so $n_1 \leq \Sigma(\varepsilon, g, M, R_1, R_2, R_3, \omega)$. $\square$

As mentioned in the introduction, Kohlenbach and Leuştean [58] extracted from a proof due to Shioji and Takahashi [88] a bound for slightly different conditions on (α_n), which also include $\alpha_n = 1/(n+2)$. If, furthermore, we restrict ourselves to a Hilbert space setting, then the metastability of the resolvent $(z_n) := (z_{1/n})$ is known from the following theorem.

Theorem 5.3.5 (Kohlenbach [52]). *Let X be a real Hilbert space and $C \subset X$ be a bounded closed convex subset with diameter $\operatorname{diam}(C) \leq M$ and $S : C \to C$ be a nonexpansive mapping. Then for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$,*

$$\exists n \leq K(\varepsilon, g) := \tilde{g}^{(\lceil M^2/\varepsilon^2 \rceil)}(0) \, \forall i, j \in [n, n + g(n)] \left(\|z_i - z_j\|\right)$$

and $\tilde{g}(n) = \max\{n, g(n)\}$.

Moreover, it is obvious that we may take $\omega = \mathrm{id}$ in this case. After lengthy, but trivial calculations, we see that for $\alpha_n = 1/(n+2)$ the rate of metastability in our case is as follows. The counterfunction g is modified to, essentially

$$f^*(k) = \frac{M^2}{\varepsilon^2} g\left(\frac{M^6 k^2}{\varepsilon^6}\right)$$

and then iterated M^3/ε^4 many times before being multiplied by M^6/ε^6. In the Addendum [59] to [58], the counterfunction is modified to essentially

$$f^*(k) = \frac{M^2}{\varepsilon^2} g\left(\frac{M^6 k^2}{\varepsilon^4}\right),$$

which is slightly better, but here f^* is iterated M^4/ε^4 times.

6 Bruck's Iteration for Lipschitzian Pseudocontractions

6.1 Introduction

In this chapter, we turn to iterations approximating fixed points of pseudocontractive mappings.

Definition 6.1.1 ([12]). Let X be a normed linear space and $S \subset X$ be a subset. A mapping $T : S \to S$ is called *pseudocontractive* if

$$\|u - v\| \leq \|(1 + \lambda)(u - v) - \lambda(Tu - Tv)\|, \quad \text{for all } u, v \in S \text{ and all } \lambda \geq 0.$$

Apart from generalizing nonexpansive mappings, pseudocontractive mappings are closely related to accretive mappings since an operator $A : X \to X$ is accretive if and only if $I - A$ is pseudocontractive, where I denotes the identity map. Therefore, any fixed point of a pseudocontractive mapping $T : X \to X$ is the zero, or root, of the accretive operator $I - T$. Accretive mappings were introduced by Friedrichs [28] and studied extensively since then as they describe certain nonlinear evolution systems, while their roots correspond to equilibrium points. The existence and approximation of these roots are mostly studied via the fixed point theory of pseudocontractive maps.

To approximate fixed points of pseudocontractions, the principal paradigm is to extend the methods and results for nonexpansive mappings. However, it is known that the Halpern iteration does not converge strongly for pseudocontractions, so Bruck proposed the following variant:

Definition 6.1.2 ([15]). Let C be a nonempty convex subset of a real normed space and let $T : C \to C$ be a pseudocontraction. Let (λ_n) and (θ_n) be sequences in $[0, 1]$ with $\lambda_n(1 + \theta_n) \leq 1$ for all $n \in \mathbb{N}$. The *Bruck iteration scheme* with starting point $x_1 \in C$ is defined as

$$x_{n+1} = (1 - \lambda_n)\, x_n + \lambda_n T x_n - \lambda_n \theta_n \,(x_n - x_1)\,.$$

This iteration combines Halpern's scheme with the so-called Krasnoselskij-Mann iteration defined for a sequence $(\theta_n) \in [0, 1]$ by

$$y_{n+1} := \theta_n y_n + (1 - \theta_n) T y_n,$$

The results of this Chapter have been published in [71] in collaboration with Prof. Ulrich Kohlenbach. The investigation was suggested by Prof. Kohlenbach along with initial remarks, while the results were worked out and written up by the author of this thesis.

which does not converge strongly even for nonexpansive mappings unless the underlying space is finite dimensional. Nonetheless, merging the two non-convergent schemes does, surprisingly, produce a convergent algorithm that approximates a fixed point of T:

Among numerous other things, Bruck showed that for bounded closed and convex subsets C of *Hilbert* spaces, this iteration converges strongly for so-called acceptably paired sequences $(\lambda_n), (\theta_n)$, see Definition 7.1.1. Moreover the limit is a fixed point of T provided that T is demicontinuous in addition to being pseudocontractive, where demicontinuity is defined as follows.

Definition 6.1.3. Let H be a Hilbert space and S a subset. A mapping $T : C \to H$ is called *demicontinuous* if it is continuous from the strong to the weak topology on H.

Bruck's theorem will be subject to quantitative analysis in the next chapter. For now, we will however treat the class of Lipschitzian pseudocontractions, since the analysis is similar to that of the Halpern iteration in that convergence and asymptotic regularity are established via the resolvent (z_t) defined by the equation $z_t = tTz_t + (1-t)x$ for fixed $x \in C$ and $t \in [0, 1)$ using similar techniques.

Chidume an Zegeye later proved [19] that Bruck's iteration is asymptotically regular, i.e.

$$\|x_n - T(x_n)\| \overset{n\to\infty}{\to} 0,$$

in *arbitrary* Banach spaces provided that T is a Lipschitzian pseudocontractive mapping, which still includes the important class of strictly pseudocontractive mappings in the sense of Browder and Petryshyn [12] (see [18]). Additionally, the somewhat artificial conditions of Bruck are replaced by the following, slightly more natural ones:

Definition 6.1.4 (Chidume and Zegeye [19])**.** The real sequences (λ_n) and (θ_n) in $(0, 1]$ are said to satisfy the Chidume-Zegeye conditions if

1. $\lim_{n\to\infty} \theta_n = 0$;
2. $\sum_{n=1}^{\infty} \lambda_n \theta_n = \infty$;
3. $\forall \varepsilon > 0 \exists m \in \mathbb{N} \forall n \geq m \, (\lambda_n \leq \theta_n \varepsilon)$;
4. $\forall \varepsilon > 0 \exists m \in \mathbb{N} \forall n \geq m \left(\frac{\left|\frac{\theta_{n-1}}{\theta_n} - 1\right|}{\lambda_n \theta_n} \leq \varepsilon \right)$;
5. $\lambda_n (1 + \theta_n) \leq 1$ for all $n \in \mathbb{N}$.

Theorem 6.1.5 (Chidume and Zegeye [19])**.** *Let C be a nonempty closed convex subset of a real Banach space X. Let $T : C \to C$ be a Lipschitz pseudocontractive map with Lipschitz constant L and $Fix(T) \neq \emptyset$. Let (x_n) be the Bruck iteration with starting point $x_1 \in C$, where the parameters (λ_n) and (θ_n) satisfy the Chidume-Zegeye conditions. Then $\|x_n - Tx_n\| \to 0$ as $n \to \infty$.*

Remark 6.1.6. Instead of $Fix(T) \neq \emptyset$ one can also assume that C is bounded.

Remark 6.1.7. We should note that both Bruck's conditions as well as those of Chidume and Zegeye include the most important choice $\lambda_n := n^{-p}$ and $\theta_n := n^{-q}$, where p and q are real numbers in $(0,1)$ such that $0 < q < \min\{p, 1-p\}$.

Theorem 6.1.5 is shown as a consequence of the fact that $\|x_n - z_{n-1}\| \to 0$, where z_n is the unique point (whose existence is guaranteed by [80]) satisfying

$$z_n = t_n T(z_n) + (1-t_n)x_1, \quad \text{where } t_n := \frac{1}{1+\theta_n}.$$

In particular, (x_n) strongly converges towards a fixed point of T provided that (z_n) does. The latter is known to be the case for example in reflexive Banach spaces X with uniformly Gâteaux differentiable norm provided that T has a fixed point or C is bounded and every nonempty bounded closed convex subset of X has the fixed point property for nonexpansive self-mappings (see [85]).

So, in particular, (z_n) and consequently (x_n) converge strongly to a fixed point of T if X is a uniformly smooth Banach space, T has a fixed point and C is closed and convex:

Theorem 6.1.8 (Corollary 11.8 in [18])**.** *Suppose X is a uniformly smooth Banach space, $C \subset X$ is closed and convex, $T : C \to C$ is a Lipschitz pseudocontractive mapping and $Fix(T)$ is nonempty. Then the Bruck iteration converges strongly to a fixed point of T.*

6.2 Quantitative Analysis

In [70], which is based on the Bachelor's Thesis of the author of this thesis [67], we extracted from the proof of Theorem 6.1.5 in [19] explicit and highly uniform rates of convergence for the asymptotic regularity $\|x_n - Tx_n\| \to 0$ and for $\|x_n - z_{n-1}\| \to 0$. Therefore, a hypothetical rate of convergence for (z_n) can be transformed into a rate of convergence for (x_n), while a hypothetical rate of metastability can be transformed into a rate of metastability for (z_n).

Recall from Chapter 5 that a uniform rate of convergence for (z_n) cannot exist even if T is nonexpansive and X is a Hilbert space. Therefore, we can again only hope to obtain a rate of metastability

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Phi(\varepsilon, g)\, \forall i, j \in [n; n+g(n)]\, \big(\|z_i - z_j\| < \varepsilon\big),$$

which we extract for the Hilbert space case. We then combine this with our rate asymptotic regularity to obtain (again for Hilbert spaces) a rate of metastability Ω for (x_n). In fact we get

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Omega(\varepsilon, g)\, \forall i, j \in [n; n+g(n)] \forall l \geq n\, \big(\|x_i - x_j\| < \varepsilon \wedge \|Tx_l - x_l\| < \varepsilon\big). \tag{6.1}$$

As guaranteed by metatheory, in addition to ε and g, the rate Ω only depends on a Lipschitz constant L for T, an upper bound $d \geq \|x_1 - p\|$ for some T-fixed point p and the moduli arising from the conditions imposed on the scalars $(\lambda_n), (\theta_n)$.

Statement (6.1) trivially implies the finitary (in the sense that only a finite initial segment of (x_n) is mentioned) statement

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Omega(\varepsilon, g) \, \forall i, j \in [n; n + g(n)] \, \big(\|x_i - x_j\| < \varepsilon \wedge \|Tx_i - x_i\| < \varepsilon \big)$$

which – in turn – trivially implies that (x_n) strongly converges to a fixed point of T as metastability ineffectively is equivalent to the usual Cauchy property.

6.2.1 Resolvent Convergence

To obtain a rate of metastability for (z_t) in Hilbert spaces, it is crucial to observe that the demicontinuity is only required to prove the existence of (z_t), and is no longer required afterwards. In particular, whenever $T : C \to C$ is a pseudocontractive selfmapping of a closed convex subset C of a Hilbert space H such that the resolvent (z_t) exists, (z_t) converges strongly to a fixed point of T, cf. Theorem 6.2.4 below.

Consequently, instead of assuming demicontinuity, we can require the existence of the sequence (z_{t_n}) for some $t_n \in [0, 1)$. The latter is a purely universal statement, so it is admissible in the metatheorem. Moreover, since the boundedness of (z_{t_n}) is provable, we do not even need to require a majorant in the sense of Definition 3.2.3. Once the extraction of the rate of metastability has been carried out, this rate is then still valid under the assumption of demicontinuity, since demicontinuity validates the existence of (z_{t_n}). So although demicontinuity is not a purely universal statement, it does not make *any* quantitative contribution to the rate of metastability.

We now prove that the path (z_t) exists whenever T is demicontinuous, a result which is closely related to results of Browder [10] and Bruck [15]. It has been shown by Lan and Wu in [72] using techniques similar to those of Browder [11]. Although Browder's proof (for the nonexpansive case) has been analyzed by Kohlenbach in [52], it is considerably more difficult to treat than our proof below which follows the ideas of [15] (which in turn is based on [38]).

Lemma 6.2.1. *Let X be a Hilbert space and $S \subset H$ be a subset. Then $T : S \to S$ is pseudocontractive if and only if*

$$\langle Tu - Tv, u - v \rangle \leq \|u - v\|^2, \quad \text{for all } u, v \in S.$$

We now assume demicontinuity of T:

Theorem 6.2.2. *Let H be a Hilbert space, $C \subseteq H$ be a nonempty bounded closed convex subset and $T : C \to C$ be a demicontinuous pseudocontraction. Then, for each $x \in C$ and $t \in [0, 1)$, there exists a unique path (z_t) in C such that $z_t = tTz_t + (1 - t)\, x$. Moreover, the strong limit*

$$\lim_{t \to 1^-} z_t = z,$$

exists and is the fixed point of T closest to x.

Proof. For each $x \in C$ and nonnegative $t < 1$, the mapping $T_t : C \to C$, $z \mapsto tTz + (1-t)\,x$ satisfies

$$\begin{aligned} \langle T_t x_1 - T_t x_2, x_1 - x_2 \rangle &= \langle tTx_1 + (1-t)\,x - tTx_2 - (1-t)\,x, x_1 - x_2 \rangle \\ &= t\,\langle Tx_1 - Tx_2, x_1 - x_2 \rangle \\ &\leq t\,\|x_1 - x_2\|^2 . \end{aligned} \tag{6.2}$$

Therefore, T_t is pseudocontractive. It is also demicontinuous: for any sequence (x_n) in C with $x_n \to x$, we have

$$\langle y, T_t x_n - T_t x \rangle = t\,\langle y, Tx_n - Tx \rangle \to 0 \quad \text{for all } y \in H$$

since T was demicontinuous. We conclude by Corollary 4 of [15] that T_t has a fixed point $z_t \in C$, *i.e.*, a point satisfying the equation

$$z_t = tTz_t + (1-t)\,x.$$

Moreover, by (6.2), T_t is even strongly pseudocontractive, so z_t is unique. To see this, suppose that z_t and z_t' are two fixed points of T_t. Then, by (6.2),

$$\left\|z_t - z_t'\right\|^2 = \left\langle z_t - z_t', z_t - z_t' \right\rangle = \left\langle T_t z_t - T_t z_t', z_t - z_t' \right\rangle \leq t\,\left\|z_t - z_t'\right\|^2 .$$

Since $t < 1$, this implies $z_t = z_t'$. That (z_t) is continuous in t follows as in [80].

Strong convergence of (z_t) will be established in the course of the proof of Theorem 6.2.4. That the strong limit is a fixed point of T follows from (here we use that C is bounded)

$$|\langle Tz_t - z_t, Tz - z \rangle| \leq \|Tz_t - z_t\| \cdot \|Tz - z\| \overset{t \to 1^-}{\to} 0$$

and that (using that T is demicontinuous)

$$\langle Tz_t - z_t, Tz - z \rangle \overset{t \to 1^-}{\to} \langle Tz - z, Tz - z \rangle.$$

We now proceed to show that the strong limit is the fixed point of T with minimal distance from x. Suppose that y is a fixed point of T. Then $y = tTy + (1-t)\,x$ for $t = 1$. Repeating the calculations leading to (6.4) further below with $z_t = y$ and $t = 1$, we obtain

$$\|y - x\|^2 \geq \|z_s - x\|^2 + \|y - z_s\|^2 , \quad \text{for all } 0 < s < 1.$$

Taking the strong limit $s \to 1$ implies

$$\|y - x\|^2 \geq \|z - x\|^2 + \|y - z\|^2$$

showing that z is the (unique) fixed point of T that is closest to x. □

We use the following Lemma:

Lemma 6.2.3 ([46]). *Let $D \in \mathbb{R}_+$ be a nonnegative real number and (a_n) be a nondecreasing sequence in the interval $[0, D]$, i.e. $0 \le a_n \le a_{n+1} \le D$. Then the following holds*

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \le \tilde{g}^{(\lceil D/\varepsilon \rceil)}(1) \, \forall i, j \in [n; n + g(n)] \, (|a_i - a_j| \le \varepsilon),$$

where $\tilde{g}(n) := n + g(n)$. Moreover, n can be taken as $\tilde{g}^{(i)}(1)$ for some suitable $i \le \lceil D/\varepsilon \rceil$.

Theorem 6.2.4. *Let X be a real inner product space and $C \subseteq X$ be a convex subset. Let $T : C \to C$ be a pseudocontraction which possesses a fixed point $v \in C$. Let $x \in C$ and assume that there exists (z_t) for x such that*

$$z_t = tTz_t + (1 - t)x, \quad t \in [0, 1).$$

Let (t_n) be a sequence in $(0, 1)$ that converges towards 1 and $h : \mathbb{N} \to \mathbb{N}$ be such that $t_n \le 1 - \frac{1}{h(n)+1}$ for all $n \in \mathbb{N}$. Set $z_n := z_{t_n}$. Then, for all $\varepsilon > 0$, all $g : \mathbb{N} \to \mathbb{N}$ and all $\mathbb{N} \ni d \ge \|v - x\|$

$$\exists n \le \Phi\left(\varepsilon, g, \chi_g, h, d\right) \forall i, j \in [n; n + g\left(n\right)] \left(\|z_i - z_j\| \le \varepsilon\right),$$

where

$$\Phi\left(\varepsilon, g, \chi_g, h, d\right) := \chi_g^M\left(g_{h,\chi_g}^{(\lceil 16d^2/\varepsilon^2 \rceil)}\left(1\right)\right)$$

with

$$g_{h,\chi_g}\left(n\right) := \max\left\{h\left(i\right) : i \le \chi_g\left(n\right) + g\left(\chi_g(n)\right)\right\}$$

and $\chi_g : \mathbb{N} \to \mathbb{N}$ is any function satisfying

$$\forall n \in \mathbb{N} \forall i \in [\chi_g(n); \tilde{g}\left(\chi_g(n)\right)] \left(|1 - t_i| \le \frac{1}{n+1}\right). \tag{6.3}$$

If (t_n) is a nondecreasing sequence in $(0, 1)$ (not necessarily converging towards 1), then the bound can be simplified to $\Psi\left(\varepsilon, g, d\right) := \tilde{g}^{(\lceil 4d^2/\varepsilon^2 \rceil)}(1)$, where $\tilde{g}(n) := n + g(n)$.

Proof. Assume that $z_t \in C$ satisfies the equation

$$z_t = tTz_t + (1 - t)\, x$$

for all $t \in [0, 1)$. For $1 > t > s > 0$, we carry out a calculation similar to [52] and [38]; Since $Tz_t = \frac{1}{t} z_t - \frac{1-t}{t} x$ and T is pseudocontractive,

$$\begin{aligned}
\|z_t - z_s\|^2 \ge \langle Tz_t - Tz_s, z_t - z_s \rangle &= \left\langle \frac{1}{t} z_t - \frac{1-t}{t} x - \frac{1}{s} z_s + \frac{1-s}{s} x, z_t - z_s \right\rangle \\
&= \left\langle \frac{1}{t} z_t - \frac{1}{t} z_s + \frac{1}{t} z_s - \frac{1}{s} z_s, z_t - z_s \right\rangle + \frac{t-s}{ts} \left\langle x, z_t - z_s \right\rangle \\
&= \frac{1}{t} \|z_t - z_s\|^2 + \left\langle \frac{s-t}{st} z_s, z_t - z_s \right\rangle + \frac{t-s}{ts} \left\langle x, z_t - z_s \right\rangle,
\end{aligned}$$

and since $0 < t < 1$,

$$\left\langle \frac{t-s}{st} z_s, z_t - z_s \right\rangle \geq \left(\frac{1}{t} - 1 \right) \|z_t - z_s\|^2 + \frac{t-s}{ts} \langle x, z_t - z_s \rangle \geq \frac{t-s}{ts} \langle x, z_t - z_s \rangle .$$

Since $s < t$, we conclude

$$\langle z_s - x, z_t - z_s \rangle \geq 0.$$

Therefore,

$$\begin{aligned} \|z_t - x\|^2 = \langle z_t - x, z_t - x \rangle &= \langle z_s - x + (z_t - z_s), z_s - x + (z_t - z_s) \rangle \\ &= \langle z_s - x, z_s - x \rangle + \langle z_t - z_s, z_t - z_s \rangle + 2 \langle z_s - x, z_t - z_s \rangle \\ &\geq \|z_s - x\|^2 + \|z_t - z_s\|^2 . \end{aligned} \tag{6.4}$$

Therefore, $(\|z_t - x\|^2)_t$ is nondecreasing (as $t \nearrow 1^-$) and

$$\|z_t - z_s\|^2 \leq | \, \|z_s - x\|^2 - \|z_t - x\|^2 \, |. \tag{6.5}$$

(z_t) is also bounded as follows from the existence of a fixed point $v \in C$ reasoning as in Proposition 2(iv) of [80]: If $v \in Fix(T)$, then

$$\begin{aligned} \|z_t - v\|^2 &= \langle tTz_t + (1-t)\, x - v, z_t - v \rangle \\ &= t \langle Tz_t - Tv, z_t - v \rangle + (1-t) \langle x - v, z_t - v \rangle \\ &\leq t\|z_t - v\|^2 + (1-t) \langle x - v, z_t - v \rangle , \end{aligned}$$

which implies

$$(1-t) \|z_t - v\|^2 \leq (1-t) \|x - v\| \cdot \|z_t - v\|.$$

Since $t < 1$, this implies that $\|z_t - v\| \leq \|x - v\|$. Hence

$$\|z_t - x\| \leq \|z_t - v\| + \|v - x\| \leq 2 \|v - x\| \leq 2d,$$

so $(\|z_t - x\|^2)_t$ is bounded by $4d^2$.

Together with Lemma 6.2.3 applied to $(\|z_{t_n} - x\|^2)_n, 4d^2$ and ε^2 and (6.5) above the theorem now follows in the case where $1 > t_{n+1} \geq t_n > 0$ for all $n \in \mathbb{N}$. For the case of a general sequence (t_n) which is assumed to converge to 1 one reasons literally as in the proof of Theorem 4.2 in [52]. □

Remark 6.2.5. Theorem 4.2 of [52] establishes the same result for nonexpansive mappings.

Remark 6.2.6. It is not hard to show that Theorem 6.2.4 also holds with the assumption $Fix(T) \neq \emptyset$ being replaced by $\forall \varepsilon > 0 \exists v_\varepsilon \in C(\|x - v_\varepsilon\| \leq d \wedge \|Tv_\varepsilon - v_\varepsilon\| \leq \varepsilon)$.

6.2.2 Asymptotic Regularity of the Bruck Iteration

Theorem 6.2.7 ([70]). *Let C be a nonempty, closed and convex subset of a real Banach space X and $x \in C$. Let $T : C \to C$ be a Lipschitzian pseudocontractive mapping with Lipschitz constant L and for some $d > 0$ assume that T possesses arbitrarily good ε-fixed points $p_\varepsilon \in C$ with $\|x - p_\varepsilon\| < d$. Let (x_n) be the Bruck iteration (Definition 6.1.2) with starting point $x_1 := x$. Let z_n be the unique element in C satisfying $z_n = t_n T(z_n) + (1 - t_n)x_1$ with $t_n := 1/(1+\theta_n)$. Given rates of convergence/divergence $R_i : (0,\infty) \to \mathbb{N}$ for the Chidume-Zegeye conditions 6.1.4, we get*

$$\forall \varepsilon > 0 \forall n \geq \Psi\left(d, L, R_1, R_2, R_3, R_4, \varepsilon\right) \left(\|x_n - Tx_n\| < \varepsilon\right)$$

and

$$\forall \varepsilon > 0 \forall n \geq \chi\left(d, L, R_1, R_2, R_3, R_4, \varepsilon\right) \left(\|x_n - z_{n-1}\| < \varepsilon\right),$$

where

$$\Psi\left(d, L, R_1, R_2, R_3, R_4, \varepsilon\right) = \max\Big\{N_2\left(C\right) + 1, R_1\left(\frac{\varepsilon}{4r}\right) + 1\Big\}$$

and

$$\chi\left(d, L, R_1, R_2, R_3, R_4, \varepsilon\right) = N_2\left(C\right) + 1$$

with

$$\begin{aligned}
N_1\left(\varepsilon\right) &:= \max\left\{R_3\left(\frac{2\varepsilon s}{3r^2}\right), R_4\left(\sqrt{\frac{\varepsilon}{r^2} + \frac{9}{4}} - \frac{3}{2}\right)\right\},\\
N_2\left(x\right) &:= R_2\left(\frac{x}{2}\right) + 1,\\
C &:= \frac{8\left(1+L\right)^2 r^2}{\varepsilon^2} + 2\left(N_1\left(\frac{\varepsilon^2}{8\left(1+L\right)^2}\right) - 1\right),\\
r &:= \max\left\{\frac{\left(2+L\right)^{R_3(d)} - 1}{1+L} d, 2d\right\},\\
s &:= \frac{1}{2\left(\frac{5}{2} + L\right)\left(2+L\right)}.
\end{aligned}$$

Proof. The first claim is Theorem 1 in [70] and the second claim follows from formula (24) in the proof of that theorem (even with ε being replaced by $\varepsilon/(2(1+L))$ in the definition of χ). $\square$

Corollary 6.2.8 ([70]). *In the situation of Theorem 6.2.7, one may drop the condition that T has arbitrarily good approximate fixed points and instead require* $\operatorname{diam}(C) \leq d$. *In this case,*

$$\begin{aligned}
&\chi(d, L, R_1, R_2, R_3, R_4, \varepsilon) = N_2(C) + 1,\\
&\Psi\left(d, L, R_1, R_2, R_3, R_4, \varepsilon\right) = \max\Big\{\chi(\varepsilon), R_1\left(\frac{\varepsilon}{2d}\right) + 1\Big\}
\end{aligned}$$

and

$$N_1(\varepsilon) := \max\left\{R_3\left(\frac{\varepsilon}{4d^2(2+L)}\right), R_4\left(\sqrt{\frac{\varepsilon}{d^2}+1}-1\right)\right\},$$
$$N_2(x) := R_2\left(\frac{x}{2}\right)+1,$$
$$C := \frac{8(1+L)^2 d^2}{\varepsilon^2} + 2\left(N_1\left(\frac{\varepsilon^2}{8(1+L)^2}\right)-1\right).$$

6.2.3 Strong Convergence of the Bruck Iteration

Theorem 6.2.9. *If, in the situation of Theorem 6.2.7, X is a Hilbert space, then (assuming w.l.o.g. $L \geq 1$) for all $\varepsilon > 0$ and all $g : \mathbb{N} \to \mathbb{N}$*

$$\exists n \leq \chi^M\left(g_{h,\chi}^{(\lceil 64d^2/\varepsilon^2\rceil)}(1)\right) + \Psi(\varepsilon) + 1\, \forall i,j \in [n; n+g(n)]\, \forall l \geq n$$
$$(\|x_i - x_j\| \leq \varepsilon \wedge \|Tx_l - x_l\| \leq \varepsilon)$$

where $h : \mathbb{N} \to \mathbb{N}$ is a function such that $h(n) \geq 1/\theta_n$ for all $n \in \mathbb{N}$ and $\chi(n) := R_1(1/n)$,

$$g'(n) := g(n+1+\Psi(\varepsilon)) + \Psi(\varepsilon) + 1, \quad g_{h,\chi}(n) := \max\left\{h(i) : i \leq \chi(n) + g'(\chi(n))\right\},$$

and R_1 and Ψ as in Corollary 6.2.8.

Proof. In Theorem 6.2.7, the resolvent z_t is instantiated with the sequence $t = t_n = \frac{1}{1+\theta_n}$ and the starting point x_1. We now show how to apply Theorem 6.2.4 to this instantiation; if we set $\chi(n) := R_1(1/n)$, then $\theta_i \leq 1/n$ for all $i \geq \chi(n)$. Since $\theta_n \in (0,1]$, this implies

$$|1 - t_i| = 1 - \frac{1}{1+\theta_i} \leq 1 - \frac{1}{1+\frac{1}{n}} = \frac{1}{n+1}, \quad \text{for all } i \geq \chi(n).$$

Since this holds for all $i \geq \chi(n)$, the function χ satisfies (6.3) independently of the counter-function g and we may set $\chi_g := \chi$ in Theorem 6.2.4.

Moreover, for all $n \in \mathbb{N}$, $h(n) \geq 1/\theta_n$ implies $1 + h(n) \geq \frac{1+\theta_n}{\theta_n}$, whence

$$\frac{1}{h(n)+1} \leq \frac{\theta_n}{1+\theta_n} = 1 - \frac{1}{1+\theta_n}.$$

Therefore,

$$t_n = \frac{1}{1+\theta_n} \leq 1 - \frac{1}{h(n)+1}, \quad \text{for all } n \in \mathbb{N}.$$

Now observe that, by Theorem 6.2.4 and Remark 6.2.6 applied to the counter-function g' and error $\varepsilon/2$, there exists an $n \leq \chi^M\left(g_{h,\chi}^{(\lceil 64d^2/\varepsilon^2\rceil)}(1)\right)$ such that

$$\|z_i - z_j\| \leq \frac{\varepsilon}{2}, \quad \text{for all } i,j \in \left[n; n+g'(n)\right]. \tag{6.6}$$

Since

$$\begin{aligned}\left[n; n + g'(n)\right] &= [n; n + 1 + \Psi(\varepsilon) + g\,(n + 1 + \Psi(\varepsilon))] \\ &\supseteq [n + \Psi(\varepsilon); n + 1 + \Psi(\varepsilon) + g\,(n + 1 + \Psi(\varepsilon))]\,,\end{aligned}$$

we conclude that if we set $n_0 := n + 1 + \Psi(\varepsilon)$, then

$$\|z_{i-1} - z_{j-1}\| \leq \frac{\varepsilon}{2}, \quad \text{for all } i, j \in [n_0; n_0 + g(n_0)]\,.$$

Since $n_0 \geq \Psi(\varepsilon)$, we conclude from (24) of [70] for all $n \geq n_0$, $\|x_n - z_{n-1}\| \leq \frac{\varepsilon}{2(1+L)} \leq \varepsilon/4$, since we may w.l.o.g. assume $L \geq 1$. Thus,

$$\|x_i - x_j\| \leq \|x_i - z_{i-1}\| + \|z_{i-1} - z_{j-1}\| + \|z_{j-1} - x_j\| \leq \varepsilon, \quad \text{for all } i, j \in [n_0; n_0 + g(n_0)]\,.$$

Moreover, we get from Theorem 6.2.7

$$\|x_n - Tx_n\| \leq \varepsilon, \quad \text{for all } n \geq \Psi(\varepsilon).$$

This completes the proof. □

Corollary 6.2.10. *If (θ_n) is nondecreasing, then for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$*

$$\exists n \leq \tilde{g}'^{(\lceil 16d^2/\varepsilon^2 \rceil)}(1) + \Psi(\varepsilon) + 1\, \forall i, j \in [n; n + g\,(n)]\, \forall l \geq n\, (\|x_i - x_j\| \leq \varepsilon \wedge \|Tx_l - x_l\| \leq \varepsilon)$$

where $\tilde{g}'(n) = g'(n) + n$ and $g'(n) = g(n + 1 + \Psi(\varepsilon)) + \Psi\,(\varepsilon) + 1$.

Proof. Since (θ_n) is nondecreasing, the second part of Theorem 6.2.4 implies that there exists an $n \leq \tilde{g}'^{(\lceil 16d^2/\varepsilon^2 \rceil)}(1)$ such that

$$\|z_i - z_j\| \leq \frac{\varepsilon}{2}, \quad \text{for all } i, j \in \left[n; n + g'(n)\right],$$

which is the analog to equation (6.6). The remainder of the proof is then the same. □

As a corollary to the proof of Theorem 6.2.9 we get the following transformation of an assumed rate of metastability for (z_n) into one for (x_n) in general Banach spaces:

Corollary 6.2.11. *In the situation of Theorem 6.2.7 (so X is not necessarily a Hilbert space), suppose that for all $g : \mathbb{N} \to \mathbb{N}$ and $\varepsilon > 0$,*

$$\exists n \leq \Omega\,(d, g, \varepsilon)\, \forall i, j \in [n; n + g(n)]\, (\|z_i - z_j\| \leq \varepsilon)\,,$$

and let $\chi^M(n) := R_1(1/n)$. Then, for all $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$,

$$\begin{aligned}&\exists n \leq \chi^M\,(\Omega(d, g, \varepsilon/2)) + \Psi(\varepsilon) + 1\, \forall i, j \in [n; n + g\,(n)]\, \forall l \geq n \\ &\qquad\qquad\qquad\qquad (\|x_i - x_j\| \leq \varepsilon \wedge \|Tx_l - x_l\| \leq \varepsilon)\,.\end{aligned}$$

Remark 6.2.12. For the canonical choice $\lambda_n = \frac{1}{(n+1)^a}$ and $\theta_n = \frac{1}{(n+1)^b}$, where $0 < b < a$ and $a + b < 1$, the bound is as stated in Corollary 6.2.10.

7 Bruck's Iteration for Demicontinuous Pseudocontractions

7.1 Introduction

We now turn to the quantitative analysis of Bruck's iteration for *demicontinuous* pseudocontractions on Hilbert spaces, i.e. pseudocontractions that are continuous from the strong to the weak topology. For this class of mappings, Bruck's iteration scheme is still suitable. Recall that for a nonempty convex subset C of a real normed space X, a pseudocontractive mapping $T : C \to C$ and real sequences $(\lambda_n), (\theta_n)$ in $[0,1]$ with $\lambda_n(1+\theta_n) \leq 1$ for all $n \in \mathbb{N}$, the *Bruck iteration scheme* with starting point $x_1 \in C$ is defined as

$$x_{n+1} = (1-\lambda_n)\, x_n + \lambda_n T x_n - \lambda_n \theta_n \,(x_n - x_1)\,.$$

Moreover, the sequences (λ_n) and (θ_n) are called acceptably paired if they satisfy the conditions of the following definition:

Definition 7.1.1 ([15]). Two sequences (λ_n) and (θ_n) in $[0,1]$ are *acceptably paired* if (θ_n) is nonincreasing, $\lim_{n\to\infty} \theta_n = 0$ and there exists a strictly increasing sequence $(f(n))_n$ of positive integers such that

1. $\liminf\limits_{n\to\infty} \theta_{f(n)} \cdot \sum\limits_{j=f(n)}^{f(n+1)} \lambda_j > 0$,
2. $\lim\limits_{n\to\infty} \left(\theta_{f(n)} - \theta_{f(n+1)}\right) \cdot \sum\limits_{j=f(n)}^{f(n+1)} \lambda_j = 0$, and
3. $\lim\limits_{n\to\infty} \sum\limits_{j=f(n)}^{f(n+1)} \lambda_j^2 = 0$.

The main focus of this section will be to give a quantitative finitary version of the following theorem:

Theorem 7.1.2 (Bruck [15]). *Let C be a nonempty bounded closed convex subset of a Hilbert space H and $T : C \to C$ be a demicontinuous pseudocontraction. If (λ_n) and (θ_n) are acceptably paired such that $\lambda_n (1+\theta_n) \leq 1$, then, for all $x_1, z \in C$, the sequence (x_n) defined by*

$$x_{n+1} = (1-\lambda_n)\, x_n + \lambda_n T x_n + \lambda_n \theta_n \,(z - x_n)$$

remains in C and converges strongly to the fixed point of T which is closest to z.

The results of this Chapter have been submitted and are currently under peer review [68].

Apart from finitizing the Cauchyness of (x_n) via a rate of metastability, we also need to finitize that the strong limit is indeed a fixed point. If T were norm-continuous, one way to do so would be to ensure that the sequence (x_n) is not only Cauchy along the interval $[n; n+g(n)]$, but also asymptotically regular:

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Phi(\varepsilon, g)\, \forall i, j \in [n; n+g(n)]\, \big(\|x_i - x_j\| < \varepsilon \,\wedge\, \|Tx_i - x_i\| < \varepsilon\big).$$

This was precisely the strategy employed in Theorem 6.2.9. By the logical equivalence of a statement to its Herbrand normal form, this implies both Cauchyness and asymptotic regularity. Cauchyness then implies that the strong limit exists, while norm-continuity and asymptotic regularity recover the fact that the limit is indeed a fixed point.

In the case at hand, however, the operator T is only demicontinuous. In fact, convergence to a fixed point is established via the *continuous* path (z_t) defined by $z_t = tTz_t + (1-t)z$, which – in turn – converges strongly to the fixed point of T closest to z. This gives rise to the following finitary version of Theorem 7.1.2:

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Phi(\varepsilon, g)\, \forall i, j \in [n; n+g(n)] \\ \big(\|x_i - x_j\| < \varepsilon \,\wedge\, \|x_i - y_i\| < \varepsilon \,\wedge\, \|y_i - Ty_i\| < \varepsilon\big), \quad (7.1)$$

where $y_i = z_{1/1+\theta_i}$. Our main theorem (Theorem 7.3.8) provides such a bound. If T is even norm uniformly continuous with modulus ω, then one can obtain a bound Δ such that (see Theorem 7.3.11)

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Delta(g, \varepsilon) \forall i, j \in [n; n+g(n)] \big(\|x_i - x_j\| \leq \varepsilon \wedge \|x_i - Tx_i\| \leq \varepsilon\big).$$

This is a generalization of Theorem 6.2.9, which required T to be Lipschitz continuous. As guaranteed by the general logical metatheorem Theorem 3.2.1, these bounds are again highly uniform in the input data; it is independent of the space and the concrete choices for the operator $T : C \to C$, the set C or the parameter sequences (λ_n) and (θ_n). Apart from the counterfunction g and the accuracy ε, the bounds only depend on an upper bound on the diameter $\mathrm{diam}(C)$, moduli for the quantitative version of acceptably pairedness (cf. Definition 7.3.1) and, in the case of Theorem 7.3.11, the modulus of uniform continuity ω.

Moreover, the new, logically transformed proof of (7.1) is finitary in that it only makes reference to a finite initial segment of (x_n) and totally elementary in that all ideal principles have been eliminated. Moreover, one can recover Bruck's original theorem using only the axiom of choice over *quantifier-free* sentences.

7.2 Analysis of Bruck's Proof

We now examine from a proof-theoretic perspective the steps into which Bruck's proof of Theorem 7.1.2 decomposes. First of all, we need to introduce the generalization of pseudocontractiveness to set-valued operators. $T \subseteq H \times H$ is pseudocontractive, if, for all $(u, x), (v, y) \in T$,

$$\langle x - y, u - v \rangle \leq \|u - v\|^2.$$

Moreover, an operator $U \subset H \times H$ is monotone if and only if $I - U$ is pseudocontractive. It is *maximal* monotone if there does not exist a monotone $U' \subset H \times H$ such that $U \subsetneq U'$.

Bruck's proof then follows the following line of argument:

(i) The monotone operator $U := I - T$ is extended to a maximal monotone, set valued operator $U^* \subset H \times H$.

(ii) There exists a unique y_θ for each $\theta > 0$ for which $0 \in \theta(y_\theta - z) + U^*(y_\theta)$.

(iii) The strong $\lim_{\theta \to 0^+} y_\theta$ exists and is the point x^* of $U^{*-1}(0)$ closest to z.

(iv) The sequence (x_n) also converges to x^*.

(v) The limit is a zero of U, and hence a fixed point of T.

The existence of a maximal monotone extension of a monotone operator $U : H \to H$ makes use of Zorn's Lemma, which is equivalent to the Axiom of Choice. However, we are, for this paper, only interested in the single-valued case. As we have seen in Theorem 6.2.2, the existence of the path $(y_t)_{t\in(0,1]}$ is also guaranteed in case $T : C \to C$ is a single-valued *demicontinuous* pseudocontraction mapping and $C \subset H$ is bounded, closed and convex since the mapping $T_t : C \to C, y \mapsto tTy + (1-t)z$ is t-strongly pseudocontractive for each $t > 0$, and thus has a unique fixed point. Once again, the mere existence of the sequence $(y_{\theta_n})_n$ makes no proof-theoretic contribution since its defining property is a purely universal statement, i.e. one with only $\forall$-quantifiers.

The convergence of (y_{θ_n}) to the fixed point of T closest to z is then carried out analogously to the multi-valued case in Bruck's proof [15]. The quantitative analysis of this step has been performed and a rate of metastability has already been extracted in Chapter 6, see Theorem 6.2.4.

The convergence $\|x_n - x^*\| \to 0$ is established via convergence of the subsequence $\|x_{f(n)} - x^*\| \to 0$, which is shown using the existence of the limit superior as a translation invariant functional $\limsup : \ell_\infty \to \mathbb{R}$ as follows: If $f : \mathbb{N} \to \mathbb{N}$ denotes the subsequence from Definition 7.1.1, then there exists a constant $\gamma \in (0,1)$ such that

$$\begin{aligned}\gamma \cdot \limsup \|x_{f(k)} - x^*\|^2 &= \gamma \cdot \limsup \|x_{f(k)} - y_{\theta_{f(k)}}\|^2 \\ &\geq \limsup \|x_{f(k+1)} - y_{\theta_{f(k)}}\|^2 \qquad (7.2) \\ &= \limsup \|x_{f(k+1)} - x^*\|^2 \\ &= \limsup \|x_{f(k)} - x^*\|^2.\end{aligned}$$

where inequality (7.2) is shown in Bruck's proof. Therefore, $\limsup \|x_{f(k)} - x^*\| = 0$, so the subsequence $(x_{f(k)})$ converges to x^*. Basic arithmetic then implies the convergence of the original sequence.

7.3 Main Results

To obtain a quantitative version of Theorem 7.1.2, we need a quantitative version of what it means for two sequences to be acceptably paired.

Definition 7.3.1. Two sequences (λ_n) and (θ_n) in $[0,1]$ are called acceptably paired with moduli $\varphi_1, \varphi_2, \varphi_3 : \mathbb{R} \to \mathbb{N}$, $f : \mathbb{N} \to \mathbb{N}$, $n_0 \in \mathbb{N}$ and $\delta > 0$ if (θ_n) is nonincreasing and the following conditions are satisfied:

1. $\forall \varepsilon > 0 \forall n \geq \varphi_1(\varepsilon) \, (\theta_n \leq \varepsilon)$,
2. $\forall n (f(n+1) \geq f(n) + 1)$,
3. $\forall n \geq n_0 \left(\theta_{f(n)} \cdot \sum\limits_{j=f(n)}^{f(n+1)} \lambda_j \geq \delta \right)$,
4. $\forall \varepsilon > 0 \forall n \geq \varphi_2(\varepsilon) \left((\theta_{f(n)} - \theta_{f(n+1)}) \cdot \sum\limits_{j=f(n)}^{f(n+1)} \lambda_j \leq \varepsilon \right)$, and
5. $\forall \varepsilon > 0 \forall n \geq \varphi_3(\varepsilon) \left(\sum\limits_{j=f(n)}^{f(n+1)} \lambda_j^2 \leq \varepsilon \right)$.

The moduli φ_i are rates of convergence of their respective sequences to 0. The numbers n_0 and δ are quantitative witnesses for the condition that the sequence $\theta_{f(n)} \cdot \sum_{i=f(n)}^{f(n+1)} \lambda_j$ stays strictly away from 0, i.e. its lim inf is greater than 0. It is also noteworthy that the function $k \mapsto k^* := \max\{n \in \mathbb{N} : f(n) \leq k\}$ is well-defined for all $k \geq f(0)$. Moreover, $(f(k))^* = k$ for all nonnegative integers k.

Remark 7.3.2 ([15]). Examples of acceptably paired sequences are:

1. $\lambda_n = 1/n$, $\theta_n = 1/\log\log n$ and $f(n) = n^n$.
2. For $0 < p < 1$ and $0 < q < \min\{p, 1-p\}$, $\lambda_n = n^{-p}$ and $\theta_n = n^{-q}$ are acceptably paired with $f(n) = \lceil n^{d/(1-p)} \rceil$ for suitable $d > 1$ (see Section 7.4.1 for details).

The corresponding moduli will be given in Section 7.4.

Lemma 7.3.3. *Suppose that X is a normed space and $(a_n) \subseteq X$ is metastable with rate $\Psi : (0,\infty) \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$. Then, for any nondecreasing $f : \mathbb{N} \to \mathbb{N}$ with $f(n) \geq n$, the sequence $(a_{f(n)})$ is metastable with rate $\tilde{\Psi}_f$ defined by $\tilde{\Psi}_f(\varepsilon, g) := \Psi(\varepsilon, g_f)$, where $g_f : \mathbb{N} \to \mathbb{N}$ is defined by $g_f(n) := f(n + g(n)) - n$.*

Proof. Since (a_n) is metastable with modulus Ψ,

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Psi(\varepsilon, g_f) \forall i,j \in [n; f(n+g(n))] \big(\|a_i - a_j\| \leq \varepsilon \big).$$

Since $f(n) \geq n$, we conclude

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Psi(\varepsilon, g_f) \forall i,j \in [f(n); f(n+g(n))] \big(\|a_i - a_j\| \leq \varepsilon \big).$$

The monotonicity of f then implies

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Psi(\varepsilon, g_f) \forall i,j \in [n; n+g(n)] \big(\|a_{f(i)} - a_{f(j)}\| \leq \varepsilon \big),$$

so $\tilde{\Psi}_f$ is a rate of metastability for $(a_{f(n)})$. $\square$

Lemma 7.3.4. *Suppose that $f : \mathbb{N} \to \mathbb{N}$ is strictly increasing and for each k, we have a statement $A(k)$. Define a function $(\cdot)^* : \{n \in \mathbb{N} : n \geq f(0)\} \to \mathbb{N}$ by $k \mapsto \max\{n \in \mathbb{N} : f(n) \leq k\}$. Then for all $g : \mathbb{N} \to \mathbb{N}$*

$$A(k) \text{ for all } k \in [n; n + \tilde{g}(n)] \Rightarrow A(k^*) \text{ for all } k \in [m; m + g(m)],$$

where $\tilde{g}(n) := (f(n) + g(f(n)))^* - n$ and $m := f(n)$.

Proof. Assume the statement $A(k)$ holds for all $k \in [n; n + \tilde{g}(n)]$. Observe that $n + \tilde{g}(n) = \big(m + g(m)\big)^*$ and $n = (f(n))^* = m^*$, so the statement $A(k)$ holds for all $k \in [m^*; (m + g(m))^*]$. Therefore,

$$A(m^*) \wedge A(m^* + 1) \wedge \ldots \wedge A((m + g(m))^*)$$

In particular, $(\cdot)^*$ is nondecreasing (since f is nondecreasing) and so

$$A(m^*) \wedge A((m + 1)^*) \wedge \ldots \wedge A((m + g(m))^*).$$

Therefore, statement $A(k^*)$ holds for all $k \in [m; m + g(m)]$. □

We now give our main results, which were obtained by logical analysis of Bruck's proof [15] using the proof-theoretic methods treated extensively in [49].

Theorem 7.3.5. *Let C be a nonempty bounded closed convex subset of a Hilbert space H with $\mathrm{diam}(C) \leq M \in \mathbb{N}$, $T : C \to C$ be a demicontinuous, single-valued pseudocontraction and $x_1, z \in C$. Suppose the sequences (λ_n) and (θ_n) are acceptably paired with moduli as in Definition 7.3.1 satisfying $\lambda_n (1 + \theta_n) \leq 1$, and the sequence (y_i) defined by*

$$y_i = \frac{1}{1 + \theta_i} T y_i + \frac{\theta_i}{1 + \theta_i} z$$

is metastable with rate $\Psi : (0, \infty) \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$. Define the sequence (x_n) by

$$x_{n+1} = (1 - \lambda_n) x_n + \lambda_n T x_n + \lambda_n \theta_n (z - x_n),$$

and a function Φ by $\Phi(\varepsilon, g, \varphi_1, \varphi_2, \varphi_3, \delta, n_0, M, f) := f(\tilde{\Psi}(\tilde{\varepsilon}, g_d) + n_1 + d + 1)$, where $\varphi_1, \varphi_2, \varphi_3 : \mathbb{R} \to \mathbb{N}$, $f : \mathbb{N} \to \mathbb{N}$, $n_0 \in \mathbb{N}$ and $\delta > 0$ are the moduli of Definition 7.3.1 and

$$\begin{aligned}
&\tilde{\Psi}_f(\varepsilon, g) := \Psi(\varepsilon, g_f), && g_f(n) := f(n + g(n)) - n \\
&g_d(n) := d + n_1 + 1 + \tilde{g}(n + n_1 + d + 1) && \tilde{g}(n) := (f(n) + g(f(n)))^* - n \\
&d := \max\{f(k_0), \lceil \log_c(\varepsilon/8M) \rceil\}, && k^* := \max\{n \in \mathbb{N} : f(n) \leq k\} \\
&n_1 := \max\{n_0, \varphi_1(\delta/2), \varphi_2(\tilde{\varepsilon}/4M^2), \varphi_3(\tilde{\varepsilon}^2/8M^2)\}, && k_0 = \max\{\varphi_2(\varepsilon^2/6M^2), \varphi_3(\varepsilon^2/12M^2)\}, \\
&c := \exp(-\delta/2)m, && \tilde{\varepsilon} = \frac{1 - c}{16} \cdot \varepsilon.
\end{aligned}$$

To simplify notation, we will omit the dependence of Φ on the moduli for the parameters (λ_n) and (θ_n) and instead write $\Phi(\varepsilon, g) := \Phi(\varepsilon, g, \varphi_1, \varphi_2, \varphi_3, \delta, n_0, M, f)$. Then,

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Phi(g, \varepsilon) \forall i, j \in [n; n + g(n)] \left(\|x_i - y_{f(i^*)}\| \leq \varepsilon \wedge \|y_{f(i^*)} - y_{f(j^*)}\| \leq \varepsilon \right)$$

Remark 7.3.6. Observe that the bound given in Theorem 7.3.5 is independent of the operator T and the space H. Moreover, it is also highly uniform with respect to the domain C (dependence only via an upper bound on the diameter $\operatorname{diam} C$) and the choice of the parameter sequences (λ_n) and (θ_n) (dependence only via the moduli $\varphi_1, \varphi_2, \varphi_3$, δ, n_0 and f).

Proof. Since T is pseudocontractive, $U := I - T$ is monotone. Moreover,

$$\begin{aligned} y_i &= \frac{1}{1+\theta_i} T y_i + \frac{\theta_i}{1+\theta_i} z \\ &= \frac{1}{1+\theta_i}(I-U)(y_i) + \frac{\theta_i}{1+\theta_i} z \\ &= \frac{1}{1+\theta_i} y_i - \frac{1}{1+\theta_i} U y_i + \frac{\theta_i}{1+\theta_i} z, \end{aligned}$$

which is equivalent to

$$0 = \left(1 - \frac{1}{1+\theta_i}\right) y_i + \frac{1}{1+\theta_i} U y_i - \frac{\theta_i}{1+\theta_i} z = \frac{\theta_i}{1+\theta_i}(y_i - z) + \frac{1}{1+\theta_i} U y_i,$$

so $\theta_i(y_i - z) + U y_i = 0$. Moreover, the Bruck iteration rewritten in terms of U reads

$$\begin{aligned} x_{n+1} &= (1-\lambda_n)x_n + \lambda_n T x_n + \lambda_n \theta_n (z - x_n) \\ &= x_n - \lambda_n \left(x_n - T x_n + \theta_n (x_n - z)\right) \\ &= x_n - \lambda_n \left(U x_n + \theta_n (x_n - z)\right). \end{aligned}$$

Therefore, for $n > i \geq 2$,

$$x_n - y_i = x_{n-1} - y_i - \lambda_{n-1}(U x_{n-1} + \theta_{n-1}(x_{n-1} - z)),$$

so

$$\begin{aligned} \|x_n - y_i\|^2 &= \Big\langle x_{n-1} - y_i - \lambda_{n-1}(U x_{n-1} + \theta_{n-1}(x_{n-1} - z)), \\ &\qquad\qquad x_{n-1} - y_i - \lambda_{n-1}(U x_{n-1} + \theta_{n-1}(x_{n-1} - z))\Big\rangle \\ &= \|x_{n-1} - y_i\|^2 - 2\lambda_{n-1}\langle x_{n-1} - y_i, U x_{n-1} + \theta_{n-1}(x_{n-1} - z)\rangle \\ &\quad + \lambda_{n-1}^2 \|U x_{n-1} + \theta_{n-1}(x_{n-1} - z)\|^2 \\ &= \|x_{n-1} - y_i\|^2 + \lambda_{n-1}^2 \|U x_{n-1} + \theta_{n-1}(x_{n-1} - z)\|^2 \\ &\quad - 2\lambda_{n-1}\theta_{n-1}\langle x_{n-1} - y_i, x_{n-1} - z\rangle \\ &\quad - 2\lambda_{n-1}\langle x_{n-1} - y_i, U x_{n-1}\rangle \\ &= \|x_{n-1} - y_i\|^2 + \lambda_{n-1}^2 \|U x_{n-1} + \theta_{n-1}(x_{n-1} - z)\|^2 \\ &\quad + 2\lambda_{n-1}(\theta_i - \theta_{n-1})\langle x_{n-1} - y_i, x_{n-1} - z\rangle \\ &\quad - 2\lambda_{n-1}\langle x_{n-1} - y_i, U x_{n-1} + \theta_i(x_{n-1} - z)\rangle. \end{aligned} \tag{7.3}$$

Since U is monotone and $\theta_i(y_i - z) + Uy_i = 0$,

$$\begin{aligned}\langle Ux_{n-1} + \theta_i(x_{n-1} - z), x_{n-1} - y_i\rangle &= \langle Ux_{n-1} + \theta_i(y_i - z), x_{n-1} - y_i\rangle + \theta_i\|x_{n-1} - y_i\|^2 \\ &= \langle Ux_{n-1} - Uy_i, x_{n-1} - y_i\rangle + \theta_i\|x_{n-1} - y_i\|^2 \\ &\geq \theta_i\|x_{n-1} - y_i\|^2.\end{aligned}$$

Equation (7.3) then implies

$$\begin{aligned}\|x_n - y_i\|^2 \leq (1 - 2\lambda_{n-1}\theta_i)\|x_{n-1} - y_i\|^2 + \lambda_{n-1}^2\|Ux_{n-1} + \theta_{n-1}(x_{n-1} - z)\|^2 \\ + 2\lambda_{n-1}(\theta_i - \theta_{n-1})\langle x_{n-1} - z, x_{n-1} - y_i\rangle.\end{aligned}$$

Observe that $\|Ux_{n-1} + \theta_{n-1}(x_{n-1} - z)\| = \|x_{n-1} - Tx_{n-1} + \theta_{n-1}(x_{n-1} - z)\|$. Since $\operatorname{diam}(C) \leq M$, we conclude

$$\|x_n - y_i\|^2 \leq \exp(-2\lambda_{n-1}\theta_i)\|x_{n-1} - y_i\|^2 + 2M^2\lambda_{n-1}(\theta_i - \theta_{n-1}) + 4M^2\lambda_{n-1}^2. \tag{7.4}$$

We show by induction on $n \geq i$ that

$$\|x_n - y_i\|^2 \leq \exp\left(-2\theta_k\sum_{j=i}^{n-1}\lambda_j\right)\|x_i - y_i\|^2 + 2M^2\sum_{j=i}^{n-1}(\theta_i - \theta_j)\lambda_j + 4M^2\sum_{j=i}^{n-1}\lambda_j^2. \tag{7.5}$$

Proof of (7.5): For $n = i$ the inequality holds with equality. Suppose that the inequality holds true for some $n \geq i$. Then (7.4) implies

$$\begin{aligned}\|x_{n+1} - y_i\|^2 &\leq \exp(-2\lambda_n\theta_i)\|x_n - y_i\|^2 + 2M^2\lambda_n(\theta_i - \theta_n) + 4M^2\lambda_n^2 \\ &\leq \exp(-2\lambda_n\theta_i)\cdot\left\{\exp\left(-2\theta_i\sum_{j=i}^{n-1}\lambda_j\right)\|x_i - y_i\|^2\right. \\ &\qquad\left. + 2M^2\sum_{j=i}^{n-1}(\theta_i - \theta_j)\lambda_j + 4M^2\sum_{j=i}^{n-1}\lambda_j^2\right\} \\ &\quad + 2M^2\lambda_n(\theta_i - \theta_n) + 4M^2\lambda_n^2 \\ &= \exp\left(-2\theta_i\sum_{j=i}^{n}\lambda_j\right)\|x_i - y_i\|^2 + 2M^2\sum_{j=i}^{n}(\theta_i - \theta_j)\lambda_j + 4M^2\sum_{j=i}^{n}\lambda_j^2,\end{aligned}$$

which is what we needed to show.

Since $\theta_i - \theta_j \leq \theta_i - \theta_n$ for $i \leq j \leq n$, (7.5) implies

$$\|x_n - y_i\|^2 \leq \exp\left(-2\theta_i\sum_{j=i}^{n-1}\lambda_j\right)\|x_i - y_i\|^2 + 2M^2(\theta_i - \theta_n)\sum_{j=i}^{n-1}\lambda_j + 4M^2\sum_{j=i}^{n-1}\lambda_j^2, \text{ for all } n \geq i. \tag{7.6}$$

Now let $f(n)$ be the subsequence of Definition 7.1.1. We now prove that $(x_{f(n)})$ is Cauchy. Taking $i = f(k)$ and $n = f(k+1)$ in (7.6), we get

$$\begin{aligned}\|x_{f(k+1)} - y_{f(k)}\|^2 \le \exp\left(-2\theta_{f(k)} \sum_{j=f(k)}^{f(k+1)} \lambda_j\right) \cdot \exp(2\theta_{f(k)}\lambda_{f(k+1)}) \cdot \|x_{f(k)} - y_{f(k)}\|^2 \\ + 2M^2(\theta_{f(k)} - \theta_{f(k+1)}) \cdot \sum_{j=f(k)}^{f(k+1)} \lambda_j + 4M^2 \sum_{j=f(k)}^{f(k+1)} \lambda_j^2. \qquad (7.7)\end{aligned}$$

Now observe that

$$\exp\left(-2\theta_{f(k)} \sum_{j=f(k)}^{f(k+1)} \lambda_j\right) \le \exp(-2\delta) < 1, \quad \text{for all } k \ge n_0.$$

Moreover, (θ_n) is a null sequence with modulus φ_1. Thus, $\exp(2\theta_{f(k)}\lambda_{f(k+1)}) \le \exp(2\theta_k) \le \exp(\delta)$ for all $k \ge \varphi_1(\delta/2)$. Furthermore, for all $k \ge \max\{\varphi_2(\tilde{\varepsilon}^2/4M^2), \varphi_3(\tilde{\varepsilon}^2/8M^2)\}$, the remainder term in (7.7) is less than $\tilde{\varepsilon}^2$. In total,

$$\|x_{f(k+1)} - y_{f(k)}\|^2 \le \exp(-\delta) \cdot \|x_{f(k)} - y_{f(k)}\|^2 + \tilde{\varepsilon}^2, \quad \text{for all } k \ge n_1$$

since $n_1 = \max\{n_0, \varphi_1(\delta/2), \varphi_2(\tilde{\varepsilon}^2/4M^2), \varphi_3(\tilde{\varepsilon}^2/2M^2)\}$. Because $c = \exp(-\delta/2)$, we then get

$$\begin{aligned}\|x_{f(k+1)} - y_{f(k)}\| &\le c \cdot \|x_{f(k)} - y_{f(k)}\| + \tilde{\varepsilon} \\ &\le c \cdot \|x_{f(k)} - y_{f(k-1)}\| + c \cdot \|y_{f(k-1)} - y_{f(k)}\| + \tilde{\varepsilon}. \qquad (7.8)\end{aligned}$$

Now observe that since (y_n) is metastable with rate Ψ, the subsequence $(y_{f(n)})$ is metastable with rate $\tilde{\Psi}$ by Lemma 7.3.3. Thus, there exists an integer $n \le \tilde{\Psi}(\tilde{\varepsilon}, g_d)$ such that $\|y_{f(k)} - y_{f(j)}\| \le \tilde{\varepsilon}$ for all $k, j \in [n; n + g_d(n)]$. Taking $n_2 := n + n_1$, we have on the one hand $n_2 \ge n_1$, and $\|y_{f(k)} - y_{f(j)}\| \le \tilde{\varepsilon}$ for all $k, j \in [n_2; n_2 + d + 1 + \tilde{g}(n_2 + d + 1)]$ on the other. Setting $j = k - 1$, we conclude

$$\|y_{f(k)} - y_{f(k-1)}\| \le \tilde{\varepsilon}, \quad \text{for all } k \in [n_2 + 1; n_2 + d + 1 + \tilde{g}(n_2 + d + 1)]. \qquad (7.9)$$

Suppose now that $k \in [n_2 + d; n_2 + d + \tilde{g}(n_2 + d + 1)]$. Then (7.9) and (7.8) yield

$$\begin{aligned}
\|x_{f(k+1)} - y_{f(k)}\| &\leq c \cdot \|x_{f(k)} - y_{f(k-1)}\| + 2\tilde{\varepsilon} \\
&\leq c \cdot (c \cdot \|x_{f(k-1)} - y_{f(k-2)}\| + 2\tilde{\varepsilon}) + 2\tilde{\varepsilon} \\
&= c^2 \cdot \|x_{f(k-1)} - y_{f(k-2)}\| + 2\tilde{\varepsilon} \cdot c + 2\tilde{\varepsilon} \\
&\leq \dots \\
&= c^{d-1} \cdot \|x_{f(k-d+2)} - y_{f(k-d+1)}\| + 2\tilde{\varepsilon} \sum_{k=0}^{d-2} c^k \\
&\leq c^d \cdot \|x_{f(k-d+1)} - y_{f(k-d)}\| + c^{d-1} \cdot \|y_{f(k-d+1)} - y_{f(k-d)}\| + 2\tilde{\varepsilon} \sum_{k=0}^{d-2} c^k \\
&\leq c^d \cdot M + 2\tilde{\varepsilon} \sum_{k=0}^{d-1} c^k \\
&\leq c^d \cdot M + 2\tilde{\varepsilon} \sum_{k=0}^{\infty} c^k \\
&= c^d \cdot M + \frac{2\tilde{\varepsilon}}{1-c}.
\end{aligned}$$

Since $d \geq \log_c(\varepsilon/8M)$ and $\tilde{\varepsilon} = \frac{1-c}{16} \cdot \varepsilon$, we have

$$\|x_{f(k+1)} - y_{f(k)}\| \leq \frac{\varepsilon}{4}, \text{ for all } k \in [n_2 + d; n_2 + d + \tilde{g}(n_2 + d + 1)].$$

Therefore, setting $n_3 := n_2 + d + 1$ and using (7.9),

$$\begin{aligned}
\|x_{f(k)} - y_{f(k)}\| \leq \|x_{f(k)} - y_{f(k-1)}\| + \|y_{f(k)} - y_{f(k-1)}\| &\leq \frac{\varepsilon}{4} + \tilde{\varepsilon} \\
&< \frac{\varepsilon}{3}, \quad \text{for all } k \in [n_3; n_3 + \tilde{g}(n_3)].
\end{aligned}$$

By Lemma 7.3.4

$$\|x_{f(k^*)} - y_{f(k^*)}\| \leq \varepsilon/3, \text{ for all } k \in [f(n_3); f(n_3) + g(f(n_3))] \tag{7.10}$$

Now, for $k \geq f(0)$, observe that k^* denotes the unique integer such that $f(k^*) \leq k < f(k^*+1)$. Take $n = k$, $i = f(k^*)$ in (7.6); since the exponential factor is less than or equal to 1,

$$\begin{aligned}
\|x_k - y_{f(k^*)}\|^2 &\leq \|x_{f(k^*)} - y_{f(k^*)}\|^2 + 2M^2(\theta_{f(k^*)} - \theta_k) \sum_{j=f(k^*)}^{k-1} \lambda_j + 4M^2 \sum_{j=f(k^*)}^{k-1} \lambda_j^2 \\
&\leq \|x_{f(k^*)} - y_{f(k^*)}\|^2 + 2M^2(\theta_{f(k^*)} - \theta_{f(k^*+1)}) \sum_{j=f(k^*)}^{f(k^*+1)} \lambda_j + 4M^2 \sum_{j=f(k^*)}^{f(k^*+1)} \lambda_j^2.
\end{aligned}$$

Observe that the latter two terms become less than $\varepsilon^2/3$ whenever $k^* \geq k_0$ since, by definition, $k_0 = \max\{\varphi_2(\varepsilon^2/6M^2), \varphi_3(\varepsilon^2/12M^2)\}$. But this is always the case whenever $k \geq f(k_0)$ since then $k^* \geq (f(k_0))^* = k_0$ by the monotonicity of $(\cdot)^*$. Therefore,

$$\|x_k - y_{f(k^*)}\|^2 \leq \|x_{f(k^*)} - y_{f(k^*)}\|^2 + \frac{2\varepsilon^2}{3}, \text{ for all } k \geq f(k_0). \tag{7.11}$$

Since $f(n_3) = f(n_2 + d + 1) \geq f(d) \geq d \geq f(k_0)$, equations (7.10), (7.11) together imply

$$\|x_k - y_{f(k^*)}\| \leq \varepsilon, \quad \text{for all } k \in [f(n_3); f(n_3) + g(f(n_3))]. \tag{7.12}$$

Now recall that $\|y_{f(i)} - y_{f(j)}\| \leq \tilde{\varepsilon}$ for all $i, j \in [n_2; n_2 + d + 1 + \tilde{g}(n_2 + d + 1)]$. Again by Lemma 7.3.4, this implies

$$\|y_{f(i^*)} - y_{f(j^*)}\| \leq \tilde{\varepsilon} \leq \varepsilon, \text{ for all } i, j \in [f(n_3), f(n_3) + g(f(n_3))]. \tag{7.13}$$

Therefore, $f(n_3) = f(n_2 + d + 1) \leq f(\tilde{\Psi}(g_d, \tilde{\varepsilon}) + n_1 + d + 1)$ satisfies the claim. □

Theorem 7.3.7. *In the situation of Theorem 7.3.5, (x_n) is metastable with rate $\Phi'(\varepsilon, g) := \Phi(\varepsilon/3, g)$.*

Proof. Since $\tilde{\varepsilon} < \varepsilon$ and $\|x_i - x_j\| \leq \|x_i - y_{f(i^*)}\| + \|x_j - y_{f(j^*)}\| + \|y_{f(i^*)} - y_{f(j^*)}\|$, equations (7.12) and (7.13) imply

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Phi(\varepsilon/3, g) \forall i, j \in [n; n + g(n)] \left(\|x_i - x_j\| < \varepsilon\right),$$

which is what we needed to show. □

Theorem 7.3.8. *In the situation of Theorem 7.3.5,*

$$\forall \varepsilon > 0\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists n \leq \Phi''(\varepsilon, g)\, \forall i, j \in [n; n + g(n)]$$
$$\big(\|x_i - x_j\| < \varepsilon \wedge \|x_i - y_i\| < \varepsilon \wedge \|y_i - Ty_i\| < \varepsilon\big),$$

where $\Phi''(\varepsilon, g) := \hat{\Phi}(\varepsilon/3,)$, and $\hat{\Phi}$ is defined like Φ in Theorem 7.3.5, but with $\hat{g}_d$, defined by

$$\hat{g}_d(n) := d + \hat{n}_1 + 1 + \hat{g}(n + d + \hat{n}_1 + 1).$$

instead of g_d, where $\hat{g}(n) := (f(n) + g(f(n)))^* - n + 1$ and $\hat{n}_1 := \max\{n_1, \varphi_1(\varepsilon/M)\}$.

Moreover, we can take $\Psi(\varepsilon, g) := \tilde{g}^{(\lceil 16d^2/\varepsilon^2 \rceil)}(1)$, where $\tilde{g}(n) = n + 1 + g(n + 1)$.

Proof. By altering the definition of g_d, the point $f(n_3)$ that satisfies the conclusion of Theorem 7.3.7 also satisfies $f(n_3) \geq n_3 = n_2 + d + 1 = n_0 + \hat{n}_1 + d + 1 \geq \varphi_1(\varepsilon, g)$. Therefore, (x_n) is metastable, and

$$\|y_i - Ty_i\| = \frac{\theta_i}{1 + \theta_i}\|Ty_i - z\| \leq \theta_i \cdot M \leq \varepsilon, \text{ for all } i \geq n.$$

It remains to verify that $\|x_i - y_i\| \leq \varepsilon$ on $[f(n_3); f(n_3) + g(f(n_3))]$. To this end, observe that

$$\|y_i - y_j\| \leq \varepsilon/3, \text{ for all } i, j \in [f(n_3); f(n_3 + \hat{g}(n_3))]. \tag{7.14}$$

Since $f(k^*+1) > k \geq f(k^*)$ for all $k \geq f(0)$, we conclude $f(n_3 + \hat{g}(n_3)) = f((f(n_3) + g(f(n_3)))^* + 1) \geq f(n_3) + g(f(n_3))$. Moreover, $f((f(n_3))^*) = f(n_3)$. Therefore, (7.14) implies

$$\|y_k - y_{f(k^*)}\| \leq \varepsilon/3, \text{ for all } k \in [f(n_3), f(n_3) + g(f(n_3))],$$

which, using (7.12), implies for all $k \in [f(n_3), f(n_3) + g(f(n_3))]$

$$\|x_k - y_k\| \leq \|x_k - y_{f(k^*)}\| + \|y_{f(k^*)} - y_k\| \leq \varepsilon/3 + \varepsilon/3 < \varepsilon.$$

That we may choose $\Psi(\varepsilon, g) := \tilde{g}^{(\lceil 16d^2/\varepsilon^2 \rceil)}(1)$ follows from Theorem 6.2.4. □

Remark 7.3.9. Observe that Theorems 7.3.5, 7.3.7 and 7.3.8 require only the demicontinuity of T. Therefore, model-theoretic approaches (cf. [40]) are not applicable, as these always require norm-continuity.

Remark 7.3.10. Suppose Ψ does not depend on g for a concrete choice of the input. Then metastability for (y_n) would read

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Psi(\varepsilon) \, \forall i, j \in [n; n + g(n)] \, \big(\|x_i - x_j\| < \varepsilon\big).$$

This is logically equivalent to

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Psi(\varepsilon) \, \forall i, j \geq n \, \big(\|x_i - x_j\| < \varepsilon\big),$$

i.e. a rate of convergence. In this case, we would get in Theorem 7.3.5 a rate of convergence $\Phi(\varepsilon) := f(\Psi(\tilde{\varepsilon}) + n_1 + d + 1)$, where, as before

$$n_1 := \max\{n_0, \varphi_1(\delta/2), \varphi_2(\tilde{\varepsilon}/4M^2), \varphi_3(\tilde{\varepsilon}^2/8M^2)\}, \quad k_0 = \max\{\varphi_2(\varepsilon^2/6M^2), \varphi_3(\varepsilon^2/12M^2)\},$$

$$d := \max\{f(k_0), \lceil \log_c(\varepsilon/8M) \rceil\}, \qquad \tilde{\varepsilon} = \frac{1-c}{16} \cdot \varepsilon,$$

$$c := \exp(-\delta/2).$$

Theorem 7.3.11. *Suppose that in the situation of Theorem 7.3.5, T is additionally uniformly continuous on C with modulus ω. For $g : \mathbb{N} \to \mathbb{N}$ define $g_b(n) := b + g(n+b)$, where $b := f((\varphi_1(\varepsilon/3M))^* + 1)$. Then*

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leq \Delta(g, \varepsilon) \forall i, j \in [n; n + g(n)] \big(\|x_i - x_j\| \leq \varepsilon \wedge \|x_i - Tx_i\| \leq \varepsilon\big),$$

where $\Delta(g, \varepsilon) := \Phi(g_b, \min\{\varepsilon/3, \omega(\varepsilon/3M)\}) + b$.

Proof. By Theorem 7.3.5, there exists a $k \leq \Phi(g_b, \min\{\varepsilon/3, \omega(\varepsilon/3M)\})$ such that for $n := k + b$,

$$\|x_i - y_{f(i^*)}\| \leq \min\{\varepsilon/3, \omega(\varepsilon/3M)\}, \text{ for all } i, j \in [n; n + g(n)].$$

Now observe that $b^* = (\varphi_1(\varepsilon/3M))^* + 1$ since $(f(k))^* = k$ for all nonnegative integers k. Therefore, $f(n^*) \geq f(b^*) = f((\varphi_1(\varepsilon/3M))^* + 1) > \varphi_1(\varepsilon/3M)$. Thus, $\theta_{f(i^*)} \leq \varepsilon/3M$ for all $i \geq n$. Consequently

$$\|y_{f(i^*)} - Ty_{f(i^*)}\| = \frac{\theta_{f(i^*)}}{1 + \theta_{f(i^*)}} \|Ty_{f(i^*)} - z\| \leq \theta_{f(i^*)} \cdot M \leq \frac{\varepsilon}{3}, \text{ for all } i \geq n.$$

Therefore,

$$\|x_i - Tx_i\| \leq \|x_i - y_{f(i^*)}\| + \|y_{f(i^*)} - Ty_{f(i^*)}\| + \|Tx_i - Ty_{f(i^*)}\| \leq \varepsilon, \text{ for all } i \in [n; n+g(n)].$$

That $\|x_i - x_j\| \leq \varepsilon$ on $[n; n+g(n)]$ follows as in Corollary 7.3.7. $\square$

7.4 Application to Concrete Instances

In this section, we compute explicitly the moduli $\varphi_1, \varphi_2, \varphi_3, n_0$ and δ for the two examples of parameter sequences of Remark 7.3.2. We then compare the bound to the one obtained in [52] for Halpern iterations of nonexpansive mappings.

7.4.1 Example 1

Suppose p and q are real numbers in $(0,1)$ such that $0 < q < \min\{p, 1-p\}$, and take $\lambda_n := n^{-p}$ and $\theta_n := n^{-q}$. Set $r = (p+q)/2$. There are two cases to consider, namely $p \geq 1/2$ and $p < 1/2$.

If $p < 1/2$, then $1/2 > p > r > q$, so we conclude $1/2 < 1-p < 1-r < 1-q$, whence $0 < 1-2p < 1-r-p < 1-p-q$. Then,

$$\frac{1-2p}{1-p} < \frac{1-r-p}{1-p} < \frac{1-p-q}{1-p} = 1 - \frac{q}{1-p},$$

and so

$$\frac{1-p}{1-2p} > \frac{1-p}{1-r-p} > \left(1 - \frac{q}{1-p}\right)^{-1}.$$

Thus, if we choose $d := \min\left\{\frac{1-p}{1-r-p}, \frac{3}{2}\left(1 - \frac{q}{1-p}\right)^{-1}\right\}$, then

$$\left(1 - \frac{q}{1-p}\right)^{-1} < d < \min\left\{\frac{1-p}{1-2p}, 2\left(1 - \frac{q}{1-p}\right)^{-1}\right\} \quad (\text{for } p < 1/2),$$

which is Bruck's condition. For $p \geq \frac{1}{2}$, we see as before that

$$\left(1 - \frac{q}{1-p}\right)^{-1} < d < 2\left(1 - \frac{q}{1-p}\right)^{-1} \quad (\text{for } p \geq 1/2).$$

An important consequence of our choice of d is that $d > 1$, which we will use throughout this section.

Now, one can take $f(n) := \lceil n^{d/(1-p)} \rceil$. To calculate the other moduli, we need the following Lemma, which is a direct consequence of Taylor's Theorem using the Lagrange remainder term.

Lemma 7.4.1. *Suppose $x, r \in \mathbb{R}$ with $x > 0$ and $r \neq 1$. Then,*

(i) there exists a real number $\xi \in (x, x+1)$ such that $(x+1)^r = x^r + r\xi^{r-1}$, and

(ii) there exists a real number $\nu \in (x-1, x)$ *such that* $(x-1)^r = x^r - r\nu^{r-1}$.

We now proceed to calculate the moduli. Observe that

$$\begin{aligned}
\theta_{f(n)} \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j &= \left\lceil n^{d/(1-p)} \right\rceil^{-q} \cdot \sum_{j=f(n)}^{f(n+1)} j^{-p} \\
&> (n^{d/(1-p)}+1)^{-q} \cdot \int_{f(n)}^{f(n+1)} j^{-p} dj \\
&= \frac{(n^{d/(1-p)}+1)^{-q}}{1-p} \left[(f(n+1))^{1-p} - (f(n))^{1-p} \right] \\
&= \frac{(n^{d/(1-p)}+1)^{-q}}{1-p} \left(\left\lceil (n+1)^{d/(1-p)} \right\rceil^{1-p} - \left\lceil n^{d/(1-p)} \right\rceil^{1-p} \right) \\
&\geq \frac{(n^{d/(1-p)}+1)^{-q}}{1-p} \left[\left((n+1)^{d/(1-p)} \right)^{1-p} - \left(n^{d/(1-p)} + 1 \right)^{1-p} \right] \\
&= \frac{(n^{d/(1-p)}+1)^{-q}}{1-p} \left[(n+1)^d - \left(n^{d/(1-p)} + 1 \right)^{1-p} \right]
\end{aligned} \tag{7.15}$$

By virtue of 7.4.1, there exists a $\xi \in (n^{d/(1-p)}, n^{d/(1-p)}+1)$ such that

$$\begin{aligned}
\left(n^{d/(1-p)} + 1 \right)^{1-p} &= n^d + (1-p)\xi^{-p} \\
&\leq n^d + (1-p) n^{-\frac{dp}{1-p}}.
\end{aligned}$$

Therefore, applying Lemma 7.4.1, there exists a $\xi \in (n, n+1)$ such that

$$\begin{aligned}
(n+1)^d - \left(n^{d/(1-p)} + 1 \right)^{1-p} &\geq (n+1)^d - n^d - (1-p) n^{-\frac{dp}{1-p}} \\
&= n^d + d\xi^{d-1} - n^d - (1-p) n^{-\frac{dp}{1-p}} \\
&\geq n^d + dn^{d-1} - n^d - (1-p) n^{-\frac{dp}{1-p}} \\
&= dn^{d-1} - (1-p) n^{-\frac{dp}{1-p}}.
\end{aligned}$$

Consequently, going back to (7.15),

$$\theta_{f(n)} \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j > \frac{(n^{d/(1-p)}+1)^{-q}}{1-p} \left(dn^{d-1} - (1-p) n^{-\frac{dp}{1-p}} \right).$$

By Lemma 7.4.1, there now exists $\xi \in (n^{d/(1-p)}, n^{d/(1-p)}+1)$ such that

$$\begin{aligned}\theta_{f(n)} \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j &> \frac{n^{\frac{-dq}{1-p}} - q\xi^{-q-1}}{1-p}\left(dn^{d-1} - (1-p)n^{-\frac{dp}{1-p}}\right) \\ &\geq \frac{n^{\frac{-dq}{1-p}} - qn^{-\frac{d(q+1)}{1-p}}}{1-p}\left(dn^{d-1} - (1-p)n^{-\frac{dp}{1-p}}\right) \\ &\geq \frac{d}{1-p}n^{d-1-\frac{dq}{1-p}} - n^{-\frac{d(p+q)}{1-p}} - \frac{dq}{1-p}n^{d-1-\frac{d(q+1)}{1-p}}.\end{aligned}$$

Now observe that $d-1-\frac{dq}{1-p} = d(1-\frac{q}{1-p}) - 1 > \frac{d}{d} - 1 = 0$ and $d - 1 - \frac{d(q+1)}{1-p} = d(1-\frac{q}{1-p}) - \frac{d}{1-p} - 1 \leq 3/2 - 2 = -1/2$. Moreover, $-\frac{d(p+q)}{1-p} < 0$, so the right-hand-side in the equation above is monotone increasing. Therefore,

$$\theta_{f(n)} \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j > \frac{d}{1-p} - 1 - \frac{dq}{1-p} = \frac{1-q}{1-p} \cdot d - 1, \quad \text{for all } n \geq 1.$$

Since $q < p$, we have $1-q > 1-p$. Moreover, $d > 1$. Thus, we may choose $n_0 := 1$ and $\delta := \frac{d(1-q)}{1-p} - 1 > 0$.

We now calculate the modulus φ_2. By Lemma 7.4.1, there exists a real number $\xi \in \left((n+1)^{d/(1-p)}, (n+1)^{d/(1-p)} + 1\right)$ such that

$$\begin{aligned}\theta_{f(n)} - \theta_{f(n+1)} &= \left\lceil n^{d/(1-p)} \right\rceil^{-q} - \left\lceil (n+1)^{d/(1-p)} \right\rceil^{-q} \\ &\leq n^{-\frac{dq}{1-p}} - \left((n+1)^{d/(1-p)} + 1\right)^{-q} \\ &= n^{-\frac{dq}{1-p}} - \left((n+1)^{-\frac{dq}{1-p}} - q \cdot \xi^{-q-1}\right) \\ &\leq n^{-\frac{dq}{1-p}} - (n+1)^{-\frac{dq}{1-p}} + q \cdot \left((n+1)^{d/(1-p)}\right)^{-q-1} \\ &= n^{-\frac{dq}{1-p}} - (n+1)^{-\frac{dq}{1-p}} + q \cdot (n+1)^{-\frac{d(1+q)}{1-p}} \\ &\leq n^{-\frac{dq}{1-p}} - (n+1)^{-\frac{dq}{1-p}} + q \cdot n^{-\frac{d(1+q)}{1-p}}.\end{aligned}$$

Applying once more Lemma 7.4.1, we see that for some $\xi \in (n, n+1)$,

$$\begin{aligned}\theta_{f(n)} - \theta_{f(n+1)} &= n^{-\frac{dq}{1-p}} - \left(n^{-\frac{dq}{1-p}} - \frac{dq}{1-p}\xi^{-\frac{dq}{1-p}-1}\right) + q \cdot n^{-\frac{d(1+q)}{1-p}} \\ &= \frac{dq}{1-p}\xi^{-\frac{dq}{1-p}-1} + q \cdot n^{-\frac{d(1+q)}{1-p}} \\ &\leq \frac{dq}{1-p}n^{-\frac{dq}{1-p}-1} + q \cdot n^{-\frac{d(1+q)}{1-p}}.\end{aligned}$$

Since $-\frac{d(1+q)}{1-p} = -\frac{dq}{1-p} - \frac{d}{1-p} \leq -\frac{dq}{1-p} - 1$,

$$\theta_{f(n)} - \theta_{f(n+1)} \leq \left(\frac{dq}{1-p} + q\right) n^{-\frac{dq}{1-p}-1} = \frac{q(d+1-p)}{1-p} n^{-\frac{dq}{1-p}-1}. \tag{7.16}$$

On the other hand,

$$\begin{aligned}\sum_{j=f(n)}^{f(n+1)} \lambda_j = \sum_{j=f(n)}^{f(n+1)} j^{-p} &\le \int_{f(n)-1}^{f(n+1)-1} j^{-p} dj \\ &= \frac{1}{1-p}\left(\left\lceil (n+1)^{d/(1-p)} - 1\right\rceil^{1-p} - \left\lceil n^{d/(1-p)} - 1\right\rceil^{1-p}\right).\end{aligned}$$

Since $1-p>0$, we conclude

$$\sum_{j=f(n)}^{f(n+1)} \lambda_j \le \frac{1}{1-p}\left((n+1)^d - \left(n^{d/(1-p)} - 1\right)^{1-p}\right).$$

By Lemma 7.4.1, there exists $\xi \in (n, n+1)$ and $\nu \in (n^{d/(1-p)}-1, n^{d/(1-p)})$ such that for $n \ge 2$

$$\begin{aligned}(1-p)\cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j &\le n^d + d\xi^{d-1} - \left(n^{d/(1-p)} - 1\right)^{1-p} \\ &\le n^d + d(n+1)^{d-1} - \left(n^{d/(1-p)} - 1\right)^{1-p} \\ &= n^d + d(n+1)^{d-1} - n^d + (1-p)\nu^{-p} \\ &\le n^d + d(n+1)^{d-1} - n^d + (1-p)\left(n^{d/(1-p)} - 1\right)^{-p} \\ &\le d(n+1)^{d-1} + (1-p)\cdot 2^p n^{-\frac{dp}{1-p}} \\ &\le d\cdot 2^{d-1} n^{d-1} + (1-p)\cdot 2^p n^{-\frac{dp}{1-p}}.\end{aligned}$$

Combining this with (7.16), we get

$$\begin{aligned}\left(\theta_{f(n)} - \theta_{f(n+1)}\right)\sum_{j=f(n)}^{f(n+1)} \lambda_j &\le \left(\frac{q(d+1-p)}{1-p} n^{-\frac{dq}{1-p}-1}\right)\frac{d\cdot 2^{d-1}n^{d-1} + (1-p)\cdot 2^p n^{-\frac{dp}{1-p}}}{1-p} \\ &= \frac{2^{d-1}dq(d+1-p)}{(1-p)^2} n^{d-2-\frac{dq}{1-p}} + \frac{2^p q(d+1-p)}{1-p} n^{-\frac{d(p+q)}{1-p}-1}.\end{aligned}$$

Now observe that $d-2-\frac{dq}{1-p} = d(1-\frac{q}{1-p}) - 2 \le \frac{3}{2}(1-\frac{q}{1-p})^{-1}(1-\frac{q}{1-p}) - 2 = -\frac{1}{2}$. Therefore,

$$\begin{aligned}\left(\theta_{f(n)} - \theta_{f(n+1)}\right)\sum_{j=f(n)}^{f(n+1)} \lambda_j &\le \frac{2^{d-1}dq(d+1-p)}{(1-p)^2} n^{-\frac{1}{2}} + \frac{2^p q(d+1-p)}{1-p} n^{-\frac{d(p+q)}{1-p}-1} \\ &\le \frac{2^{d-1}dq(d+1-p)}{(1-p)^2} n^{-\frac{1}{2}} + \frac{2^p q(d+1-p)}{(1-p)^2} n^{-\frac{1}{2}} \\ &\le \frac{2^d dq(d+1-p)}{(1-p)^2} n^{-\frac{1}{2}}.\end{aligned}$$

Therefore, we may take $\varphi_2(\varepsilon) := \left\lceil \left(\frac{2^d dq(d+1-p)}{\varepsilon(1-p)^2} \right)^2 \right\rceil$.

Now, we calculate the modulus φ_3. To this end, observe that

$$\sum_{j=f(n)}^{f(n+1)} \lambda_j^2 = \sum_{j=f(n)}^{f(n+1)} j^{-2p} \leq \int_{j=f(n)-1}^{f(n+1)-1} j^{-2p} dj$$
$$= \begin{cases} \frac{1}{1-2p} \left(\left\lceil (n+1)^{d/(1-p)} - 1 \right\rceil^{1-2p} - \left\lceil n^{d/(1-p)} - 1 \right\rceil^{1-2p} \right) & \text{, if } p \neq \frac{1}{2}, \\ \log \left\lceil (n+1)^{d/(1-p)} - 1 \right\rceil - \log \left\lceil n^{d/(1-p)} - 1 \right\rceil & \text{, if } p = \frac{1}{2}. \end{cases}$$

We have to distinguish the cases $p > 1/2$, $p = 1/2$ and $p < 1/2$. For $p > 1/2$, we use the estimate

$$\sum_{j=f(n)}^{f(n+1)} \lambda_j^2 \leq \frac{\left(n^{d/(1-p)} - 1\right)^{1-2p}}{2p-1} \leq \varepsilon, \quad \text{for all } n \geq \left(\left((2p-1)\varepsilon\right)^{1/(1-2p)} + 1 \right)^{(1-p)/d}.$$

For $p = 1/2$, we see that

$$\sum_{j=f(n)}^{f(n+1)} \lambda_j^2 \leq \log \frac{(n+1)^{d/(1-p)}}{\left\lceil n^{d/(1-p)} - 1 \right\rceil} \leq \log \frac{(n+1)^{d/(1-p)}}{n^{d/(1-p)} - 1}.$$

Now observe that $(n+1)^{d/(1-p)} = n^{d/(1-p)} + \frac{d}{1-p} \xi^{\frac{d}{1-p}-1}$ for some $\xi \in (n, n+1)$. Therefore,

$$\sum_{j=f(n)}^{f(n+1)} \lambda_j^2 \leq \log \left(\frac{n^{d/(1-p)} + \frac{d}{1-p}(n+1)^{\frac{d}{1-p}-1}}{n^{d/(1-p)} - 1} \right)$$
$$= \log \left(1 + \frac{\frac{d}{1-p}(n+1)^{\frac{d}{1-p}-1} + 1}{n^{d/(1-p)} - 1} \right)$$
$$\leq \log \left(1 + 2^{d/(1-p)} \cdot 2^{d/(1-p)} \cdot \frac{\frac{d}{1-p} n^{\frac{d}{1-p}-1}}{n^{d/(1-p)}} \right)$$
$$= \log \left(1 + \frac{d \cdot 2^{\frac{2d}{1-p}}}{1-p} \cdot \frac{1}{n} \right) \leq \varepsilon, \quad \text{for all } n \geq \frac{d \cdot 2^{\frac{2d}{1-p}}}{(1-p)(\exp(\varepsilon) - 1)}.$$

For $p < 1/2$, we see that there exists $\nu \in (n^{d/(1-p)} - 1, n^{d/(1-p)})$ and $\xi \in (n, n+1)$ such that

$$
\begin{aligned}
(1-2p)\cdot\sum_{j=f(n)}^{f(n+1)}\lambda_j^2 &= \left\lceil (n+1)^{d/(1-p)}-1\right\rceil^{1-2p} - \left\lceil n^{d/(1-p)}-1\right\rceil^{1-2p} \\
&\le (n+1)^{\frac{d(1-2p)}{1-p}} - \left(n^{d/(1-p)}-1\right)^{1-2p} \\
&= (n+1)^{\frac{d(1-2p)}{1-p}} - n^{\frac{d(1-2p)}{1-p}} + (1-2p)\nu^{-2p} \\
&\le (n+1)^{\frac{d(1-2p)}{1-p}} - n^{\frac{d(1-2p)}{1-p}} + (1-2p)(n^{d/(1-p)}-1)^{-2p} \\
&= n^{\frac{d(1-2p)}{1-p}} + \frac{d(1-2p)}{1-p}\xi^{\frac{d(1-2p)}{1-p}-1} - n^{\frac{d(1-2p)}{1-p}} + (1-2p)(n^{d/(1-p)}-1)^{-2p} \\
&\le \frac{d(1-2p)}{1-p} n^{\frac{d(1-2p)}{1-p}-1} + (1-2p)(n^{d/(1-p)}-1)^{-2p}
\end{aligned}
$$

Observe that $\frac{d(1-2p)-1+p}{1-p} = \frac{d(1-2p)}{1-p} - 1 < 0$ since $d < \frac{1-p}{1-2p}$. Therefore,

$$
\begin{aligned}
\sum_{j=f(n)}^{f(n+1)}\lambda_j^2 &\le \frac{d}{1-p} n^{\frac{d(1-2p)}{1-p}-1} + (n^{d/(1-p)}-1)^{-2p} \\
&\le \varepsilon, \quad \text{for all } n \ge \max\left\{ \left(\frac{2d}{(1-p)\varepsilon}\right)^{\frac{1-p}{d(1-2p)-1+p}}, \left((2/\varepsilon)^{\frac{1}{2p}}+1\right)^{(1-p)/d} \right\}.
\end{aligned}
$$

Observing that $\theta_n = n^{-q}$ converges to 0 with modulus $\sqrt[q]{1/\varepsilon}$, we summarize the moduli for this choice of the parameter sequences.

1. $n_0 := 1$ and $\delta := \frac{d(1-q)}{1-p} - 1$,
2. $f(n) := \lceil n^{d/(1-p)}\rceil$, where $d := \min\left\{\frac{1-p}{1-r-p}, \frac{3}{2}\left(1-\frac{q}{1-p}\right)^{-1}\right\}$,
3. $\varphi_1(\varepsilon) := \sqrt[q]{1/\varepsilon}$,
4. $\varphi_2(\varepsilon) := \left\lceil \left(\frac{2^d dq(d+1-p)}{\varepsilon(1-p)^2}\right)^2 \right\rceil + 1$,
5. $\varphi_3(\varepsilon) := \begin{cases} ((2p-1)\varepsilon)^{\frac{1-p}{d(1-2p)}} + 1 & \text{for } p > \frac{1}{2}, \\ \frac{d\cdot 2^{\frac{2d}{1-p}}}{(1-p)(\exp(\varepsilon)-1)} + 1 & \text{for } p = \frac{1}{2}, \\ \max\left\{\left(\frac{2d}{(1-p)\varepsilon}\right)^{\frac{1-p}{d(1-2p)-1+p}}, \left((2/\varepsilon)^{\frac{1}{2p}}+1\right)^{(1-p)/d}\right\} + 1 & \text{for } p < \frac{1}{2}. \end{cases}$

7.4.2 Example 2

We begin with the following well-known inequality, whose proof we include for completeness.

Lemma 7.4.2. *For all* $x \geq 0$, $\log(1+x) \leq \frac{x}{\sqrt{1+x}}$.

Proof. Define $f : [0, \infty] \to \mathbb{R}$ by $f(x) := \frac{x}{\sqrt{1+x}} - \log(1+x)$. Then

$$f'(x) = \frac{\sqrt{1+x} - \frac{x}{2\sqrt{1+x}}}{1+x} - \frac{1}{1+x} = \frac{\frac{2+2x-x}{\sqrt{1+x}} - 1}{1+x} = \frac{2+x-\sqrt{1+x}}{(1+x)^{3/2}} \geq 0.$$

Moreover, $f(0) = 0$, so $f(x) \geq 0$ for all $x \geq 0$, whence the claim follows. □

Set $\lambda_n = 1/n$ and $\theta_n = 1/\log\log n$ for $n \geq 3$ and $\lambda_1 = \lambda_2 = \theta_1 = \theta_2 = 0$ (see [15]). Then, we may take $n_0 := 3$, $\delta := 1/2$, $f(n) := n^n$, $\varphi_1(\varepsilon) := \exp\exp(1/\varepsilon)$, $\varphi_2 := \max\{e^4, \exp((1/\varepsilon)^2 - 1) - 1\}$ and $\varphi_3 := \max\{3, \log(2/\varepsilon + 1)\}$. That φ_1 is as required is immediate. Moreover, by Example 1 of [15],

$$\theta_{f(n)} \sum_{j=f(n)}^{f(n+1)} \lambda_j \geq \frac{\log n}{\log n + \log\log n} \geq \frac{\log n}{2\log n} = \frac{1}{2}, \quad \text{for all } n \geq 3,$$

so n_0 and δ are as required.

Again from [15],

$$1 + \log(n+1) \geq \sum_{j=f(n)}^{f(n+1)} \lambda_j \geq \log n, \quad \text{for all } n \geq 3.$$

Moreover, for $n \geq 3$,

$$\begin{aligned}\log\log(n+1)^{n+1} - \log\log n^n &\leq \log\log \frac{(n+1)^{n+1}}{n^n} = \log\log\left(\left(\frac{n+1}{n}\right)^n (n+1)\right) \\ &\leq \log\log(e(n+1)).\end{aligned}$$

Consequently

$$\begin{aligned}\theta_{f(n)} - \theta_{f(n+1)} &= \frac{1}{\log\log n^n} - \frac{1}{\log\log(n+1)^{n+1}} \\ &= \frac{\log\log(n+1)^{n+1} - \log\log n^n}{(\log\log n^n) \cdot (\log\log(n+1)^{n+1})} \\ &\leq \frac{\log\log(e(n+1))}{(\log\log n^n) \cdot (\log\log(n+1)^{n+1})}.\end{aligned}$$

Now $e^4 \log e^4 = 4e^4 = 3e^4 + e^4 > e^5 + e = e(e^4+1)$, so $\log(n \log n) \geq \log(e(n+1))$ for all $n \geq e^4$. Consequently,

$$\begin{aligned}\left(\theta_{f(n)} - \theta_{f(n+1)}\right) \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j &\leq \frac{1+\log(n+1)}{\log(n\log n)} \cdot \frac{\log\log(e(n+1))}{\log((n+1)\log(n+1))} \\ &\leq \frac{\log(e(n+1))}{\log(n\log n)} \cdot \frac{\log(1+\log(n+1))}{\log(n+1)} \\ &\leq 1 \cdot \frac{\log(1+\log(n+1))}{\log(n+1)}.\end{aligned}$$

Now, we apply Lemma 7.4.2

$$
\begin{aligned}
\left(\theta_{f(n)} - \theta_{f(n+1)}\right) \cdot \sum_{j=f(n)}^{f(n+1)} \lambda_j &\leq \frac{\log(n+1)}{\log(n+1)\sqrt{1+\log(n+1)}} \\
&= \frac{1}{\sqrt{1+\log(n+1)}} \\
&\leq \varepsilon, \quad \text{for all } n \geq \max\{e^4, \exp((1/\varepsilon)^2 - 1) - 1\}.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
\sum_{j=f(n)}^{f(n+1)} \lambda_j^2 \leq \int_{j=f(n)-1}^{f(n+1)-1} j^{-2} dj &= -2\left(\left((n+1)^{(n+1)} - 1\right)^{-1} - (n^n - 1)^{-1}\right) \\
&= \frac{2}{n^n - 1} - \frac{2}{(n+1)^{(n+1)} - 1} \\
&\leq \frac{2}{n^n - 1} \leq \varepsilon, \quad \text{for all } n \geq \max\{3, \log(2/\varepsilon + 1)\}.
\end{aligned}
$$

8 The Hybrid Steepest Descent Method

8.1 Introduction

For a real Hilbert space H and a mapping $\Theta : H \to \mathbb{R}$, the convex optimization problem for Θ over some closed convex set C consists in finding a point x^* that minimizes Θ over C. Solving this minimization problem is equivalent to solving the *Variational Inequality Problem* for the gradient $\mathcal{F} := \Theta'$ over S, which is defined as follows:

$$\text{Find } u^* \in S \text{ such that } \langle v - u^*, \mathcal{F}u^* \rangle \geq 0 \text{ for all } v \in S. \tag{VIP($\mathcal{F}, S$)}$$

Apart from their connection to the convex optimization problem, variational inequalities have numerous applications and have therefore been widely studied in the literature [3, 25, 108, 109]. For an overview of some applications of variational inequalities and the hybrid steepest descent method in particular, we refer the reader to [104, 106]

Apart from existence and uniqueness of solutions, considerable effort has also been put into devising explicit algorithms to compute solutions. The following observation is of central importance for the latter as it transforms the Variational Inequality Problem into a fixed point problem:

Proposition 8.1.1 (VIP as a fixed point problem)**.** Given a mapping $\mathcal{F} : H \to H$ and a nonempty closed and convex set S, the following three statements are equivalent.

(i) $u^* \in C$ is a solution to VIP$(\mathcal{F}, S)$, i.e.

$$\langle v - u^*, \mathcal{F}(u^*) \rangle \geq 0, \quad \text{for all } v \in S.$$

(ii) For any $\mu > 0$,

$$\langle v - u^*, (u^* - \mu\mathcal{F}(u^*)) - u^* \rangle \leq 0, \quad \text{for all } v \in S.$$

(iii) For any $\mu > 0$,

$$u^* \in Fix(P_S(I - \mu\mathcal{F})).$$

Whenever the mapping $I - \mu\mathcal{F}$ becomes a strict contraction for some $\mu > 0$, the map $P_S(I - \mu\mathcal{F})$ also becomes a contraction, so the Variational Inequality Problem in this situation has a unique solution, and a natural candidate to approximate this solution is the Picard iteration:

$$x_{n+1} := P_S(x_n - \mu\mathcal{F}(x_n)).$$

This algorithm is referred to as the projected gradient method [36, 76] and converges strongly for all $\mu > 0$ such that $I - \mu\mathcal{F}$ is a strict contraction. It is known [104] that, if $\mathcal{F}$ is κ-Lipschitzian and η-strongly monotone, i.e.

$$\langle \mathcal{F}x - \mathcal{F}y, x - y \rangle \geq \eta \|x - y\|^2, \quad \text{for all } x, y \in H,$$

then $I - \mu\mathcal{F}$ is a strict contraction with Lipschitz constant $\tau := 1 - \sqrt{1 - \mu(2\eta - \mu\kappa^2)}$ for all $\mu \in (0, 2\eta/\kappa^2)$. The main drawback of this approach is that it requires a closed-form expression for the projection P_S onto S, which is not always available.

The hybrid steepest descent method [101, 104], HSDM for short, avoids the use of the projection P_S. This method only requires that the set S is the set of fixed points $Fix(T)$ of some nonexpansive mapping $T : H \to H$:

Theorem 8.1.2 (Yamada [104]). *Let $T : H \to H$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. Suppose that a mapping $\mathcal{F} : H \to H$ is κ-Lipschitzian and η-strongly monotone over $T(H)$. Then, for any $u_0 \in H$, any $\mu \in (0, 2\eta/\kappa^2)$ and any sequence $(\lambda_n) \subset (0, 1]$ satisfying*

$$\text{1. } \lim_{n\to\infty} \lambda_n = 0, \qquad \text{2. } \sum_{n=1}^{\infty} \lambda_n \text{ diverges, and} \qquad \text{3. } \lim_{n\to\infty} \frac{\lambda_n - \lambda_{n+1}}{\lambda_n^2} = 0,$$

the sequence (u_n) generated by

$$u_{n+1} := T(u_n) - \lambda_{n+1}\mu\mathcal{F}(T(u_n))$$

converges to the unique solution of VIP$(\mathcal{F}, Fix(T))$.

Another possibility is that the projection P_S is not known, but $S = \bigcap_{n=1}^{N} S_n$, where the individual projections P_{S_n} are simple enough to have known closed form expressions [104]. This case is covered by the following Theorem.

Theorem 8.1.3 (Yamada [104]). *For $n = 1, \ldots, N$, let $T_n : H \to H$ be nonexpansive mappings that satisfies $S := \bigcap_{n=1}^{N} Fix(T_i) \neq \emptyset$, and assume that*

$$F = Fix(T_N \cdots T_1) = Fix(T_1 T_N \cdots T_2) = \cdots = Fix(T_{N-1} T_N - 2 \cdots T_1 T_N). \qquad (+)$$

Suppose that the mapping $\mathcal{F} : H \to H$ is κ-Lipschitzian and η-strongly monotone. Then, for any $u_0 \in H$, any $\mu \in (0, 2\eta/\kappa^2)$ and any sequence $(\lambda_n) \subset (0, 1]$ satisfying

$$\text{1. } \lim_{n\to\infty} \lambda_n = 0, \qquad \text{2. } \sum_{n=1}^{\infty} \lambda_n \text{ diverges, and} \qquad \text{3. } \sum_{n=1}^{\infty} |\lambda_n - \lambda_{n+N}| < \infty,$$

the sequence (u_n) generated by

$$u_{n+1} := T_{[n+1]}(u_n) - \lambda_{n+1}\mu\mathcal{F}(T_{[n+1]}(u_n))$$

converges strongly to the unique solution of VIP$(\mathcal{F}, S)$*, where $[n] := n \mod N$.*

It should be remarked that Theorem 8.1.3 admits $\lambda_n := 1/n$, while Theorem 8.1.2 only allows for $\lambda_n := 1/n^\rho$ for $0 < \rho < 1$. However, since one can choose $N = 1$ in Theorem 8.1.3, the choice $\lambda_n := 1/n$ is also covered for the case of a single nonexpansive mapping T with $Fix(T) = S$. Moreover, it is interesting to note that the Bauschke condition (+) introduced in [4] is always satisfied whenever $T_n = P_{S_n}$ for closed convex sets S_n with nonempty intersection.

8.2 Relation to Moudafi's Viscosity Approximation Method

Roughly at the same time as Yamada, Moudafi [81] independently proposed the Viscosity Approximation Method, which is given for nonexpansive $T : C \to C$ and strictly contractive $f : C \to C$ by

$$x_{n+1} := \lambda_{n+1} f(x_n) + (1 - \lambda_n) T x_n.$$

Now observe that Yamada's iteration scheme, the Hybrid Steepest Descent Method, can be rearranged as follows:

$$\begin{aligned} u_{n+1} &= T u_n - \lambda_{n+1} \mu \mathcal{F}(T u_n) \\ &= (1 - \lambda_{n+1}) T u_n + \lambda_{n+1} (I - \mu \mathcal{F})(T u_n). \end{aligned}$$

Therefore, Yamada's iteration scheme is a special case of the Viscosity Approximation Method if one chooses the contraction $f := (I - \mu \mathcal{F}) \circ T$. However, Yamada's proof establishing convergence of the HSDM under the proposed conditions is easily reformulated to accomodate for the prima facie more general Viscosity Approximation Method. Moreover, the bounds proposed in this chapter also hold for the Viscosity Approximation Method, as the reader may readily verify.

Moreover, both Yamada's and Bruck's conditions imposed on (λ_n) do not include the important case $\lambda_n = 1/(n+1)$. Xu [103] later showed that the Viscosity Approximation Method converges for $\lambda_n = 1/(n+1)$ by proving convergence under Wittmann's conditions (cf. also Chapter 5)

$$\text{(i)} \lim_{n\to\infty} \alpha_n = 0, \qquad \text{(ii)} \sum_{n=0}^{\infty} \alpha_n = \infty, \qquad \text{(iii)} \lim_{n\to\infty} (\alpha_n - \alpha_{n-1})/\alpha_n = 0$$

However, one should note that Yamada's Theorem for *finitely many mappings* T_i (Theorem 8.1.3) imposes precisely these conditions on (λ_n) for the case $N = 1$.

Convergence of the Viscosity Approximation Method for finitely many mappings

$$x_{n+1} := \lambda_{n+1} f(x_n) + (1 - \lambda_{n+1}) T_{[n+1]}(x_n), \quad [n] := n \mod N \tag{8.1}$$

was later shown by Jung [42]. One again easily verifies that the bounds provided for the Hybrid Steepest Descent Method for the case of a finite family of nonexpansive mappings T_i also holds for the corresponding Viscosity Approximation Method.

Finally, one should observe that the HSDM for a finite family of mappings is, in fact, not a special case of (8.1). In fact, rearranging the HSDM as before, one obtains

$$\begin{aligned} u_{n+1} &= T_{[n+1]}(u_n) - \lambda_{n+1} \mu \mathcal{F}(T_{[n+1]}(u_n)) \\ &= (1 - \lambda_{n+1}) T u_n + \lambda_{n+1} (I - \mu \mathcal{F})(T_{[n+1]} u_n) \end{aligned}$$

Since the contraction $(I - \mu \mathcal{F}) \circ T_{[n+1]}$ now depends on n, it is not permitted in the Viscosity Approximation Method.

8.3 Quantitative Versions of the Convergence of the HSDM

Once again, effective rates on the strong convergence of (x_n) are generally ruled out, since the Hybrid Steepest Descent Method approximates in particular a fixed point of T. Therefore, we again can only attempt to find a rate of metastability $\Phi : (0, \infty) \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$:

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Phi(\varepsilon, g) \, \forall i, j \in [n; n + g(n)] \, \big(\|u_i - u_j\| < \varepsilon \big).$$

Quantitative, finitary versions of all of Theorems 8.1.2 and 8.1.3, however, should not only finitize the Cauchyness of (u_n), but also that the strong limit is indeed a solution to the variational inequality problem: For all $\varepsilon \in (0, 1]$ and all $g : \mathbb{N} \to \mathbb{N}$ there exists an $n \leq \Xi(\varepsilon, g)$ and an $\varepsilon' > 0$ such that, for all $i, j \in [n, n + g(n)]$ and $v \in Fix(T)$,

(i) $\|u_i - u_j\| \leq \varepsilon$ for all $i, j \in [n; n + g(n)]$

(ii) $\|Tv - v\| \leq \varepsilon'$ implies $\langle \mathcal{G} u_n - u_n, v - u_n \rangle \leq \varepsilon$, where $\mathcal{G} = I - \mu \mathcal{F}$ for suitable $\mu > 0$.

The new, logically transformed proof of (i) and (ii) is totally elementary in that all ideal principles have been eliminated; one can recover Yamada's original theorem using only the axiom of choice over *quantifier-free* formulas.

8.4 A Quantitative Solution to the VIP

We now examine the structure of the proof of Theorem 8.1.2 from a proof-theoretic perspective. Given a nonexpansive mapping $T : H \to H$ and a κ-Lipschitzian and η-strongly monotone mapping $\mathcal{F} : H \to H$, fix an arbitrary $\mu \in (0, 2\eta/\kappa^2)$. Then, the mapping $T^{(\lambda)} : H \to H$ defined by $T^{(\lambda)}(x) := T(x) - \lambda \mu \mathcal{F}(Tx)$ is a strict contraction for all $\lambda \in (0, 1]$. As such, given a sequence $(\lambda_n) \subset (0, 1]$, there exists for each nonnegative integer n a unique solution v_n to the equation

$$v_n = T^{(\lambda_n)}(v_n) = Tv_n - \lambda_n \mu \mathcal{F}(Tv_n). \tag{8.2}$$

Next, Yamada shows using weak sequential compactness that v_n converges weakly to the unique fixed point u^* of T that solves the Variational Inequality Problem VIP$(\mathcal{F}, Fix(T))$. Since, moreover, $\|v_n - T(v_n)\|$ converges to zero, the demiclosedness principle then implies that the weak limit u^* is a fixed point of T. This, in turn, is used to prove using constructive reasoning that (v_n) converges strongly to u^*. The final step is then a constructive proof of $\|u_n - v_n\| \to 0$, where (u_n) is the iteration proposed in Theorem 8.1.2.

Structurally, the proof of $v_n \to u^*$ is reminiscent of the proof of the following classical result due to Browder, which we recall from Chapter 5 (cf. Theorem 5.1.3)

Theorem 8.4.1 ([11]). *Let H be a Hilbert space and $T : H \to H$ be a nonexpansive mapping that maps a bounded, closed and convex subset C of H into itself. Let v_0 be an arbitrary point of C, and for each k with $0 < k < 1$, let $U_k(x) := kU(x) + (1 - k)v_0$.*

Then U_k is a strict contraction of H, U_k has a unique fixed point u_k in C, and u_k converges as $k \to 1$ strongly to a fixed point u_0 of U in C. The fixed point u_0 in C is uniquely specified as the fixed point of U in C closest to v_0.

The nonconstructive part of the proof of this theorem also consists of weak sequential compactness, and the unique existence of a point that solves a variational inequality. The latter in this case is the variational inequality that characterizes the metric projection $P_{Fix(T)}(v_0)$ of the point v_0 onto $Fix(T)$, which reads

$$\langle v_0 - P_{Fix(T)}(v_0), v - P_{Fix(T)}(v_0)\rangle \leq 0, \quad \text{for all } v \in Fix(T).$$

An extensive proof-theoretic analysis of this proof has already been carried out by Kohlenbach [52]. By virtue of the complete modularity of the logical machinery employed therein, one can reuse the quantitative versions of the use of weak sequential compactness and the demiclosedness principle.

The unique existence of the solution u^* to the Variational Inequality Problem, on the other hand, is substantially more difficult to constructivize than the existence of the metric projection $P_{Fix(T)}(v_0)$. To make sense of this, observe first that for neither of the two proofs the exact point is needed, but only an ε-approximation. For the projection, this corresponds to finding for all $\varepsilon > 0$ a point $u \in C$ such that

$$Tu = u \wedge \forall v \in C\big(Tv = v \to \|u - v_0\| \leq \|v - v_0\| + \varepsilon\big). \tag{8.3}$$

The correct form of a quantitative version, i.e. the Dialectica interpretation combined with negative translation outlined in Chapter 2, of this statement is the one given in the following Lemma:

Lemma 8.4.2 (Lemma 2.6 of [52]). *Let* $v_0 \in C$ *such that* $\mathrm{diam}(C) \leq d$. *Let* $\varepsilon \in (0,1], t \in [0,1], \Delta : C \times (C \to (0,1]) \to (0,1]$ *and* $V : C \times (C \to (0,1]) \to C$. *Then, one can construct* $u := u_{v_0,T}(t,\varepsilon,\Delta,V) \in C$ *and* $\varphi := \varphi_{v_0,T}(t,\varepsilon,\Delta,V) : C \to (0,1]$ *such that*

$$\|u - Tu\| < \Delta(u,\varphi)$$

and

$$\|TV(u,\varphi) - V(u,\varphi)\| < \varphi(V(u,\varphi)) \to \|v_0 - u\|^2 \leq \|v_0 - V^t(u,\varphi)\|^2 + \varepsilon, \tag{8.4}$$

where $V^t(u,\varphi) := (1-t)u + tV(u,\varphi)$. *In fact,* u, φ *can be defined explicitly as functionals in* ε, Δ *and* V *in addition to* v_0 *and* T.

Remark 8.4.3. We will sometimes call the functionals Δ and V *counterfunctions* in the style of the no-counterexample-interpretation due to Kreisel [62, 63].

Let us now turn to formulating an analogue of this lemma in the context of the VIP. To be able to reuse as much as possible from the previous analysis, it is convenient to reformulate the iteration (8.2) as a convex combination: For $\mathcal{G} := I - \mu\mathcal{F}$, where $\mu \in (0, 2\eta/\kappa^2)$, one can re-write (8.2) as

$$v_n = (1 - \lambda_n)Tv_n + \lambda_n\mathcal{G}(Tv_n),$$

and the iteration proposed in Theorem 8.1.2 as

$$u_{n+1} = (1 - \lambda_{n+1})Tu_n + \lambda_{n+1}\mathcal{G}(Tu_n).$$

As remarked earlier, for any choice $\mu \in (0, 2\eta/\kappa^2)$, the mapping $\mathcal{G}$ is a strict contraction with Lipschitz constant $\tau := \sqrt{1 - \mu(2\eta - \mu\kappa^2)} < 1$. From now on, we simply assume that we are given an *arbitrary* τ-contraction $\mathcal{G}$, making no reference to $\mathcal{F}$.

Now, the operators T and $\mathcal{G}$ need only be defined as self-maps on a closed and convex subset C of H. To be able to apply Lemma 8.4.2, we also assume that C is bounded with $\mathrm{diam}(C) \leq d$. This condition, however, is no real restriction, as we will show later on, so that $C = H$ is still admissible for our results (see Corollaries 8.6.12 and 8.7.10).

Observe that the characterization stated in Proposition 8.1.1(iii) of the solution to the VIP is formalized by

$$\exists u^* \in C\Big(Tu^* = u^* \wedge \forall v \in C\big(Tv = v \to \|u^* - \mathcal{G}u^*\| \leq \|v - \mathcal{G}u^*\|\big)\Big).$$

As already mentioned, we only need the weakened, ε-version of this statement. Analogously to the case of the ε-metric projection (8.3), this corresponds to

$$\forall \varepsilon > 0 \exists u^* \in C\Big(Tu^* = u^* \wedge \forall v \in C\big(Tv = v \to \|u^* - \mathcal{G}u^*\| \leq \|v - \mathcal{G}u^*\| + \varepsilon\big)\Big).$$

The same tools that were used to transform (8.3) now tell us that our task is to solve the following problem:

Problem 8.4.4. Suppose C is a closed, bounded, convex subset of a Hilbert space H with $\mathrm{diam}(C) \leq d$ for some nonnegative integer d, $T : C \to C$ is nonexpansive and $\mathcal{G} : C \to C$ is τ-contractive. For $\varepsilon \in (0, 1], t \in [0, 1], \Delta : C \times (C \to (0, 1]) \to (0, 1]$ and $V : C \times (C \to (0, 1]) \to H$, solve for u^* and φ in the formula

$$\begin{aligned} &\forall \varepsilon \in (0, 1] \forall \Delta : C \times (C \to (0, 1]) \forall V : C \times (C \to (0, 1]) \to C \\ &\quad \exists u^* \in C \exists \varphi : C \to (0, 1]\Big(\|Tu^* - u^*\| \leq \Delta(u^*, \varphi) \wedge \\ &\qquad \|TV(u^*, \varphi) - V(u^*, \varphi)\| < \varphi(V(u^*, \varphi)) \to \|\mathcal{G}u^* - u^*\|^2 < \|\mathcal{G}u^* - V^t(u^*, \varphi)\|^2 + \varepsilon\Big), \end{aligned} \tag{8.5}$$

where, as before, $V^t(u^*, \varphi) := (1 - t)u^* + tV(u^*, \varphi)$.

By Proposition 8.1.1, the unique point $u^* \in Fix(P_{Fix(T)} \circ \mathcal{G})$ will solve the VIP. The quantitative version of this step is given by the following Lemma.

Lemma 8.4.5 (Lemma 2.7 of [52])**.** *Let $u^*, u, v \in H$ such that $\|u^* - v\| \leq d$. For $t \in [0, 1]$, define $w_t := (1 - t)u^* + tv$. Then*

$$\forall \varepsilon \in (0, 1](\|u^* - u\|^2 \leq \frac{\varepsilon^2}{2d^2} + \|u - w_{\frac{\varepsilon}{3d^2}}\|^2 \to \langle u - u^*, v - u^*\rangle < \varepsilon).$$

To solve Problem 8.4.4, recall that by Proposition 8.1.1, u^* is the unique fixed point of the mapping $x \mapsto P_{Fix(T)}(\mathcal{G}x)$. Since the metric projection is nonexpansive and $\mathcal{G}$ is, for proper choice of μ, a strict contraction, this mapping is also a strict contraction. Thus, the Picard iteration, starting with an arbitrary point p, converges strongly to u^*:

$$u^* = \lim_{n \to \infty} (P_{Fix(T)} \circ \mathcal{G})^{(n)}(p).$$

In view of this, it is not surprising that a quantitative version of the existence of u^* will iterate the solution functionals of Lemma 8.4.2.

Before we proceed, we need the following variant of Lemma 8.4.2, as it turns out later that we need to win against two counterfunction pairs (Δ_1, V_1) and (Δ_2, V_2), simultaneously:

Lemma 8.4.6. *Let $v_0 \in C$ such that $v_0 - v \leq d$ for some $v \in Fix(T)$. Let $\varepsilon \in (0,1], t_1, t_2 \in [0,1], \Delta_1, \Delta_2 : C \times (C \to (0,1]) \to (0,1]$ and $V_1, V_2 : C \times (C \to (0,1]) \to C$. Then, one can construct a $u := u'_{v_0,T}(t_1, t_2, \varepsilon, \Delta_1, \Delta_2, V_1, V_2) \in C$ and a $\varphi := \varphi'_{v_0,T}(t_1, t_2, \varepsilon, \Delta_1, \Delta_2, V_1, V_2) : C \to (0,1]$ such that for $i = 1, 2$,*

$$\|u - Tu\| < \Delta_i(u, \varphi)$$

and

$$\begin{aligned}&\|TV_i(u,\varphi) - V_i(u,\varphi)\| < \varphi(V_i(u,\varphi))\\ &\qquad \to \|v_0 - u\|^2 \leq \|v_0 - V_i^{t_i}(u,\varphi)\|^2 + \varepsilon.\end{aligned}$$

Proof. Given $\Delta_1, \Delta_2 : C \times (C \to (0,1]) \to (0,1]$ and $V_1, V_2 : C \times (C \to (0,1]) \to C$, define $\Delta : C \times (C \to (0,1]) \to (0,1]$ by

$$\Delta(u,\varphi) := \min\{\Delta_1(u,\varphi), \Delta_2(u,\varphi)\},$$

and $V : C \times (C \to (0,1]) \to C$ by

$$V(u,\varphi) := \begin{cases} V_1(u,\varphi), & \text{if } \|v_0 - V_1^{t_1}(u,\varphi)\| \leq \|v_0 - V_2^{t_2}(u,\varphi)\|,\\ V_2(u,\varphi), & \text{otherwise.}\end{cases}$$

Using the solution operators of Lemma 8.4.2, we define $u'_{v_0,T}(t, \varepsilon, \Delta_1, \Delta_2, V_1, V_2) := u_{v_0,T}(t, \varepsilon, \Delta, V)$ and $\varphi'_{v_0,T}(t, \varepsilon, \Delta_1, \Delta_2, V_1, V_2) := \varphi_{v_0,T}(t, \varepsilon, \Delta, V)$. $\square$

Returning to the original problem, we start with an arbitrary point p in C and use Lemma 2.4 of [52] to obtain a point u_0 and a functional φ_0 which together solve the quantitative version (according to Lemma 8.4.2) of the ε-projection of p onto $Fix(T)$ for suitable counterfunctions Δ_0 and V_0. We then repeat this procedure for $\mathcal{G}u_0$, obtaining a point u_1, and so on. In total, we obtain points u_i and functionals φ_i such that

$$\begin{aligned}&\|Tu_i - u_i\| \leq \Delta_i(u_i,\varphi_i) \wedge \\ &\quad \Big(\|TV_i(u_i,\varphi_i) - V_i(u_i,\varphi_i)\| < \varphi_i(V_i(u_i,\varphi_i))\\ &\qquad\qquad \to \|\mathcal{G}u_{i-1} - u_i\|^2 < \|\mathcal{G}u_{i-1} - V_i^t(u_i,\varphi_i)\|^2 + \varepsilon\Big),\end{aligned} \tag{8.6}$$

for suitable counterfunctions Δ_i and V_i which depend on the counterfunctions Δ and V from statement (8.5). (As before, $V_i^t(u_i,\varphi_i) := (1-t)\mathcal{G}u_i + tV_i(u_i,\varphi_i)$.)

The key in solving Problem 8.4.4 will be the observation that u_i is the ε-projection of $\mathcal{G}u_{i-1}$ with respect to counterfunctions V_i and Δ_i. Therefore, the points u_i are an

ε-version of the Picard-iteration of the contractive mapping $P_{Fix(T)} \circ \mathcal{G}$. As such, the distance $\|u_i - u_{i-1}\|$ can be made arbitrarily small for a sufficiently large i, given that we choose our counterfunctions in the correct way – in this case counterfunctions that ensure that the ε-projection is ε-nonexpansive with respect to the involved points. The simple observation

$$\begin{aligned}\|\mathcal{G}u_i - u_i\|^2 &\leq \|\mathcal{G}u_{i-1} - u_i\|^2 + \tau^2\|u_i - u_{i-1}\|^2 + 2\tau \cdot \|\mathcal{G}u_{i-1} - u_i\| \cdot \|u_i - u_{i-1}\| \\ &\overset{(8.6)}{<} \|\mathcal{G}u_{i-1} - V_i^t(u_i, \varphi_i)\|^2 + \varepsilon + \tau^2\|u_i - u_{i-1}\|^2 + 2d\tau\|u_i - u_{i-1}\| \\ &\leq \left(\|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\| + \|\mathcal{G}u_i - \mathcal{G}u_{i-1}\|\right)^2 + \varepsilon + \tau^2\|u_i - u_{i-1}\|^2 + 2d\tau\|u_i - u_{i-1}\| \\ &\leq \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 2\tau^2\|u_i - u_{i-1}\|^2 + 4d\tau\|u_i - u_{i-1}\| + \varepsilon \qquad (8.7)\end{aligned}$$

then tells us that we may take $u^* = u_i$ if the integer i is large enough to ensure the distance between u_i and u_{i-1} is small enough.

Our task is now to analyze the following proof that the metric projection is nonexpansive. For $x, y \in H$, denote by Px and Py their projections onto an arbitrary convex set. Then

$$\langle Px - x, Px - Py\rangle \leq 0 \quad \text{and} \quad \langle Py - y, Py - Px\rangle \leq 0.$$

Summing up these two inequalities yields $\langle Px - Py + y - x, Px - Py\rangle \leq 0$, which implies

$$\|Px - Py\| \leq \langle x - y, Px - Py\rangle \leq \|x - y\| \cdot \|Px - Py\|,$$

which implies the claim.

Quantitatively, this translates as follows (for later convenience already instantiated with the points u_i and projection onto $Fix(T)$). Suppose that we have for some $\tilde{\varepsilon} > 0$ to be specified later on that

$$\|u_{i+1} - \mathcal{G}u_i\|^2 \leq \frac{\tilde{\varepsilon}^4}{8d^2} + \left\|\left(1 - \frac{\tilde{\varepsilon}^2}{6d^2}\right)u_{i+1} + \frac{\tilde{\varepsilon}^2}{6d^2}u_i - \mathcal{G}u_i\right\|^2, \qquad (8.8)$$

$$\|u_i - \mathcal{G}u_{i-1}\|^2 \leq \frac{\tilde{\varepsilon}^4}{8d^2} + \left\|\left(1 - \frac{\tilde{\varepsilon}^2}{6d^2}\right)u_i + \frac{\tilde{\varepsilon}^2}{6d^2}u_{i+1} - \mathcal{G}u_{i-1}\right\|^2. \qquad (8.9)$$

For notational simplicity later on, we write

$$A(\tilde{\varepsilon}, u, v, p) :\equiv \|u - p\|^2 \leq \frac{\tilde{\varepsilon}^4}{8d^2} + \left\|\left(1 - \frac{\tilde{\varepsilon}^2}{6d^2}\right)u + \frac{\tilde{\varepsilon}^2}{6d^2}v - p\right\|^2.$$

Then (8.8) $\equiv A(\tilde{\varepsilon}, u_{i+1}, u_i, \mathcal{G}u_i)$ and (8.9) $\equiv A(\tilde{\varepsilon}, u_i, u_{i+1}, \mathcal{G}u_{i-1})$ for $i \geq 1$. By Lemma 8.4.5, $A(\tilde{\varepsilon}, u_{i+1}, u_i, \mathcal{G}u_i)$ and $A(\tilde{\varepsilon}, u_i, u_{i+1}, \mathcal{G}u_{i-1})$ together imply

$$\begin{aligned}\langle u_{i+1} - \mathcal{G}u_i, u_{i+1} - u_i\rangle &< \tilde{\varepsilon}^2/2, \\ \langle u_i - \mathcal{G}u_{i-1}, u_i - u_{i+1}\rangle &< \tilde{\varepsilon}^2/2.\end{aligned}$$

Thus, since $\mathcal{G}$ is a τ-contraction,

$$\begin{aligned}\|u_{i+1} - u_i\|^2 &< \|\mathcal{G}u_i - \mathcal{G}u_{i-1}\| \cdot \|u_{i+1} - u_i\| + \tilde{\varepsilon}^2 \\ &\leq \tau\|u_i - u_{i-1}\| \cdot \|u_{i+1} - u_i\| + \tilde{\varepsilon}^2. \qquad (8.10)\end{aligned}$$

Now, the problem is that, when we divide the inequality by $\|u_{i+1} - u_i\|$ (if it is strictly greater than 0), the term $\tilde{\varepsilon}/\|u_{i+1} - u_i\|$ becomes unbounded for small $\|u_{i+1} - u_i\|$. However, we want to make $\|u_{i+1} - u_i\|$ small anyway, so this is not a problem, and it gives rise to the following case distinction.

(i) For $\|u_{i+1} - u_i\| < \tilde{\varepsilon}$, we immediately get $\|u_{i+1} - u_i\| < \tilde{\varepsilon} \le \tau\|u_i - u_{i-1}\| + \tilde{\varepsilon}$

(ii) For $\|u_{i+1} - u_i\| \ge \tilde{\varepsilon}$, we get $\|u_{i+1} - u_i\| < \tau\|u_i - u_{i-1}\| + \tilde{\varepsilon}$ by dividing (8.10) by $\|u_{i+1} - u_i\|$.

Thus, we have shown for all integers $i \ge 0$ that $A(\tilde{\varepsilon}, u_{i+1}, u_i, \mathcal{G}u_i)$ and $A(\tilde{\varepsilon}, u_i, u_{i+1}, \mathcal{G}u_{i-1})$ imply $\|u_{i+1} - u_i\| < \tau\|u_i - u_{i-1}\| + \tilde{\varepsilon}$. Now suppose $\|u_{i+1} - u_i\| < \tau\|u_i - u_{i-1}\| + \tilde{\varepsilon}$ for all nonnegative integers $k \le i$, then

$$\begin{aligned}
\|u_{i+1} - u_i\| &< \tau\|u_i - u_{i-1}\| + \tilde{\varepsilon} \\
&< \tau^2\|u_{i-1} - u_{i-2}\| + \tau\tilde{\varepsilon} + \tilde{\varepsilon} \\
&< \dots \\
&< \tau^{i+1} \cdot \|u_0 - p\| + \tilde{\varepsilon} \cdot \sum_{k=0}^{i-1} \tau^k \\
&< \tau^{i+1} d + \frac{\tilde{\varepsilon}}{1-\tau}
\end{aligned}$$

Going back to (8.7) with $\tilde{\varepsilon}/3$ instead of ε, we see that for $d \ge 1$

$$\begin{aligned}
\|\mathcal{G}u_i - u_i\|^2 &< \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 2\tau^2\|u_i - u_{i-1}\|^2 + 4d\tau\|u_i - u_{i-1}\| + \frac{\tilde{\varepsilon}}{3} \\
&< \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 2d^2\tau^{2i+2} + 4d^2\tau^{i+1} + \frac{2\tau^2\tilde{\varepsilon}^2}{(1-\tau)^2} + \frac{4d\tau^2\tilde{\varepsilon}}{1-\tau} + \frac{\tilde{\varepsilon}}{3} \\
&\le \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 2d^2\tau^{i+1}(\tau^{i+1} + 2) + \frac{2\tilde{\varepsilon}^2}{(1-\tau)^2} + \frac{4d\tilde{\varepsilon}}{1-\tau} + \frac{\tilde{\varepsilon}}{3} \\
&\le \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 2d^2\tau^{i+1}(\tau^{i+1} + 2) + \frac{3+4d}{(1-\tau)^2}\tilde{\varepsilon} \\
&\le \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + 3d^2\tau^{i+1} + \frac{3+4d}{(1-\tau)^2}\tilde{\varepsilon} \\
&\le \|\mathcal{G}u_i - V_i^t(u_i, \varphi_i)\|^2 + \varepsilon,
\end{aligned}$$

for $i \geq i_0 := \lceil \log_\tau(\tilde{\varepsilon}/6d^2) - 1 \rceil$ and $\tilde{\varepsilon} := \frac{(1-\tau)^2}{6+8d}\varepsilon$. In other words

$$
\begin{aligned}
&\Big\{ A(\tilde{\varepsilon}, u_{i_0}, u_{i_0-1}, \mathcal{G}u_{i_0-1}) \wedge A(\tilde{\varepsilon}, u_0, u_1, p) \wedge \bigwedge_{k=1}^{i_0-1} \Big(\tilde{\varepsilon}, A(u_i, u_{i+1}, \mathcal{G}u_{i-1}) \wedge A(\tilde{\varepsilon}, u_i, u_{i-1}, \mathcal{G}u_{i-1} \Big) \\
&\quad \wedge \Big(\|TV(u_{i_0}, \varphi_{i_0}) - V(u_{i_0}, \varphi_{i_0}) < \varphi_{i_0}(V(u_{i_0}, \varphi_{i_0})) \\
&\quad\quad \to \|u_{i_0} - \mathcal{G}u_{i_0-1}\|^2 \leq \frac{\tilde{\varepsilon}}{3} + \|(1-t)u_{i+1} + tV(u_{i_0}, \varphi_{i_0}) - \mathcal{G}u_{i_0-1}\|^2 \Big) \Big\} \\
&\quad\quad\quad \to \Big\{ \Big(\|TV(u_{i_0}, \varphi_{i_0}) - V(u_{i_0}, \varphi_{i_0})\| < \varphi_{i_0}(V(u_{i_0}, \varphi_{i_0})) \\
&\quad\quad\quad\quad \to \|\mathcal{G}u_{i_0} - u_{i_0}\|^2 \leq \|\mathcal{G}u_{i_0} - V^t_{i_0}(u_{i_0}, \varphi_{i_0})\|^2 + \varepsilon \Big) \Big\}.
\end{aligned}
\tag{8.11}
$$

Therefore, we need to construct the finite sequence $(u_i)_{0 \leq i \leq i_0}$ satisfying (8.8) and (8.9). Then u_{i_0} and φ_{i_0} will solve Problem 8.4.4.

From these considerations, it is clear that for $i = 0$, we will need to win against a convex combination (of known weight and accuracy, see (8.8) and (8.9)) of ourselves, i.e. u_0, and the subsequent point u_1, the ε-projection of $\mathcal{G}u_0$, which we anticipate as the outcome of the iterative process with reference point $\mathcal{G}u_0$ given by Lemma 8.4.2. Namely, we choose as counterfunction the anticipated next point. But this anticipated next point needs to win both against convex combinations of its predecessor (the point we are trying to construct right now!) and its successor, cf. (8.9).

Notation 8.4.7. Suppose that $t(x)$ is a mathematical expression that depends on a variable x. Then $\lambda x.t(x)$ denotes the function mapping x to $t(x)$. For example, $\lambda n.n$ for integers n denotes the identity on the integers. Likewise, $\lambda x.x^2$ for real numbers x denotes the square-function on the reals. This notation will prove highly convenient in the sequel.

For $0 \leq i \leq i_0$, the considerations mentioned above give rise to the counterfunctions (using the previously introduced notation)

$$
V_i(u,\varphi) := \begin{cases} V(u,\varphi), & \text{for } i = i_0, \\ u'_{\mathcal{G}u,T}(t, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_{i+1}, \lambda v \lambda \psi.\varphi(v), V_{i+1}, \lambda v \lambda \psi.u), & \text{for } i = i_0 - 1, \\ U(\Delta_{i+1}, \lambda v \lambda \psi.\varphi(v), V_{i+1}, \lambda v \lambda \psi.u, \mathcal{G}u), & \text{for } i \leq i_0 - 2, \end{cases}
$$

and

$$
\Delta_i(u,\varphi) := \begin{cases} \Delta(u,\varphi), & \text{for } i = i_0, \\ \varphi'_{\mathcal{G}u,T}(t, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_{i+1}, \lambda v \lambda \psi.\varphi(v), V_{i+1}, \lambda v \lambda \psi.u)(u), & \text{for } i = i_0 - 1, \\ \Phi(\Delta_{i+1}, \lambda v \lambda \psi.\varphi(v), V_{i+1}, \lambda v \lambda \psi.u, \mathcal{G}u)(u), & \text{for } i \leq i_0 - 2, \end{cases}
$$

where

$$
\begin{aligned}
U(\Delta, \Delta', V, V', u) &:= u'_{\mathcal{G}u,T}(\tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta, \Delta', V, V'), \\
\Phi(\Delta, \Delta', V, V', u) &:= \varphi'_{\mathcal{G}u,T}(\tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta, \Delta', V, V'),
\end{aligned}
$$

and u', φ' are the solution functionals of Lemma 8.4.6. Moreover, V and Δ are the original counterfunctions of Problem 8.4.4, i.e. of the original problem. Now set

$$u_i := \begin{cases} u_{p,T}(\tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_0, V_0), & \text{for } i = 0,\\ U(\Delta_i, \lambda v \lambda \psi.\varphi_{i-1}(v), V_i, \lambda v \lambda \psi x.u_{i-1}, \mathcal{G}u_{i-1}), & \text{for } 1 \leq i \leq i_0 - 1,\\ u'_{\mathcal{G}u_{i-1},T}(t, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_i, \lambda v \lambda \psi.\varphi_{i-1}(v), V_i, \lambda v \lambda \psi.u_{i-1}), & \text{for } i = i_0, \end{cases}$$

and

$$\varphi_i := \begin{cases} \varphi_{p,T}(\tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_0, V_0), & \text{for } i = 0,\\ \Phi(\Delta_i, \lambda v \lambda \psi.\varphi_{i-1}(v), V_i, \lambda v \lambda \psi.u_{i-1}, \mathcal{G}u_{i-1}), & \text{for } 1 \leq i \leq i_0 - 1,\\ \varphi'_{\mathcal{G}u_{i-1},T}(t, \tilde{\varepsilon}^2/6d^2, \tilde{\varepsilon}^4/8d^2, \Delta_i, \lambda v \lambda \psi.\varphi_{i-1}(v), V_i, \lambda v \lambda \psi.u_{i-1}), & \text{for } i = i_0, \end{cases}$$

for some arbitrary point $p \in C$. We now show that these counterfunctions are as required.

For $0 \leq i \leq i_0 - 2$, the points u_i and the functions φ_i satisfy by Lemma 8.4.6

$$\begin{aligned} \|u_i - Tu_i\| &< \Delta_i(u_i, \varphi_i)\\ &= \Phi(\Delta_{i+1}, \lambda v \lambda \psi.\varphi_i(v), V_{i+1}, \lambda v \lambda \psi.u_i, \mathcal{G}u_i)(u_i)\\ &= \varphi_{i+1}(u_i). \end{aligned}$$

Similarly,

$$\begin{aligned} \|u_{i_0-1} - Tu_{i_0-1}\| &< \Delta_{i_0-1}(u_{i_0-1}, \varphi_{i_0-1})\\ &= \varphi'_{\mathcal{G}u_{i_0-1},T}(t, \varepsilon^2/24d^2, \varepsilon^4/32d^2, \Delta_{i_0}, \lambda v \lambda \psi.\varphi_{i_0-1}(v), V_{i_0}, \lambda v \lambda \psi.u_{i_0-1})(u_{i_0-1})\\ &= \varphi_{i_0}(u_{i_0-1}). \end{aligned}$$

Consequently, $\|u_i - Tu_i\| < \varphi_{i+1}(u_i)$ for all $0 \leq i \leq i_0 - 1$. Moreover, for $1 \leq i \leq i_0$,

$$\begin{aligned} \|u_i - Tu_i\| &< (\lambda v \lambda \psi.\varphi_{i-1}(v))(u_i, \varphi_i)\\ &= \varphi_{i-1}(u_i). \end{aligned}$$

Furthermore, $\Delta_{i_0} = \Delta$, so $\|u_{i_0} - Tu_{i_0}\| < \Delta(u_{i_0}, \varphi_{i_0})$.

Now recall that, for notational simplicity, we denoted formula (8.8) by the formula $A(\tilde{\varepsilon}, u_i, u_{i+1}, \mathcal{G}u_{i-1})$. The second part of Lemma 8.4.6 then reads for $1 \leq i \leq i_0 - 1$

$$\begin{aligned} &\|TV_i(u_i, \varphi_i) - V_i(u_i, \varphi_i)\| < \varphi_i(V_i(u_i, \varphi_i)) \to A(\tilde{\varepsilon}, u_i, V_i(u_i, \varphi_i), \mathcal{G}u_{i-1}), \text{ and}\\ &\|Tu_{i-1} - u_{i-1}\| < \varphi_i\big((\lambda v \lambda \psi.u_{i-1})(u_i, \varphi_i)\big) \to A(\tilde{\varepsilon}, u_i, (\lambda v \lambda \psi.u_{i-1})(u_i, \varphi_i), \mathcal{G}u_{i-1}). \end{aligned}$$

But observe that $V_i(u_i, \varphi_i) = U(\Delta_{i+1}, \lambda v \lambda \psi.\varphi_i(v), V_{i+1}, \lambda v \lambda \psi.u_i, \mathcal{G}u_i) = u_{i+1}$ and, regarding the second implication, $(\lambda v \lambda \psi.u_{i-1})(u_i, \varphi_i) = u_{i-1}$. Thus, the above implications read

$$\begin{aligned} &\|Tu_{i+1} - u_{i+1}\| < \varphi_i(u_{i+1}) \to A(\tilde{\varepsilon}, u_i, u_{i+1}, \mathcal{G}u_{i-1}), \text{ and}\\ &\|Tu_{i-1} - u_{i-1}\| < \varphi_i(u_{i-1}) \to A(\tilde{\varepsilon}, u_i, u_{i-1}, \mathcal{G}u_{i-1}), \text{ for } 1 \leq i \leq i_0 - 1. \end{aligned}$$

Since the $V_{i_0} = V$ and $V_0(u_0, \varphi_0) = u_1$, we also get

$$\|TV(u_{i_0}, \varphi_{i_0}) - V(u_{i_0}, \varphi_{i_0})\| < \varphi_{i_0}(V(u_{i_0}, \varphi_{i_0}))$$
$$\to \|u_{i_0} - \mathcal{G}u_{i_0-1}\|^2 \leq \frac{\varepsilon^4}{32d^2} + \|(1-t)u_{i+1} + tV(u_{i_0}, \varphi_{i_0}) - \mathcal{G}u_{i_0-1}\|^2$$

and $\|Tu_1 - u_1\| < \varphi_0(u_1) \to A(\tilde{\varepsilon}, u_0, u_1, p)$. Applying the modus ponens and using (8.11), we then see that u_{i_0} and φ_{i_0} are, in fact, solutions for Problem 8.4.4.

8.5 Majorizing the Solution Functionals

We now extend the definition of majorization, i.e. Definition 3.2.3, to the types of the functionals involved in our solution to Problem 8.4.4.

Definition 8.5.1. (i) We say that a function $\varphi : C \to (0, 1]$ is majorized by $k \in \mathbb{N}^*$ if $1/k \leq \varphi(v)$ for all $v \in C$. In this case, we write $k \gtrsim \varphi$.

(ii) We say that a function $\Delta : C \times (C \to (0, 1]) \to (0, 1]$ is majorized by $f : \mathbb{N}^* \to \mathbb{N}^*$ if, for all $\varphi : C \to (0, 1]$ and $k \in \mathbb{N}^*$,

$$k \gtrsim \varphi \to f(k) \gtrsim \lambda v.\Delta(v, \varphi).$$

(iii) We say that the solution operator Φ of Problem 8.4.4 and Lemma 8.4.2 (suppressing dependence on the parameters ε and t) is majorized by $\Phi^* : (\mathbb{N}^* \to \mathbb{N}^*) \to \mathbb{N}^*$ if, for all V, Δ and f as before,

$$f \gtrsim \Delta \to \Phi^*(f) \gtrsim \Phi(\Delta, V).$$

(iv) Similarly, the solution operator Φ' of Lemma 8.4.6 is majorized by Φ'^* if, for all V_1, V_2, Δ_1, Δ_2 and f_1, f_2 as before,

$$f_1 \gtrsim \Delta_1 \wedge f_2 \gtrsim \Delta_2 \to \Phi'^*(f_1, f_2) \gtrsim \Phi(\Delta_1, \Delta_2, V_1, V_2).$$

We should remark that there is a certain disparity between Definition 3.2.3 and this extension; While Definition 3.2.3 corresponds to the extension of *strong* majorization to the abstract types, Definition 8.5.1 is an extension of the *regular* majorization to make it more readable. However, the differences are negligible for the solution operators for Problem 8.4.4 as we are only interested in the validity in the full set theoretic model as opposed to the model of strongly majorizable set-theoretic functions. For details, see [49].

We now show how to majorize the solution operator Ψ of Problem 8.4.4. To do so, we first need to majorize the solution operator of Lemma 8.4.2, which can be stated explicitly as follows [52]: For $i \leq n_\varepsilon := \lceil d^2/\varepsilon \rceil$ we define $\psi_i : C \to (0, 1]$ and $u_i \in C$ inductively by

$$\psi_1(\Delta, V) := \lambda v.1 \qquad u_1(\Delta, V) := \hat{u} \in Fix(T)$$
$$\psi_{i+1}(\Delta, V) := \lambda v.\Delta'(v, \psi_i(\Delta, V)) \qquad u_{i+1}(\Delta, V) := V'(u_i(\Delta, V), \psi_{n_\varepsilon - i-1}(\Delta, V)),$$

where

$$\Delta'(u, \psi) := \min\{\Delta(u, \psi^u), \psi^u(V(u, \psi^u))\},$$
$$V'(u, \psi) := (1-t)u + t(V(u, \psi^u)),$$
$$\psi^u(v) := \frac{\psi((1-t)u + tv)^2}{16d}.$$

Then, for some $i \leq n_\varepsilon$, we have that $u_i(\Delta, V)$ and $\psi^{u_i}_{n_\varepsilon - i}(\Delta, V)$ satisfy the claim. We write $u_{v0,T}(t, \varepsilon, \Delta, V) := u_i$ and $\varphi_{v0,T}(t, \varepsilon, \Delta, V) := \psi^{u_i}_{n_\varepsilon - i}$, where i is the least index such that $u_i, \psi^{u_i}_{n_\varepsilon - i}$ satisfy the claim of Lemma 8.4.2.

Notation 8.5.2. Given any function $f : \mathbb{N}^* \to \mathbb{N}^*$, define the function $f^M : \mathbb{N}^* \to \mathbb{N}^*$ by $f^M(n) := \max\{f(i) : 1 \leq n\}$. Observe that f^M is monotone for any given f.

Lemma 8.5.3. *(i) The functional $\varphi^* : \mathbb{N}^{*\mathbb{N}^*} \to \mathbb{N}^*$ defined by $\varphi^*(f) := \max\{\psi^*_i(f) : 1 \leq i \leq n_\varepsilon\}$, where $\psi^*_i : \mathbb{N}^{*\mathbb{N}^*} \to \mathbb{N}^*$ is defined recursively by*

$$\psi^*_1(f) := 1,$$
$$\psi^*_{i+1}(f) := \max\{f(16d \cdot \psi^*_i(f)^2), 16d \cdot \psi^*_i(f)^2\},$$

is a majorant to the solution operator φ of Lemma 8.4.2, i.e. $\varphi^* \gtrsim \varphi$.

(ii) The functional $\tilde{\varphi}^* : \mathbb{N}^{*\mathbb{N}^*} \to \mathbb{N}^*$ defined by $\tilde{\varphi}^*(f) := \tilde{f}^{(n_\varepsilon)}(1)$ is also a majorant to the solution operator φ of Lemma 8.4.2, where $\tilde{f} : \mathbb{N}^* \to \mathbb{N}^*$ is defined by $\tilde{f}(n) := \max\{f^M(16dn^2), 16dn^2\}$.

(iii) If f is a nondecreasing function, then $\tilde{\varphi}^*(f) = \varphi^*(f)$, and $f^M = f$.

(iv) Given any majorant φ^* to the solution operator of Lemma 8.4.2, the function $\varphi^+ : \mathbb{N}^{*\mathbb{N}^*} \times \mathbb{N}^{*\mathbb{N}^*} \to \mathbb{N}^*$ defined by $\varphi^+(f_1, f_2) := \psi^*(\max\{f_1, f_2\})$ majorizes the solution operator φ' of Lemma 8.4.6. Here $\max\{f_1, f_2\}$ denotes the pointwise maximum of the two functions f_1 and f_2.

Proof. (i) We first show that $\psi^*_i \gtrsim \psi_i$ by induction on i. For $i = 1$, the claim is trivial since $\psi_0(\Delta, V)(v) = 1$ for all $v \in C$. Now, suppose that $\psi^*_i \gtrsim \psi_i$ for some positive integer i and $f \gtrsim \Delta$. Then,

a) on the one hand, using the induction hypothesis and the definition of majorization, $\psi^*_i(f) \gtrsim \psi_i(\Delta, V)$. The observation

$$k \gtrsim \psi \to 16dk^2 \gtrsim \psi^u, \quad \text{for all } u \in C$$

then implies $16d \cdot \psi^*_i(f) \gtrsim \psi^u_i(\Delta, V)$.

b) On the other hand, $f \gtrsim \Delta$ by definition implies

$$k \gtrsim \psi^u_i(\Delta, V) \to f(k) \gtrsim \lambda v.\Delta(v, \psi^u_i(\Delta, V))$$

But the induction hypothesis implies as before $16d \cdot \psi^*_i(f) \gtrsim \psi^u_i(\Delta, V)$, so $f(16d \cdot \psi^*_i(f)) \gtrsim \lambda v.\Delta(v, \psi^u_i(\Delta, V))$.

In total, $\psi^*_{i+1} \gtrsim \psi_{i+1}$. That φ^* is a common majorant for all ψ_i, where $i \leq n_\varepsilon$, follows from Lemma 6.4 of [49]. Therefore $\varphi^* \gtrsim \varphi$.

(ii) First, we show by induction on i that $\tilde{f}^{(i)}(1) \geq \psi^*_i(f)$. For $i = 1$, the statement holds with equality. Moreover,

$$\begin{aligned}\psi^*_{i+1}(f) &= \max\{f(16d \cdot \psi^*_i(f)^2), 16d \cdot \psi^*_i(f)^2\} \\ &\leq \max\{f^M(16d \cdot \psi^*_i(f)^2), 16d \cdot \psi^*_i(f)^2\}\end{aligned}$$

By the monotonicity of f^M and the induction hypothesis, we conclude

$$\begin{aligned}\psi^*_{i+1}(f) &\leq \max\left\{f^M\left(16d \cdot (\tilde{f}^{(i)}(1))^2\right), 16d \cdot \left(\tilde{f}^{(i)}(1)\right)^2\right\} \\ &= \tilde{f}\left(\tilde{f}^{(i)}(1)\right) \\ &= \tilde{f}^{(i+1)}(1).\end{aligned}$$

Therefore, $\tilde{f}^{(i)}(1) \geq \psi^*_i(f)$ for all i. Since $\tilde{f}$ is monotone, $\tilde{f}^{(n_\varepsilon)}(1) \geq \tilde{f}^{(i)}(1)$ for all $i \leq n_\varepsilon$, so the claim follows from part (i).

(iii) $\varphi^*(f) = \tilde{\varphi}^*(f)$ for nondecreasing f is shown as in the previous part with equality throughout.

(iv) Suppose $f_i \gtrsim \Delta_i$ for $i = 1, 2$. Then $\max\{f_1, f_2\} \gtrsim \Delta_i$ as well, so we conclude that $\max\{f_1, f_2\} \gtrsim \lambda u \lambda\varphi. \min\{\Delta_1(u, \varphi), \Delta_2(u, \varphi)\}$. Consequently, since $\varphi^* \gtrsim \varphi$ by hypothesis, we obtain $\varphi^+ \gtrsim \varphi'$.

□

Lemma 8.5.4. *Given a majorant $f \gtrsim \Delta$, define a function $f_i : \mathbb{N}^* \to \mathbb{N}^*$ by $f_i(k) := \tilde{f}^{(n^i_{\tilde{\varepsilon}})}(k)$, where $\tilde{f}$ is defined as in Lemma 8.5.3 and $n_{\tilde{\varepsilon}} := \lceil 8d^4/\varepsilon^4 \rceil$. Then $f_i \gtrsim \Delta_i$ for $0 \leq i \leq i_0$.*

Proof. We show by (backward) induction on n that for any majorant φ^* of φ, the functions $\hat{f}_i$

$$\begin{cases} \hat{f}_{i_0} := \tilde{f} \gtrsim \Delta_{i_0}, \text{ and} \\ \hat{f}_i := \lambda k.\varphi^*(\max\{f_{i+1}, \lambda n.k\}) \gtrsim \Delta_i, \text{ for } i \leq i_0 - 1, \end{cases} \tag{8.12}$$

majorize Δ_i, respectively. By definition, $f_{i_0} = f \gtrsim \Delta = \Delta_{i_0}$, completing the induction base. Now recall that, by definition, $f_i \gtrsim \Delta_i$ if and only $k \gtrsim \varphi \to f_i(k) \gtrsim \lambda v.\Delta_i(v, \varphi)$ for $0 \leq i \leq i_0 - 1$. So suppose $f_{i+1} \gtrsim \Delta_{i+1}$ and $k \gtrsim \varphi$. Then, $\lambda n.k \gtrsim \lambda\psi.\varphi$. Thus, the induction hypothesis $f_{i+1} \gtrsim \Delta_{i+1}$ implies using the last part of Lemma 8.5.3

$$\lambda k.\varphi^*(\max\{f_{i+1}, \lambda n.k\}) \gtrsim \Delta_i,$$

Completing the proof of (8.12).

We now prove by induction on i that $\hat{f}_i(k) \leq f_i(k)$ for all i and k, which will complete the proof of the lemma. The induction start $i = 0$ is trivial. For notational simplicity, we write $g_{i,k}(n) := \max\{\hat{f}_i(n), k\}$. Now observe that, since $\tilde{f}$ is monotone and satisfies $f(n) \geq n$ for all positive integers n, so does $\hat{f}_i$ for each i. Therefore, parts (ii) and (iii) of Lemma 8.5.3 imply

$$\begin{aligned}\hat{f}_{i_0-i-1}(k) &= g_{i_0-i,k}^{(n_{\bar{\varepsilon}})}(1) = g_{i_0-i,k}^{(n_{\bar{\varepsilon}}-1)}(g_{i_0-i,k}(1)) = g_{i_0-i,k}^{(n_{\bar{\varepsilon}}-1)}(\max\{\hat{f}_{i_0-i}(1), k\}) \\ &= g_{i_0-i,k}^{(n_{\bar{\varepsilon}}-1)}(\max\{\hat{f}_{i_0-i}(1), k\}) = \hat{f}_{i_0-i}^{(n_{\bar{\varepsilon}}-1)}(\max\{\hat{f}_{i_0-i}(1), k\}).\end{aligned}$$

Using the induction hypothesis and the monotonicity of $\tilde{f}$, we then see that

$$\hat{f}_{i_0-i-1}(k) = \max\left\{\hat{f}_{i_0-i}^{(n_{\bar{\varepsilon}})}(1), \hat{f}_{i_0-i}^{(n_{\bar{\varepsilon}}-1)}(k)\right\} \leq \hat{f}_{i_0-i}^{(n_{\bar{\varepsilon}})}(k) \leq \tilde{f}_{i_0-i}^{(n_{\bar{\varepsilon}})}(k) = \tilde{f}^{\left(n_{\bar{\varepsilon}}^{i}\cdot n_{\bar{\varepsilon}}\right)}(k) = f_{i_0-i-1}(k).$$

□

Lemma 8.5.5. *Suppose $f : \mathbb{N}^* \to \mathbb{N}^*$ is monotone, satisfies $f(n) \geq n$ for all positive integers n and $f \gtrsim \Delta$.*

Proof. Define for each nonnegative integer $i \leq i_0$ the integer k_i by

$$k_0 := \tilde{f}_0^{(n_{\bar{\varepsilon}})}(1), \quad k_{i+1} := \tilde{f}_i^{(n_{\bar{\varepsilon}})}(k_i).$$

We first show that $k_i \gtrsim \varphi_i$ for all $0 \leq i \leq i_0$ by induction on i. The base case follows from (iii) of Lemma 8.5.3 using the fact that f_i is monotone. The induction step follows from Lemma 8.5.3 and (we write $g_{i,k}(n) := \max\{\tilde{f}_i(n), \lambda n.k_{i-1}\}$)

$$g_{i,k}^{(n_{\bar{\varepsilon}})}(1) = g_{i,k}^{(n_{\bar{\varepsilon}}-1)}(\max\{\tilde{f}_i(1), k_{i-1}\}) = g_{i,k}^{(n_{\bar{\varepsilon}}-1)}(k_{i-1}) = \tilde{f}_i^{(n_{\bar{\varepsilon}}-1)}(k_{i-1}) \leq \tilde{f}_i^{(n_{\bar{\varepsilon}})}(k_{i-1}) = k_i.$$

□

We can now state the solution to Problem 8.4.4:

Theorem 8.5.6. *Suppose C is a closed, bounded, convex subset of a Hilbert space H with* $\mathrm{diam}(C) \leq d$ *for some nonnegative integer d, $T : C \to C$ is nonexpansive and $\mathcal{G} : C \to C$ is τ-contractive. For $\varepsilon \in (0,1], t \in [0,1], \Delta : C \times (C \to (0,1]) \to (0,1]$ and $V : C \times (C \to (0,1]) \to H$, one can construct $u := u_{v_0,T}(t, \varepsilon, \Delta, V) \in C$ and $\varphi := \varphi_{v_0,T,t,\varepsilon}(\Delta, V) : C \to (0,1]$ such that*

$$\|u - Tu\| < \Delta(u, \varphi)$$

and

$$\begin{aligned}&\|TV(u,\varphi) - V(u,\varphi)\| < \varphi(V(u,\varphi)) \\ &\qquad \to \|\mathcal{G}u - u\|^2 < \|(1-t)\mathcal{G}u - tV(u,\varphi)\|^2 + \varepsilon\end{aligned}$$

In fact, u, φ can be defined explicitly as functionals in Δ, V. Moreover, if we define a mapping $K : \mathbb{N}^{*\mathbb{N}^*} \to \mathbb{N}^*$ by $K(f) := k_{i_0}(\tilde{f})$, then

$$K \gtrsim \varphi,$$

where $f_i(k) := \tilde{f}^{(n^i_{\tilde{\varepsilon}})}(k)$ and

$$\begin{aligned} k_0(\tilde{f}) &:= \tilde{f}_0^{(n_{\tilde{\varepsilon}})}(1), & k_{i+1}(f) &:= \tilde{f}_i^{(n_{\tilde{\varepsilon}})}(k_i(f)) \\ \tilde{f}(n) &:= \max\{f^M(16dn^2), 16dn^2\}, & n_{\tilde{\varepsilon}} &:= \lceil 8d^4/\tilde{\varepsilon}^4 \rceil, \\ i_0 &:= \lceil \log_\tau(\tilde{\varepsilon}/6d^2) - 1 \rceil, & \tilde{\varepsilon} &:= \frac{(1-\tau)^2}{6+8d}\varepsilon. \end{aligned}$$

8.6 Strong Convergence Results

In this section, we prove our main results for the case of a single nonexpansive mapping $T : C \to C$. We start by giving a quantitative version of the convergence of the resolvent (v_n), where v_n is defined for each nonnegative integer n as the unique point satisfying the equation

$$v_n = (1 - \lambda_n)Tv_n + \lambda_n \mathcal{G} T v_n. \tag{8.13}$$

and $(\lambda_n) \subset (0, 1]$ is a null sequence.

Lemma 8.6.1 (cf. [104])**.** *The mapping $T^{(\lambda)} : C \to C$ defined by $T^{(\lambda)}(x) := (1-\lambda)Tx + \lambda\mathcal{G}(Tx)$ is a strict contraction with Lipschitz constant $(1 - \lambda(1 - \tau))$.*

First of all, we need the following lemma, which is similar to [52]:

Lemma 8.6.2. *Suppose $\lambda_n \in (0, 1]$, $u^*, v \in C$, $d \in \mathbb{N}$ and $h : \mathbb{N} \to \mathbb{N}$ satisfy $\lambda_n \geq \frac{1}{h(n)}$ and $\|v_n - u^*\| \leq d$. Then*

(i) $\|Tu^* - u^*\| \leq \dfrac{\varepsilon^2}{9d(1-\tau)\cdot h(n)}$, *(ii)* $\langle \mathcal{G}u^* - u^*, v_n - v\rangle \leq \dfrac{\varepsilon^2}{3(1-\tau)}$, *and*
(iii) $\langle \mathcal{G}u^* - u^*, v - u^*\rangle \leq \dfrac{\varepsilon^2}{3(1-\tau)}$

imply $\|v_n - u^*\| \leq \varepsilon$.

Proof. Observe that

$$\begin{aligned} &(1 - \lambda_n)(v_n - u^* - Tv_n + Tu^*) + \lambda_n(v_n - u^* - \mathcal{G}Tv_n + \mathcal{G}Tu^*) \\ &\quad = v_n - u^* - Tv_n + Tu^* - \lambda_n(v_n - u^*) + \lambda_n Tv_n - \lambda_n Tu^* + \lambda_n(v_n - u^*) \\ &\qquad - \lambda_n \mathcal{G}Tv_n + \lambda_n \mathcal{G}Tu^* \\ &\quad = v_n - Tv_n + \lambda_n Tv_n - \lambda_n \mathcal{G}Tv_n - u^* + Tu^* - \lambda_n Tu^* + \lambda_n \mathcal{G}Tu^* \\ &\quad = Tu^* - u^* + \lambda_n(\mathcal{G}Tu^* - Tu^*). \end{aligned} \tag{8.14}$$

Moreover,

$$\begin{aligned}\langle v_n - u^* - Tv_n + Tu^*, v_n - u^*\rangle &= \|v_n - u^*\|^2 - \langle Tv_n - Tu^*, v_n - u^*\rangle \\ &\geq \|v_n - u^*\|^2 - \|Tv_n - Tu^*\| \cdot \|v_n - u^*\| \\ &\geq \|v_n - u^*\|^2 - \|v_n - u^*\|^2 = 0, \end{aligned} \tag{8.15}$$

and

$$\begin{aligned}&\lambda_n \langle v_n - u^* - \mathcal{G}Tv_n + \mathcal{G}Tu^*, v_n - u^*\rangle \\ &\quad = \lambda_n\|v_n - u^*\|^2 - \lambda_n \langle \mathcal{G}Tv_n - \mathcal{G}Tu^*, v_n - u^*\rangle \\ &\quad \geq \lambda_n\|v_n - u^*\|^2 - \lambda_n\|\mathcal{G}Tv_n - \mathcal{G}Tu^*\| \cdot \|v_n - u^*\| \\ &\quad \geq \lambda_n(1-\tau)\|v_n - u^*\|^2. \end{aligned} \tag{8.16}$$

Combining (8.14), (8.15) and (8.16),

$$\begin{aligned}\lambda_n(1-\tau)\|v_n - u^*\|^2 &\leq \lambda_n \langle v_n - u^* - \mathcal{G}Tv_n + \mathcal{G}Tu^*, v_n - u^*\rangle \\ &\leq \lambda_n \langle v_n - u^* - \mathcal{G}Tv_n + \mathcal{G}Tu^*, v_n - u^*\rangle \\ &\qquad + (1-\lambda_n) \langle v_n - u^* - Tv_n + Tu^*, v_n - u^*\rangle \\ &\leq \langle Tu^* - u^*, v_n - u^*\rangle + \lambda_n \langle \mathcal{G}Tu^* - Tu^*, v_n - u^*\rangle \\ &\leq \|Tu^* - u^*\| \cdot \|v_n - u^*\| + \lambda_n \langle \mathcal{G}Tu^* - Tu^*, v_n - u^*\rangle \\ &= d \cdot \|Tu^* - u^*\| + \lambda_n \langle \mathcal{G}Tu^* - \mathcal{G}u^*, v_n - u^*\rangle \\ &\qquad + \lambda_n \langle \mathcal{G}u^* - u^*, v_n - u^*\rangle + \lambda_n \langle u^* - Tu^*, v_n - u^*\rangle \\ &\leq d(1 + \tau\lambda_n + \lambda_n) \cdot \|Tu^* - u^*\| + \lambda_n\langle \mathcal{G}u^* - u^*, v_n - u^*\rangle \\ &= d(1 + \tau\lambda_n + \lambda_n) \cdot \|Tu^* - u^*\| + \lambda_n \langle \mathcal{G}u^* - u^*, v_n - v\rangle \\ &\qquad + \lambda_n\langle \mathcal{G}u^* - u^*, v - u^*\rangle. \end{aligned}$$

Therefore,

$$(1-\tau)\|v_n - u^*\|^2 \leq \frac{3d}{\lambda_n} \cdot \|Tu^* - u^*\| + \langle \mathcal{G}u^* - u^*, v_n - v\rangle + \langle \mathcal{G}u^* - u^*, v - u^*\rangle.$$

Now, the claim follows from the assumptions (i), (ii) and (iii). □

Corollary 8.6.3. *If we instantiate $v := v_n$, then (ii) becomes true with '$= 0$' instead of '$\leq \varepsilon^2/3(1-\tau)$', so we get that*

$$\|Tu^* - u^*\| \leq \frac{\varepsilon^2}{6d(1-\tau) \cdot h(n)}, \quad \textit{and} \quad \langle \mathcal{G}u^* - u^*, v_n - u^*\rangle \leq \frac{\varepsilon^2}{2(1-\tau)}$$

imply $\|v_n - u^*\| \leq \varepsilon$.

From here on, we follow except for a few minor details the argumentation of [52]. For the sake of completeness, we adapt the proof to our situation.

Lemma 8.6.4. *For $t \in (0,1]$, denote by z_t the unique point satisfying $z_t = (1-t)Tz_t + t\mathcal{G}Tz_t$. Then $\|z_t - Tz_t\| < \varepsilon$ for all $\varepsilon > 0$ and $0 < t < \varepsilon/d$.*

Proof. Follows from

$$\|z_t - Tz_t\| = t\|Tz_t - \mathcal{G}Tz_t\| \leq td < \varepsilon.$$

□

Lemma 8.6.5 (Lemma 2.9 of [52]). *Let X be a normed linear space. Then the following holds:*

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \forall u \in X \forall (v_n) \subset X \, \forall m \in \mathbb{N} \left(\|v_{g_{u,\varepsilon}(m)} - u\| \leq \varepsilon/2 \to \|v_{g(m)} - v_m\| \leq \varepsilon \right),$$

where

$$g_{u,\varepsilon}(m) := \begin{cases} g(m), & \text{if } \|v_{g(m)} - u\| > \varepsilon/2, \\ m, & \text{otherwise.} \end{cases}$$

Lemma 8.6.6 (Lemma 2.13 of [52]). *Let $\chi : \mathbb{N} \to \mathbb{N}$ be a rate of convergence of (λ_n) towards 0, i.e. $\lambda_i \leq \frac{1}{n+1}$ for all nonnegative integers n and all $i \geq \chi(n)$. Then, for (v_n) as defined in (8.13) and $\tilde{g}_{u,\varepsilon}$ defined as in Lemma 8.6.5 (but with $\tilde{g}(n) := \max\{n, g(n)\}$),*

$$\forall \varepsilon \in (0,1] \, \forall g : \mathbb{N}^* \to \mathbb{N}^* \, \forall \varphi : C \to (0,1] \, \forall u \in C \, \forall k \gtrsim \varphi \\ \left(\|Tv_{\tilde{g}_{u,\varepsilon}(\chi(d \cdot k))} - v_{\tilde{g}_{u,\varepsilon}(\chi(d \cdot k))}\| < \varphi(v_{\tilde{g}_{u,\varepsilon}(\chi(d \cdot k))}) \right).$$

Theorem 8.6.7. *Let H be a real Hilbert space, $d \in \mathbb{N}^*$ and $C \subset H$ be a bounded closed convex subset with $d \geq \operatorname{diam} C$. Let $T : C \to C$ be nonexpansive and $\mathcal{G} : C \to C$ be a strict contraction with Lipschitz constant $\tau < 1$. Let (λ_n) be a sequence in $(0,1]$ that converges towards 0 and $h : \mathbb{N} \to \mathbb{N}^*$ such that $\lambda_n \geq \frac{1}{h(n)}$ for all $n \in \mathbb{N}$. Let $\chi : \mathbb{N} \to \mathbb{N}$ be a rate of convergence of (λ_n) towards 0, i.e. $\lambda_i \leq \frac{1}{n+1}$ for all nonnegative integers n and all $i \geq \chi(n)$. Denote by v_n the unique solution to the equation*

$$v_n = (1-\lambda_n)Tv_n + \lambda_n \mathcal{G}Tv_n.$$

Then, for all $\varepsilon \in (0,1]$ and $g : \mathbb{N} \to \mathbb{N}^*$

$$\exists j \leq \Xi(\varepsilon, g, \chi, h, d) \left(\|v_j - v_{\tilde{g}(j)}\| \leq \varepsilon \right),$$

where $\tilde{g}(n) := \max\{n, g(n)\}$ and

$$\Xi(\varepsilon, g, \chi, h, d) := \chi\left(d \cdot k_{i_0}(\tilde{f})\right),$$

where $i_0 := \lceil \log_\tau(\tilde{\varepsilon}/6d^2) - 1 \rceil$ *and*

$$f(n) := \left\lceil \frac{6d(1-\tau)h^M(\tilde{g}^M(\chi(d\cdot n)))}{(\varepsilon/2)^2} \right\rceil,$$

$$\tilde{f}(n) := \max\{f^M(16dn^2), 16dn^2\}, \qquad \tilde{g}(n) := \max\{n, g(n)\},$$

$$k_0(\tilde{f}) := \tilde{f}_0^{(n_{\tilde{\varepsilon}})}(1), \qquad k_{i+1}(\tilde{f}) := \tilde{f}_i^{(n_{\tilde{\varepsilon}})}(k_i(f)),$$

$$f_i(k) := \tilde{f}^{\left(n^i_{\tilde{\varepsilon}}\right)}(k), \qquad \tilde{f}_i(n) := \max\{f_i^M(16dn^2), 16dn^2\},$$

$$n_{\tilde{\varepsilon}} := \lceil 8d^4/\tilde{\varepsilon}^4 \rceil, \qquad \tilde{\varepsilon} := \frac{(1-\tau)^2}{6+8d}\varepsilon_d,$$

$$\varepsilon_d := \frac{(\varepsilon/2)^4}{8(1-\tau)^2 d^2}.$$

Proof. For $\varepsilon \in (0,1]$ and $g : \mathbb{N} \to \mathbb{N}$ define analogously to [52] a functional $J_{\varepsilon,g} : C \times (C \to (0,1]) \to \mathbb{N}$ by

$$J_{\varepsilon,g}(u,\varphi) := \begin{cases} \min\left\{ j \in \mathbb{N} : \|T(v_{\tilde{g}_{u,\varepsilon}(j)}) - v_{\tilde{g}_{u,\varepsilon}(j)}\| < \varphi(v_{\tilde{g}_{u,\varepsilon}(j)}) \right\}, & \text{if such a } j \text{ exists,} \\ 0, & \text{otherwise,} \end{cases}$$

where $\tilde{g}_{u,\varepsilon}$ is defined as $g_{u,\varepsilon}$ in Lemma 8.6.5 with $\tilde{g}$ instead of g. Observe that by Lemma 8.6.6, we are always in the first case of the definition of $J_{\varepsilon,g}$ whenever $\varphi : C \to (0,1]$ is majorizable. Moreover, we have

$$k \gtrsim \varphi \to \forall u \in C\big(J_{\varepsilon,g}(u,\varphi) \leq \chi(d\cdot k)\big). \tag{8.17}$$

Now define $V_{\varepsilon,g} : C \times (C \to (0,1]) \to C$ by $V_{\varepsilon,g}(u,\varphi) := v_{\tilde{g}_{u,\varepsilon}(J_{\varepsilon,g}(u,\varphi))}$, $\varepsilon_d := \frac{(\varepsilon/2)^4}{8(1-\tau)^2 d^2}$ and $t := \frac{(\varepsilon/2)^2}{6(1-\tau)^2 d^2}$. Moreover, if we define $\Delta_{\varepsilon,g}(u,\varphi) := \frac{(\varepsilon/2)^2}{6d(1-\tau)\cdot h(\tilde{g}_{u,\varepsilon}(J_{\varepsilon,g}(u,\varphi)))}$, then, given a majorant $k \gtrsim \varphi$, (8.17) implies $J_{\varepsilon,g}(u,\varphi) \leq \chi(d\cdot k)$ for all $u \in C$. Hence

$$k \gtrsim \varphi \to \forall u \in C\left(\frac{(\varepsilon/2)^2}{6d(1-\tau)\cdot h(\tilde{g}_{u,\varepsilon}(J_{\varepsilon,g}(u,\varphi)))} \geq \left\lceil \frac{(\varepsilon/2)^2}{6d(1-\tau)h^M\left(\tilde{g}^M\left(\chi(d\cdot k)\right)\right)} \right\rceil \right). \tag{8.18}$$

Therefore, $f \gtrsim \Delta_{\varepsilon,g}$. We write $\tilde{u} := U(\varepsilon_d, t, \Delta_{\varepsilon,g}, V_{\varepsilon,g})$ and $\tilde{\varphi} := \Phi(\varepsilon_d, t, \Delta_{\varepsilon,g}, V_{\varepsilon,g})$ to simplify notation, where Φ and U are the solution functionals to Problem 8.4.4. By Theorem 8.5.6, we then get $k_{i_0}(f) \gtrsim \tilde{\varphi}$, whence (8.17) implies $J_{\varepsilon,g}(\tilde{u},\tilde{\varphi}) \leq \Xi(\varepsilon, g, \chi, h, d)$.

Then, for $j := J_{\varepsilon,g}(\tilde{u},\tilde{\varphi})$ and $v := V_{\varepsilon,g}(\tilde{u},\tilde{\varphi})$

$$\|T\tilde{u} - \tilde{u}\| < \frac{(\varepsilon/2)^2}{6d(1-\tau)\cdot h(\tilde{g}_{\tilde{u},\varepsilon}(j))} \tag{8.19}$$

and

$$\|Tv - v\| < \tilde{\varphi}(v) \to \|\mathcal{G}\tilde{u} - \tilde{u}\|^2 < \|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tv\|^2 + \varepsilon_d.$$

But $\|Tv - v\| < \tilde{\varphi}(v)$ by construction of $J_{\varepsilon,g}$, so

$$\|\mathcal{G}\tilde{u} - \tilde{u}\|^2 < \|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tv\|^2 + \varepsilon_d. \tag{8.20}$$

Lemma 8.4.5 then yields $\langle \mathcal{G}\tilde{u} - \tilde{u}, v - \tilde{u} \rangle < \frac{(\varepsilon/2)^2}{(1-\tau)}$, so Corollary 8.6.3 implies

$$\|v - \tilde{u}\| \leq \varepsilon/2. \tag{8.21}$$

From Lemma 8.6.5, and the definitions of v and $J_{\varepsilon,g}$, we conclude $\|v_{\tilde{g}(j)} - v_j\| \leq \varepsilon$. □

Corollary 8.6.8. *For all $\varepsilon \in (0,1]$ and $g : \mathbb{N} \to \mathbb{N}$, there exists an $n \leq \Xi(\varepsilon/2, \lambda n.n + g(n), \chi, h, d)$ such that*

$$\|v_i - v_j\| \leq \varepsilon, \quad \textit{for all } i,j \in [n; n+g(n)].$$

Proof. Follows as in [52]. □

Lemma 8.6.9 (Modulus of Continuity for the VIP)**.** *Suppose $u, v, w \in C$ satisfy $\|u - v\| \leq \frac{\varepsilon}{2d(2+\tau)}$ and $\langle \mathcal{G}u - u, w - u \rangle \leq \varepsilon/2$. Then $\langle \mathcal{G}v - v, w - v \rangle \leq \varepsilon$.*

Proof. Follows from

$$\begin{aligned}
\langle \mathcal{G}v - v, w - v \rangle &= \langle \mathcal{G}v - v, w - u \rangle + \langle \mathcal{G}v - v, u - v \rangle \\
&= \langle \mathcal{G}u - u, w - u \rangle + \langle \mathcal{G}v - \mathcal{G}u, w - u \rangle + \langle u - v, w - u \rangle + \langle \mathcal{G}v - v, u - v \rangle \\
&\leq \frac{\varepsilon}{2} + d(2+\tau)\|u - v\| \leq \varepsilon.
\end{aligned}$$

□

Theorem 8.6.10. *In the situation of Theorem 8.6.7, suppose that $\phi_1, \phi_2 : (0,\infty) \to \mathbb{N}$ satisfy*

(i) $\sum_{i=1}^{\phi_1(k)} \lambda_i \geq k$ *for all nonnegative integers k, and*

(ii) $\frac{|\lambda_n - \lambda_{n+1}|}{\lambda_{n+1}^2} \leq \varepsilon$ *for all $\varepsilon > 0$ and $n \geq \phi_2(\varepsilon)$.*

Define the sequence (u_n) by $u_{n+1} := (1-\lambda_{n+1})Tu_n + \lambda_{n+1}\mathcal{G}Tu_n$ for an arbitrary starting point $u_0 \in C$. Then, for all $\varepsilon \in (0,1]$, $g : \mathbb{N} \to \mathbb{N}$ and $v \in C$, there exists an $n \leq \Xi(\varepsilon/6, g_c, \chi, h, d) + c$, where $g_c(n) := n + c + g(n+c)$ and $c := \phi_1\left(\frac{(\phi_2((1-\tau)\varepsilon/6d) + \log(6d/\varepsilon))}{1-\tau}\right)$ such that

$$\|u_i - u_j\| \leq \varepsilon \quad \textit{for all } i,j \in [n; n+g(n)].$$

Proof. If we define $p := \phi_2((1-\tau)\varepsilon/6d)$, Equation (29) of [104] implies (see also the remarks preceding Lemma 8.6.1)

$$\|u_n - v_n\| \leq \|u_p - v_p\| \prod_{i=p+1}^{n} (1 - \lambda_i(1-\tau)) + \varepsilon/6, \text{ for all } n \geq p. \tag{8.22}$$

Now, for $n \geq \phi_1\left(\frac{(p + \log(6d/\varepsilon))}{1-\tau}\right)$, we have $\sum_{i=1}^{n} \lambda_i(1-\tau) \geq p + \log(6d/\varepsilon)$, so

$$\sum_{i=p+1}^{n} \lambda_i(1-\tau) \geq -p(1-\tau) + \sum_{i=1}^{n} \lambda_i(1-\tau) \geq \log(6d/\varepsilon).$$

Therefore, since $-x > \log(1-x)$ for all $0 < x < 1$,

$$\frac{\varepsilon}{6d} \geq \exp\left(-\sum_{i=p+1}^{n} \lambda_i(1-\tau)\right) > \exp\left(\sum_{i=p+1}^{n} \log(1-\lambda_i(1-\tau))\right)$$
$$= \exp\left(\log\left(\prod_{i=p+1}^{n} (1-\lambda_i(1-\tau))\right)\right).$$

Now observe that $\phi_1(k) \geq k$ for all nonnegative integers k, so $\phi_1\left(\frac{(p+\log(6d/\varepsilon))}{1-\tau}\right) \geq p$. Going back to (8.22), we therefore see that

$$\|u_n - v_n\| \leq \varepsilon/3, \text{ for all } n \geq \phi_1\left(\frac{(\phi_2((1-\tau)\varepsilon/6d) + \log(6d/\varepsilon))}{1-\tau}\right). \tag{8.23}$$

Moreover, by Corollary 8.6.8, there exists an $n \leq \Xi(\varepsilon/6, \tilde{g}, \chi, h, d)$ such that

$$\|v_i - v_j\| \leq \varepsilon/3, \quad \text{for all } i, j \in [n, n+c+g(n+c)].$$

Thus, $\tilde{n} := n + c$ satisfies

$$\|u_i - u_j\| \leq \|u_i - v_i\| + \|u_j - v_j\| + \|v_i - v_j\| \leq \varepsilon, \quad \text{for all } i, j \in [\tilde{n}, \tilde{n} + g(\tilde{n})].$$

□

Theorem 8.6.11. *In the situation of Theorem 8.6.10, the following holds: For all $\varepsilon \in (0,1]$, $g : \mathbb{N} \to \mathbb{N}$ and $x \in C$, there exists an $n \leq \Xi(\delta/6, g_c, \chi, h, d) + c$, where $\delta = \frac{\varepsilon}{2d(2+\tau)}$, $g_c(n) := n + c + g(n+c)$ and $c := \phi_1\left(\frac{1}{\tau}(\phi_2(\delta/6) + \log(6d/\delta))\right)$ such that*

(i) $\|u_i - u_j\| \leq \varepsilon$ for all $i, j \in [n; n+g(n)]$, and

(ii) if $x \in C$ satisfies $\|Tx - x\| \leq \varepsilon'$, then $\langle \mathcal{G}u_n - u_n, x - u_n \rangle \leq \varepsilon$,

where $\varepsilon' := k'_{i_0}(f)$ and $k'_{i_0}(f)$ is defined as $k_{i_0}(f)$ in Theorem 8.6.7, but with $\delta/6$ instead of ε.

Proof. We first prove that for all $\varepsilon \in (0,1]$, $x \in C$ and $g : \mathbb{N} \to \mathbb{N}$, there exists a nonnegative integer $j \leq \Xi(\varepsilon, g, \chi, h, d)$ and a $\tilde{u}' \in C$ such that

$$\|v_j - v_{\tilde{g}(j)}\| \leq \varepsilon, \text{ and } \quad \|Tx - x\| \leq \varepsilon' \to \langle \mathcal{G}\tilde{u}' - \tilde{u}', x - \tilde{u}' \rangle \leq \frac{(\varepsilon/2)^2}{2}, \tag{8.24}$$

where $\varepsilon' := \alpha^{(i_0 \cdot (n_\varepsilon - 1))}(\max\{\varphi^*(f), \alpha(1)\})$.

In the proof of Theorem 8.6.7, after equation (8.18), one can alter the counterfunction $V_{\varepsilon,g}$ to $V'_{\varepsilon,g} : C \times (C \to (0,1]) \to C$ defined by

$$V'_{\varepsilon,g}(u,\varphi) := \begin{cases} V_{\varepsilon,g}(u,\varphi), & \text{if } \|\mathcal{G}u - V^t_{\varepsilon,g}(u,\varphi)\| \leq \|\mathcal{G}u - (1-t)u - tx\|, \\ x, & \text{otherwise.} \end{cases}$$

For $\tilde{u}' := U(\varepsilon_d, t, \Delta_{\varepsilon,g}, V'_{\varepsilon,g})$ and $\tilde{\varphi}' := \Phi(\varepsilon_d, t, \Delta_{\varepsilon,g}, V'_{\varepsilon,g})$ we then get for $j := J_{\varepsilon,g}(\tilde{u}', \tilde{\varphi}')$ and $v := V'_{\varepsilon,g}(\tilde{u}', \tilde{\varphi}')$ as before

$$\|T\tilde{u}' - \tilde{u}'\| < \frac{(\varepsilon/2)^2}{6d \cdot h(\tilde{g}_{\tilde{u},\varepsilon}(j))} \tag{8.25}$$

and

$$\|Tv - v\| < \tilde{\varphi}(v) \to \|\mathcal{G}\tilde{u}' - \tilde{u}'\|^2 < \|\mathcal{G}\tilde{u}' - (1-t)\tilde{u}' - tv\|^2 + \varepsilon_d.$$

Then, by construction of $V'_{\varepsilon,g}(u, \varphi)$, we now get two implications; As before,

$$\begin{aligned}\|TV_{\varepsilon,g}(\tilde{u}', \varphi') - V_{\varepsilon,g}(\tilde{u}', \varphi')\| &< \tilde{\varphi}(V_{\varepsilon,g}(\tilde{u}', \varphi')) \\ &\to \|\mathcal{G}\tilde{u}' - \tilde{u}'\|^2 < \|\mathcal{G}\tilde{u}' - (1-t)\tilde{u}' - tV_{\varepsilon,g}(\tilde{u}', \varphi')\|^2 + \varepsilon_d,\end{aligned} \tag{8.26}$$

and, additionally,

$$\|Tx - x\| < \tilde{\varphi}'(x) \to \|\mathcal{G}\tilde{u}' - \tilde{u}'\|^2 < \|\mathcal{G}\tilde{u}' - (1-t)\tilde{u}' - tx\|^2 + \varepsilon_d. \tag{8.27}$$

Now, (8.25) and (8.26) imply as before $\|v_j - v_{\tilde{g}(j)}\| \le \varepsilon$. Moreover,

$$\|Tx - x\| < \tilde{\varphi}'(x) \to \langle \mathcal{G}\tilde{u}' - \tilde{u}', x - \tilde{u}' \rangle < \frac{(\varepsilon/2)^2}{2}.$$

Now observe that the majorant of the solution operator Φ is independent of the counterfunction $V'_{\varepsilon,g}$; therefore, we may take the same majorant for $\tilde{\varphi}'$ as we took for $\tilde{\varphi}$. Therefore

$$\|Tx - x\| < k_{i_0}(f) \to \langle \mathcal{G}\tilde{u}' - \tilde{u}', x - \tilde{u}' \rangle < \frac{(\varepsilon/2)^2}{2}. \tag{8.28}$$

This completes the proof of (8.24).

Thus, we get in Theorems 8.6.7 and 8.6.10 also the additional conclusion (8.28). Thus, as before in Theorem 8.6.10, we get an $n \le \Xi(\delta/6, g_c, \chi, h, d) + c$ such that

$$\|u_i - u_j\| \le \delta \le \varepsilon, \quad \text{for all } i, j \in [n; n + g(n)].$$

Moreover, as in the situation of Theorem 8.6.7, we get $\|v - \tilde{u}'\| \le \frac{\delta/6}{2}$ (compare (8.21)). Similarly, we get as in Theorem 8.6.10 that $\|u_n - v_n\| \le \delta/3$ (compare (8.23)). Observe also that $\tilde{g}_{\tilde{u},\varepsilon}(n)$ is either n or $\tilde{g}(n)$, so $v = v_{\tilde{g}_{\tilde{u},\varepsilon}(n)}$ implies

$$\begin{aligned}\|u_n - \tilde{u}'\| &\le \|u_n - v_n\| + \|v_n - v\| + \|v - \tilde{u}'\| \\ &\le \frac{\delta}{3} + \|v_n - v_{\tilde{g}(n)}\| + \frac{\delta}{12} \le \delta = \frac{\varepsilon}{2d(2+\tau)}.\end{aligned} \tag{8.29}$$

The claim follows from (8.29) and (8.28) using Lemma 8.6.9. □

Corollary 8.6.12. *For all of the above results, one can drop the condition of C being bounded with $\operatorname{diam}(C) \le d$ in favor of T having a fixed point v such that $\|u_0 - v\| \le d/2$, $\|v - \mathcal{G}v\| \le \frac{d(1-\tau)}{4}$ and $\|v - w\| \le \frac{d}{4(1+\tau)}$, where w is the unique fixed point of $\mathcal{G}$.*

Proof. By Lemma 8.6.1, we have for all nonnegative integers n

$$\begin{aligned}\|u_{n+1} - v\| &\leq \|T^{(\lambda_{n+1})}(u_n) - T^{(\lambda_{n+1})}(v)\| + \|T^{(\lambda_{n+1})}(v) - v\| \\ &\leq (1 - \lambda_{n+1}(1-\tau))\|u_n - v\| + \lambda_{n+1}(1-\tau) \cdot \frac{\|\mathcal{G}v - v\|}{1-\tau}.\end{aligned}$$

Since $\|u_0 - v\| \leq d/2$, we conclude by induction that $\|u_n - v\| \leq d/2$ for all nonnegative integers n.

Moreover, observe that Lemma 8.6.1 implies

$$\begin{aligned}\|v_n - v\| &\leq \|T^{(\lambda_n)}(v_n) - T^{(\lambda_n)}(v)\| + \|T^{(\lambda_n)}(v) - v\| \\ &\leq (1 - \lambda_n(1-\tau))\|v_n - v\| + \lambda_n(1-\tau) \cdot \frac{\|v - \mathcal{G}v\|}{1-\tau},\end{aligned}$$

so, since λ_n is strictly positive, $\|v_n - v\| \leq \frac{\|\mathcal{G}v - v\|}{1-\tau} \leq d/4 \leq d/2$. Moreover,

$$\|\mathcal{G}v_n - v\| \leq \|\mathcal{G}v_n - \mathcal{G}w\| + \|w - v\| \leq \tau\|v_n - w\| + \|v - w\| \leq \tau\|v_n - v\| + (1+\tau)\|v - w\|.$$

Therefore, the sequences (v_n), $(\mathcal{G}v_n)$ and (u_n) remain in the ball of radius $d/2$ (and therefore diameter d) around v. Since the estimate $\mathrm{diam}(C) \leq d$ was only ever used for elements of the sequences (v_n), $(\mathcal{G}v_n)$ and (u_n), and convex combinations of those elements, the claim follows.

□

8.7 Finite Families

For the rest of this section, let C be a closed and convex subset of H. Suppose that $T_1, \ldots, T_N : C \to C$ are nonexpansive mappings with a common fixed point $p \in C$ which satisfy $\bigcap_{i=1}^N Fix(T_i) = Fix(T_N \cdots T_1)$. Then a function $\hat{\rho} : \mathbb{N} \times (0,\infty) \to (0,\infty)$ is a modulus for this property if, for all nonnegative integers d, all $x \in C$ and all $\varepsilon > 0$

$$\|x - p\| \leq d \text{ and } \|T_N \cdots T_1 x - x\| < \hat{\rho}(d,\varepsilon) \text{ imply } \|T_i x - x\| < \varepsilon, \text{ for all } 1 \leq i \leq N. \tag{8.30}$$

It is clear that one can, without loss of generality, assume that $\hat{\rho}$ is monotone in ε and satisfies $\hat{\rho}(d,\varepsilon) \leq \varepsilon$ for all $\varepsilon > 0$ and all $d \in \mathbb{N}$, which we do from now on.

In [104], Yamada actually assumes that

$$\bigcap_{i=1}^N Fix(T_i) = Fix(T_N \cdots T_1) = Fix(T_{N-1} \cdots T_1 T_N) = \ldots = Fix(T_1 \cdots T_N),$$

which is the well-known Bauschke condition [4]. In [93], however, Suzuki showed[1] that the Bauschke condition is already implied by the case for e.g. $T_N \cdots T_1$. We now give a quantitative account of this:

[1] The author is most greatful to Prof. Genaro López Acedo for pointing out this result.

Theorem 8.7.1. *Supppose C is a bounded closed convex subset of a Hilbert space H with diameter* $\operatorname{diam}(C) \leq d$*, and the nonexpansive mappings $T_1, \ldots, T_N$ satisfy* (8.30)*. Then, if*

$$\|T_{N-k} \cdots T_1 T_N \cdots T_{N-k+1} x - x\| < \hat{\rho}\left(d, \frac{\varepsilon}{2N+1}\right)$$

holds for some $k \in \{1, \ldots, N-1\}$, then

$$\|T_i x - x\| < \varepsilon, \quad \text{for all } i \in \{1, \ldots, N\}.$$

Proof. Suppose $\|T_{N-k} \cdots T_1 T_N \cdots T_{N-k+1} x - x\| < \hat{\rho}(d, \delta)$, where we write $\delta := \varepsilon/(2N+1)$. Then, since $T_N \cdots T_{N-k+1}$ is nonexpansive,

$$\|T_N \cdots T_1 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| < \hat{\rho}(d, \delta)$$

By hypothesis (8.30), this implies

$$\|T_i T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| < \delta, \quad \text{for all } i \in \{1, \ldots, N\}. \tag{8.31}$$

Therefore,

$$\begin{aligned}
\|T_N \cdots T_{N-k+1} x - x\| &\leq \|T_{N-k} \cdots T_1 T_N \cdots T_{N-k+1} x - x\| \\
&\quad + \|T_{N-k} \cdots T_1 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&< \hat{\rho}(d, \delta) + \|T_{N-k} \cdots T_2 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\quad + \|T_{N-k} \cdots T_2 T_1 T_N \cdots T_{N-k+1} x - T_{N-k} \cdots T_2 T_N \cdots T_{N-k+1} x\| \\
&\leq \hat{\rho}(d, \delta) + \|T_{N-k} \cdots T_2 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\quad + \|T_1 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\leq \hat{\rho}(d, \delta) + \delta + \|T_{N-k} \cdots T_2 T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\leq \cdots \\
&\leq \hat{\rho}(d, \delta) + (N-k-1)\delta + \|T_{N-k} T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\leq \hat{\rho}(d, \delta) + (N-k)\delta \\
&\leq \hat{\rho}(d, \delta) + (N-1)\delta \leq N\delta.
\end{aligned}$$

Using once more (8.31), we then get

$$\begin{aligned}
\|T_i x - x\| &\leq \|T_i x - T_i T_N \cdots T_{N-k+1} x\| + \|T_i T_N \cdots T_{N-k+1} x - T_N \cdots T_{N-k+1} x\| \\
&\quad + \|T_N \cdots T_{N-k+1} x - x\| \\
&< (2N+1)\delta = \varepsilon, \quad \text{for all } i \in \{1, \ldots, N\}.
\end{aligned}$$

□

Notation 8.7.2. We write $C(N)$ for the set of permutations $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$ that are of the form

$$\pi(n) = n + k \mod N$$

for some $k \in \mathbb{N}$.

Now, if we are given a modulus $\hat{\rho}$ satisfying (8.30), we define $\rho : \mathbb{N} \times (0, \infty) \to \infty$ by $\rho(d, \varepsilon) := \hat{\rho}(d, \varepsilon/(2N+1))$. In light of the previous theorem, this new modulus ρ will then satisfy for any $\pi \in C(N)$ the implication

$$\|x - p\| \le d \text{ and } \|T_{\pi(N)} \cdots T_{\pi(1)} x - x\| < \rho(d, \varepsilon) \text{ imply } \|T_i x - x\| < \varepsilon, \text{ for all } 1 \le i \le N. \tag{8.32}$$

Observe that if all T_i are also strongly quasi-nonexpansive (SQNE) in the sense of Bruck [16], then one can transform an SQNE-modulus in the sense of Kohlenbach [54] into a function ρ satisfying (8.30):

Proposition 8.7.3 (see [54]). *Let (X, d) be a metric space and $S \subseteq X$ be a subset. Let $T_1, \ldots, T_N$ be SQNE-mappings with SQNE-moduli $\omega_1, \ldots, \omega_N$, respectively, with respect to some common fixed point $p \in S$ of $T_1, \ldots, T_N$ and let $d \in N$. Assume that $T_1, \ldots, T_N$ are uniformly continuous on $S_d := \{x \in S : d(x, p) \le d\}$ with modulus of continuity $\alpha : (0, \infty) \to (0, \infty)$, i.e. for all $\varepsilon > 0$ and all $y, y' \in S_d$,*

$$d(y, y') < \alpha(\varepsilon) \textit{ implies } d(T_i y, T_i y') < \varepsilon \textit{ for all } 1 \le i \le N.$$

For $\omega(d, \varepsilon) := \min_{1 \le i \le N} \omega_i(d, \varepsilon)$, define

$$\chi_d(0, \varepsilon) := \min\{\alpha(\varepsilon/2), \varepsilon\}, \quad \chi_d(n+1, \varepsilon) := \min\left\{\omega\left(d, \frac{1}{2}\chi_d(n, \varepsilon)\right), \frac{1}{2}\chi_d(n, \varepsilon)\right\}.$$

Then $\rho(d, \varepsilon) := \chi_d(N-1, \varepsilon)$ satisfies for all $x \in C$ and all $\varepsilon > 0$

$$d(x, p) \le d \textit{ and } d(T_N T_{N-1} \cdots T_1 x, x) < \rho(d, \varepsilon) \textit{ imply } d(T_i x, x) < \varepsilon \textit{ for all } 1 \le i \le N.$$

Observe that, if the T_i are SQNE and nonexpansive, then the identity on $(0, \infty)$ is a modulus of continuity α in the sense of the proposition above.

Consider now the following iteration scheme (see e.g. [104])

$$u_{n+1} := T_{[n+1]}^{(\lambda_{n+1})}(u_n) := (1 - \lambda_{n+1}) T_{[n+1]}(u_n) + \lambda_{n+1} \mathcal{G} T_{[n+1]}(u_n), \tag{8.33}$$

where $(\lambda_n) \subset (0, 1]$ and $[n] := n \mod N$.

Lemma 8.7.4. *Suppose the closed convex set $C \subseteq H$ of a Hilbert space H is bounded with $\mathrm{diam}(C) \le d$ for all nonnegative integers n and $\chi : \mathbb{N} \to \mathbb{N}$ is a rate of convergence for (λ_n) to 0, i.e. $\lambda_n \le 1/k$ for all nonnegative integers $n \ge \chi(k)$. Then*

$$\|u_{n+1} - T_{[n+1]}(u_n)\| \le \frac{1}{k}, \quad \textit{for all nonnegative integers } n \ge \chi(d \cdot k).$$

Proof. Follows immediately from (8.33). □

Theorem 8.7.5. *Suppose C is bounded with $\mathrm{diam}\, C \le d$ and χ is as before. Given moduli $\phi_3 : (0, \infty) \times \mathbb{N} \to \mathbb{N}$ and $\phi_4 : (0, \infty) \to \mathbb{N}$ such that*

1. *$\phi_3(\varepsilon, n) \ge n$ for all $\varepsilon > 0$ and all $n \in \mathbb{N}$,*

2. $\prod_{i=n}^{m}(1-\lambda_i(1-\tau)) \leq \varepsilon$ *for all nonnegative integers* n, m *with* $m \geq \phi_3(\varepsilon, n)$*, and*
3. $\sum_{i=\phi_4(k)}^{\infty} |\lambda_{i+N} - \lambda_i| \leq \varepsilon$ *for all* $\varepsilon > 0$.

Then, for all $\varepsilon > 0$ *and all* $n \geq \hat{\chi}(\varepsilon) := \max\{\phi_3(\varepsilon/2d, \phi_4(\varepsilon/4d)), \chi(\lceil Nd/2\varepsilon \rceil)\}$*,*

$$\|u_n - T_{[n+N]} \cdots T_{[n+1]}(u_n)\| \leq \varepsilon.$$

Using Theorem 8.7.1, this theorem immediately implies the asymptotic regularity of (u_n) with respect to the mapping $T_{\pi(N)} \cdots T_{\pi(1)}$ for each $\pi \in C(N)$:

Corollary 8.7.6. *In the situation of Theorem 8.7.5, for all* $\pi \in C(n)$*, all* $\varepsilon > 0$ *and all* $n \geq \hat{\chi}(\rho(d, \varepsilon/N))$

$$\|u_n - T_{\pi(N)} \ldots T_{\pi(1)} u_n\| \leq \varepsilon.$$

Proof. Inequality (37) of [104] reads (see also the remarks preceding Lemma 8.6.1)

$$\|u_{n+N} - u_n\| \leq d \sum_{k=m+1}^{n} |\lambda_{k+N} - \lambda_k| + \|u_{m+N} - u_m\| \prod_{k=m+1}^{n} (1 - \lambda_{k+N}(1-\tau)),$$

for all $n > m \geq 0$.

Therefore, for $m = \phi_4(\varepsilon/2d) - 1$, we get for all $n \geq \phi_4(\varepsilon/2d) - 1$

$$\begin{aligned}\|u_{n+N} - u_n\| &\leq d \sum_{k=\phi_4(\varepsilon/2d)}^{\infty} |\lambda_{k+N} - \lambda_k| + d \prod_{k=\phi_4(\varepsilon/2d)}^{n} (1 - \lambda_{k+N}(1-\tau)) \\ &\leq \frac{\varepsilon}{2} + d \prod_{k=\phi_4(\varepsilon/2d)}^{n} (1 - \lambda_{k+N}(1-\tau)).\end{aligned}$$

Therefore, $\|u_{n+N} - u_n\| \leq \varepsilon$ for all $n \geq \phi_3(\varepsilon/2d, \phi_4(\varepsilon/2d))$.

Now observe that

$$\begin{aligned}u_{n+N} - T_{[n+N]} \cdots T_{[n+1]}(u_n) = &\sum_{k=1}^{N-1} \Big(T_{[n+N]} \cdots T_{[n+N-k+1]}(u_{n+N-k}) \\ &\qquad - T_{[n+N]} \cdots T_{[n+N-k]}(u_{n+N-k-1})\Big) \\ &+ u_{n+N} - T_{[n+N]}(u_{n+N-1}).\end{aligned}$$

Therefore, since each T_i is nonexpansive

$$\|u_{n+N} - T_{[n+N]} \cdots T_{[n+1]}(u_n)\| \leq \sum_{k=0}^{N-1} \|u_{n+N-k} - T_{[n+N-k]}(u_{n+N-k-1})\|.$$

Consequently, using Lemma 8.7.4, for all $n \geq \max\{\phi_3(\varepsilon/2d, \phi_4(\varepsilon/4d)), \chi(\lceil Nd/2\varepsilon \rceil)\}$,

$$\|u_n - T_{[n+N]} \cdots T_{[n+1]}(u_n)\| \leq \|u_n - u_{n+N}\| + \|u_{n+N} - T_{[n+N]} \cdots T_{[n+1]}(u_n)\| \leq \varepsilon.$$

$\square$

We will need the following fact:

Lemma 8.7.7 (see e.g. Fact 2.13(a) of [104])**.** *For any real sequence* $(\lambda_n) \subset [0,1]$ *and nonnegative integers* n *and* m *such that* $n \geq m$,

$$\sum_{i=m}^{n}\left(\lambda_i \prod_{j=i+1}^{n}(1-\lambda_j)\right) \leq 1.$$

Lemma 8.7.8. *Suppose as before that* C *is bounded with* $\operatorname{diam} C \leq d$*, where* $d \in \mathbb{N}$*. Suppose moreover that* $\lambda_n \in [0,1]$*,* $u^* \in C$*,* $\phi_3 : \mathbb{N} \to \mathbb{N}$ *and* $\chi : \mathbb{N} \to \mathbb{N}$ *satisfy*

1. $\lambda_n \leq 1/k$ *for all nonnegative integers* $n \geq \chi(k)$*, and*
2. $\prod_{i=n}^{m}(1-\lambda_i(1-\tau)) \leq \varepsilon$ *for all* $\varepsilon > 0$ *and all nonnegative integers* n, m *with* $m \geq \phi_3(\varepsilon, n)$*.*

Then, for all $\varepsilon > 0$*,* $n_0 \in \mathbb{N}$ *and all* $g : \mathbb{N} \to \mathbb{N}$*,*

(i) $n_0 \geq \chi\left(\left\lceil \dfrac{12d^2}{\varepsilon^2(1-\tau)} \right\rceil\right)$,

(ii) $\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle \leq \dfrac{\varepsilon^2(1-\tau)}{6}$ *for all* $n \in [n_0; \tilde{g}(\phi_3(\varepsilon^2/3d^2), n_0) - 1]$*, and*

(iii) $\|T_{[n+1]}(u^*) - u^*\| \leq \Omega_d(\varepsilon, g, n_0) := \dfrac{\varepsilon^2}{18d\tilde{g}(\phi_3(\varepsilon^2/3d, n_0) - n_0)}$ *for all* $n \in [n_0; \tilde{g}(\phi_3(\varepsilon^2/3d^2, n_0)) - 1]$

imply $\|u_{\tilde{g}(\phi_3(\varepsilon^2/3d, n_0))} - u^*\| \leq \varepsilon$*, where* $\tilde{g}(n) := \max\{n, g(n)\}$*.*

Proof. Observe that

$$\begin{aligned}
\|u_{n+1} - u^*\|^2 &= \left\langle T_{[n+1]}^{(\lambda_{n+1})}(u_n) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u_n) - u^* \right\rangle \\
&= \left\langle T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u_n) - u^*) \right\rangle \\
&\quad + \left\langle T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u_n) - u^* \right\rangle \\
&= \left\| T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) \right\|^2 + \left\langle T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u_n) - u^* \right\rangle \\
&\quad + \left\langle T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\
&= \left\| T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) \right\|^2 \\
&\quad + \left\langle 2T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\
&= \left\| T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) \right\|^2 \\
&\quad + \lambda_{n+1} \left\langle 2\mathcal{G}T_{[n+1]}(u_n) - \mathcal{G}T_{[n+1]}(u^*) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\
&\quad + (1 - \lambda_{n+1}) \left\langle 2T_{[n+1]}(u_n) - T_{[n+1]}(u^*) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\
&= \left\| T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) \right\|^2 + \underbrace{2 \left\langle T_{[n+1]}(u_n) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle}_{t_1:=} \\
&\quad + \underbrace{\left\langle u^* - T_{[n+1]}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle}_{t_2:=} \\
&\quad + \underbrace{\lambda_{n+1} \left\langle 2\mathcal{G}T_{[n+1]}(u_n) - 2T_{[n+1]}(u_n) + T_{[n+1]}(u^*) - \mathcal{G}T_{[n+1]}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle}_{t_3:=}.
\end{aligned}$$

Observe that $T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* = \lambda_{n+1}(\mathcal{G}T_{[n+1]}(u^*) - u^*) + (1 - \lambda_{n+1})(T_{[n+1]}(u^*) - u^*)$. Therefore,

$$\begin{aligned}
t_1 &= 2 \left\langle T_{[n+1]}(u_n) - u^*, T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\
&= 2\lambda_{n+1} \left\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}T_{[n+1]}(u^*) - u^* \right\rangle \\
&\quad + 2(1 - \lambda_{n+1})\langle T_{[n+1]}(u_n) - u^*, T_{[n+1]}(u^*) - u^* \rangle \\
&\leq 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}T_{[n+1]}(u^*) - \mathcal{G}u^* \rangle + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle \\
&\quad + 2d(1 - \lambda_{n+1}) \cdot \|T_{[n+1]}(u^*) - u^*\| \\
&\leq 2d\lambda_{n+1}\tau\|T_{[n+1]}(u^*) - u^*\| + 2d(1 - \lambda_{n+1})\|T_{[n+1]}(u^*) - u^*\| \\
&\quad + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle \\
&\leq 2d \cdot \|T_{[n+1]}(u^*) - u^*\| + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle. \qquad (8.34)
\end{aligned}$$

Moreover,

$$\begin{aligned} t_2 &= \left\langle u^* - T_{[n+1]}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\ &\leq d \cdot \|u^* - T_{[n+1]}(u^*)\|. \end{aligned} \tag{8.35}$$

For the term t_3, we get the estimate

$$\begin{aligned} t_3 &= \lambda_{n+1} \left\langle 2\mathcal{G}T_{[n+1]}(u_n) - 2T_{[n+1]}(u_n) + T_{[n+1]}(u^*) - \mathcal{G}T_{[n+1]}(u^*), T_{[n+1]}^{(\lambda_{n+1})}(u^*) - u^* \right\rangle \\ &\leq 2\lambda_{n+1}^2 \left\langle \mathcal{G}T_{[n+1]}(u_n) - T_{[n+1]}(u_n), \mathcal{G}T_{[n+1]}(u^*) - u^* \right\rangle \\ &\quad + 2d\lambda_{n+1}(1-\lambda_{n+1})\|T_{[n+1]}(u^*) - u^*\| \\ &\quad + \lambda_{n+1}^2 \langle T_{[n+1]}(u^*) - \mathcal{G}T_{[n+1]}(u^*), \mathcal{G}T_{[n+1]}(u^*) - u^* \rangle \\ &\quad + d\lambda_{n+1}(1-\lambda_{n+1})\|T_{[n+1]}(u^*) - u^*\| \\ &\leq 3d^2\lambda_{n+1}^2 + 3d\lambda_{n+1}(1-\lambda_{n+1})\|T_{[n+1]}(u^*) - u^*\| \\ &\leq 3d^2\lambda_{n+1}^2 + 3d\lambda_{n+1}\|T_{[n+1]}(u^*) - u^*\|. \end{aligned} \tag{8.36}$$

Combining the estimates for the terms t_1, t_2 and t_3, we obtain

$$\begin{aligned} \|u_{n+1} - u^*\|^2 \leq{}& \left\| T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(u^*) \right\|^2 + (3d\lambda_{n+1} + 3d)\|T_{[n+1]}(u^*) - u^*\| \\ &+ 3d^2\lambda_{n+1}^2 + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle. \end{aligned}$$

Therefore, by Lemma 8.6.1

$$\begin{aligned} \|u_{n+1} - u^*\|^2 &\leq (1 - \lambda_{n+1}(1-\tau))^2\|u_n - u^*\|^2 + (3d\lambda_{n+1} + 3d)\|T_{[n+1]}(u^*) - u^*\| \\ &\quad + 3d^2\lambda_{n+1}^2 + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle. \\ &\leq (1 - \lambda_{n+1}(1-\tau))^2\|u_n - u^*\|^2 + 6d \cdot \|T_{[n+1]}(u^*) - u^*\| \\ &\quad + 3d^2\lambda_{n+1}^2 + 2\lambda_{n+1}\langle T_{[n+1]}(u_n) - u^*, \mathcal{G}u^* - u^* \rangle. \end{aligned}$$

Using now the hypotheses, we see that, for all $n \in [n_0; \tilde{g}(\phi_3(\varepsilon^2/3d^2), n_0) - 1]$,

$$\|u_{n+1} - u^*\|^2 \leq (1 - \lambda_{n+1}(1-\tau))\|u_n - u^*\|^2 + 6d \cdot \Omega_d(\varepsilon, g, n_0) + \lambda_{n+1}(1-\tau) \cdot \frac{\varepsilon^2}{3}.$$

Using induction, we then get (using the abbreviation $\tilde{n}_0 = \tilde{g}(\phi_3(\varepsilon^2/3d^2), n_0)$) by Lemma 8.7.7

$$\begin{aligned} \|u_{\tilde{n}_0} - u^*\|^2 &\leq \|u_{n_0} - u^*\|^2 \cdot \prod_{i=n_0+1}^{\tilde{n}_0} (1 - \tau(1-\lambda_i)) + 6d(\tilde{n}_0 - n_0)\Omega_d(\varepsilon, g, n_0) \\ &\quad + \frac{\varepsilon^2}{3} \sum_{i=n_0+1}^{\tilde{n}_0} \left(\lambda_i(1-\tau) \prod_{j=i+1}^{\tilde{n}_0} (1 - \lambda_j(1-\tau)) \right) \\ &\leq \|u_{n_0} - u^*\|^2 \cdot \prod_{i=n_0+1}^{\tilde{n}_0} (1 - \tau(1-\lambda_i)) + \frac{\varepsilon^2}{3} + 6d(\tilde{n}_0 - n_0)\Omega_d(\varepsilon, g, n_0) \\ &\leq \frac{\varepsilon^2}{3} + \frac{\varepsilon^2}{3} + \frac{\varepsilon^2}{3} = \varepsilon^2. \end{aligned}$$

□

Theorem 8.7.9. *Suppose C is bounded with $\operatorname{diam}(C) \leq d$, and suppose that $T_1, \ldots, T_N : C \to C$ are nonexpansive mappings with a common fixed point $p \in C$ that satisfy $\bigcap_{i=1}^N Fix(T_i) = Fix(T_N \cdots T_1)$. Suppose $\rho : \mathbb{N} \times (0, \infty) \to (0, \infty)$ satisfies* (8.32) *and let the moduli χ, $\hat{\chi}$, ϕ_3 and ϕ_4 be as before. Then, for any τ-contraction $\mathcal{G} : C \to C$, the iteration given by* (8.33) *is metastable with rate $\Xi(\varepsilon, g; \chi, \phi_3, \phi_4, d, \tau)$, i.e.*

$$\forall \varepsilon > 0 \, \forall g : \mathbb{N} \to \mathbb{N} \, \exists n \leq \Xi(\varepsilon, g; \chi, \phi_3, \phi_4, d, \tau) \left(\|u_n - u_{\tilde{g}(n)}\| \leq \varepsilon \right)$$

where $\tilde{g}(n) := \max\{n, g(n)\}$ and Ξ is defined by

$$\Xi(\varepsilon, g; \chi, \phi_3, \phi_4, d, \tau) := \phi_3' \left(\varepsilon^2/12d^2, \max \left\{ n_0, \hat{\chi} \left(\rho \left(d, \frac{1}{k_{i_0}(\tilde{f})N} \right) \right) \right\} \right),$$

where $\phi_3'(\varepsilon, i) := \max\{n, \max\{\phi_3(\varepsilon, i) : i \leq n\}\}$ and $k_{i_0}(\tilde{f})$ is as in Theorem 8.6.7 except for f and ε_d, which are now defined as

$$\begin{aligned} f(k) &:= \rho \left(d, \Omega_d^M(\varepsilon/2, \tilde{g}^M, \max\{n_0, \hat{\chi}(\rho(d, 1/Nk))\} \right) \\ n_0 &:= \max\{\chi(\lceil 96d/(1-\tau)\varepsilon^2 \rceil), \chi(\lceil 48d^2/(1-\tau)\varepsilon^2 \rceil)\} \\ \varepsilon_d &:= \frac{((1-\tau)\varepsilon^2/96)^2}{2d^2} \\ \Omega_d^M(\varepsilon, g, n) &:= \max \left\{ \Omega_d(\varepsilon, g, i) : i \leq n \right\}. \end{aligned}$$

Proof. Define $T := T_N \cdots T_1$ and

$$J_\varepsilon(\varphi) := \min \left\{ l \geq n_0 : \|Tu_k - u_k\| \leq \varphi(v) \text{ for all } v \in C \text{ and } k \geq l \right\}.$$

For majorizable $\varphi : C \to (0, 1]$, this is well-defined, and by Corollary 8.7.6,

$$k \gtrsim \varphi \to J_\varepsilon(\varphi) \leq \max \left\{ n_0, \hat{\chi}(\rho(d, 1/Nk) \right\}. \tag{8.37}$$

Now define the counterfunction $V_{\varepsilon,g}(u, \varphi) = u_{i(u)+1}$, where $i(u)$ is defined as the least index $i \in [J_\varepsilon(\varphi) - 1, \tilde{g}_{u,\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\varphi))) - 2]$ such that for all integers $k \in [J_{\varepsilon(\varphi)} - 1, \tilde{g}_{u,\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\varphi)) - 2]$

$$\|\mathcal{G}u - (1-t)u - tu_i\| \leq \|\mathcal{G}u - (1-t)u - tu_k\|.$$

Moreover, if we define $\Delta_{\varepsilon,g}(u, \varphi) := \rho \left(d, \Omega_d(\varepsilon/2, \tilde{g}_{u,\varepsilon}, J_\varepsilon(\varphi)) \right)$, then, given a majorant $k \gtrsim \varphi$, (8.37) implies $\rho \left(d, \Omega_d^M \left(\varepsilon/2, \tilde{g}^M, \max\{n_0, \hat{\chi} \left(\frac{1}{k} \right) \} \right) \right) \geq \Delta_{\varepsilon,g}(u, \varphi)$ for all $u \in C$. Therefore, $f \gtrsim \Delta_{\varepsilon,g}$. We again write $\tilde{u} := U(\varepsilon_d, t, \Delta_{\varepsilon,g}, V_{\varepsilon,g})$ and $\tilde{\varphi} := \Phi(\varepsilon_d, t, \Delta_{\varepsilon,g}, V_{\varepsilon,g})$, where $t := \frac{((1-\tau)\varepsilon^2/96)^2}{3d^2}$.

Then,

$$\|T\tilde{u} - \tilde{u}\| < \rho \left(d, \Omega_d(\varepsilon/2, \tilde{g}_{\tilde{u},\varepsilon}, J_\varepsilon(\tilde{\varphi})) \right) \tag{8.38}$$

and

$$\|Tu_{i(\tilde{u})+1} - u_{i(\tilde{u})+1}\| < \tilde{\varphi}(u_{i(\tilde{u})+1}) \to \|\mathcal{G}\tilde{u} - \tilde{u}\|^2 < \|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tu_{i(\tilde{u})+1}\|^2 + \varepsilon_d$$

By construction, $\|Tu_{i(\tilde{u})+1} - u_{i(\tilde{u})+1}\| < \tilde{\varphi}(u_{i(\tilde{u})+1})$, so we conclude $\|\mathcal{G}\tilde{u} - \tilde{u}\|^2 < \|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tu_{i(\tilde{u})+1}\|^2 + \varepsilon_d$. By construction, we also have $\|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tu_{i(\tilde{u})+1}\| \leq \|\mathcal{G}\tilde{u} - (1-t)\tilde{u} - tu_{k+1}\|$ for all $k \in [J_\varepsilon(\tilde{\varphi}), \tilde{g}_{\tilde{u},\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\tilde{\varphi})) - 1]$. Therefore,

$$\|\mathcal{G}\tilde{u}-\tilde{u}\|^2 < \|\mathcal{G}\tilde{u}-(1-t)\tilde{u}-tu_{k+1}\|^2+\varepsilon_d, \quad \text{for all } k \in [J_\varepsilon(\tilde{\varphi}), \tilde{g}_{\tilde{u},\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\tilde{\varphi}))-1].$$

Therefore, Lemma 8.4.5 implies $\langle \mathcal{G}\tilde{u} - \tilde{u}, u_{k+1} - \tilde{u}\rangle \leq (1-\tau)\varepsilon^2/96$ for all nonnegative integers $k \in [J_\varepsilon(\tilde{\varphi}), \tilde{g}_{\tilde{u},\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\tilde{\varphi})) - 1]$. Consequently, since by construction $J_\varepsilon(\tilde{\varphi}) \geq \max\left\{\chi(\lceil 96d/(1-\tau)\varepsilon^2\rceil), \chi(\lceil 48d^2/(1-\tau)\varepsilon^2\rceil)\right\}$,

$$\langle \mathcal{G}\tilde{u} - \tilde{u}, T_{[k+1]}(u_k) - \tilde{u}\rangle \leq \langle \mathcal{G}\tilde{u} - \tilde{u}, u_{k+1} - \tilde{u}\rangle + d \cdot \|T_{[k+1]}(u_k) - u_{k+1}\| \leq \frac{(\varepsilon/2)^2(1-\tau)}{12},$$

and

$$J_\varepsilon(\tilde{\varphi}) \geq \chi\left(\left\lceil \frac{12}{(\varepsilon/2)^2(1-\tau)}\right\rceil\right).$$

Moreover, (8.32) and (8.38) imply $\|T_{[k+1]}(\tilde{u}) - \tilde{u}\| < \Omega_d^M(\varepsilon/2, \tilde{g}_{\tilde{u},\varepsilon}, J_\varepsilon(\tilde{\varphi}))$ for all nonnegative integers k. Therefore, Lemma 8.7.8 implies

$$\|u_{\tilde{g}_{\tilde{u},\varepsilon}(\phi_3(\varepsilon^2/12d^2, J_\varepsilon(\tilde{\varphi}))} - \tilde{u}\| \leq \frac{\varepsilon}{2}.$$

Therefore, for $k := \phi_3(\varepsilon^2/12d^2, J_\varepsilon(\tilde{\varphi}))$, Lemma 8.6.5 yields

$$\|u_k - u_{\tilde{g}(k)}\| \leq \varepsilon.$$

□

As before, one can weaken the assumption that C is bounded as follows:

Corollary 8.7.10. *For all of the results in this section, one can drop the condition of C being bounded with* $\operatorname{diam}(C) \leq d$ *in favor of* $\|u_0 - v\| \leq d/4$, $\|v - \mathcal{G}v\| \leq \frac{d(1-\tau)}{4}$ *and* $\|v - w\| \leq \frac{d}{4(1+\tau)}$, *where v is a common fixed point of the T_i and w is the unique fixed point of $\mathcal{G}$.*

Proof. Similarly to the situation before, Lemma 8.6.1 implies for all nonnegative integers n

$$\begin{aligned}\|u_{n+1} - v\| &\leq \|T_{[n+1]}^{(\lambda_{n+1})}(u_n) - T_{[n+1]}^{(\lambda_{n+1})}(v)\| + \|T_{[n+1]}^{(\lambda_{n+1})}(v) - v\| \\ &\leq (1-\lambda_{n+1}(1-\tau))\|u_n - v\| + \lambda_{n+1}(1-\tau)\cdot\frac{\|\mathcal{G}v - v\|}{1-\tau}.\end{aligned}$$

Since $\|u_0 - v\| \leq d/4$, we conclude by induction that $\|u_n - v\| \leq d/4$. Moreover,

$$\|\mathcal{G}u_n - v\| \leq \|\mathcal{G}u_n - \mathcal{G}w\| + \|w - v\| \leq \tau\|u_n - w\| + \|v - w\| \leq \tau\|u_n - v\| + (1+\tau)\|v - w\|.$$

Consequently, both (u_n) and $(\mathcal{G}u_n)$ remain in the closed ball of radius $d/2$, and hence of diameter d, centered at v. Since all points for which the condition $\operatorname{diam}(C)$ were either elements of the sequences (u_n) and $(\mathcal{G}u_n)$, or convex combinations thereof, the claim follows. □

9 Sunny Nonexpansive Retracts

9.1 Introduction

The purpose of this chapter is twofold. First and foremost, we would like to generalize Theorem 5.3.5 from Hilbert spaces to uniformly smooth and uniformly convex Banach spaces X. As we discussed in Chapter 4, this would present a significant generalization and would include, for example, the L^p spaces for $1 < p < \infty$. To this end, consider a nonempty closed convex subset $C \subseteq X$, a nonexpansive mapping $T : C \to C$ with nonempty fixed-point set $Fix(T)$ and an anchor $u \in C$. In view of Theorem 5.3.4 it would be sufficient to find a rate of metastability for the resolvent (z_n), defined by the equation

$$z_n = \left(1 - \frac{1}{n}\right) Tz_n + \frac{1}{n}u, \quad \text{for all positive integers } n.$$

In Hilbert spaces, rates of metastability for (z_n) are known from Kohlenbach [52]. Therein, two different strategies are employed to obtain two distinct such rates. One of the rates of metastability is extracted from an elementary proof due to Halpern. The proof, however, relies heavily on the properties of Hilbert spaces and has, so far, not been extended to Banach spaces. The other rate of metastability presented in [52] was extracted from a proof of Browder [11], who showed that (z_n) converges to the metric projection $P_{Fix(T)}(u)$ of the anchor u onto the fixed-point set $Fix(T)$ of T.

In Hilbert spaces, the metric projection is characterized by the variational inequality

$$\langle u - P_{Fix(T)}(u), v - P_{Fix(T)}(u)\rangle \leq 0, \quad \text{for all } v \in Fix(T).$$

In fact, as Browder already realized in the aforementioned paper [11], the only part of his proof that hinges on Hilbert space theory is precisely this characterization of the metric projection. More precisely, if X is a uniformly convex and uniformly smooth Banach space, C, T, u and (z_n) are as before and we *assume* that there exists a point $u_0 \in Fix(T)$ such that

$$\langle u - u_0, J(v - u_0)\rangle \leq 0 \quad \text{for all } v \in Fix(T), \tag{9.1}$$

then (z_n) converges precisely to this point u_0. However, it is known that if X is not a Hilbert space, $u_0 \neq P_{Fix(T)}(u)$ in general. In fact, de Figueiredo and Karlovitz [22] showed that the radial projection mapping T on any real normed space X of dimension $\dim(X) \geq 3$, defined by

$$T(x) := \begin{cases} x & \text{if } \|x\| \leq 1, \\ x/\|x\| & \text{if } \|x\| \geq 1, \end{cases}$$

is nonexpansive if and only if X is a Hilbert space. On the other hand, one easily verifies that in uniformly smooth Banach spaces, there exists for every $u \in C$ at most one point u_0 that satisfies (9.1), and that the mapping $r : C \to Fix(T)$ which assigns to each u the unique point u_0 satisfying (9.1) is nonexpansive whenever it is well-defined. So in particular, $r \neq P_{Fix(T)}$.

Consequently, one needs a different approach to obtain the point u_0. However, once one has shown the existence of such a point u_0, Bruck's proof generalizes essentially unchanged to the Banach space case. Therefore, once one has extended the part of the proof-theoretic analysis in [52] corresponding to the existence of the metric projection $P_{Fix(T)}$ to the existence of the point u_0, one will immediately be able to give a rate of metastability for (z_n).

We provide a partial solution to this problem by analyzing a highly non-effective proof of the existence of u_0 due to Bruck, split among the papers [17], [13] and [14]. While the original proof is well beyond the reach of existing metatheorems through the use of Zorn's Lemma and Tychonoff's Theorem, a thorough inspection from a proof-theoretic standpoint immediately suggests an alternative proof that eliminates the use of Tychonoff's Theorem and Zorn's Lemma while maintaining the general idea of the theorem in favor of the much less non-effective weak sequential compactness of bounded, closed and convex subsets of uniformly smooth and uniformly convex Banach spaces. A proof that uses at most the axiom of dependent choice is by no means a new result in itself, see Reich [85], Dominguez, López, Xu [24] and Aleyner, Reich [1]. However, our approach is more suitable to proof mining since the argument has a more pointwise character.

Although we reduced the use of ideal principles drastically, it is not yet known whether the weak sequential compactness can be formalized in the system $\mathcal{A}^\omega[X, \|\cdot\|, \ldots]$. Nevertheless, we will argue using Theorem 4.2.10 that the proof for Hilbert spaces, which is formalizable in $\mathcal{A}^\omega$ (see Kohlenbach [51]), can be extended to uniformly smooth and uniformly convex Banach spaces. Therefore, this approach brings us considerably closer to a rate of metastability for (z_n). So far, the research in this vein has already enabled us to answer the question whether any space with a norm-to-norm uniformly continuous duality selection mapping is uniformly smooth in the affirmative, a question left open by Kohlenbach and Leuştean in [58]. This theorem is shown in Section 9.2 and constitutes the second main result of this chapter.

9.2 Semi-Inner-Product Spaces and the Duality Map

Definition 9.2.1 (Lumer [79]). Let X be a real vector space. We shall say that a real *semi-inner-product* is defined on X, if to any $x, y \in X$ there corresponds a real number $[x, y]$ such that the following properties hold:

(i) $[x + y, z] = [x, z] + [y, z]$ and $[\lambda x, y] = \lambda [x, y]$ for all $x, y, z \in X$ and real λ,

(ii) $[x, x] > 0$ for $x \neq 0$ and

(iii) $[x, y]^2 \leq [x, x] [y, y]$.

We then call X a real *semi-inner-product space.*

Theorem 9.2.2 (Lumer [79])**.** *A semi-inner-product space is a normed linear space with the norm $\|x\| := [x, x]^{1/2}$. Every normed linear space can be made into a semi-inner-product space (in general, in infinitely many ways).*

Definition 9.2.3 (Giles [32])**.** A semi-inner-product space has the *homogeneity property* when the semi-inner-product additionally satisfies

(iv) $[x, \lambda y] = \lambda\,[x, y]$.

Theorem 9.2.4 (Giles [32])**.** *Every normed linear vector space can be represented as a semi-inner-product space with the homogeneity property.*

Definition 9.2.5 (Giles [32])**.** Let X be a real semi-inner-product space with unit sphere S. Then

1. X is a real *continuous semi-inner-product space* if

$$\lim_{\lambda\to 0}[y, x + \lambda y] = [y, x], \quad \text{for all } x, y \in S;$$

2. X is a real *uniformly continuous semi-inner-product space* if the above limit is approached uniformly in $x, y \in S$.

Remark 9.2.6. In [32], Giles assumes throughout the paper that all semi-inner-product spaces possess the homogeneity property. However, his proof for the implication from left to right of the next theorem does not require the homogeneity property. We therefore include the proof in this thesis for the sake of completeness.

Theorem 9.2.7 (Giles [32])**.** *A real semi-inner-product space X is uniformly continuous if and only if the corresponding vector space X with norm $\|x\| = [x, x]^{1/2}$ is uniformly smooth.*

Proof. Let X be a real semi-inner-product space. For $x, y \in X$ on the unit sphere and real $\lambda > 0$ (we can without loss of generality assume $x + \lambda y \neq 0$ since $\|x\| = \|y\| = 1$ and we are interested in t tending to zero),

$$\begin{aligned}
\frac{\|x + \lambda y\| - \|x\|}{\lambda} &= \frac{\|x + \lambda y\|\,\|x\| - \|x\|^2}{\lambda\|x\|} \\
&\geq \frac{[x + \lambda y, x] - \|x\|^2}{\lambda\|x\|} \\
&= \frac{\|x\|^2 + \lambda[y, x] - \|x\|^2}{\lambda\|x\|} \\
&= \frac{[y, x]}{\|x\|}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}\frac{\|x+\lambda y\|-\|x\|}{\lambda} &= \frac{\|x+\lambda y\|^2-\|x\|\|x+\lambda y\|}{\lambda\|x+\lambda y\|}\\ &\leq \frac{[x+\lambda y, x+\lambda y]-[x, x+\lambda y]}{\lambda\|x+\lambda y\|}\\ &= \frac{[x, x+\lambda y]+\lambda[y, x+\lambda y]-[x, x+\lambda y]}{\lambda\|x+\lambda y\|}\\ &= \frac{[y, x+\lambda y]}{\|x+\lambda y\|}.\end{aligned}$$

In total,

$$\frac{[y,x]}{\|x\|} \leq \frac{\|x+\lambda y\|-\|x\|}{\lambda} \leq \frac{[y, x+\lambda y]}{\|x+\lambda y\|},$$

whence the uniform continuity implies the uniform smoothness.

Conversely, suppose that X is a semi-inner-product space such that X endowed with the norm $\|x\| := [x,x]^{1/2}$ is uniformly smooth. Then the normalized duality mapping J on X is single valued. On the other hand, $[\cdot, x]$ is a linear operator on X for each $x \in X$ such that:

(i) $[x,x] = \|x\|^2$ for all $x \in X$.

(ii) $\sup_{\|y\|\leq 1}[y,x] \leq \|x\|^2$ by (iii) of Definition 9.2.1, where equality is attained for $y = x/\|x\|$. Thus, $\|x\|^2 = \|[\,\cdot\,, x]\|$ for each $x \in X$.

Consequently $[\,\cdot\,, x] \in J(x)$ for each $x \in X$. Since J is single valued, $[\,\cdot\,, x]$ is the unique duality selection mapping, which is uniformly continuous in uniformly smooth Banach spaces. □

Remark 9.2.8. There is a close connection to the normalized duality mapping (cp. Definition 4.2.3); in fact, $\langle\cdot, j(\cdot)\rangle$ is a semi-inner-product, where $j(x) \in J(x)$ for all $x \in X$; condition (i) is true by definition and (ii) is clear since $\langle x, j(x)\rangle = \|x\|^2 > 0$ for all $j(x) \in J(x)$ and all $x \neq 0$. Finally, (iii) holds by definition.

Moreover, Lemma 4.2.4 ensures that we can even choose a duality selection mapping j in such a way that the associated semi-inner-product is homogeneous.

Remark 9.2.9. Suppose that J is a norm-to-norm uniformly continuous duality selection map, cf. Definition 4.2.9. Then, the induced semi-inner-product (in the sense of Remark 9.2.8) is uniformly continuous: For all $x, y \in X$,

$$\lim_{\lambda\to 0}[y, x+\lambda y] = \lim_{\lambda\to 0}\langle y, J(x+\lambda y)\rangle = \langle y, J(x)\rangle = [y,x].$$

Theorem 9.2.10. *Let X be a normed space with norm $\|\cdot\|$. If X is a space with a norm-to-norm uniformly continuous duality selection map, it is uniformly smooth.*

Proof. Suppose X is a space with a norm-to-norm uniformly continuous duality selection map J. Then X is a semi-inner-product space with semi-inner-product $\langle \cdot, J(\cdot) \rangle$ by Remark 9.2.8. Moreover, the semi-inner-product is norm-to-norm uniformly continuous in the sense of Definition 9.2.5 by Remark 9.2.9. Hence, by Theorem 9.2.7, X equipped with the norm $\langle x, Jx \rangle^{1/2}$ is uniformly smooth. But since $\langle x, Jx \rangle = \|x\|^2$, the space X is uniformly smooth with respect to the original norm $\| \cdot \|$. $\square$

Corollary 9.2.11. *If X has a norm-to-norm uniformly continuous duality selection mapping, then the normalized duality map is single valued.*

9.3 Sunny Nonexpansive Retracts

Definition 9.3.1 (Bruck [13])**.** Let $C \subseteq X$ be a closed convex subset of a Banach space X and $F \subset C$ be nonempty closed subset of C. A mapping $Q : C \to F$

1. is a *retraction* if $Qx = x$ for all $x \in C$;
2. is *sunny* if, for all $x \in C$ and $t \geq 0$ such that $Qx + t(x - Qx) \in C$,

$$Q\left(Qx + t\left(x - Qx\right)\right) = Qx.$$

Moreover, C is called a *sunny nonexpansive retract* of X if there exists a sunny nonexpansive retraction $Q : X \to C$.

Remark 9.3.2. Bruck originally used the term 'nonexpansive projection' instead of 'sunny nonexpansive retraction'. However, the latter term is now the established one.

Remark 9.3.3. A sunny nonexpansive retraction $Q : C \to F$ retracts C onto F along rays.

Definition 9.3.4 (Birkhoff [7])**.** Let X be a normed vector space. For $x, y \in X$,

1. x is said to be *orthogonal* to y if

$$\|x + ty\| \geq \|x\|, \quad \text{for all } t \in \mathbb{R}$$

2. x is said to be *acute* to y if

$$\|x + ty\| \geq \|x\|, \quad \text{for all } t \geq 0$$

Remark 9.3.5. If X is a real Banach space, we immediately conclude from Lemma 4.2.5 that x is acute to y if and only if there exists $j(x) \in J(x)$ with $\langle y, j\left(x\right) \rangle \geq 0$.

Moreover, in smooth Banach spaces, x is orthogonal to y if and only if $\langle y, J(x) \rangle = 0$, see Theorem 2 of Giles [32].

Definition 9.3.6 (Bruck [13])**.** Let $C \subseteq X$ be a closed convex subset of a Banach space X and $F \subset C$ be nonempty closed subset of C. A retraction $r : C \to F$ is *orthogonal* if for each $p \in C$ and $y \in F$, $r\left(p\right) - y$ is acute to $p - r\left(p\right)$:

$$\|(1-t)\, r(p) + tp - y\| \geq \|r(p) - y\|, \quad \text{for all } t \geq 0.$$

As Bruck shows, the notion of orthogonal retractions is slightly more general than that of sunny nonexpansive retractions.

Theorem 9.3.7 (Bruck [13])**.** *Let $C \subseteq X$ be a closed convex subset of a Banach space X and $F \subset C$ be nonempty closed subset of C. Suppose $r : C \to F$ is a retraction. Then each of the following conditions implies the next:*

1. *r is sunny nonexpansive;*
2. *r is firmly nonexpansive, i.e.*

$$\|r(x) - r(y)\| \leq \|(1-t)(r(x) - r(y)) + t(x-y)\| \quad \textit{for all } x, y \in C \textit{ and } t \geq 0;$$

3. *r is an orthogonal retraction.*

Moreover, if X is smooth, these conditions are equivalent and there can exist at most one sunny nonexpansive retraction of C onto F.

Theorem 9.3.8 (Bruck [13])**.** *If X is a strictly convex and uniformly smooth Banach space with $C \subset X$ closed and convex, $T : C \to C$ is nonexpansive, and $Fix(T)$ is nonempty, then there exists a sunny nonexpansive retraction of C onto $Fix(T)$.*

Remark 9.3.9. Suppose that X is smooth, and $r : C \to F$ is a nonexpansive retraction. Then r is sunny nonexpansive if and only if

$$\langle p - r(p), J(y - r(p)) \rangle \leq 0, \quad \text{for all } y \in Fix(T).$$

This follows from the following argument: Since X is smooth, there can be at most one sunny nonexpansive retraction of C onto $Fix(T)$, and this sunny nonexpansive retract is the unique orthogonal retraction by Theorem 9.3.7. By Definition 9.3.6, this is to say that $r(p) - y$ is acute to $p - r(p)$. Hence, by Remark 9.3.5, the claim follows.

Remark 9.3.10. Suppose that X is smooth and $C \subset X$. We can then also show that there can be at most one sunny nonexpansive retraction by the following (pointwise) argument: Let $x \in X$ and suppose Px and Qx are sunny nonexpansive retracts of x onto C. Then, by the preceding remark,

$$\langle x - Px, J(Qx - Px) \rangle \leq 0, \quad \text{and } \langle x - Qx, J(Px - Qx) \rangle \leq 0.$$

Since J is homogeneous, adding these two inequalities yields

$$\|Qx - Px\|^2 = \langle Qx - Px, J(Qx - Px) \rangle \leq 0,$$

which implies $Qx = Px$.

We now exhibit the full structure of the proof of Theorem 9.3.8, starting with the existence of a nonexpansive retraction onto the fixed point set. It should be remarked that Bruck wrongly defined the order "$\leq$" in his original proof. This mistake was tacitly corrected by Goebel and Kirk [35] to coincide with the definition stated below.

Proof. We split the proof into three parts:

Step 1: We first show that $Fix(T)$ is a nonexpansive retract of C, laying out the details of the proof of Bruck found in [14, 17]. Let

$$N(Fix(T)) := \{f : C \to C |\, f \text{ is nonexpansive and } Fix(T) \subset Fix(f)\},$$

and define an preorder on $N(Fix(T))$ by setting $f \leq g$ if $\|fx - fy\| \leq \|gx - gy\|$ for all $x, y \in C$.

Now define a function $\varphi : N(Fix(T)) \to N(Fix(T))$ by

$$f \mapsto \frac{1}{2}\,\mathrm{id} + \frac{1}{2} f,$$

where $\mathrm{id} : C \to C$ denotes the identity on C. We first check well-definedness of φ before proving several properties of φ: Suppose $f \in N(Fix(T))$. Then f is nonexpansive and hence for all $u, v \in C$,

$$\begin{aligned}
\|\varphi(f)(u) - \varphi(f)(v)\| &= \|\frac{1}{2}(u - v) + \frac{1}{2}(f(u) - f(v))\| \\
&\leq \frac{1}{2}\|u - v\| + \frac{1}{2}\|f(u) - f(v)\| \\
&\leq \frac{1}{2}\|u - v\| + \frac{1}{2}\|u - v\| \\
&= \|u - v\|.
\end{aligned}$$

Thus $\varphi(f)$ is nonexpansive whenever f is. Moreover, if $f(u) = u$, then $\varphi(f)(u) = \frac{1}{2}u + \frac{1}{2}f(u) = u$, so $Fix(\varphi(f)) \supseteq Fix(f) \supseteq Fix(T)$, so φ is well-defined. Now suppose that for $u \neq v$ we have

$$\|\varphi(f)(u) - \varphi(f)(v)\| = \|u - v\|.$$

Then since $\varphi(f)$ is nonexpansive for all $f \in N(Fix(T))$,

$$\begin{aligned}
\|u - v\| &= \|\varphi(f)(u) - \varphi(f)(v)\| \\
&= \|\frac{1}{2}(u - v) + \frac{1}{2}(f(u) - f(v))\| \\
&\leq \frac{1}{2}\|u - v\| + \frac{1}{2}\|f(u) - f(v)\| \\
&\leq \frac{1}{2}\|u - v\| + \frac{1}{2}\|u - v\| \\
&= \|u - v\|,
\end{aligned}$$

so equality holds throughout. Consequently $\|f(u) - f(v)\| = \|u - v\| = \|\frac{1}{2}(u - v) + \frac{1}{2}(f(u) - f(v))\|$. But by the strict convexity of X, we conclude $f(u) - f(v) = u - v$. Thus,

$$\varphi(f)(u) - \varphi(f)(v) = f(u) - f(v) = u - v.$$

Moreover if $f \in N(Fix(T))$, $\varphi(f)(v) = \frac{1}{2}v + \frac{1}{2}f(v) = v$ if and only if $f(v) = v$. Thus, f and $\varphi(f)$ have the same fixed points.

Finally, for all $f, g \in N(Fix(T))$,

$$\begin{aligned}\|(\varphi(f)\circ g)(u) - (\varphi(f)\circ g)(v)\| &= \|\varphi(f)(g(u)) - \varphi(f)(g(v))\| \\ &\leq \|g(u) - g(v)\|, \quad \text{for all } u, v \in C,\end{aligned}$$

so $\varphi(f)\circ g \leq g$. We gather the properties of φ as a summary.

For all $f, g \in N(Fix(T))$ and $u, v \in C$, the following hold:

(i) $Fix(\varphi(f)) = Fix(f) \supseteq Fix(T)$.

(ii) $\|\varphi(f)(u) - \varphi(f)(v)\| = \|u - v\| \Rightarrow \varphi(f)(u) - \varphi(f)(v) = u - v$.

(iii) $\varphi(f)\circ g \leq g$.

In particular, if g is minimal, (iii) implies $\varphi(f)\circ g \sim g$ for all $f \in N(Fix(T))$, i.e. $\varphi(f)\circ g \leq g$ and $g \leq \varphi(f)\circ g$, which reads

$$\|(\varphi(f)\circ g)(u) - (\varphi(f)\circ g)(v)\| = \|g(u) - g(v)\|, \quad \text{for all } u, v \in C.$$

Therefore, by (ii),

$$(\varphi(f)\circ g)(u) - (\varphi(f)\circ g)(v) = g(u) - g(v).$$

Taking $v \in Fix(T)$, we conclude $(\varphi(f)\circ g)(u) = g(u)$, so $\varphi(f)\circ g$ and g define the same function, i.e.

$$\varphi(f)\circ g = g. \tag{9.2}$$

Setting $f = g$, we conclude that $Fix(g) \subseteq R(g) \subseteq Fix(\varphi(g)) = Fix(g)$, so $Fix(g) = R(g)$, where $R(g)$ denotes the range of g. Setting $f = T$, we conclude $Fix(g) \subseteq Fix(T)$. Thus

$$Fix(T) = Fix(g) = R(g),$$

so any minimal element of $N(Fix(T))$ is a nonexpansive retraction of C onto $Fix(T)$.

It is noteworthy to observe that Bruck only used minimality of g is only needed with respect $\varphi(g)\circ g$ and $\varphi(T)\circ g$, i.e. minimality is used to show the existence of a mapping g such that

$$\|g(x) - v\| \leq \left\|\frac{1}{2}(g(x) - v) + \frac{1}{2}(g(g(x)) - v)\right\|$$

and

$$\|g(x) - v\| \leq \left\|\frac{1}{2}(g(x) - v) + \frac{1}{2}(T(g(x)) - v)\right\|$$

for some fixed point v of T. The former is used to show that $R(g) = Fix(g)$, while the latter is used to show that $Fix(g) = Fix(T)$. Later on, we will see that this is only used to show that $g(g(x_n)) = g(x_n)$ for all $n \in \mathbb{N}$.

Observe that this perforated use of Zorn's lemma can be reduced even further. In fact, (9.2) is even only needed for $f = T$. In fact, one then gets the identity

$$Tg(x) = g(x), \quad \text{for all } x \in C. \tag{9.3}$$

As before, this implies that any fixed point of g is a fixed point of T, so $Fix(g) = Fix(T)$. Once this has been established, (9.3) also implies that $g(x)$ is a fixed point of g, so $g(g(x)) = g(x)$, i.e. $Fix(T) = Fix(g) = R(g)$. Therefore, any g that is minimal with respect to the counterfunction $\varphi(T) \circ g$ is already a nonexpansive retraction of C onto $Fix(T)$

Step 2: We now prove that a minimal element of $N(Fix(T))$ actually exists, once again following Bruck [14, 17] closely. To this end, we first show that $N(Fix(T))$ is weakly sequentially compact. Fix $x_0 \in Fix(T)$ and for $x \in C$ define

$$C_x = \{y \in C : \|y - x_0\| \leq \|x - x_0\|\}.$$

Then for each $x \in C$ and f in $N(Fix(T))$, $f(x) \in C_x$ since $f(C) \subseteq C$ and

$$\|f(x) - x_0\| = \|f(x) - f(x_0)\| \leq \|x - x_0\|.$$

Thus $N(Fix(T))$ is a subset of the Cartesian product $P = \prod_{x \in C} C_x$, where each C_x is bounded. Since X is uniformly smooth, it is in particular reflexive, so the bounded, closed and convex subsets C_x are weakly compact. Then P is compact in the product topology, i.e. the topology of weak pointwise convergence. Moreover, $N(Fix(T))$ is closed in P. Suppose that $\{g_\lambda : \lambda \in \Lambda\}$ is a net in $N(Fix(T))$ which converges to $g \in P$. Then $g_\lambda(u) = u$ for all $u \in Fix(T)$, so $g(u) = \text{w-lim}\, g_\lambda(u) = u$. Moreover, Lemma 4.2.2, i.e. the weak lower semicontinuity of the norm, implies

$$\begin{aligned}
\|g(x) - g(y)\| &= \|\,\text{w-lim}_\lambda (g_\lambda(x) - g_\lambda(y))\| \\
&\leq \liminf_\lambda \|g_\lambda(x) - g_\lambda(y)\| \\
&\leq \|x - y\|.
\end{aligned}$$

Thus $g \in N(Fix(T))$, hence $N(Fix(T))$ is closed in P. Thus $N(Fix(T))$ is weakly sequentially compact.

Now define for each $f \in N(Fix(T))$ the initial segment $\text{Is}(f)$ to be $\{g \in N(Fix(T)) : g \leq f\}$. We show that $\text{Is}(f)$ is closed in $N(Fix(T))$, and hence compact. Suppose that $\{g_\lambda : \lambda \in \Lambda\}$ is a net in $\text{Is}(f)$ converging to g. Then $g_\lambda(u) = u$ for all $u \in Fix(T)$, so $g(u) = \text{w-lim}\, g_\lambda(u) = u$. Moreover, by the weak lower semicontinuity of the norm (see Lemma 4.2.2),

$$\begin{aligned}
\|g(x) - g(y)\| &= \|\,\text{w-lim}_\lambda (g_\lambda(x) - g_\lambda(y))\| \\
&\leq \liminf_\lambda \|g_\lambda(x) - g_\lambda(y)\| \\
&\leq \|f(x) - f(y)\|.
\end{aligned}$$

Thus $g \in \text{Is}(f)$. Being closed in the weakly sequentially compact set $N(Fix(T))$, also $\text{Is}(f)$ is compact.

Now, if $\{g_\lambda : \lambda \in \Lambda\}$ is linearly ordered, then $\text{Is}(g_\lambda)$ is also linearly ordered by inclusion. Since each $\text{Is}(g_\lambda)$ is nonempty and compact, there exists a $g \in \bigcap_\lambda \text{Is}(g_\lambda)$. Therefore, by Zorn's Lemma, there exists a minimal element g in $N(Fix(T))$. We have already seen that g is then a nonexpansive retraction of C onto $Fix(T)$.

Step 3: From now on, we follow another paper of Bruck [13]. Let f be a nonexpansive retraction of C onto $Fix(T)$, i.e. $f : C \to Fix(T)$ such that f is nonexpansive and $f(x^*) = x^*$ for all fixed points x^* of T. Let $(\lambda_n) \subset [0, 1)$ be a sequence converging to 1 and $p \in C$ be arbitrary. Define the sequence (x_n) by the implicit scheme

$$x_n = \lambda_n f(x_n) + (1 - \lambda_n)p.$$

The sequence is well-defined in view of Banach's fixed point theorem since $x \mapsto \lambda_n f(x) + (1 - \lambda_n)p$ is a strict contraction for each nonnegative integer n. Now observe that

$$x_n - x^* = \lambda_n(f(x_n) - x^*) + (1 - \lambda_n)(p - x^*), \quad \text{for all } x^* \in Fix(T). \tag{9.4}$$

Then, (x_n) is bounded since, for any $x^* \in Fix(T) = Fix(f)$,

$$\|x_n - x^*\| = \|\lambda_n(f(x_n) - f(x^*)) + (1 - \lambda_n)(p - x^*)\| \leq \lambda_n\|x_n - x^*\| + (1 - \lambda_n)\|p - x^*\|,$$

so $\|x_n - v\| \leq \|p - v\|$.

Moreover, by rearranging (9.4), we see that

$$\begin{aligned} p - x^* &= \frac{1}{1 - \lambda_n}(x_n - x^*) - \frac{\lambda_n}{1 - \lambda_n}(f(x_n) - x^*) \\ &= (1 + s_n)(x_n - x^*) - s_n(f(x_n) - x^*), \end{aligned}$$

where $s_n := \lambda_n/(1 - \lambda_n)$. Observe that $s_n \in [0, \infty)$. Now, for all $t > 0$,

$$\begin{aligned} \|t(p - x^*) + (1 - t)(x_n - x^*)\| &= \big\|t(1 + s_n)(x_n - x^*) - ts_n(f(x_n) - x^*) \\ &\qquad + (1 - t)(x_n - x^*)\big\| \\ &= \|(1 + ts_n)(x_n - x^*) - ts_n(f(x_n) - x^*)\| \\ &\geq \big|(1 + ts_n)\|x_n - x^*\| - ts_n\|f(x_n) - x^*\|\big| \\ &= \big|\|x_n - x^*\| + ts_n(\|x_n - x^*\| - \|f(x_n) - x^*\|)\big|. \end{aligned}$$

Since f is nonexpansive and x^* is a fixed point of f, $\|x_n - x^*\| \geq \|f(x_n) - x^*\|$, so

$$\|t(p - x^*) + (1 - t)(x_n - x^*)\| \geq \|x_n - x^*\|, \quad \text{for all } x^* \in Fix(T) \text{ and } t > 0. \tag{9.5}$$

Since (x_n) is bounded and f is nonexpansive, $(f(x_n))$ is also bounded. Therefore, because $x_n - f(x_n) = (1 - \lambda_n)(p - f(x_n))$ and (λ_n) converges to 1, $\|x_n - f(x_n)\|$ converges to 0.

Equation (9.5) together with Lemma 4.2.4(iii) (with $x = x_n - x^*$ and $y = t(p - x_n)$) also implies

$$\begin{aligned} \|x_n - x^* + t(p - x_n)\|^2 &\geq \|x_n - x^*\|^2 \\ &\geq \|x_n - x^* + t(p - x_n)\|^2 - 2t\langle p - x_n, J(x_n - x^* + t(p - x_n))\rangle, \end{aligned}$$

and so $\langle p - x_n, J(x_n - x^* + t(p - x_n))\rangle \geq 0$ since $t > 0$. Letting $t \to 0^+$ then implies using the norm-to-norm continuity of J that

$$\langle x_n - p, J(x_n - x^*)\rangle \leq 0, \quad \text{for all } x^* \in Fix(T). \tag{9.6}$$

Now observe that, if we denote $f(x_n)$ by y_n,

$$\begin{aligned}\langle y_n - p, J(y_n - y_m)\rangle - \langle x_n - p, J(x_n - y_m)\rangle &= \langle p, J(x_n - y_m) - J(y_n - y_m)\rangle \\ &\quad + \langle y_n, J(y_n - y_m)\rangle - \langle x_n, J(x_n - y_m)\rangle \\ &= \langle p, J(x_n - y_m) - J(y_n - y_m)\rangle \\ &\quad + \langle y_n - x_n, J(y_n - y_m)\rangle \\ &\quad + \langle x_n, J(y_n - y_m) - J(x_n - y_m)\rangle \\ &= \langle p - x_n, J(x_n - y_m) - J(y_n - y_m)\rangle \\ &\quad + \langle y_n - x_n, J(y_n - y_m)\rangle .\end{aligned}$$

Since J is norm-to-norm uniformly continuous on bounded subsets,

$$\lim_{n\to\infty} \langle y_n - p, J(y_n - y_m)\rangle - \langle x_n - p, J(x_n - y_m)\rangle = 0, \quad \text{for all } m \in \mathbb{N}. \tag{9.7}$$

uniformly in $m \in \mathbb{N}$. But $\langle x_n - p, J(x_n - y_m)\rangle \leq 0$ by (9.6) since $y_m \in Fix(T)$. Since $\langle x_n - p, J(x_n - y_m)\rangle \leq 0$, by (9.7), there exists a sequence $t_n \to 0^+$ such that

$$\langle y_n - p, J(y_n - y_m)\rangle \leq t_n, \quad \text{for all } m, n \in \mathbb{N}.$$

Similarly, interchanging the roles of m and n yields,

$$\langle y_m - p, J(y_m - y_n)\rangle \leq t_m, \quad \text{for all } m, n \in \mathbb{N}.$$

Adding these two inequalities yields

$$\|y_n - y_m\|^2 = \langle y_n - p + p - y_m, J(y_n - y_m)\rangle \leq t_n + t_m,$$

and so

$$\begin{aligned}\|x_n - x_m\| &\leq \|x_n - y_n\| + \|y_n - y_m\| + \|y_m - x_m\| \\ &\leq \|x_n - f(x_n)\| + \|x_m - f(x_m)\| + \sqrt{t_n + t_m}.\end{aligned}$$

Hence, $(f(x_n))$ and (x_n) are Cauchy sequences. Let x^* be the strong limit of $(f(x_n))$. Then

$$\|x_n - x^*\| \leq \|x_n - f(x_n)\| + \|f(x_n) - x^*\| ,$$

so x_n converges strongly to x^* as well. Moreover,

$$\begin{aligned}\|f(x^*) - x^*\| &\leq \|f(x^*) - f(x_n)\| + \|f(x_n) - x^*\| \\ &\leq \|x^* - x_n\| + \|f(x_n) - x^*\| ,\end{aligned}$$

so x^* is a fixed point of f and hence of T. Moreover, since J is norm-to-norm continuous, equation (9.6) implies

$$\langle x^* - p, J(x^* - z)\rangle \leq 0, \quad \text{for all } z \in Fix(T).$$

□

Remark 9.3.11. Observe that inequality (9.6) can also be shown as follows. For all $v \in Fix(T)$,

$$\begin{aligned}\langle x_n - p, J(x_n - v)\rangle &= \lambda_n\langle f(x_n) - p, J(x_n - v)\rangle \\ &= \lambda_n\langle f(x_n) - f(v), J(x_n - v)\rangle + \lambda_n\langle f(v) - p, J(x_n - v)\rangle \\ &\leq \lambda_n\|x_n - v\|^2 + \lambda_n\langle f(v) - v, J(x_n - v)\rangle \\ &\quad + \lambda_n\langle v - x_n, J(x_n - v)\rangle + \lambda_n\langle x_n - p, J(x_n - v)\rangle \\ &= \lambda_n\langle f(v) - v, J(x_n - v)\rangle + \lambda_n\langle x_n - p, J(x_n - v)\rangle.\end{aligned}$$

Since $\lambda_n < 1$ and $v \in Fix(T)$ implies $v \in Fix(f)$, we conclude

$$\langle x_n - p, J(x_n - v)\rangle \leq 0.$$

Remark 9.3.12. Recall that we observed in Step 2 of the proof that minimality of g is only required to prove $g \leq \varphi(T) \circ g$. As we will see in Section 9.5, if we define the sequence (g_n) by

$$g_0 := \mathrm{id}, \quad g_{n+1} := \varphi(T) \circ g_n$$

then $(g_n(x))$ converges weakly to a fixed point of T and hence of $\varphi(T)$ for all $x \in C$. Moreover, the mapping $g : x \mapsto \text{w-lim}_{n\to\infty} g_n(x)$ is nonexpansive and invariant on $Fix(T)$. Consequently, $\varphi(T) \circ g \leq g$ and $g \in N(Fix(T))$, i.e. as required.

Moreover, weak convergence of $(g_n(x))$ only requires weak sequential compactness of bounded, closed and convex subsets of uniformly smooth and uniformly convex Banach spaces instead of compactness of $P = \prod_{x\in C} C_x$. Observe that the latter was shown using Tychonoff's Theorem, which is equivalent to the Axiom of Choice, while the former is much weaker, as we will discuss in Section 9.7.

Before we come to the alternative proof, let us first sketch the functional interpretation of the existence of minimal elements f of $N(Fix(T))$.

9.4 The Functional Interpretation of the Existence of Minimal $f \in N(Fix(T))$

Observe that the existence of a minimal $f \in N(Fix(T))$ formalizes as follows:

$$\begin{aligned}&\exists f : C \to C\Big(f \in N(Fix(T)) \\ &\qquad \to \forall g : C \to C\big(g \in N(Fix(T)) \to \forall x, y \in C\,(\|fx - fy\| \leq \|gx - gy\|)\big)\Big).\end{aligned}$$

We now sketch the Shoenfield functional interpretation of this statement, which was shown coincide with the combination of the Krivine negative translation and the Dialectica interpretation by Streicher and Kohlenbach [91].

$$\begin{aligned}&\exists f : C \to C\Big(\forall x, y \in C\ \big(\|fx - fy\| \le \|x - y\| \wedge \big(\|Tx - x\| = 0 \to \|fx - x\| = 0\big)\big)\\&\quad \wedge \forall g : C \to C\Big(\forall x, y \in C\ \big(\|gx - gy\| \le \|x - y\| \wedge \big(\|Tx - x\| = 0 \to \|gx - x\| = 0\big)\big)\\&\qquad \to \forall x, y \in C\,(\|fx - fy\| \le \|gx - gy\|)\Big)\Big).\end{aligned}$$

When examining the proof, one notices that, in fact, we only need minimality with respect to the counterfunctional φ:

$$\begin{aligned}&\exists f : C \to C\Big(\forall x, y \in C\ \big(\|fx - fy\| \le \|x - y\| \wedge \big(\|Tx - x\| = 0 \to \|fx - x\| = 0\big)\big)\\&\quad \wedge \Big(\forall x, y \in C\ \big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| \wedge \big(\|Tx - x\| = 0 \to \|\varphi(f)x - x\| = 0\big)\big)\\&\qquad \to \forall x, y \in C\,(\|fx - fy\| \le \|\varphi(f)x - \varphi(f)y\|)\Big)\Big).\end{aligned}$$

Which is equivalent to

$$\begin{aligned}&\exists f : C \to C\Big(\forall x, y \in C\ \Big(\forall n_1\, \|fx - fy\| \le \|x - y\| + 2^{-n_1}\\&\qquad\qquad \wedge \big(\forall k_1\, \|Tx - x\| \le 2^{-k_1} \to \forall l_1\, \|fx - x\| \le 2^{-l_1}\big)\Big)\\&\quad \wedge \forall x, y \in C\ \Big(\big(\forall n_2 \|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2}\\&\qquad\qquad \wedge \big(\forall k_2\, \|Tx - x\| \le 2^{-k_2} \to \forall l_2\, \|\varphi(f)x - x\| \le 2^{-l_2}\big)\big)\\&\qquad \to \forall x, y \in C \forall m\,(\|fx - fy\| \le \|\varphi(f)x - \varphi(f)y\| + 2^{-m})\Big)\Big),\end{aligned}$$

which is equivalent to

$$\begin{aligned}&\exists f : C \to C\Big(\forall x, y \in C\, \forall l_1 \exists k_1 \exists n_1\\&\qquad \Big(\|fx - fy\| \le \|x - y\| + 2^{-n_1} \wedge \big(\|Tx - x\| \le 2^{-k_1} \to \|fx - x\| \le 2^{-l_1}\big)\Big)\\&\quad \wedge \forall x, y \in C\ \Big(\forall n_2 \forall k_2 \exists l_2\, \big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2}\\&\qquad\qquad \wedge \big(\|Tx - x\| \le 2^{-k_2} \to \|\varphi(f)x - x\| \le 2^{-l_2}\big)\big)\\&\qquad \to \forall x, y \in C \forall m\,(\|fx - fy\| \le \|\varphi(f)x - \varphi(f)y\| + 2^{-m})\Big)\Big),\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\exists f : C \to C\Big(\forall x, y \in C\, \forall l_1 \exists k_1 \exists n_1 \\
&\quad \Big(\|fx - fy\| \le \|x - y\| + 2^{-n_1} \wedge \big(\|Tx - x\| \le 2^{-k_1} \to \|fx - x\| \le 2^{-l_1}\big)\Big) \\
&\quad \wedge \forall x, y, u, v \in C \forall m \exists n_2, k_2 \forall l_2 \\
&\quad\quad \Big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2} \wedge \big(\|Tx - x\| \le 2^{-k_2} \to \|\varphi(f)x - x\| \le 2^{-l_2}\big) \\
&\quad\quad\quad \to \|fx - fy\| \le \|\varphi(f)x - \varphi(f)y\| + 2^{-m}\Big)\Big),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\exists f : C \to C\Big(\forall x, y \in C\, \forall l_1 \exists k_1 \exists n_1 \Big(\|fx - fy\| \le \|x - y\| + 2^{-n_1} \wedge \big(\|Tx - x\| \le 2^{-k_1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \to \|fx - x\| \le 2^{-l_1}\big)\Big) \\
&\quad \wedge \forall x, y, u, v \in C \forall m \forall L_2 : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \exists n_2, k_2 \Big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2} \\
&\qquad\qquad \wedge \big(\|Tx - x\| \le 2^{-k_2} \to \|\varphi(f)x - x\| \le 2^{-L_2(n_2,k_2)}\big) \\
&\qquad \to \|fu - fv\| \le \|\varphi(f)u - \varphi(f)v\| + 2^{-m}\Big)\Big),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\exists f : C \to C \forall u, v, w, x, y, z \in C \forall l_1, m \forall L_2 : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \exists n_1, n_2, k_1, k_2 \\
&\quad \Big(\|fw - fz\| \le \|w - z\| + 2^{-n_1} \wedge \big(\|Tw - w\| \le 2^{-k_1} \to \|fw - w\| \le 2^{-l_1}\big)\Big) \\
&\quad\quad \wedge \Big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2} \\
&\quad\quad\quad \wedge \big(\|Tx - x\| \le 2^{-k_2} \to \|\varphi(f)x - x\| \le 2^{-L_2(n_2,k_2)}\big) \\
&\quad\quad\quad\quad \to \|fu - fv\| \le \|\varphi(f)u - \varphi(f)v\| + 2^{-m}\Big),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\exists f : C \to C \exists N_1, N_2, K_1, K_2 : C^6 \times \mathbb{N}^2 \times (\mathbb{N} \times \mathbb{N} \to \mathbb{N}) \to \mathbb{N} \\
&\quad \forall u, v, w, x, y, z \in C \forall l_1, m \forall L_2 : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \\
&\quad\quad \Big(\|fw - fz\| \le \|w - z\| + 2^{-n_1} \wedge \big(\|Tw - w\| \le 2^{-k_1} \to \|fw - w\| \le 2^{-l_1}\big)\Big) \\
&\quad\quad\quad \Big(\|\varphi(f)x - \varphi(f)y\| \le \|x - y\| + 2^{-n_2} \\
&\quad\quad\quad\quad \wedge \big(\|Tx - x\| \le 2^{-k_2} \to \|\varphi(f)x - x\| \le 2^{-L_2(n_2,k_2)}\big) \\
&\quad\quad\quad\quad\quad \to \|fu - fv\| \le \|\varphi(f)u - \varphi(f)v\| + 2^{-m}\Big),
\end{aligned}$$

with $n_i := N_i(u, v, w, x, y, z, l_1, m, L_2)$ and $k_i := K_i(u, v, w, x, y, z, l_1, m, L_2)$ for $i \in \{1, 2\}$. A final application of QF-AC then results in a formula $\forall\exists A_{qf}$. We do not carry out this step since the resulting formula would become too large (each instance of e.g. u would have to be replaced by $U(f, N_1, N_2, K_1, K_2)$).

This is a prime example of how the types can rise through the use of functional interpretation. We do not intend to solve this problem; The functional interpretation of the statement $f \in N(Fix(T))$ is mainly included to illustrate the explosion of the involved types. Instead, we now propose an alternative approach.

9.5 An Alternative Proof

Suppose that X is uniformly convex and uniformly smooth (see Definition 4.1.6), $C \subset X$ is closed and convex and subset and $T : C \to C$ be a nonexpansive mapping with $Fix(T) \neq \emptyset$. Define a sequence of functions $f_k : C \to C$ iteratively by

$$f_0(x) := x,$$
$$f_{k+1}(x) := \frac{f_k(x) + Tf_k(x)}{2}.$$

Then $f_k(x)$ is the k-th Krasnoselskij iterate for T starting with $x \in C$ with parameter $1/2$. Moreover,

$$\|f_{k+1}(x) - f_{k+1}(y)\| \leq \frac{1}{2}\left(\|f_k(x) - f_k(y)\| + \|Tf_k(x) - Tf_k(y)\|\right) \leq \|f_k(x) - f_k(y)\|,$$

so each f_k is nonexpansive. As a result of Reich [84], the Krasnoselskij iteration converges weakly to a fixed point of T. Then, the mapping $f : C \to C$ defined by $f(x) :=$ w-lim $f_k(x)$ is nonexpansive:

$$\begin{aligned}
\|f(x) - f(y)\|^2 &= \langle f(x) - f(y), J(f(x) - f(y))\rangle \\
&= \lim_{k\to\infty} \langle f_k(x) - f_k(y), J(f(x) - f(y))\rangle \\
&\leq \langle x - y, J(f(x) - f(y))\rangle \\
&\leq \|x - y\| \cdot \|f(x) - f(y)\|. \qquad (9.8)
\end{aligned}$$

Now, observe that for all $x \in C$,

$$\|f_{k+1}(x) - x\| \leq \frac{1}{2}\left(\|f_k(x) - x\| + \|Tf_k(x) - Tx\| + \|Tx - x\|\right) \leq \|f_k(x) - x\| + \frac{\|Tx - x\|}{2}.$$

Thus, for all nonnegative integers k and all $x \in C$,

$$\|f_k(x) - x\| \leq k \cdot \frac{\|Tx - x\|}{2}.$$

In particular, if x is a fixed point of T, then it is a fixed point of f. Moreover, since the Krasnoselskij iteration is asymptotically regular and weakly convergent, the demiclosedness principle implies that the weak limit is a fixed point. Therefore, f is a nonexpansive retract.

9.6 Analysis of the Alternative Proof

Observe that Bruck showed that any nonexpansive retract $f : C \to Fix(T)$ can be transformed into a sunny nonexpansive retract $f' : C \to Fix(T)$. We now exhibit the finitary combinatorial core of the transformation.

Lemma 9.6.1. *Suppose* $\mathrm{diam}(C) \leq M$ *and fix an element* $p \in C$*. Given any* $f : C \to C$*, define for each positive integer* n *a function* $f_n(x) := \frac{1}{n}p + \frac{n-1}{n}f(x)$*. Moreover, for each* k*, denote by* $x_n^{(k)}$ *the* k*-th Picard iterate of* f_n *starting with* p*, i.e.* $x_n^{(k)} := f_n^{(k)}(p)$*. Given* $\varepsilon > 0$ *define* $n_0 := \max\left\lceil \frac{2M}{\omega(\varepsilon/5,M)}, \frac{5M^2}{\varepsilon}\right\rceil + 1$*, where* ω *is a modulus of norm-to-norm uniform continuity on bounded subsets for the duality mapping* J*, see Definition 4.2.9. Suppose that for some positive integer* k

(i) $\|f_{n_0}(x_{n_0}^{(k)}) - x_{n_0}^{(k)}\| \leq \dfrac{1}{2Mn_0}$

(ii) $\|f(x^*) - x^*\| \leq \frac{\varepsilon}{5Mn_0}$ *and*

(iii) $\|f(x_{n_0}^{(k)}) - f(x^*)\| \leq \|x_{n_0}^{(k)} - x^*\| + \frac{\varepsilon}{5M}$.

Then

$$\langle f(x_{n_0}^{(k)}) - p, J(f(x_{n_0}^{(k)}) - x^*)\rangle \leq \varepsilon.$$

Proof. Observe that for all integers $n \geq 2$,

$$x_n^{(k)} - f(x_n^{(k)}) = \frac{n}{n-1}\left(x_n^{(k)} - f_n(x_n^{(k)})\right) + \frac{1}{n-1}\left(p - x_n^{(k)}\right). \tag{9.9}$$

Therefore, for all $x^* \in C$,

$$\begin{aligned}
\|x_n^{(k)} - x^*\|^2 &= \langle x_n^{(k)} - x^*, J(x_n^{(k)} - x^*)\rangle \\
&= \langle f(x^*) - x^*, J(x_n^{(k)} - x^*)\rangle + \langle f(x_n^{(k)}) - f(x^*), J(x_n^{(k)} - x^*)\rangle \\
&\quad + \langle x_n^{(k)} - f(x_n^{(k)}), J(x_n^{(k)} - x^*)\rangle \\
&\leq M \cdot \|f(x^*) - x^*\| + \|f(x_n^{(k)}) - f(x^*)\| \cdot \|x_n^{(k)} - x^*\| \\
&\quad + \frac{n}{n-1}\langle x_n^{(k)} - f_n(x_n^{(k)}), J(x_n^{(k)} - x^*)\rangle + \frac{1}{n-1}\langle p - x_n^{(k)}, J(x_n^{(k)} - x^*)\rangle.
\end{aligned}$$

Consequently

$$\begin{aligned}
\langle x_n^{(k)} - p, J(x_n^{(k)} - x^*)\rangle &\leq (n-1)M \cdot \|f(x^*) - x^*\| + nM \cdot \|x_n^{(k)} - f_n(x_n^{(k)})\| \\
&\quad + \|x_n^{(k)} - x^*\|\left(\|f(x_n^{(k)}) - f(x^*)\| - \|x_n^{(k)} - x^*\|\right).
\end{aligned} \tag{9.10}$$

Therefore, using (9.9),

$$\begin{aligned}\langle f(x_n^{(k)}) - p, J(f(x_n^{(k)}) - x^*)\rangle &\le \langle f(x_n^{(k)}) - x_n^{(k)}, J(f(x_n^{(k)}) - x^*)\rangle \\ &\quad + \langle x_n^{(k)} - p, J(f(x_n^{(k)}) - x^*\rangle \\ &\le M \cdot \|f(x_n^{(k)}) - x_n^{(k)}\| + \langle x_n^{(k)} - p, J(x_n^{(k)} - x^*)\rangle \\ &\quad + \langle x_n^{(k)} - p, J(f(x_n^{(k)}) - x^*) - J(x_n^{(k)} - x^*)\rangle \\ &\le \frac{n}{n-1} \cdot M \cdot \|f_n(x_n^{(k)}) - x_n^{(k)}\| + \frac{M^2}{n-1} \\ &\quad + \langle x_n^{(k)} - p, J(x_n^{(k)} - x^*)\rangle \\ &\quad + \langle x_n^{(k)} - p, J(f(x_n^{(k)}) - x^*) - J(x_n^{(k)} - x^*)\rangle \\ &\le \frac{n}{n-1} \cdot M \cdot \|f_n(x_n^{(k)}) - x_n^{(k)}\| + \langle x_n^{(k)} - p, J(x_n^{(k)} - x^*)\rangle \\ &\quad + \langle x_n^{(k)} - p, J(f(x_n^{(k)}) - x^*) - J(x_n^{(k)} - x^*)\rangle + \frac{M^2}{n-1}.\end{aligned}$$

Combining this with (9.10) yields (since $\frac{n}{n-1} + n \le 2n$)

$$\begin{aligned}\langle f(x_n^{(k)}) - p, J(f(x_n^{(k)}) - x^*)\rangle \le 2Mn \cdot \|f_n(x_n^{(k)}) - x_n^{(k)}\| \\ + \frac{M^2}{n-1} + (n-1)M \cdot \|f(x^*) - x^*\| \\ + M \cdot \Big(\|f(x_n^{(k)}) - f(x^*)\| - \|x_n^{(k)} - x^*\|\Big) \\ + \langle x_n^{(k)} - p, J(f(x_n^{(k)}) - x^*) - J(x_n^{(k)} - x^*)\rangle. \quad (9.11)\end{aligned}$$

Now observe that by (9.9)

$$\begin{aligned}\|f(x_n^{(k)}) - x^* - (x_n^{(k)} - x^*)\| &= \|f(x_n^{(k)}) - x_n^{(k)}\| \\ &\le \frac{n}{n-1}\|x_n^{(k)} - f_n(x_n^{(k)})\| + \frac{1}{n-1}\|p - x_n^{(k)}\| \\ &\le 2 \cdot \|x_n^{(k)} - f_n(x_n^{(k)})\| + \frac{M}{n-1}.\end{aligned}$$

Therefore, since $\frac{1}{n_0-1} \le \frac{\omega(\varepsilon/5,M)}{2M}$ and $\|x_{n_0}^{(k)} - f_{n_0}(x_{n_0}^{(k)})\| \le \frac{\omega(\varepsilon/5,M)}{4}$, this implies

$$\langle x_{n_0}^{(k)} - p, J(f(x_{n_0}^{(k)}) - x^*) - J(x_{n_0}^{(k)} - x^*)\rangle \le \frac{\varepsilon}{5}.$$

Going back to (9.11), we obtain using our hypotheses

$$\langle f(x_{n_0}^{(k)}) - p, J(f(x_{n_0}^{(k)}) - x^*)\rangle \le \varepsilon.$$

□

Lemma 9.6.2. *Suppose* $\mathrm{diam}(C) \le M$ *and fix an element* $p \in C$*. Given any* $f : C \to C$*, define for each positive integer* n *a function* $f_n(x) := \frac{1}{n}p + \frac{n-1}{n}f(x)$*. Moreover, for each*

k, denote by $x_n^{(k)}$ the k-th Picard iterate of f_n starting with p, i.e. $x_n^{(k)} := f_n^{(k)}(p)$. Given $\varepsilon > 0$ and any integer $n \geq 2$, suppose

$$\|f(x_n^{(k)}) - f(x_n^{(k-1)})\| \leq \|x_n^{(k)} - x_n^{(k-1)}\| + \frac{\varepsilon}{2n}, \quad \text{for all } k \leq \lceil \log_{\frac{n-1}{n}}(\varepsilon/2M) \rceil.$$

Then, for $k_0 := \lceil \log_{\frac{n-1}{n}}(\varepsilon/2M) \rceil$,

$$\|x_n^{(k_0)} - f_n(x_n^{(k_0)})\| \leq \varepsilon, \quad \text{for all positive integers } n.$$

Proof. Observe that, by definition,

$$\begin{aligned}
\|x_n^{(k+1)} - x_n^{(k)}\| = \|f_n(x_n^{(k)}) - f_n(x_n^{(k-1)})\| &= \frac{n-1}{n} \cdot \|f(x_n^{(k)}) - f(x_n^{(k-1)})\| \\
&\leq \frac{n-1}{n} \left(\|x_n^{(k)} - x_n^{(k-1)}\| + \frac{\varepsilon}{2n} \right)
\end{aligned}$$

Therefore, since $1 - \frac{n-1}{n} = \frac{1}{n}$,

$$\begin{aligned}
\|f_n(x_n^{(k_0)}) - x_n^{(k_0)}\| &\leq \left(\frac{n-1}{n}\right)^{k_0} \cdot M + \frac{\varepsilon}{2n} \cdot \sum_{k=0}^{k_0} \left(\frac{n-1}{n}\right)^k \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2n} \cdot n \left(1 - \left(\frac{n-1}{n}\right)^{k_0}\right) \\
&\leq \varepsilon,
\end{aligned}$$

which completes the proof. □

Corollary 9.6.3. *Fix an element $p \in C$. Given any $f : C \to C$, define for each positive integer n a function $f_n(x) := \frac{1}{n}p + \frac{n-1}{n}f(x)$. Moreover, for each k, denote by $x_n^{(k)}$ the k-th Picard iterate of f_n starting with p, i.e. $x_n^{(k)} := f_n^{(k)}(p)$. Given $\varepsilon > 0$ define $n_0 := \max\left\lceil \frac{2M}{\omega(\varepsilon/5,M)}, \frac{5M^2}{\varepsilon} \right\rceil + 1$, $\varepsilon' := \frac{1}{2Mn_0}$ and $k_0 := \lceil \log_{\frac{n_0-1}{n_0}}(\varepsilon'/2M) \rceil$. Suppose for some $x^* \in C$,*

(i) $\|f(x_{n_0}^{(k)}) - f(x_{n_0}^{(k-1)})\| \leq \|x_{n_0}^{(k)} - x_{n_0}^{(k-1)}\| + \frac{\varepsilon'}{2n_0}$ *for all* $k \leq k_0$*,*

(ii) $\|f(x^*) - x^*\| \leq \frac{\varepsilon}{5Mn_0}$ *and*

(iii) $\|f(x_{n_0}^{(k)}) - f(x^*)\| \leq \|x_{n_0}^{(k)} - x^*\| + \frac{\varepsilon}{5M}$*.*

Then

$$\langle f(x_{n_0}^{(k_0)}) - p, J(f(x_{n_0}^{(k_0)}) - x^*) \rangle \leq \varepsilon.$$

Proof. Assumption (i) implies the first condition of Lemma 9.6.1 because of Lemma 9.6.2. Assumptions (ii) and (iii) and the claim are identical to those in Lemma 9.6.1 with $k := k_0$. □

Remark 9.6.4. If $f : C \to C$ is a nonexpansive retract of C onto $Fix(T)$ and $x^* \in Fix(T)$, then the conditions of Lemma 9.6.3 are all satisfied for any $\varepsilon > 0$.

Remark 9.6.5. The use of the nonexpansive retract has been reduced to all instances for which its defining properties were used, i.e. all points appearing in (i) to (iii) above. In fact, Corollary 9.6.3 and the preceding Lemmas still remain true if we replace each occurrence of $f(x_n^{(k)})$ by $y_n^{(k)}$, each occurrence of $f(x^*)$ by y^* and each occurrence of $f_n(x_n^(k))$ by $\frac{1}{n}p + \frac{n-1}{n}f(x)$. The use of the function f in the proofs is purely for convenience and increased readability.

9.7 Weak Sequential Compactness

In what follows, we examine the non-constructive core of our alternate proof, which consists in the existence of the function $f : C \to C$ defined by $f(x) :=$ w-lim $f_k(x)$. The proof may be subdivided into the following steps:

(i) Being closed, bounded and convex, C is weakly sequentially compact. Thus, $(f_k(x))$ possesses a weakly convergent subsequence (f_{k_j}) with weak limit w.

(ii) $(f_k(x))$ is asymptotically regular, i.e. $\|Tf_k(x) - f_k(x)\| \to 0$.

(iii) By Browder's demiclosedness principle [11], any subsequential limit is a fixed point of T.

(iv) Using (ii) and (iii), Browder shows that $(f_k(x))$ has at most one weak cluster point.

First, it is noteworthy to recall that we chose $f(x) :=$ w-lim $f_k(x)$ since this is a particular nonexpansive retraction onto $Fix(T)$. But by (iii) above, any weak *cluster* point is a fixed point of T. Moreover, going back to (9.8), we see that if v and w are weak cluster points of $(f_k(x))$ and $(f_k(y))$ respectively, then $\|v - w\| \le \|x - y\|$. Hence, while the actual weak convergence of $(f_k(x))$ holds true, it is not necessary to carry out the proof, and therefore does not need to be analyzed further.

Therefore, only (i) to (iii) need to be considered further. A rate of asymptotic regularity for $(f_k(x))$, i.e. part (ii), is well known, while (iii) has been carried out in [53] for the Hilbert space case. However, the result generalizes immediately to the case of uniformly smooth Banach spaces. The challenging part, therefore, is to carry out the analysis of (i).

The first challenge is to find a suitable proof of (i). For Hilbert spaces, the classical proof of the weak compactness of the unit ball needs the following ingredients:

1. Sequential compactness of the Hilbert cube.
2. The Riesz Representation Theorem applied to the separable space $\mathcal{L}$.
3. Steps 1 and 2 imply the sequential compactness of the separable closed linear subspace $\mathcal{L} := \overline{\mathrm{Span}}\{x_n : n \in \mathbb{N}\}$.

4. The orthogonal projection of points of the Hilbert space onto $\mathcal{L}$ then implies the claim.

For spaces more general than Hilbert spaces, the proofs found in the literature usually take the detour via weak-* compactness to conclude weak compactness from reflexivity. This approach, however, is unsuitable from a proof mining perspective.

However, it seems feasible to generalize the procedure outlined above to uniformly convex and uniformly smooth Banach spaces; Theorem 4.2.10 provides a representation theorem. Moreover, Giles proof given in [32] seems to be a particularly suitable candidate for analysis as it directly extends the Hilbert space argument.

Moreover, in uniformly convex and uniformly smooth Banach spaces, every nonempty closed subspaces possesses a normal vector, see e.g. Giles [32], i.e. a vector that is orthogonal in the sense of Definition 9.3.4 to all elements of the closed subspace.

10 Conclusion and Future Research

10.1 Conclusion

In this thesis, we presented numerous applications of the proof mining machinery to concrete proofs of theorems in mathematics, more precisely fixed point theory and nonlinear convex optimization. The logical tools were laid out, along with their advantages and limitations. We then examined key analytical concepts and their compatibility with respect to existing metatheorems.

We then applied this machinery to a convergence result for Halpern iterations, which can be seen as nonlinear generalizations of von Neumann's Mean Ergodic Theorem. We first extracted a rate of convergence for the asymptotic regularity in arbitrary normed spaces (Theorem 5.3.1). We then give a rate of metastability for Halpern iterations under Xu's conditions in uniformly smooth Banach spaces relative to a rate of metastability of the resolvent of the underlying operator (Theorem 5.3.4). This was then combined with the rate of metastability for Hilbert spaces (see the discussion following Theorem 5.3.5), which is known from [52].

We then shifted our focus to convergence results for Bruck's iteration, which approximates fixed points of pseudocontractive mappings. Fixed points of these mappings correspond to zeros of accretive operators, which, in turn, are the stability points of nonlinear evolution equations. For the class of Lipschitzian pseudocontractions, convergence results are shown in a fashion similar to corresponding theorems for the subclass of nonexpansive mappings. We first provided a rate of metastability for the resolvent in Hilbert spaces (Theorem 6.2.4) that is similar to the rate for nonexpansive mappings. We then combined this with an earlier result [70] to provide a full, quantitative version of the convergence of Bruck's iteration for Lipschitzian pseudocontractions in Hilbert spaces (Theorem 6.2.9).

Next, we developed the quantitative theory of pseudocontractions further by generalizing the results from Lipschitzian pseudocontractions to demicontinuous pseudocontractions (Theorems 7.3.5 and 7.3.11). This class is significantly more general since demicontinuity does not even imply norm continuity of the operator. We discussed interesting logical phenomena that stem from the lack of continuity and hence the lack of extensionality. The shift to the substantially weaker continuity condition, however, lead to more complex rates. In particular, the analysis of Bruck's proof only yielded a metastable version even for the asymptotic regularity.

In Chapter 6, we applied proof mining methods to a convex optimization problem that is reduced to a fixed point problem. We provided a full, quantitative analysis of the resulting algorithms in the proof mining sense (Theorems 8.6.11 and 8.7.9). The analysis of Yamada's proof [104] allowed for the opportunity to witness interesting logical phenomena in mathematical practice. Firstly, the transformed, finitary proof exhibited the increase in types predicted by the metatheorems. This in particular demonstrated the mechanics of majorization, and, closely related, the complete modularity of the Dialectica interpretation with respect to the modus ponens.

In the final chapter, we worked towards a rate of metastability for the resolvent of non-expansive operators in uniformly smooth and uniformly convex Banach spaces. The solution thereof would allow one to immediately obtain a rate of metastability for Halpern's iteration as well by means of the results we presented in Chapter 5. While our results in this vein have only been partial, we outlined a strategy to tackle the problem in the future.

10.2 Future Research

The canonical next step is to verify that the alternative proof given in Section 9.5 is indeed formalizable in a suitable system $\mathcal{A}^\omega[X, \|\cdot\|, \ldots]$. Should that be the case, it will be interesting to carry out the extraction in order to examine the complexity of the bound.

In Chapter 8, we introduced the Hybrid Steepest Descent Method due to Yamada [104] and gave a brief motivation of its importance in convex optimization. As pointed out by Yamada and Ogura [105] and Yamada, Yukawa and Yamagishi [107], numerous further problems can be solved by the Hybrid Steepest Descent Method if the domain of the Variational Inequality Problem is the fixed point set of a quasi-nonexpansive mapping instead of a nonexpansive mapping. Yamada and Ogura [105] extended the convergence of the Hybrid Steepest Descent Method to quasi-shrinking maps, which are a subclass of the quasi-nonexpansive mappings, and their fixed point sets characterize the level sets of certain continuous convex mappings, see [105]. Therefore, extending the results of Chapter 8 to quasi-shrinking mappings presents an interesting proof-mining idea for the future.

Bibliography

[1] A. Aleyner and S. Reich. An explicit construction of sunny nonexpansive retractions in Banach spaces. *Fixed Point Theory Appl.*, (3):295–305, 2005.

[2] J. Avigad, P. Gerhardy, and H. Towsner. Local stability of ergodic averages. *Transactions of the American Mathematical Society*, 362:261–288, 2010.

[3] V. Barbu and T. Precupanu. *Convexity and optimization in Banach spaces*. Springer Monographs in Mathematics. Springer, Dordrecht, fourth edition, 2012.

[4] H. H. Bauschke. The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space. *Journal of Mathematical Analysis and Applications*, 202(1):150 – 159, 1996.

[5] G. Bellin. Ramsey interpreted: a parametric version of Ramsey's theorem. In *Logic and computation (Pittsburgh, PA, 1987)*, volume 106 of *Contemporary Mathematics*, pages 17–37. American Mathematical Society, 1990.

[6] U. Berger. Program extraction from normalization proofs. In *Typed lambda calculi and applications (Utrecht, 1993)*, volume 664 of *Lecture Notes in Computer Science*, pages 91–106. Springer, 1993.

[7] G. Birkhoff. Orthogonality in linear metric spaces. *Duke Mathematical Journal*, 1:169–172, 1935.

[8] G. Birkhoff. The mean ergodic theorem. *Duke Mathematical Journal*, 5(1):19–20, 1939.

[9] E. Bishop. *Foundations of constructive analysis*. McGraw-Hill, 1967.

[10] F. E. Browder. Existence of periodic solutions for nonlinear equations of evolution. *Proceedings of the National Academy of Sciences*, 53:1100–1103, 1965.

[11] F. E. Browder. Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces. *Archive for Rational Mechanics and Analysis*, 24:82–90, 1967.

[12] F. E. Browder. Nonlinear mappings of nonexpansive and accretive type in Banach spaces. *Bulletin of the American Mathematical Society*, 73:875–882, 1967.

[13] R. E. Bruck. Nonexpansive projections on subsets of Banach spaces. *Pacific Journal of Mathematics*, 47:341–355, 1973.

[14] R. E. Bruck. Properties of fixed-point sets of nonexpansive mappings in Banach spaces. *Transactions of the American Mathematical Society*, 179:251–262, 1973.

[15] R. E. Bruck. A strongly convergent iterative solution of $0 \in U(x)$ for a maximal monotone operator U in Hilbert space. *Journal of Mathematical Analysis and Applications*, 48:114–126, 1974.

[16] R. E. Bruck. Random products of contractions in metric and Banach spaces. *Journal of Mathematical Analysis and Applications*, 88(2):319–332, 1982.

[17] R. E. Bruck, Jr. Nonexpansive retracts of Banach spaces. *Bulletin of the American Mathematical Society*, 76:384–386, 1970.

[18] C. Chidume. *Geometric Properties of Banach Spaces and Nonlinear Iterations.* Springer Lecture Notes in Mathematics, 2009.

[19] C. E. Chidume and H. Zegeye. Approximate fixed point sequences and convergence theorems for Lipschitz pseudocontractive maps. *Proceedings of the American Mathematical Society*, 132:831–840, 2000.

[20] J. A. Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.

[21] D. Dalen. *Logic and Structure.* Universitext (1979). Springer, 2004.

[22] D. G. de Figueiredo and L. A. Karlovitz. On the radial projection in normed spaces. *Bulletin of the American Mathematical Society*, 73:364–368, 1967.

[23] C. N. Delzell. Kreisel's unwinding of Artin's proof. In P. Odifreddi, editor, *Kreiseliana*, pages 113–246. A K Peters, Wellesley, MA, 1996.

[24] T. Dominguez Benavides, G. López Acedo, and H.-K. Xu. Construction of sunny nonexpansive retractions in Banach spaces. *Bull. Austral. Math. Soc.*, 66(1):9–16, 2002.

[25] I. Ekeland and R. Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999.

[26] S. Feferman. Theories of finite type related to mathematical practice. In J. Barwise, editor, *Handbook of Mathematical Logic*, volume 90 of *Studies in Logic and the Foundations of Mathematics*, pages 913 – 971. Elsevier, 1977.

[27] S. Feferman. Kreisel's "unwinding" program. In P. Odifreddi, editor, *Kreiseliana*, pages 247–273. A K Peters, Wellesley, MA, 1996.

[28] K. O. Friedrichs. Symmetric positive linear differential equations. *Communications on Pure and Applied Mathematics*, 11:333–418, 1958.

[29] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49(4):431–446, 2008.

[30] P. Gerhardy. Proof mining in practice. In *Logic Colloquium 2007*, volume 35 of *Lecture Notes in Logic*, pages 82–91. Association for Symbolic Logic, 2010.

[31] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360(5):2615–2660, 2008.

[32] J. R. Giles. Classes of semi-inner-product spaces. *Transactions of the American Mathematical Society*, 129:436–446, 1967.

[33] J.-Y. Girard. *Proof theory and logical complexity*, volume 1 of *Studies in Proof Theory. Monographs.* Bibliopolis, Naples, 1987.

[34] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958.

[35] K. Goebel and W. Kirk. *Topics in Metric Fixed Point Theory.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1990.

[36] A. A. Goldstein. Convex programming in Hilbert space. *Bulletin of the American Mathematical Society*, 70:709–710, 1964.

[37] D. Günzel and U. Kohlenbach. Logical metatheorems for abstract spaces axiomaaxiom in positive bounded logic. *Advances in Mathematics*, 290:503–551, 2016.

[38] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73:957–961, 1967.

[39] O. Hanner. On the uniform convexity of L^p and ℓ^p. *Arkiv för Matematik*, 3(3):239–244, 1956.

[40] C. W. Henson and J. Iovino. Ultraproducts in analysis. In *Analysis and logic (Mons, 1997)*, volume 262 of *London Mathematical Society Lecture Note Series.*, pages 1–110. Cambridge University Press, Cambridge, 2002.

[41] R. C. James. Orthogonality and linear functionals in normed linear spaces. *Transactions of the American Mathematical Society*, 61(2):265–292, 1947.

[42] J. S. Jung. Viscosity approximation methods for a family of finite nonexpansive mappings in Banach spaces. *Nonlinear Analysis*, 64(11):2536–2552, 2006.

[43] T. Kato. Nonlinear semigroups and evolution equations. *Journal of the Mathematical Society Japan*, 19:508–520, 1967.

[44] W. A. Kirk and B. Sims, editors. *Handbook of metric fixed point theory.* Kluwer Academic Publishers, Dordrecht, 2001.

[45] U. Kohlenbach. Analysing proofs in analysis. In *Logic: from foundations to applications (Staffordshire, 1993)*, Oxford Science Publications, pages 225–260. Oxford Univ. Press, New York, 1996.

[46] U. Kohlenbach. Arithmetizing proofs in analysis. In *Logic Colloquium '96 (San Sebastián)*, volume 12 of *Lecture Notes Logic*, pages 115–158. Springer, Berlin, 1998.

[47] U. Kohlenbach. On the no-counterexample interpretation. *Journal of Symbolic Logic*, 64(4):1491–1511, 1999.

[48] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.

[49] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics.* Springer Monographs in Mathematics, 2008.

[50] U. Kohlenbach. Effective uniform bounds from proofs in abstract functional analysis. In S. B. Cooper, B. Löwe, and A. Sorbi, editors, *New computational paradigms*, pages 223–258. Springer, New York, 2008.

[51] U. Kohlenbach. On the logical analysis of proofs based on nonseparable Hilbert space theory. In *Proofs, Categories and Computations. Essays in Honor of Grigori Mints*, pages 131–143. College Publications, 2010.

[52] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226:2764–2795, 2011.

[53] U. Kohlenbach. A uniform quantitative form of sequential weak compactness and Baillon's nonlinear ergodic theorem. *Communications in Contemporary Mathematics*, 14(1):1250006, 20, 2012.

[54] U. Kohlenbach. On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces. *To appear in: Israel Journal of Mathematics*, 2016.

[55] U. Kohlenbach and L. Leuştean. A quantitative mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory and Dynamical Systems*, 29(6):1907–1915, 2009.

[56] U. Kohlenbach and L. Leuştean. Asymptotically nonexpansive mappings in uniformly convex hyperbolic spaces. *Journal of the European Mathematical Society*, 12(1):71–92, 2010.

[57] U. Kohlenbach and L. Leuştean. Effective metastability of Halpern iterates in CAT(0) spaces. *Advances in Mathematics*, 231:2526–2556, 2012.

[58] U. Kohlenbach and L. Leuştean. On the computational content of convergence proofs via Banach limits. *Philosophical Transactions of the Royal Society. Series A. Mathematical, Physical and Engineering Sciences*, 370:pp. 3449–3463, 2012.

[59] U. Kohlenbach and L. Leuştean. Addendum to [57]. *Advances in Mathematics*, 250:650–651, 2014.

[60] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative results on Fejer monotone sequences. *ArXiv e-prints*, 2014.

[61] U. Kohlenbach and P. Safarik. Fluctuations, effective learnability and metastability in analysis. *Annals of Pure and Applied Logic*, 165(1):266–304, 2014.

[62] G. Kreisel. On the interpretation of non-finitist proofs, part I. *Journal of Symbolic Logic*, 16:241–267, 1951.

[63] G. Kreisel. On the interpretation of non-finitist proofs, part II: Interpretation of number theory. Applications. *Journal of Symbolic Logic*, 17:43–58, 1952.

[64] G. Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting, editor, *Constructivity in Mathematics*, pages 101–128. Amsterdam, North-Holland Pub. Co., 1959.

[65] G. Kreisel. On weak completeness of intuitionistic predicate logic. *Journal of Symbolic Logic*, 27(2):139–158, 06 1962.

[66] S. Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Mathematical Journal*, 2:35–47, 1951.

[67] D. Körnlein. Logical extraction of bounds on asymptotic regularity for pseudocontractions. Bachelor's thesis, Technische Universität Darmstadt, 2010.

[68] D. Körnlein. Quantitative results for Bruck iterations of demicontinuous pseudocontractions. *submitted*, 2015.

[69] D. Körnlein. Quantitative results for Halpern iterations of nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 428(2):1161–1172, 2015.

[70] D. Körnlein and U. Kohlenbach. Effective rates of convergence for Lipschitzian pseudocontractive mappings in general Banach spaces. *Nonlinear Analysis*, 74:5253–5267, 2011.

[71] D. Körnlein and U. Kohlenbach. Rate of metastability for Bruck's iteration of pseudocontractive mappings in Hilbert space. *Numerical Functional Analysis and Optimization*, 35(1):20–31, 2014.

[72] K. Lan and J. Wu. Convergence of approximants for demicontinuous pseudocontractive maps in Hilbert spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 49(6):737–746, June 2002.

[73] L. Leuştean. Rates of asymptotic regularity for Halpern iterations of nonexpansive mappings. *Journal of Universal Computer Science*, 13:1680–1691, 2007.

[74] L. Leuştean and A. Nicolae. Effective results on nonlinear ergodic averages in CAT spaces. *Ergodic Theory and Dynamical Systems*, pages 1–22 (electronic), 1 2016.

[75] L. Leuştean. Nonexpansive iterations in uniformly convex W-hyperbolic spaces. In *Nonlinear analysis and optimization I. Nonlinear analysis*, volume 513 of *Contemporary Mathematics*, pages 193–210. American Mathematical Society, 2010.

[76] E. Levitin and B. Polyak. Constrained minimization methods. *Computational Mathematics and Mathematical Physics*, 6(5):1 – 50, 1966.

[77] H. Luckhardt. Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. *Journal of Symbolic Logic*, 54(1):234–263, 1989.

[78] H. Luckhardt. Bounds extracted by Kreisel from ineffective proofs. In P. Odifreddi, editor, *Kreiseliana*, pages 289–300. A K Peters, 1996.

[79] G. Lumer. Semi-inner-product spaces. *Transactions of the American Mathematical Society*, 100:29–43, 1961.

[80] C. H. Morales and J. S. Jung. Convergence of paths for pseudocontractractive mappings in Banach spaces. *Proceedings of the American Mathematical Society*, 128:3411–3419, 2000.

[81] A. Moudafi. Viscosity approximation methods for fixed-points problems. *Journal of Mathematical Analysis and Applications*, 241(1):46–55, 2000.

[82] E. Neumann. Computational problems in metric fixed point theory and their Weihrauch degrees. *Logical Methods in Computer Science*, 11(4):1–44, 2015.

[83] W. V. Petryshyn. A characterization of strict convexity of Banach spaces and other uses of duality mappings. *Journal of Functional Analysis*, 6:282–291, 1970.

[84] S. Reich. Weak convergence theorems for nonexpansive mappings in Banach spaces. *Journal of Mathematical Analysis and Applications*, 67(2):274–276, 1979.

[85] S. Reich. Strong convergence theorems for resolvents of accretive operators in Banach spaces. *Journal of Mathematical Analysis and Applications*, 75:287–292, 1980.

[86] K. C. Schade. Proof mining for Halpern iterations in CAT(0) spaces. Master's thesis, TU Darmstadt, 2012.

[87] K. C. Schade and U. Kohlenbach. Effective metastability for modified Halpern iterations in CAT(0) spaces. *Fixed Point Theory and Applications*, 2012.

[88] N. Shioji and W. Takahashi. Strong convergence of approximated sequences for nonexpansive mappings in Banach spaces. *Proceedings of the American Mathematical Society*, 125:3641–3645, 1997.

[89] J. R. Shoenfield. *Mathematical logic*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1967.

[90] C. Spector. Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics. In *Proc of Symposia in Pure Mathematics, Vol. V*, pages 1–27. American Mathematical Society, 1962.

[91] T. Streicher and U. Kohlenbach. Shoenfield is Gödel after Krivine. *Mathematical Logic Quarterly*, 53(2):176–179, 2007.

[92] G. Sundholm. Constructions, proofs and the meaning of logical constants. *Journal of Philosophical Logic*, 12(2):151–172, 1983.

[93] T. Suzuki. Some notes on Bauschke's condition. *Nonlinear Anal.*, 67(7):2224–2231, 2007.

[94] W. Takahashi. *Nonlinear functional analysis*. Yokohama Publishers, Yokohama, 2000.

[95] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.

[96] T. Tao. *Structure and Randomness: Pages from Year One of a Mathematical Blog.* American Mathematical Society, 2008.

[97] A. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis.* Lecture Notes in Mathematics. Springer, 1973.

[98] A. S. Troelstra and D. van Dalen. *Constructivism in mathematics. Vol. I and II*, volume 121 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, 1988.

[99] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences*, 18:70–82, 1932.

[100] R. Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58:486–491, 1992.

[101] H. Xu and T. Kim. Convergence of hybrid steepest-descent methods for variational inequalities. *Journal of Optimization Theory and Applications (JOTA)*, 119(1):185–201, 2003.

[102] H.-K. Xu. Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66:240–256, 2002.

[103] H.-K. Xu. Viscosity approximation methods for nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 298(1):279–291, 2004.

[104] I. Yamada. The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In Y. C. Dan Butnariu and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 473 – 504. Elsevier, 2001.

[105] I. Yamada and N. Ogura. Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings. *Numerical Functional Analysis and Optimization*, 25(7-8):619–655, 2005.

[106] I. Yamada, N. Ogura, and N. Shirakawa. A numerically robust hybrid steepest descent method for the convexly constrained generalized inverse problems. In *Inverse problems, image analysis, and medical imaging (New Orleans, LA, 2001)*, volume 313 of *Contemporary Mathematics*, pages 269–305. American Mathematical Society, 2002.

[107] I. Yamada, M. Yukawa, and M. Yamagishi. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, chapter Minimizing the Moreau Envelope of Nonsmooth Convex Functions over the Fixed Point Set of Certain Quasi-Nonexpansive Mappings, pages 345–390. Springer New York, New York, NY, 2011.

[108] E. Zeidler. *Nonlinear functional analysis and its applications. III.* Springer-Verlag, New York, 1985.

[109] E. Zeidler. *Nonlinear functional analysis and its applications. II/B.* Springer-Verlag, New York, 1990.

Wissenschaftlicher Werdegang

1990 *Geboren am* 7. Juli 1990.

2007–2010 *Technische Universität Darmstadt*, BSc. Mathematics with Computer Science.

2010–2011 *London School of Economics*, MSc. Financial Mathematics.

2011–2016 *Technische Universität Darmstadt*, Wissenschaftlicher Mitarbeiter am Fachbereich Mathematik, Promotion in Angewandter Beweistheorie.