TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik

Proof-Theoretical Aspects of Nonlinear and Set-Valued Analysis

Vom Fachbereich Mathematik
der Technischen Universität Darmstadt
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte Dissertation

von

Nicholas Norbert Pischke, M.Sc.

aus Berlin

Referent:	Prof. Dr. Ulrich Kohlenbach
1. Korreferent:	Prof. Laurenţiu Leuştean, Ph.D.
2. Korreferent:	Prof. Adriana Nicolae, Ph.D.
Tag der Einreichung:	10.10.2023
Tag der mündlichen Prüfung:	19.01.2024

Darmstadt 2024
D 17

Pischke, Nicholas Norbert : Proof-Theoretical Aspects of Nonlinear and Set-Valued Analysis
Darmstadt, Technische Universität Darmstadt,
Jahr der Veröffentlichung der Dissertation auf TUprints: 2024
URN: urn:nbn:de:tuda-tuprints-265849
Tag der mündlichen Prüfung: 19.01.2024

Auf einsamen Begriff gestellt
Ragt ein Gebäude steil hinauf:
Und fügt sich an den Sternenhauf
Von ferner Göttlichkeit durchhellt.

HERMANN BROCH
Mathematisches Mysterium, 1913

אם אין דעת אין בינה
אם אין בינה אין דעת

AVOT 3:17

Acknowledgments

The only person who could be mentioned first in the acknowledgments is my advisor, Prof. Dr. Ulrich Kohlenbach. In 2019, after taking his class on applied proof theory, my mathematical life has been influenced in the greatest way. He supervised both my Bachelor and Master thesis, each time being incredibly generous regarding ideas for potential research. The work on this thesis has been no different. But even beyond the mathematical advice as well as the incredibly precise and insightful comments that are characteristic for him, I have benefited also from all personal conversations with him which are among the most cherished moments of my recent years. I could surely have not imagined a better supervisor.

Beyond my advisor, I want to thank Pedro Pinto who I am lucky to call both a co-author as well as a good friend.

In terms of good friends, I want to thank Patrick Uftring for making the university a kind of intellectual haven that I have always dreamed of and for inspiring me to work more than I presumably should.

This list would not be complete without also thanking Davide Manca and Anton Freund for many interesting conversations around proof theory and especially beyond (most importantly during lunch time).

Further, I want to thank Prof. Laurenţiu Leuştean, Ph.D. as well as Prof. Adriana Nicolae, Ph.D. for refereeing this thesis and for providing a lot of valuable feedback. Also, I want to thank Prof. Dr. Thomas Streicher and Prof. Dr. Stefan Ulbrich for agreeing to participate in my thesis committee.

I also want to thank the Deutsche Forschungsgesellschaft for their support which I received while working on this thesis under the grant DFG KO 1737/6-2.

Last, but in no way least, I want to thank Swenja, arguably the most important person in my life. Without her, not a single word in this thesis would ever have been written. I put myself into this work, so there is also something of you in it.

Abstract

This thesis is concerned with extending the underlying logical approach as well as the breadth of applications of the proof mining program to various (mostly previously untreated) areas of nonlinear analysis and optimization, with a particular focus being placed on topics which involve set-valued operators.

For this, we extend the current logical methodology of proof mining by new systems and corresponding so-called logical metatheorems that cover these more involved areas of nonlinear analysis. Most of these systems crucially rely on the use of intensional methods, treating sets with potentially high quantifier complexity in the defining matrix via characteristic functions and axioms that describe only their properties and do not completely characterize the elements of the sets.

The applicability of all of these metatheorems is then substantiated by a range of case studies for the respective areas which in particular also highlight the naturalness of the use of intensional methods in the design of the corresponding systems.

The first new area covered thereby is the theory of nonlinear semigroups induced by corresponding evolution equations for accretive operators. In that context, we present (besides an initial foray into the area from 2015) essentially the first applications of proof mining to the theory of partial differential equations. Concretely, we provide quantitative versions of four central results on the asymptotic behavior of solutions to such equations.

The second new area unlocked in this thesis is that of the continuous dual of a Banach space and its norm (which are also approached via intensional methods). This in particular relies on a proof-theoretically tame treatment of suprema over (certain) bounded sets in this intensional context which is further exploited later on. These systems, which give access to this until now untreated fundamental notion from func-

tional analysis, are then used to provide further substantial extensions to treat various notions from convex analysis like the Fréchet derivative of a convex function, Fenchel conjugates, Bregman distances and monotone operators on Banach spaces in the sense of Browder.

These systems are then utilized to provide applications in the context of Picard- and Halpern-style iterations of so-called Bregman strongly nonexpansive mappings where we provide both new quantitative and qualitative results.

Lastly, we discuss the key notion of extensionality of a set-valued operator and its relation to set-theoretic maximality principles in more depth (which was already singled out – to some degree – in previous work). We thereby exhibit an issue arising with treating full extensionality in the context of these intensional approaches to set-valued operators and present useful fragments of the full extensionality statement where these issues are avoided.

Corresponding to these fragments, we discuss a range of uniform continuity statements for set-valued operators beyond the usual notion involving the Hausdorff-metric. In particular, in that context, we utilize the previous tame treatment of suprema over bounded sets to also provide the first proof-theoretic treatment of that Hausdorff-metric in the context of systems for proof mining.

The applicability of this treatment of the Hausdorff-metric is then in particular substantiated by a last case study where we provide quantitative information for a Mann-type iteration of set-valued mappings which are nonexpansive w.r.t. the Hausdorff-metric.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit Erweiterungen des Proof Mining Programms, sowohl in Bezug auf die zugrunde liegenden logischen Ansätze als auch in Bezug auf die Breite der Anwendungen auf (meist vorher unbehandelte) Bereiche der nichtlinearen Analysis und Optimierung, in beiden Fällen mit einem besonderen Fokus auf Themen welche sich auf mengenwertige Operatoren beziehen.

Dafür erweitern wir die aktuellen logischen Methoden des Proof Minings durch neue Systeme und zugehörige sogenannten logische Metatheoreme, welche diese recht involvierten Bereiche der nichtlinearen Analysis behandeln. Die meisten dieser hier entwickelten Systeme beruhen dabei in essenzieller Weise auf dem Ausnutzen von sogenannten intensionalen Methoden, das heißt der Behandlung von Mengen mit möglicherweise hoher Quantoren-Komplexität in der definierenden Matrix durch charakteristische Funktionen und Axiome welche nur die essenziellen Eigenschaften dieser Mengen beschreiben und nicht vollständig deren Elemente charakterisieren.

Die Anwendbarkeit all dieser neuen Metatheoreme wird dann durch eine Reihe von Fallstudien für die entsprechenden Bereiche begründet, welche insbesondere auch die Natürlichkeit der intensionalen Methoden als gewählten Ansatz für die entsprechenden Systeme hervorheben.

Der erste neue Bereich, welcher damit erschlossen wird, ist die Theorie der nichtlinearen Halbgruppen, induziert durch zugehörige Evolutionsgleichungen für akkretive Operatoren, in dessen Kontext wir in gewissem Sinne (neben einem initialen Vorstoß aus dem Jahr 2015) die ersten Anwendungen des Proof Minings allgemein auf die Theorie der partiellen Differentialgleichungen liefern. Konkret präsentieren wir quantitative Versionen von vier zentralen Resultaten über das asymptotische Verhalten von Lösungen solcher Gleichungen.

Der zweite neue Anwendungsbereich, welcher durch die vorliegende Thesis erschlossen wird, ist der des stetigen Dualraums eines Banachraums und der dazugehörigen Norm (welche auch durch intensionale Methoden angegangen werden). Dies beruht insbesondere auf einer beweistheoretisch-milden Behandlung von Suprema über (gewissen) beschränkten Mengen, welche auch noch später weiter angewandt wird. Jene Systeme für diese bis jetzt nicht behandelten grundlegenden Begriffe der Funktionalanalysis werden dann weiter ausgebaut um verschiedene andere Begriffe aus der konvexen Analysis wie Fréchet-Ableitungen einer konvexen Funktion, Fenchel-Konjugate, Bregman-Distanzen und monotone Operatoren auf Banachräumen im Sinne von Browder zu behandeln.

Diese Systeme werden dann eingesetzt, um Proof Mining Anwendungen im Kontext von Picard- und Halpern-artigen Iterationen von sogenannten Bregman-stark-nichtexpansiven Abbildungen abzuleiten. In diesem Zuge liefern diese Anwendungen sowohl neue quantitative als auch neue qualitative Resultate.

Zuletzt diskutieren wir in dieser Arbeit den Schlüsselbegriff der Extensionalität eines mengenwertige Operators und dessen Verhältnis zu mengentheoretischen Maximalitätsprinzipien in weiterer Tiefe (welches schon in vorherigen Arbeiten in einem gewissen Rahmen herausgestellt wurde) . Dabei stellen wir ein Problem heraus, welches mit der Behandlung der vollen Extensionalität im Kontext von diesem intensionalen Ansatz zur Behandlung von mengenwertigen Operatoren generell auftritt, und präsentieren Fragmente des Extensionalitätsprinzips welche diese Probleme vermeiden.

Korrespondierend zu diesen Fragmenten diskutieren wir neue Stetigkeitsbegriffe für mengenwertige Operatoren, welche neben dem klassichen Begriff der gleichmäßigen Stetigkeit im Sinne der Hausdorff-Metrik liegen. Insbesondere benutzen wir hier wieder den vorherigen Ansatz zur beweistheoretisch-milden Behandlung von Suprema über beschränkten Mengen, um den ersten beweistheoretischen Ansatz für die Behandlung der Hausdorff-Metrik im Kontext von Systemen des Proof Minings zu entwickeln.

Die Anwendbarkeit dieser Behandlung der Hausdorff-Metrik wird dann insbesondere durch die letzte Fallstudie herausgestellt, in welcher wir quantitative Informationen für Mann-artige Iterationen von mengenwertigen Abbildungen liefern, welche nichtexpansiv im Sinne der Hausdorff-Metrik sind.

Contents

1 Introduction

1.1 A brief history of Proof Mining

In a way, proof theory started in the school of David Hilbert in Göttingen in the 1920s. In particular, a driving force behind these early developments was what is today known as Hilbert's program, a project in early mathematical logic which aimed to show that so-called *ideal principles* [78] (which may be of non-constructive, set-theoretic or of infinitary nature) used in proofs of concrete so-called *real* statements could be (at least in principle) eliminated. In a modern view, this program is often subsumed by stating that the goal was to prove the consistency of powerful theories containing such ideal principles in certain finitistic theories.[1]

As is well-known, Gödel's second incompleteness theorem [72] already rules out the provability of the consistency of the theory in itself, let alone some finitistic fragment. While Hilbert's program in this general sense is therefore impossible, research into the relation between mathematical logic and "ordinary" core mathematics in the years since then nevertheless yielded that this reductive perspective is still largely correct. Concretely, evidence for this comes from the field of Proof Mining, in which this thesis is situated.

The origins of proof mining start with Georg Kreisel in the 1950's (see in particular the early works [125, 126] and see [107] for a detailed discussion on the influence of Kreisel's work on modern proof mining). Before Kreisel's work, the focus of most strands of research in proof theory was on establishing methods that allowed for a relativized version of Hilbert's program to be carried out where one aims at providing

[1] If real statements are to be understood as universal statements, then establishing the consistency of a suitable formal theory in some suitable finitistic fragment would actually suffice to establish conservativity for real statements as formal consistency is provably equivalent to a reflection principle for universal sentences (see [195]).

relative consistency proofs between different theories, i.e. at reducing the consistency of one theory T_1 to another theory T_2, where it may be considered less philosophically problematic to accept the latter as being consistent than the former.

In many cases such relative consistency proofs proceed by employing a general proof-theoretic device called a proof interpretation which transforms sentences of T_1 into sentences of T_2 and is sound w.r.t. the provability structure, i.e. if a sentence is provable in T_1, its translation is provable in T_2. In most situations, these transformations preserve the falsum of a theory so that if $T_1 \vdash \bot$, then the soundness of the interpretation yields $T_2 \vdash \bot$. In other words, if T_1 is inconsistent, so is T_2.

Kreisel realized that these subtly constructed methods perform a much deeper task than just preserving the provability of falsum, often eliminating existential quantifiers in a formula for concrete terms whose complexities relate to the principle used in the proof, and called for a "shift of emphasis": instead of applying the interpretations to sentences of which anyhow no proof is expected to exist (i.e. $\bot$), one should apply them to concrete mathematical statements with actual proofs with the projected gain being that it is to be expected that such a treatment would reveal further information on the statement thus proven, along the lines of the now famous leitmotif of Kreisel [129]:

> [...] formulate what more we know about a formally derived theorem F than if we merely know that F is true.

Or, in other words: taking into account the principles used in the proof, can we infer further information on the statement thus proven?

Before we delve into what methods may be or are employed in such an endeavor, we shortly discuss some of the various kinds of "further information" that can be considered in relation to the quantifier complexity of a statement. E.g. for statements of the form $\forall x F_{qf}(x)$ where F_{qf} is quantifier-free (which we previously called a real statement), no additional information can sensibly be given: the statement $F_{qf}(x)$ is just true for all x.

Moving to the case of one existential quantifier $\exists x F_{qf}(x)$, multiple kinds of further information immediately come to mind: one could hope for actual witnesses t, i.e. $F_{qf}(t)$, a list of potential witnesses $t_1, \ldots, t_k$ (kin to the conclusion of Herbrand's theorem), i.e. $F_{qf}(t_1) \vee \cdots \vee F_{qf}(t_k)$, or bounds t on actual witnesses, i.e. $\exists x \leqslant t\ F_{qf}(x)$, among some others.

Further moving up in the quantifier ranks, for statements of the form $\forall x \exists y F_{qf}(x, y)$ one could hope again for precise witnesses, i.e. functions f with $\forall x F_{qf}(x, fx)$, or for bounding functions f, i.e. $\forall x \exists y \leqslant fx F_{qf}(x, y)$, among others.

The last example that we want to consider is the case of a statement of the form $\forall x \exists y \forall z F_{qf}(x, y, z)$. The immediate kind of further information that could be considered is a witness/bounding function on y in terms of x. The immediate issue that arises here is that of the complexity of such a function. While in the previous case of $\forall x \exists y F_{qf}(x, y)$, already the truth of the statement, in the natural numbers say, guarantees the existence of a computable function providing a witness for y in terms of x just by unbounded search[2] through all y, it can be rather immediately seen that there are instances of $F_{qf}(x, y, z)$ where even a function f bounding witnesses for y in terms of x, i.e. $\forall x \exists y \leqslant fx \forall z F_{qf}(x, y, z)$, can never be computable.[3]

However, if one is interested in computable information, one can in this case move to the Herbrand normal form of the statement (where we for simplicity assume that we can quantify also over functions operating on the ground variables, i.e. g is a function variable):

$$\forall x \forall g \exists y F_{qf}(x, y, gy).$$

Then, another type of further information would be a function $\Phi(x, g)$ that witnesses $\exists y$ in terms of x and g, which we can recognize as being equivalent to finding a solution to the so-called *no-counterexample interpretation* of $\forall x \exists y \forall z F_{qf}(x, y, z)$ as formulated by Kreisel in [125, 126]. In the case where $\forall x \exists y \forall z F_{qf}(x, y, z)$ represents a convergence statement, upper bounds on such witnesses (for a slightly modified statement) are now commonly called rates of metastability or quasi-rates for the convergence, where the former name was coined by Terence Tao (see [203, 204]) who rediscovered this corresponding (non-effectively equivalent) reformulation of a convergence statement in the

[2]While this is a computable solution, there is of course no complexity information available for this function. In that way, it can still be considered unsatisfactory in regard to the previously discussed leitmotif as an analysis of an actual proof of the statement $\forall x \exists y F_{qf}(x, y)$ might provide a much more tailored witness function whose complexity will be in proportion to the complexity of the principles used in the proof.

[3]Consider e.g. $F_{qf}(x, y, z) = T(x, x, y) \vee \neg T(x, x, z)$ where $T(a, b, c)$ is the so-called Kleene T-predicate expressing that the Turing machine with code a run on b halts with runtime code c. Thus F_{qf} expresses that the Truing machine with code x run on x either halts with runtime code y or does not halt with runtime code z. The statement $\forall x \exists y \forall z F_{qf}(x, y, z) \equiv \forall x(\exists y T(x, x, y) \vee \forall z \neg T(x, x, z))$, expressing that every Turing machine with code x either does or does not halt on input x, is true just by classical logic but any computable function f with $\forall x \exists y \leqslant fx \forall z F_{qf}(x, y, z)$ would allow one to decide the special halting problem which is of course not possible.

course of his interest in “finitary analysis”.

In any way, the program thus created to apply proof-theoretic methods to concrete mathematical theorems with the aim of extracting new results was dubbed “unwinding of proofs”. In the era of unwinding of proofs, the main emphasis was placed on methods from structural proof theory like epsilon-substitution, cut-elimination and the extraction and analysis of Herbrand-terms from Herbrand’s theorem (see the discussions in e.g. [60, 125, 126, 137, 138, 156]).

Besides a few highlights during this era, in particular with the notable work of Luckhardt [137] on effective bounds for Roth’s theorem on exceptionally good rational approximations of algebraic-irrational numbers, the era of unwinding of proofs was rather sparsely populated regarding applications until it was essentially revived by the work of Ulrich Kohlenbach starting with his doctoral thesis [86]. At that time, the unwinding program saw a shift of focus both in methods and in areas of applications and was soon to be “rebranded” under the name of “proof mining” (at the suggestion of Dana Scott) which has since then been steadily expanded with new applications.

Concretely, Gödel’s functional interpretation [74] (also called Dialectica interpretation after the journal it was published in) became the methodological focus based on its well-behavedness w.r.t. to negative translations to treat classical proofs (by interpreting Markov’s principle) and because of its compositionality when treating the modus ponens (compared in particular to the no-counterexample interpretation). With the use of the Dialectica interpretation, one in particular conveniently also moves to systems of arithmetic and analysis in all finite types instead of first- or second-order systems which also brings with it the benefit that they allow one to avoid some coding issues that exist in low-type systems. Other methods that are used are similarly proof interpretations in the sense discussed above and in particular include negative translations, to deal with classical logic via a reduction to intuitionistic logic as mentioned before, and Kreisel’s *modified realizability* [127, 128] for semi-intuitionistic proofs. However, the main tool employed in modern proof mining arises as an ingenious combination of the Dialectica interpretation and Howard’s notion of majorizability [79] as introduced in Kohlenbach’s work beginning in [86, 87] and later given the name of *monotone functional interpretation* (see also in particular [91]). This interpretation only asks for the construction of computable majorants for the Dialectica interpretation and allows for potentially noncomputable witnesses.[4] The immediate benefits are that the plethora of statements with computable interpretations increases dramatically, with e.g. statements like *weak Kőnig's lemma* WKL having a (trivial) computable monotone functional interpretation (as every – potentially uncomputable – path of a binary tree, coded as a 0-1-sequence, can be majorized by the constant-1 function). The immediate drawback is of course that via this interpretation, no precise witnesses but only computable majorants are constructed for the interpretation of the respective theorem (which in the case of the natural numbers amounts to upper bounds).

However, there are areas of mathematics where this drawback is often superficial and where thus this combination of the Dialectica interpretation and majorizability proves to be particularly rewarding. One area where this in particular is the case is analysis where many statements are naturally monotone and so a bound is as good as a witness. Further, analysis also seems to be particularly rewarding as many modern proofs seem to be mainly of a geometrical nature which seem to avoid the Gödelian phenomena that in principle could arise in these circumstances. Even further, in this context, the availability of WKL is very convenient for formalizing a range of common compactness arguments, being equivalent to many such central results from functional analysis as known from the reverse mathematics program. So it is not surprising that the main focus of proof mining since the 1990s has been centered on and around analysis. In particular, early examples of applications in this modern age of proof mining include the applications presented by Ulrich Kohlenbach (also together with Paulo Oliva) on best approximation theory [89, 90, 117] (note also the very recent work on best approximations [193]).

Nevertheless, in the first period of proof mining in it its modern form, the systems employed in the pursuit of applications essentially centered around systems of arithmetic in all finite types (which will be more precisely defined later on) like WE-PA^ω + WKL + QF-AC for extracting primitive recursive majorants (in the sense of Gödel [74] and Hilbert [78]) or WE-PA^ω + DC + QF-AC (adding the strong principle of dependent choice DC) in which case one can only guarantee the extractability of majorants which are bar-recursive in the sense of the seminal work of Spector [198].

[4] This monotone functional interpretation and the use of majorizability has subsequently lead to other interpretations where this boundedness character is further infused into the interpretation, most notably the bounded functional interpretation by Ferreira and Oliva [65] which has subsequently been further developed in various ways (see e.g. [59, 62, 63]) and recently has found use in logical aspects of the proof mining program [64].

In that way, applications were naturally restricted to the context of Polish metric spaces which are representable in Baire space, i.e. $\mathbb{N}^{\mathbb{N}}$ equipped with the metric d defined by

$$d(f,g) = \begin{cases} 2^{-\min n[f(n)\neq g(n)]} & \text{if } \exists n \in \mathbb{N}(f(n) \neq g(n)), \\ 0 & \text{otherwise,} \end{cases}$$

for $f, g \in \mathbb{N}^{\mathbb{N}}$.

This restriction was lifted in the second main paradigm shift in the modern age of proof mining, starting with Ulrich Kohlenbach's seminal work [95]. Namely, in the years leading up to this work, the crucial observation for some notable applications was made that bounds and proofs which were obtained from arguments in some representable class of spaces also naturally hold in larger classes of spaces where the separability assumption is dropped and, moreover, the bounds were also very uniform in the parameters of these spaces, only depending on some simple upper bounds on metric distances despite the absence of any (relative) compactness assumptions. The question of whether this phenomenon was purely coincidental or was an instance of a deeper logical reason immediately arose.

The techniques that lead to a logical explanation of this phenomenon at the same time broadened the proof mining framework in many crucial ways that are today characteristic for its success. Concretely, these applications and their uniformities can be logically explained by systems in all finite types which additionally include new abstract base types by which one gains the ability to talk abstractly about certain classes of spaces which do not have to be separable. After extending the notion of majorizability to such classes of spaces (which we for simplicity assume to only contain metric spaces) by majorizing objects of type X by a natural number n bounding $d(a,x)$ for some reference point a, the application of the respective extension of the monotone functional interpretation extracts computable majorants in this extended sense from corresponding proofs which therefore are uniform exactly in the way described before, depending e.g. on elements from X only via upper bounds on metric distances. The macros obtained by an application of functional interpretation (together with a negative translation) combined with majorization are commonly dubbed "general logical metatheorems (on bound extraction)" or "bound extraction theorems" and these metatheorems thus guarantee the existence of uniform and computable majorants for statements provable in the associated theories which in particular may still use full classical logic and a wide range of other "non-constructive" principles. Further, besides merely guaranteeing the existence of such additional information, the metatheorems allow for an a priori estimation of their complexity (which can be as elementary as polynomials) and they provide an algorithmic approach towards actually extracting the quantitative information. In particular, these metatheorems also further elucidate the extent of the phenomenon of so-called proof-theoretic tameness of modern (nonlinear) analysis as already shortly discussed before, i.e. the empirical fact that most proofs in e.g. analysis, although in principle being subject to well-known Gödelian phenomena, nevertheless "seem to be tame in the sense of allowing for the extraction of bounds of rather low complexity" [103] (see also [139, 140] for further discussions of these types of phenomenas and their implications for logic and mathematics).

The range of classes of spaces and of objects on them that can be treated with this approach is rather broad: In general, if we are given a class of spaces and objects with corresponding defining axioms, potentially using constants from an extended language, such that all of them have a monotone functional interpretation in this extended sense (which e.g. trivially holds if the axioms are universal) and if all corresponding additional constants can be majorized, then the methodology immediately applies in that context as well and allows one to also derive bound extraction theorems for such spaces and objects. Examples of metatheorems derived in this spirit may be found in [71, 76, 95, 115, 132, 133, 165, 192], as well as [96], for metatheorems obtained via (modifications of) Gödel's Dialectica interpretation, and in [63, 64] for metatheorems obtained via the related bounded functional interpretation [65]. The spaces treated so far in particular include general metric and normed spaces, so-called W-hyperbolic spaces, CAT(0)-spaces, uniformly convex as well as uniformly smooth Banach spaces and Hilbert spaces, among many others.

Beyond this small introduction, and the further formal details that will be discussed throughout this thesis, we in general refer to the monograph [96] where the whole development of proof mining up to 2008 is detailed comprehensively. Further discussions on early developments can be found in the survey [116] and more recent progress, with a focus on nonlinear analysis and optimization, is surveyed in [100, 102].

1.2 The contents of this thesis

This thesis is now concerned with extending the underlying logical approach as well as the breadth of applications of proof mining to various (mostly previously untreated) areas of nonlinear analysis and optimization, with a particular focus being placed on topics which involve set-valued operators. Such set-valued operators are one of the main objects of concern of many of the recent proof mining applications like in [101] in the context of Bauschke's solution [6] to the zero displacement conjecture, in [108] for abstract Cauchy problems, in [120] for iteration schemes using set-valued operators or in particular like in the case of the proximal point algorithm (see [145, 183]) and its adaptations and extensions as treated in [55, 56, 104, 105, 106, 112, 134, 161].

For this, we always first extend the current logical methodology of proof mining by new systems and corresponding metatheorems that cover these more involved areas of nonlinear analysis. Most of the methods developed in the course of this crucially rely on the use of intensional methods, treating sets with potentially high quantifier complexity in the defining sentences via characteristic functions and axioms that describe only their properties and do not characterize their elements completely. The applicability of all of these metatheorems is then substantiated by a range of case studies for the respective areas which in particular also highlight the naturalness of the intensional methods in the design of the systems.

Concretely, in Chapter 2, we first sketch the definition of one of the main modern systems employed in proof mining in general and in this thesis in particular which provides a treatment of normed linear spaces using abstract types as discussed before. Further, we in that context in particular give the essential logical preliminaries to systems for arithmetic in all finite types together with a primer on representations of real numbers in these systems. Besides that, we sketch the main result for this system treating abstract normed spaces, the general logical metatheorem, which underlies essentially all of the logical contributions made in this thesis.

Then, in Chapter 3, we provide a recap of the main results from the author's Master Thesis [168] and the resulting logical contributions published in [165] on set-valued accretive and monotone operators in Banach and Hilbert spaces and their treatment via intensional methods that (together with some new material) form the "spiritual" base for many parts of the thesis.[5] In particular, we provide a more detailed sketch of the proof of the corresponding logical metatheorems of the resulting systems, augmenting the rather brief discussion from Chapter 2.

In Chapter 4, we extend the logical systems presented in Chapter 3 so that they become applicable to nonlinear semigroups induced by corresponding evolutions equations for accretive operators. These logical results are contained in the pre-print [163].

In Chapters 5, 6 and 7, we provide four applications of these metatheorems to a range of results on the asymptotic behavior of these semigroups. These applications are contained in the pre-print [163] as well as the articles [162, 167], respectively, with the joint work with Pedro Pinto [162] worded by myself.

In Chapter 8, we change the setting and provide proof mining metatheorems via intensional methods for the continuous dual of a Banach space as well as various notions from convex analysis like the Fréchet derivative of a convex function, Fenchel conjugates and Bregman distances. This in particular relies on a proof-theoretically tame treatment of suprema over (certain) bounded sets which is also exploited later on.

In Chapter 9, we give applications of the preceding metatheorems to Picard- and Halpern-style iterations of Bregman strongly nonexpansive mappings where we in particular provide both new quantitative and qualitative results. This in particular also yields the corresponding results for proximal point type variants of these methods for monotone operators over Banach spaces in the sense of Browder. Parts of this chapter (concerning Theorems 9.3.14 and 9.4.1 as well as Proposition 9.4.6 and Lemma 9.5.7) utilized sketches communicated to me by Ulrich Kohlenbach.

In Chapter 10, we transfer the results from Chapter 3 to the setting of monotone operators over Banach spaces in the sense of Browder, as mentioned before, and their relativized resolvents in the sense of Eckstein as well as Bauschke, Borwein and Combettes. This in particular also provides a firm logical basis for the applications to the proximal point type methods for these operators studied in the previous Chapter 9.

In Chapter 11, we initially exhibit an issue arising with treating full extensional-

[5] As such, the presentation and formulation of Chapter 3 is largely taken from the works [165, 168].

ity in the context of these intensional approaches to set-valued operators and, in that vein, present useful fragments of the full extensionality statement where these issues are avoided. Further, we extend the logical considerations from Chapter 3 on the extensionality principle for set-valued operators and its relation to the set-theoretic maximality principles of such operators by showing that this characteristic equivalence also extends to these fragments, pointing to a rather robust phenomenon. Further, we study the continuity principles associated with these fragments of extensionality and show how they can be introduced in the logical systems from the preceding chapters. In the course of this, we also employ the tame treatment of suprema over bounded sets developed in Chapter 8 to provide a logical treatment of the Hausdorff-metric.

In Chapter 12, we provide an application of proof mining to the Mann-iteration of set-valued mappings which are nonexpansive w.r.t. the Hausdorff-metric, illustrating the applicability of the previous logical considerations.

2 Preliminaries

In the following, we will now sketch the definitions of the main systems employed in this thesis and the statement of the corresponding general logical metatheorem which underlies essentially all of the logical contributions made in this thesis. In this presentation, we mostly follow the notation and presentation given in [96].

Before that, we however fix some general notation: We write $\mathbb{N}$ for the set of all natural numbers *including* 0 and write $\mathbb{N}^*$ for the set of natural numbers *excluding* 0. Further, we write $\mathbb{R}$ for the set of real numbers and sometimes we write $\mathbb{R}_{>0}$ for the interval $(0, \infty)$.

We define $\dot{-}$ on $\mathbb{N}$ by $n \dot{-} m = \max\{0, n - m\}$. Further, we use the interval notation

$$[r; s] := [r, s] \cap \mathbb{N}.$$

Lastly, in a metric space (X, d) and given $r > 0$ and $x \in X$, we write $\overline{B}_r(x)$ for the closed ball of radius r around on x and $B_r(x)$ for the open ball, respectively.

2.1 Finite type arithmetic

We begin with the basic systems for arithmetic in all finite types WE-HA$^\omega$ and WE-PA$^\omega$: Over the collection of all so-called finite types T defined by[1]

$$0 \in T,\ \rho, \tau \in T \to \tau(\rho) \in T,$$

we consider a many-sorted language containing variables and quantifiers for every type $\tau \in T$ as well as some suitable functionally complete set of propositional connectives, which we for simplicity assume to be $\wedge, \vee$ and $\to$. The language of WE-HA$^\omega$ now

[1]Following [96], we denote the function type of two types ρ, τ by $\tau(\rho)$, representing the type of all functions mapping objects of type ρ to objects of type τ. Other common notations of this type in the literature include e.g. $\rho \to \tau$.

additionally contains the constants 0 for zero (of type 0) and S for successor (of type 0(0)) as well as constants $\Pi_{\rho,\tau}$ (of type $\rho(\tau)(\rho)$) and $\Sigma_{\delta,\rho,\tau}$ (of type $\tau\delta(\rho\delta)(\tau\rho\delta)$) for the combinators of Schönfinkel [188] (which were later used extensively by Curry and Howard, see [80] for the latter) and lastly the constants $\underline{R}_{\underline{\rho}} = (R_1)_{\underline{\rho}}, \dots, (R_k)_{\underline{\rho}}$ for simultaneous primitive recursion in the sense of Hilbert [78] and Gödel [74] where R_i has the type

$$\rho_i(\rho_k 0 \underline{\rho}^t) \dots (\rho_1 0 \underline{\rho}^t)\underline{\rho}^t 0$$

for $\underline{\rho} = (\rho_1) \dots (\rho_k)$ and where we write $\underline{\rho}^t = (\rho_k) \dots (\rho_1)$. In the above and also in the following, we mostly use the conventions for saving parentheses in types used in [96]. Further, we stratify the types in T by their degree $\deg(\tau)$, defined recursively via

$$\deg(0) := 0, \ \deg(\tau(\rho)) := \max\{\deg(\tau), \deg(\rho) + 1\},$$

and we denote pure types by natural numbers via

$$0(n) := n + 1.$$

The only way to form new terms is by application: if t is a term of type $\tau(\rho)$ and s is a term of type ρ, then $t(s)$ is a term of type τ. The only primitive predicate in the language is $=_0$ for equality at type 0 and equality at higher types is introduced as an abbreviation by recursion on the type via

$$t =_{\tau(\rho)} s := \forall x^\rho \, (tx =_\tau sx) \, .$$

The theory WE-HA$^\omega$ now arises by extending intuitionistic logic, formulated for the many-sorted language (see e.g. [96, 205]), by the usual equality axioms for $=_0$, the usual axioms for the successor constant S, the axioms specifying the combinators and recursors (see [96] for details on all of this) and the induction axiom

$$F(0) \wedge \forall x^0 \, (F(x) \rightarrow F(Sx)) \rightarrow \forall x^0 F(x) \tag{IA}$$

where $F(x^0)$ is any formula from the language. The last thing added to WE-HA$^\omega$ is the quantifier-free extensionality rule of Spector [198]

$$\frac{F_0 \rightarrow s =_\rho t}{F_0 \rightarrow r[s/x^\rho] =_\tau r[t/x^\rho]} \tag{QF-ER}$$

where F_0 is a quantifier-free *formula* and t, s are terms of type ρ and r is a term of type τ. Note that using this rule, we can actually derive the seemingly stronger

$$\frac{\exists y^\sigma F_0(y) \rightarrow s =_\xi t}{\exists y^\sigma F_0(y) \rightarrow r[s/x^\xi] =_\tau r[t/x^\xi]} \tag{Σ_1-ER}$$

with F_0, s, t and ξ, τ as before and σ an additional finite type but where we assume that y is not free in r, s, t. To see this, note that

$$\exists y^\sigma F_0(y) \to s =_\xi t \equiv \forall y^\sigma (F_0(y) \to s =_\xi t)$$

and the latter implies $F_0(y) \to s =_\xi t$. Now, using QF-ER applied to this (where it is important that F_0 in the formulation may have free variables), we get $F_0(y) \to r[s/x^\xi] =_\tau r[t/x^\xi]$ and universal generalization yields

$$\forall y^\sigma (F_0(y) \to r[s/x^\xi] =_\tau r[t/x^\xi]) \equiv \exists y^\sigma F_0(y) \to r[s/x^\xi] =_\tau r[t/x^\xi],$$

which is as required.

We denote the system which is obtained by instead adding the full axioms of extensionality

$$\forall z^{\tau(\rho)}, x^\rho, y^\rho \, (x =_\rho y \to zx =_\tau zy) \tag{$E_{\rho,\tau}$}$$

for all types ρ, τ by E-HA$^\omega$.

The classical systems WE-PA$^\omega$ and E-PA$^\omega$ are now just defined as WE-HA$^\omega$ or E-HA$^\omega$, respectively, augmented with the law of excluded middle $F \vee \neg F$ for all formulas F in the language.

Note also that through the combinators Π and Σ, the theory WE-HA$^\omega$ has λ-abstraction (see e.g. [96]): for every term t of type τ, one can construct a term $\lambda x^\rho.t$ of type $\tau(\rho)$ such that $\text{free}(\lambda x^\rho.t) = \text{free}(t) \backslash \{x\}$ and

$$\text{WE-HA}^\omega \vdash (\lambda x^\rho.t)(s) =_\tau t[s/x]$$

for any term s of type ρ where we write $t[s/x]$ for the term arising from t by simultaneously substituting s for all occurrences of x.

The main finite type system considered here, denoted by $\mathcal{A}^\omega$, now arises from WE-PA$^\omega$ by, for one, adding the quantifier-free axiom of choice

$$\forall \underline{x} \exists \underline{y} F_0(\underline{x}, \underline{y}) \to \exists \underline{Y} \forall \underline{x} F_0(\underline{x}, \underline{Y\,x}) \tag{QF-AC}$$

where F_0 is quantifier-free but the types of the variable tuples $\underline{x}$, $\underline{y}$ are arbitrary and where we use the notation $\underline{Y\,x}$ to abbreviate $Y_1\underline{x}, \dots, Y_k\underline{x}$ if $\underline{Y} = Y_1, \dots, Y_k$, and, for another, adding the schema of dependent choice $\text{DC} = \{\text{DC}^{\underline{\rho}} \mid \underline{\rho} \subseteq T\}$ with

$$\forall x^0, \underline{y}^{\underline{\rho}} \exists \underline{z}^{\underline{\rho}} F(x, \underline{y}, \underline{z}) \to \exists \underline{f}^{\underline{\rho}(0)} \forall x^0 F(x, \underline{f}(x), \underline{f}(S(x))) \tag{DC$^{\underline{\rho}}$}$$

where $\underline{f}^{\underline{\rho}(0)}$ stands for $f_1^{\rho_1(0)}, \dots, f_k^{\rho_k(0)}$ and F may now be arbitrary.

In the language of WE-PA$^\omega$ and its extensions, we will later rely on a chosen representation of the real numbers and in that context, we follow the definitions and conventions given in [96]. The following paragraphs only discuss the details which are crucial for the proofs carried out later.

As usual, rational numbers are represented using pairs of natural numbers and for that it will be convenient to fix a paring function j where we follow the choice made in [95]:

$$j(n^0, m^0) := \begin{cases} \min u \leqslant_0 (n+m)^2 + 3n + m[2u =_0 (n+m)^2 + 3n + m] & \text{if existent,} \\ 0^0 & \text{otherwise.} \end{cases}$$

The arithmetical operations $+_\mathbb{Q}, \cdot_\mathbb{Q}, (\cdot)^{-1}_\mathbb{Q}$ can then be introduced through primitive recursive terms operating on such codes and the relations $=_\mathbb{Q}$, $<_\mathbb{Q}$ are quantifier-free definable.

The chosen representation of real numbers now relies on fast converging Cauchy sequences of rational numbers (i.e. reals are coded as objects of type 1) with a fixed Cauchy modulus 2^{-n} (see [96] for details) and we consider $\mathbb{N}$ and $\mathbb{Q}$ as being embedded in that representation via the constant sequences. Similarly as to $\mathbb{Q}$, the usual arithmetical operations like $+_\mathbb{R}, \cdot_\mathbb{R}, |\cdot|_\mathbb{R}$ are definable using closed terms and the relations $=_\mathbb{R}/<_\mathbb{R}$ on type 1 objects are represented by formulas in the underlying language. Naturally, these relations are not decidable anymore but are given by Π^0_1/Σ^0_1-formulas, respectively. An arithmetical operation where some care is needed in the context of this formal treatment of real numbers is the reciprocal $(\cdot)^{-1}$: In fact, there is no closed term of type $1(1)$ in WE-PA$^\omega$ which represents γ^{-1} correctly for all $\gamma \neq 0$. We deal with this as in [93] by using a binary term $(\cdot)^{-1}_\cdot$ of type $1(1)(0)$ such that $(\gamma)^{-1}_l$ correctly represents γ^{-1} for all $|\gamma| > 2^{-l}$. An expression like γ^{-1} is then dealt with by working with an additional parameter l of type 0 and using $(\gamma)^{-1}_l$ together with the additional implicative assumption $|\gamma|_\mathbb{R} >_\mathbb{R} 2^{-l}$. In practice, this can be mostly ignored and we thus mainly use γ^{-1} freely without highlighting the additional parameter.

In this thesis, we in general omit the index of $\mathbb{R}$ for arithmetical operations to make everything more readable. In proofs, we will almost always omit all types as to not distract from the general ideas and patterns.

In the context of representing reals, we will later rely on an operator $\widehat{\cdot}$ which allows for an implicit quantification over all fast-converging Cauchy sequences of rationals.

Following [96], we define this operator via

$$\widehat{x}n := \begin{cases} xn & \text{if } \forall k <_0 n \left(|xk -_{\mathbb{Q}} x(k+1)|_{\mathbb{Q}} <_{\mathbb{Q}} 2^{-k-1} \right), \\ xk & \text{for } k <_0 n \text{ least with } |xk -_{\mathbb{Q}} x(k+1)|_{\mathbb{Q}} \geqslant_{\mathbb{Q}} 2^{-k-1} \text{ otherwise,} \end{cases}$$

for x of type 1 and we refer to [96] for any further discussions of its properties.

For establishing the metatheorems, we will need to canonically select a Cauchy sequence representation for a given real number. For non-negative real numbers, following [95], this can be formally achieved by a function $(\cdot)_\circ$ which selects a representative $(r)_\circ \in \mathbb{N}^{\mathbb{N}}$ via

$$(r)_\circ(n) := j(2k_0, 2^{n+1} - 1),$$

where

$$k_0 := \max k \left[\frac{k}{2^{n+1}} \leqslant r \right].$$

Naturally, such an association will be non-effective. However, it will suffice that the operation behaves well-enough w.r.t. the notion of majorization. For this, we will in particular rely on the following properties of $(\cdot)_\circ$:[2]

Lemma 2.1.1 ([95])**.** *Let $r \in [0, \infty)$. Then:*

1. *$(r)_\circ$ is a representation of r in the sense of the above (see also [96]).*
2. *For $s \in [0, \infty)$, if $r \leqslant s$, then $(r)_\circ \leqslant_{\mathbb{R}} (s)_\circ$ and also $(r)_\circ \leqslant_1 (s)_\circ$.*
3. *$(r)_\circ$ is nondecreasing (as a type 1 function).*

However, later we will need an extension of this function $(\cdot)_\circ$ to all real numbers such that we retain the nice properties mentioned above regarding majorizability. For this, if $r < 0$, we define

$$(r)_\circ(n) = j(2\bar{k}_0 \dot{-} 1, 2^{n+1} - 1)$$

where

$$\bar{k}_0 := \max k \left[\frac{k}{2^{n+1}} \leqslant |r| \right].$$

Then $(r)_\circ(n) = -_{\mathbb{Q}}(|r|)_\circ(n)$ and we get the following lemma containing exactly the properties that we later need for this notion to be useful in the context of majorizability.

Lemma 2.1.2. *Let $r \in \mathbb{R}$. Then:*

[2] Here, we write $f \leqslant_1 g$ for two objects f^1, g^1 if $fn \leqslant_0 gn$ for all n^0.

1. *$(r)_\circ$ is a representation of r in the sense of the above (see also [96]).*

2. *For $s \in [0,\infty)$, if $|r| \leqslant s$, then $(r)_\circ \leqslant_1 (s)_\circ$.*

3. *$(r)_\circ$ is nondecreasing (as a type 1 function).*

Proof. That $(r)_\circ$ is a representation of r is immediate and clearly $(r)_\circ$ is nondecreasing as j is monotone. For item (2), let $|r| \leqslant s$. If $r \geqslant 0$, the result is contained in the above Lemma 2.1.1. If $r < 0$, write $\bar{k}_0$ for the value corresponding to $|r|$ and k_0 for the value corresponding to s. Then we have

$$\bar{k}_0 = \max k \left[\frac{k}{2^{n+1}} \leqslant |r|\right] \leqslant \max k \left[\frac{k}{2^{n+1}} \leqslant s\right] = k_0$$

so that

$$(r)_\circ(n) = j(2\bar{k}_0 \dot{-} 1, 2^{n+1} - 1) \leqslant j(2k_0, 2^{n+1} - 1) = (s)_\circ(n)$$

using the monotonicity of j. $\square$

Lastly, given a sequence $\alpha \in \mathbb{N}^\mathbb{N}$, we write r_α for the unique real represented by $\widehat{\alpha}$ and we sometimes write $[\alpha](n)$ for the n-th element of that sequence for better readability.

2.2 Abstract types and systems for normed spaces

As motivated in Chapter 1, the main extension of $\mathcal{A}^\omega$ is obtained by adding a new abstract base type X as originally considered by Kohlenbach in [95]. The type allows us to deal with abstract spaces that cannot necessarily be represented in $\mathcal{A}^\omega$/WE-PA$^\omega$. Define the extended set of types T^X as follows:

$$0, X \in T^X, \quad \xi, \tau \in T^X \Rightarrow \tau(\xi) \in T^X.$$

The theory $\mathcal{A}^\omega$ can then be formulated over the resulting extended language by extending the constants (if appropriate) to take arguments and produce values in those new types and by trivially extending the axiom schemes and rules to allow formulas from the new language (see [71, 95, 96] for details on all of this).

The main extension used here will be the theory $\mathcal{A}^\omega[X, \|\cdot\|]$ for real normed vector spaces, obtained by first extending the language of $\mathcal{A}^\omega$ (formulated over T^X) by new constants $0_X, 1_X$ of type X, $+_X$ of type $X(X)(X)$, $-_X$ of type $X(X)$, $\cdot_X$ of type

$X(X)(1)$ and $\|\cdot\|_X$ of type $1(X)$. It should be noted that $=_0$ is still the only primitive relation and in particular, identity on X is treated as a defined predicate via[3]

$$x^X =_X y^X := \|x -_X y\|_X =_{\mathbb{R}} 0$$

which is, by the previous discussion on the representation of the reals, a Π^0_1-formula and not decidable. To form $\mathcal{A}^\omega[X, \|\cdot\|]$, we then add the relevant defining axioms stating that X with these operations is a real normed vector space with 1_X such that $\|1_X\|_X =_{\mathbb{R}} 1$ and $-_X x$ being the additive inverse of x (see [95]):[4]

1. The usual vector space axioms formulated for $+_X$, $-_X$, $\cdot_X$, 0_X and $=_X$,
2. $\forall x^X \, (\|x -_X x\|_X =_{\mathbb{R}} 0)$,
3. $\forall x^X, y^X \, (\|x -_X y\|_X =_{\mathbb{R}} \|y -_X x\|_X)$,
4. $\forall x^X, y^X, z^X \, (\|x -_X z\|_X \leqslant_{\mathbb{R}} \|x -_X y\|_X + \|y -_X z\|_X)$,
5. $\forall \alpha^1, x^X, y^X \, (\|\alpha x -_X \alpha y\|_X =_{\mathbb{R}} |\alpha| \cdot \|x -_X y\|_X)$,
6. $\forall \alpha^1, \beta^1, x^X \, (\|\alpha x -_X \beta x\|_X =_{\mathbb{R}} |\alpha - \beta| \cdot \|x\|_X)$,
7. $\forall x^X, y^X, u^X, v^X \, (\|(x +_X y) -_X (u +_X v)\|_X \leqslant_{\mathbb{R}} \|x -_X u\|_X + \|y -_X v\|_X)$,
8. $\forall x^X, y^X \, (\|(-_X x) -_X (-_X y)\|_X =_{\mathbb{R}} \|x -_X y\|_X)$,
9. $\forall x^X, y^X \, (| \, \|x\|_X - \|y\|_X \, | \leqslant_{\mathbb{R}} \|x -_X y\|_X)$,
10. $\|1_X\|_X =_{\mathbb{R}} 1$.

Further, extensionality of all those operations is provable in $\mathcal{A}^\omega[X, \|\cdot\|]$.[5]

Derived from $\mathcal{A}^\omega[X, \|\cdot\|]$ is the theory $\mathcal{A}^\omega[X, \langle\cdot,\cdot\rangle]$ for real inner product spaces, extending the former by the parallelogram law

$$\forall x^X, y^X \left(\|x +_X y\|_X^2 + \|x -_X y\|_X^2 =_{\mathbb{R}} 2 \left(\|x\|_X^2 + \|y\|_X^2 \right) \right).$$

[3]Here, and in the following, we write $x -_X y$ as an abbreviation for $x +_X (-_X x)$.

[4]Here, and in the following, we will omit the types $X, \mathbb{R}$ from the operations $\cdot_X, \cdot_{\mathbb{R}}$ or omit $\cdot_X, \cdot_{\mathbb{R}}$ altogether to improve the readability of the formulas.

[5]The easy provability of extensionality for the new constants stands behind the choice of the above norm axioms, see Chapter 8 for a further discussion.

As is well-known, any inner product space satisfies this law and conversely, any normed space satisfying it actually admits an inner product which can then be defined via the norm with

$$\langle x^X, y^X \rangle_X :=_1 \frac{1}{4}\left(\|x +_X y\|_X^2 - \|x -_X y\|_X^2\right).$$

Also here, extensionality of the defined operation is provable in the system.

2.3 A general logical metatheorem

We now state the main result for the system $\mathcal{A}^\omega[X, \|\cdot\|]$, the general logical metatheorem on the extraction of computable bounds from proofs established in [71, 95]. We do not dive into the precise details of the proof just yet and rather postpone these to Chapter 3 (and, beyond this discussion in Chapter 3 and the rest of this thesis, we refer to [71, 95] as well as [96]). Also, we postpone any precise discussions of the involved notions to Chapter 3 where we will be concerned with the main extensions of $\mathcal{A}^\omega[X, \|\cdot\|]$ that underly many of the central parts of this thesis. In that vein, we now only state the theorem for $\mathcal{A}^\omega[X, \|\cdot\|]$ so that the central aspects of the enterprise of the metatheorems (which are one of the main concerns of this thesis) can be appreciated. Namely, the metatheorem guarantees, as discussed in Chapter 1, the existence of further information in the sense of the previous chapter on provable sentences in the theory $\mathcal{A}^\omega[X, \|\cdot\|]$ which are essentially of a $\forall\exists$-form. The main features of the bounds guaranteed by the metatheorem are that they are computable as well as very uniform, depending not on the space or any concrete objects but only on majorants thereof. As discussed before, this notion of majorizability goes back to Howard [79] but in the form used here is an extension developed in [71, 95] of the strong majorizability notion of Bezem [16]. This notion of strongly majorizable functionals over an abstract normed space is defined in tandem with an associated model $\mathcal{M}^{\omega,X}$ for the language of $\mathcal{A}^\omega[X, \|\cdot\|]$. On a high level, the structure of the proof of the metatheorem is now as follows: For theorems of a $\forall\exists$-form, witnesses for the existential quantifiers (in terms of the universal quantifiers) are extracted using Gödel's Dialectica interpretation as mentioned before (see also the precise definitions in Chapter 3). These witnesses have types from T^X and using majorization, corresponding bounds with types from T are constructed for the witnesses which are initially validated in the corresponding model $\mathcal{M}^{\omega,X}$. If the types are low enough, which we call admissible (see Chapter 3), one can recover to the ordinary truth in a model based on the usual full set-theoretic type structure $\mathcal{S}^{\omega,X}$ defined through a given normed space X by interpreting the additional constants of $\mathcal{A}^\omega[X, \|\cdot\|]$ accordingly (using the operator $(\cdot)_\circ$ to choose representations of real numbers). Lastly, besides the existence of bounds, the metatheorems also guarantee an estimation of their computational complexity which, in the presence of the axiom of dependent choice DC, can be as complex as bar-recursive in the sense of Spector [198] but if this principle is not needed, bounds that are primitive recursive in the sense of Gödel can be guaranteed (and this extends to further stratifications[6]). Lastly, notice that full classical logic is permitted in the systems. If classical logic does not (or only minimally) feature in the proof, then the results of the theorem can be strengthened as we will also discuss in Chapter 3.

Theorem 2.3.1 ([71]). *Let ρ be admissible and let $B_\forall(x,u)/C_\exists(x,v)$ be purely universal/existential, respectively, where the types of the quantifiers are admissible and such that they only contain $x,u/x,v$ freely. Assume that*

$$\mathcal{A}^\omega[X, \|\cdot\|] \vdash \forall x^\rho \left(\forall u^0 B_\forall(x,u) \to \exists v^0 C_\exists(x,v)\right).$$

Then there exists a partial functional $\Phi : S_{\widehat{\rho}} \rightharpoonup \mathbb{N}$ which is defined on all strongly majorizable elements of $S_{\widehat{\rho}}$, where the corresponding restriction to these elements is bar-recursively computable and where the following holds in all non-trivial real normed vector spaces $(X, \|\cdot\|)$: for all $x \in S_\rho$ and $x^* \in S_{\widehat{\rho}}$, if $x^* \gtrsim x$, then

$$\mathcal{S}^{\omega,X} \models \forall u \leqslant_0 \Phi(x^*) B_\forall(x,u) \to \exists v \leqslant_0 \Phi(x^*) C_\exists(x,v).$$

Here, $\gtrsim$ is the extension due to [71, 95] of the strong majorizability relation of Bezem and $\widehat{\rho} \in T$ is the type of the majorants of objects of type $\rho \in T^X$ and $S_\rho, S_{\widehat{\rho}}$ are the sets of all set-theoretic functionals of type $\rho, \widehat{\rho}$, respectively.

[6] In particular, note [92] where a sequence of theories $G_n A^\omega$ is defined whose provably total function correspond to the n-th level of the Grzegorczyk hierarchy [75] and correspondingly (if used in a version extended to the abstract type X) guarantee bounds of such complexity.

3 Proof Mining with Set-Valued Operators

3.1 Introduction

This chapter summarizes the main logical results on the treatment of certain classes of set-valued operators in systems amenable for proof mining presented in the Master Thesis of the author [168] and published (in a slightly revised form) in the paper [165]. In that vein, the formulations are largely taken from these works.

Concretely, the thesis [168] introduced formal systems that allow for the application of methods from proof mining to proofs from *accretive* and *monotone operator theory*, central branches of nonlinear functional analysis which constitute the abstract study of certain prominent classes of set-valued mappings between linear spaces. In particular, in this work we established general logical metatheorems in the spirit of Chapter 2 (recall also Chapter 1) that guarantee the existence and quantify the complexity of the computational content of theorems pertaining to accretive and monotone set-valued operators and, further, allow for the extraction of this content.

Besides the proofs of the logical metatheorems for the resulting systems, which we present in more detail (complementing the brief discussion from Chapter 2), the summary given here in general omits most proofs which can be found in [165] and instead focuses on the main ideas behind the chosen representations for the analytical objects in question on which the rest of the thesis crucially relies. In particular, we want to emphasize that the chosen approach is very suitable for applications since, as discussed extensively in [165] already, the resulting systems allow for the convenient formalization of large classes of theorems and proofs involving abstract accretive and monotone set-valued operators and their (total) resolvents. In particular, the systems have already led to entirely new case studies (see e.g. [166]).

Further, we also discuss some of the other main theoretical considerations made in [165], like the characterization of the key property of an operator being maximal by equivalent notions involving formal extensionality of the operator as well as the treatment of range conditions and the notion of majorizability for set-valued operators.

3.2 Set-valued operators and nonexpansive maps in Banach and Hilbert spaces

In this section, we survey the basic notions and results for accretive and monotone operators as well as for nonexpansive maps over normed and inner product spaces which are essential for large parts of this thesis. For this, let $(X, \|\cdot\|)$ be a real normed space.

3.2.1 Nonexpansive functions

We begin with nonexpansive functions and their relatives on normed and inner product spaces where we follow the definitions of [31]. Let $D \subseteq X$ be non-empty and let $T : D \to X$ be a function. Then T is called

1. *nonexpansive* if
$$\forall x, y \in D \left(\|Tx - Ty\| \leqslant \|x - y\|\right),$$
2. *firmly nonexpansive* if
$$\forall x, y \in D \forall r > 0 \left(\|Tx - Ty\| \leqslant \|r(x - y) + (1 - r)(Tx - Ty)\|\right),$$

There is a useful equivalent reformulation of the notion of firm nonexpansivity when we pass to inner product spaces $(X, \langle\cdot,\cdot\rangle)$: then T is firmly nonexpansive if, and only if
$$\forall x, y \in D \left(\langle x - y, Tx - Ty\rangle \geqslant \|Tx - Ty\|^2\right).$$

3.2.2 Set-valued operators

A set-valued operator on a space X is simply a mapping $A : X \to 2^X$. Set-theoretically, such an A is nothing else but its graph $\mathrm{gra}A := \{(x, y) \mid y \in Ax\}$ and we correspondingly use the notations $y \in Ax$, $(x, y) \in A$ and $(x, y) \in \mathrm{gra}A$ interchangeably.

For a set-valued operator A, we define $\mathrm{dom}A := \{x \in X \mid Ax \neq \emptyset\}$ and $\mathrm{ran}A := \bigcup_{x \in X} Ax$. We write A^{-1} for the inverse operator defined by $x \in A^{-1}u$ iff $u \in Ax$. We set λA by $(\lambda A)x := \{\lambda u \mid u \in Ax\}$. If B is another set-valued operator on X, we define $A + B$ via $(A + B)x := \{u + v \mid u \in Ax \text{ and } v \in Bx\}$.

The main classes of set-valued operators which we want to consider first are the analytically motivated accretive and monotone operators. Besides the references cited in the following, we in particular refer to the standard references [4, 202] for further exposition on the theory of accretive operators in Banach spaces and to [11] for the theory of monotone operators in Hilbert spaces.

Definition 3.2.1 ([84])**.** Let $(X, \|\cdot\|)$ be a normed space. A set-valued operator A is called *accretive* if

$$\forall (x,u),(y,v) \in \mathrm{gra}A, \lambda > 0 \, (\|x - y + \lambda(u - v)\| \geqslant \|x - y\|)$$

and A is called *m-accretive* if $\mathrm{ran}(Id + \gamma A) = X$ for all $\gamma > 0$.

Now, for an inner product space, there is the following equivalent characterization of accretivity which is commonly called monotonicity.

Definition 3.2.2 (essentially [146, 147])**.** Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. A set-valued operator A is called *monotone* if

$$\forall (x,u),(y,v) \in \mathrm{gra}A \, (\langle x - y, u - v \rangle \geqslant 0)$$

and A is called *maximally monotone* if it is monotone and $\mathrm{gra}A \subsetneq \mathrm{gra}B$ implies that B is not monotone, i.e. the graph of A is not properly contained in the graph of another monotone operator.

3.2.3 Resolvents and correspondence results

In (nonlinear) functional analysis, one of the main tools for studying many classes of (set-valued) operators A is their corresponding *resolvent* J^A_γ, defined as follows for $\gamma > 0$:

$$J^A_\gamma := (Id + \gamma A)^{-1}.$$

In particular, J^A_γ is by its definition at first a set-valued map with $\mathrm{dom}J^A_\gamma = \mathrm{ran}(Id + \gamma A)$ as well as $\mathrm{ran}J^A_\gamma \subseteq \mathrm{dom}A$ and where the following defining equivalence holds:

$$p \in J^A_\gamma x \text{ iff } \gamma^{-1}(x - p) \in Ap.$$

However, if the operator in question is accretive, then its resolvent is characterized by a further collection of convenient properties.

Theorem 3.2.3 (essentially [4, 31]). *Let A be a set-valued operator on a normed space X. Then the following are equivalent:*

(a) A is accretive,

(b) J^A_γ is single-valued and firmly nonexpansive (on its domain) for all $\gamma > 0$,

(c) J^A_γ is single-valued and firmly nonexpansive (on its domain) for some $\gamma > 0$,

(d) J^A_γ is single-valued and nonexpansive (on its domain) for all $\gamma > 0$.

In the case of monotonicity, this further extends to the following:

Theorem 3.2.4 (essentially [4, 31, 147], see also [11]). *Let X be a Hilbert space and A a set-valued operator.*

1. *Items (a) - (d) of Theorem 3.2.3 are equivalent to*

   (e) A is monotone.

2. *A is maximally monotone if and only if J^A_γ is single-valued, firmly nonexpansive and $\mathrm{ran}(\mathrm{Id} + \gamma A) = X$ for some/any $\gamma > 0$.*

The last statement is known as Minty's theorem [147]. We already see that maximality conditions are linked with the totality of the resolvent, a result which sets a characteristic theme in the correspondence theory of operators and their resolvents as it extends to various other classes besides monotone operators (see e.g. [13, 14, 15]).

We here just want to note that this correspondence between totality of the resolvent and set-theoretic maximality does not extend to accretive operators on normed spaces (as first asked in [51] and then answered in [37, 49] negatively). The one direction that remains valid is the following:

Lemma 3.2.5 (essentially [51]). *Let A be accretive.*

1. *If $\mathrm{ran}(\mathrm{Id} + \gamma A) = X$ for some $\gamma > 0$, then A has no proper accretive extension.*

2. *If $\mathrm{ran}(\mathrm{Id} + \gamma A) = X$ for some $\gamma > 0$, then $\mathrm{ran}(\mathrm{Id} + \gamma A) = X$ for all $\gamma > 0$.*

In particular, m-accretivity implies maximal accretivity.

3.3 Logical systems for operators and their resolvents

We now introduce the systems for operators and their resolvents defined in [165]. For this, akin to [165], we split the treatment between whether the resolvents of the operator are assumed to be total or partial.

Before we consider the resolvents, we have to take a look at set-valued operators. These set-valued operators $A : X \to 2^X$ are now modeled using a constant χ_A of type $0(X)(X)$ which represents A via a function that takes an argument x from X and returns a characteristic function for Ax. We write $y \in Ax$ or $(x, y) \in A$ or $(x, y) \in \mathrm{gra}A$ for $\chi_A xy =_0 0$.

3.3.1 Formal systems for total resolvents

Regarding operators with total resolvents, we begin with the system for m-accretive operators, i.e. accretive operators with total resolvents. To define this system, we first add the constant χ_A as discussed before.

Now, over this extended language, we have to introduce the resolvent. As discussed in Theorem 3.2.3, the resolvents of accretive or monotone operators are always single-valued. So, in the context of m-accretive operators, we can infuse this single-valuedness and totality of the resolvent already into the type and in that vein add a constant J^{χ_A} of type $X(X)(1)$. The output of type X shall be seen as the (unique) value of the resolvent and the input of type 1 represents the real parameter $\gamma > 0$. In that vein, we write J^A_γ for $J^{\chi_A}\gamma$ where γ is of type 1. This approach via a constant of such a type is of course only of use if feasible axioms can now be presented so that bound extraction results can be obtained and common proofs from the literature can be formalized, i.e. over this language we now need to suitably represent the defining equality $J^A_\gamma = (Id + \gamma A)^{-1}$. Naively, over this extended language, this equality can be expressed formally via

$$\forall x^X, p^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to p =_X J^A_\gamma x \leftrightarrow \gamma^{-1}(x -_X p) \in Ap\right).$$

This statement, by virtue of the biimplication and the universal quantifier hidden in $=_X$, has too high quantifier-complexity to a priori guarantee that metatheorems for proof mining extend to systems where it is assumed as an axiom. However, focusing on the problematic direction

$$\forall x^X, p^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \wedge p =_X J^A_\gamma x \to \gamma^{-1}(x -_X p) \in Ap\right),$$

we can, instead of requiring the inclusion for all extensionally equal representations p, move to the intensional version

$$\forall x^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right).$$

This statement, as inclusions in the graph of A are quantifier-free statements, is universal and thus a priori does not hinder bound extraction results. Even more, as discussed already in [165] (and as will be further substantiated by this thesis), this axiom is indeed the right axiom to choose for the resolvents (note in particular that the other direction, which already is universal and thus a priori unproblematic, will be provable in the system later defined, see Proposition 3.3.2).

For majorization of the resolvent later on, the systems considered in [165] actually contain three further constants besides χ_A and J^{χ_A}: $\widetilde{\gamma}$ of type 1, $m_{\widetilde{\gamma}}$ of type 0 and c_X of type X. These are used for majorization of the resolvent constant J^{χ_A} later on in the sense that a bound for $\left\|x - J^A_\gamma x\right\|$ for *some* x (designated by c_X) and *some* $\gamma > 0$ (designated by $\widetilde{\gamma}$ and where $\gamma > 0$ is witnessed using $m_{\widetilde{\gamma}}$) will suffice for constructing a majorant of J^{χ_A} (see the the proof of Lemma 3.7.7 later on). With these further constants, we now consider the following system:

Definition 3.3.1 ([165])**.** The theory $\mathcal{V}^\omega$ is defined as the extension of the theory $\mathcal{A}^\omega[X, \|\cdot\|]$ with the above constants and corresponding axioms

(I) $\forall x^X, y^X (\chi_A xy \leqslant_0 1)$,

(II) $\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right)$,

(III) $\begin{cases} \forall x^X, y^X, u^X, v^X, \lambda^1\big(u \in Ax \wedge v \in Ay \\ \qquad\qquad \to \|x -_X y +_X |\lambda|(u -_X v)\|_X \geqslant_{\mathbb{R}} \|x -_X y\|_X\big), \end{cases}$

(IV) $\widetilde{\gamma} \geqslant_{\mathbb{R}} 2^{-m_{\widetilde{\gamma}}}$.

Note that the behavior of J^A_γ for $\gamma \leqslant_{\mathbb{R}} 0$ is left undefined.

The system $\mathcal{V}^\omega$ is strong enough to formalize large parts of the theory of m-accretive operators and we will see examples of some essential theorems on the operator and resolvent that $\mathcal{V}^\omega$ proves in Proposition 3.3.2 later on.

Before that, we introduce the formal system that accommodates inner product spaces and corresponding maximal monotone operators (or monotone operators with total resolvents): As monotonicity and accretivity are equivalent for inner product spaces (see Theorem 3.2.4), we can utilize the previous system $\mathcal{V}^\omega$. Hence, adding the axioms (I)-(IV) from before to $\mathcal{A}^\omega[X,\langle\cdot,\cdot\rangle]$ (or, in other words, adding the parallelogram law to $\mathcal{V}^\omega$) results in a corresponding system for monotone operators with total resolvents which we denote by $\mathcal{T}^\omega$.

We now turn to some of the central properties of the resolvent that the above systems can prove:

Proposition 3.3.2 ([165])**.** $\mathcal{V}^\omega$ *proves:*

1. J^A_γ *is unique for any* $\gamma > 0$*, i.e.*

$$\forall \gamma^1, p^X, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \gamma^{-1}(x -_X p) \in Ap \to p =_X J^A_\gamma x\right).$$

2. J^A_γ *is firmly nonexpansive for any* $\gamma > 0$*, i.e.*

$$\forall \gamma^1, r^1, x^X, y^X \Big(\gamma >_{\mathbb{R}} 0 \wedge r >_{\mathbb{R}} 0 \to \left\|J^A_\gamma x -_X J^A_\gamma y\right\|_X \\ \leqslant_{\mathbb{R}} \left\|r(x -_X y) +_X (1-r)(J^A_\gamma x -_X J^A_\gamma y)\right\|_X\Big).$$

3. J^A_γ *is nonexpansive for any* $\gamma > 0$*, i.e.*

$$\forall \gamma^1, x^X, y^X \left(\gamma >_{\mathbb{R}} 0 \to \|x -_X y\|_X \geqslant_{\mathbb{R}} \left\|J^A_\gamma x -_X J^A_\gamma y\right\|_X\right).$$

4. J^{χ_A} *is extensional in both arguments:*

$$\forall \gamma^1 >_{\mathbb{R}} 0, \gamma'^1 >_{\mathbb{R}} 0, x^X, x'^X \left(x =_X x' \wedge \gamma =_{\mathbb{R}} \gamma' \to J^A_\gamma x =_X J^A_{\gamma'} x'\right).$$

5. $\forall \gamma^1, \lambda^1, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \lambda >_{\mathbb{R}} 0 \to J^A_\lambda x =_X J^A_\gamma \left(\frac{\gamma}{\lambda} x +_X (1 - \frac{\gamma}{\lambda}) J^A_\lambda x\right)\right).$

6. $\forall \gamma^1, \lambda^1, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \lambda >_{\mathbb{R}} 0 \to \left\|x -_X J^A_\gamma x\right\|_X \leqslant_{\mathbb{R}} \left(2 + \frac{\gamma}{\lambda}\right) \left\|x -_X J^A_\lambda x\right\|_X\right).$

$\mathcal{A}^\omega[X,\langle\cdot,\cdot\rangle]$ *proves:*

7. $\forall x^X, y^X (\langle x, y\rangle_X \leqslant_{\mathbb{R}} 0 \leftrightarrow \forall \alpha^1 (\|x\|_X \leqslant_{\mathbb{R}} \|x -_X |\alpha| y\|_X)).$

Further, $\mathcal{T}^\omega$ *proves:*

8. *A is monotone, i.e.*

$$\forall x^X, y^X, u^X, v^X \left(u \in Ax \wedge v \in Ay \to \langle x -_X y, u -_X v\rangle_X \geqslant_{\mathbb{R}} 0\right).$$

9. J^A_γ satisfies the alternative notion of firm nonexpansivity for any $\gamma > 0$, i.e.

$$\forall \gamma^1, x^X, y^X \left(\gamma >_{\mathbb{R}} 0 \to \langle x -_X y, J^A_\gamma x -_X J^A_\gamma y\rangle_X \geqslant_{\mathbb{R}} \left\|J^A_\gamma x -_X J^A_\gamma y\right\|^2_X\right).$$

As discussed already in [165], some subtleties arise when dealing with reciprocals like in the axioms and theorems presented above and we want to indicate what these subtleties are and how they can be formally addressed. As mentioned in the discussion of real arithmetic in Chapter 2, formulas containing reciprocal expressions like, e.g., the resolvent axiom

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right)$$

are just seen as abbreviations for extended versions which make the necessary dependency on a parameter l^0 with $|\gamma| >_{\mathbb{R}} 2^{-l}$ explicit, i.e. in the above example, one actually considers

$$\forall \gamma^1, x^X, l^0 \left(\gamma >_{\mathbb{R}} 2^{-l} \to (\gamma)^{-1}_l(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right)$$

where $(\cdot)^{-1}_l$ is the previously discussed closed term representing the reciprocal correctly for arguments α^1 satisfying $|\alpha| >_{\mathbb{R}} 2^{-l}$.

In most situations, like, e.g., in the formal theorems presented above (and their proofs as presented in [165]), these details can be neglected without resulting in any issues (see e.g. the discussion in [165]).

Other important objects like e.g. the so-called *Yosida approximate* A_γ (which is ubiquitous in the literature, see, e.g., [4]), can also be treated in the context of this system. Concretely, the Yosida approximate is defined as

$$A_\gamma := \frac{1}{\gamma}(Id - J^A_\gamma)$$

and thus can be treated by λ-abstraction as

$$\lambda l^0, \gamma^1, x^X. \left((\gamma)^{-1}_l(x -_X J^A_\gamma x)\right).$$

The additional parameter l is induced here again through the subtleties with reciprocals. As before, we will continue to be vague about this issue and essentially treat $A_\gamma x$ like an abbreviation for $\gamma^{-1}(x -_X J^A_\gamma x)$.

3.3.2 Formal systems for partial resolvents

Some applications of accretive or monotone operators do not require full maximality but only impose certain so-called *range conditions* on the operator which force the domains of the resolvents to be "large enough" (which will be discussed in some detail also later on). To accommodate for such operators, we now discuss how the previous approach needs to be modified to treat partial resolvents.

We opt for the following strategy as in [165]: we still use a constant J^{χ_A} of type $X(X)(1)$. Instead of specifying the behavior of this constant on any point x as, e.g., done by

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right),$$

we only specify it on its domain in the sense of

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\gamma) \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right). \tag{†}$$

For this to be a priori admissible in the context of bound extraction theorems, the statement $x \in \mathrm{dom}J^A_\gamma$ has to have a suitable representation in the language of the underlying system such that the resulting axiom has a monotone functional interpretation.

For this note that, as discussed in Section 3.2.3, the domain satisfies

$$\mathrm{dom}J^A_\gamma = \mathrm{ran}(Id + \gamma A)$$

and inclusion of an x in the latter is definable by an existential statement

$$\exists y^X \left(\frac{1}{\gamma}(x -_X y) \in Ay\right)$$

which may be used in the premise of the above sentence (†) to form a universal sentence (which thus has a trivial monotone functional interpretation). In that vein, we will in the following use the abbreviation

$$x \in \mathrm{dom}(J^A_\gamma) := \exists y^X \left(\gamma^{-1}(x -_X y) \in Ay\right)$$

and, with that choice, we obtain the theories $\mathcal{V}^\omega_p$ and $\mathcal{T}^\omega_p$ from the previous ones by replacing the axiom

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right) \tag{II}$$

from before with

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \exists y^X \left(\gamma^{-1}(x -_X y) \in Ay\right) \to \gamma^{-1}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right), \qquad \text{(II}'\text{)}$$

instantiating (†) with the above definition for $x \in \mathrm{dom}(J^A_\gamma)$.

The constant c_X, which was previously only used to designate an arbitrary anchor point for majorization, is now used to actually designate a common element of the domains of all J^A_γ and for that we add the corresponding defining axiom

$$\forall \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(c_X -_X J^A_\gamma c_X) \in A(J^A_\gamma c_X)\right). \qquad \text{(V)}$$

This assumption that $\bigcap_{\gamma>0} \mathrm{dom} J^A_\gamma \neq \emptyset$ is easily satisfiable in many applications as any nontrivial operator A has a non-empty domain and it is often assumed that the operator satisfies a range condition like

$$\mathrm{dom} A \subseteq \bigcap_{\gamma>0} \mathrm{ran}(Id + \gamma A),$$

which, as mentioned before, will be further discussed later on.

We obtain the following proposition as an immediate generalization of the previous Proposition 3.3.2.

Proposition 3.3.3 ([165]). *$\mathcal{V}^\omega_p$ proves:*

1. *J^A_γ is unique for any $\gamma > 0$, i.e.*

$$\forall \gamma^1, p^X, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \gamma^{-1}(x -_X p) \in Ap \to p =_X J^A_\gamma x\right).$$

2. *J^A_γ is firmly nonexpansive for any $\gamma > 0$ (on its domain), i.e.*

$$\forall \gamma^1, r^1, x^X, y^X \Big(\gamma >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\gamma) \wedge y \in \mathrm{dom}(J^A_\gamma) \wedge r >_{\mathbb{R}} 0 \\ \to \left\|J^A_\gamma x -_X J^A_\gamma y\right\|_X \leq_{\mathbb{R}} \left\|r(x -_X y) +_X (1-r)(J^A_\gamma x -_X J^A_\gamma y)\right\|_X\Big).$$

3. *J^A_γ is nonexpansive for any $\gamma > 0$ (on its domain), i.e.*

$$\forall \gamma^1, x^X, y^X \Big(\gamma >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\gamma) \wedge y \in \mathrm{dom}(J^A_\gamma) \\ \to \|x -_X y\|_X \geq_{\mathbb{R}} \left\|J^A_\gamma x -_X J^A_\gamma y\right\|_X\Big).$$

4. *J^A is extensional in both arguments (on its domain), i.e.*

$$\begin{cases}\forall \gamma^1 >_{\mathbb{R}} 0, x^X, x'^X \big(x \in \mathrm{dom}(J^A_\gamma) \\ \qquad\qquad \wedge x' \in \mathrm{dom}(J^A_\gamma) \wedge x =_X x' \to J^A_\gamma x =_X J^A_{\gamma'} x'\big), \\ \forall \gamma^1 >_{\mathbb{R}} 0, \gamma'^1 >_{\mathbb{R}} 0, x^X \big(x \in \mathrm{dom}(J^A_\gamma) \\ \qquad\qquad \wedge x \in \mathrm{dom}(J^A_{\gamma'}) \wedge \gamma =_{\mathbb{R}} \gamma' \to J^A_\gamma x =_X J^A_{\gamma'} x\big).\end{cases}$$

5. $$\begin{cases}\forall \gamma^1, \lambda^1, x^X \big(\gamma >_{\mathbb{R}} 0 \wedge \lambda >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\lambda) \\ \qquad\qquad \to J^A_\lambda x =_X J^A_\gamma \left(\frac{\gamma}{\lambda} x +_X \left(1 - \frac{\gamma}{\lambda}\right) J^A_\gamma x\right)\big).\end{cases}$$

6. $$\begin{cases}\forall \gamma^1, \lambda^1, x^X \Big(\gamma >_{\mathbb{R}} 0 \wedge \lambda >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\gamma) \wedge x \in \mathrm{dom}(J^A_\lambda) \\ \qquad\qquad \to \left\|x -_X J^A_\gamma x\right\|_X \leq_{\mathbb{R}} \left(2 + \frac{\gamma}{\lambda}\right) \left\|x -_X J^A_\lambda x\right\|_X \Big).\end{cases}$$

Further, $\mathcal{T}^\omega_p$ proves:

7. *A is monotone, i.e.*

$$\forall x^X, y^X, u^X, v^X \left(u \in Ax \wedge v \in Ay \to \langle x -_X y, u -_X v\rangle_X \geq_{\mathbb{R}} 0\right).$$

8. *J^A_γ satisfies the alternative notion of firm nonexpansivity for any $\gamma > 0$ (on its domain), i.e.*

$$\forall \gamma^1, x^X, y^X \Big(\gamma >_{\mathbb{R}} 0 \wedge x \in \mathrm{dom}(J^A_\gamma) \wedge y \in \mathrm{dom}(J^A_\gamma) \\ \to \langle x -_X y, J^A_\gamma x -_X J^A_\gamma y\rangle_X \geq_{\mathbb{R}} \left\|J^A_\gamma x -_X J^A_\gamma y\right\|^2_X \Big).$$

3.4 Extensionality and maximality

The whole enterprise of proof mining of course prominently features issues with extensionality as one of the main theoretical problems around the extraction of computational information from non-constructive proofs and as such, issues with extensionality in fact lie at the heart of any such approach. For a deeper discussion of this, we refer to [96].

In our case, as already mentioned above, these issues feature most prominently in the fact that no system which enjoys bound extraction results (akin to the ones established later) and which allows for discontinuous operators A (as the previously introduced systems do) can prove the extensionality of the operator A.[1]

A central theoretical result from [165] is now the connection between the extensionality of A and the maximality statement for A as well as to the previously mentioned stronger version of the resolvent axiom

$$\forall \gamma^1, x^X, p^X \left(\gamma >_{\mathbb{R}} 0 \land p =_X J^A_\gamma x \to \gamma^{-1}(x -_X p) \in Ap\right).$$

Theorem 3.4.1 ([165])**.** *Over $\mathcal{V}^\omega$, the following are equivalent:*

1. *Extensionality of A, i.e.*

$$\forall x^X, y^X, x'^X, y'^X \left(x =_X x' \land y =_X y' \to \chi_A xy =_0 \chi_A x'y'\right).$$

2. *The strong resolvent axiom, i.e.*

$$\forall x^X, p^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \land p =_X J^A_\gamma x \to \gamma^{-1}(x -_X p) \in Ap\right).$$

3. *Maximal accretivity of A, i.e.*

$$\forall x^X, u^X \Big(\forall y^X, v^X, \lambda^1 \Big(v \in Ay \to \|x -_X y +_X |\lambda|(u -_X v)\|_X \geq_{\mathbb{R}} \|x -_X y\|_X \Big) \to u \in Ax\Big).$$

4. *Closure of the graph of A, i.e.*

$$\forall x^X, y^X, x_{(\cdot)}^{X(0)}, y_{(\cdot)}^{X(0)} \Big(x_n \to_X x \land y_n \to_X y \land \forall n^0 (y_n \in Ax_n) \to y \in Ax\Big)$$

where $x_n \to_X x$ is short for

$$\forall k^0 \exists N^0 \forall m \geq_0 N \left(\|x_m -_X x\|_X \leq_{\mathbb{R}} 2^{-k}\right)$$

and similar for $y_n \to_X y$.

Over $\mathcal{T}^\omega$, items (1) - (4) are additionally equivalent to

5. *maximal monotonicity of A, i.e.*

$$\forall x^X, u^X \left(\forall y^X, v^X \left(v \in Ay \to \langle x -_X y, u -_X v\rangle_X \geq_{\mathbb{R}} 0\right) \to u \in Ax\right).$$

[1] In fact, stronger results are possible which will be discussed in Chapter 11.

So, in particular, the maximality of A can not be provable in any of the previous systems, albeit being a property of any m-accretive operator. However, the systems do recognize the set-theoretic maximality of A in the following weakened way:

Theorem 3.4.2. *The system $\mathcal{V}^\omega$ proves the following intensional maximality principle:*

$$\forall x^X, u^X \Big(\forall y^X, v^X, \lambda^1 \Big(v \in Ay \to \|x -_X y +_X |\lambda|(u -_X v)\|_X \geq_{\mathbb{R}} \|x -_X y\|_X\Big) \\ \to \exists x'^X, u'^X \left(x =_X x' \wedge u =_X u' \wedge u' \in Ax'\right)\Big).$$

So, if a proof uses the set-theoretic maximality of A to infer $u \in Ax$ but the rest of the proof is extensional in x and u, then this application of maximality of A can be treated by the system $\mathcal{V}^\omega$ (mitigated through the use of the resolvent).

Note also that full extensionality is admissible in a rule form as we still have the following weak rule of A-extensionality

$$\frac{\exists y^\sigma F_0(y) \to s =_X s' \quad \exists y^\sigma F_0(y) \to t =_X t'}{\exists y^\sigma F_0(y) \to (s \in At \leftrightarrow s' \in At')}$$

for a quantifier-free formula F_0 as a special case of the extensionality rule Σ_1-ER of $\mathcal{V}^\omega$.

Lastly, we want to note that over the partial systems, as explored in [165], there also exists a classification of extensionality of A via an extensional formulation of the definition of the domain but we do not discuss this here any further.

3.5 Range conditions

As mentioned before in the context of the systems for partial resolvents, instead of requiring that all resolvents are total, a more minimal assumption is often made in the literature in the form of a condition ensuring that the domains of the resolvents are large enough relative to an application, e.g. such that some particular iteration scheme is well-defined. Such assumptions are called range conditions and we here just briefly sketch the discussion from [165] on how the following (rather canonical) case can be treated in the previous systems (with more variants of such conditions discussed later on):

$$\mathrm{dom}A \subseteq \bigcap_{\gamma>0}(Id + \gamma A)(\mathrm{dom}A).$$

When we naively formalize the above range condition, we end up with the sentence

$$\forall \gamma^1, x^X, y^X \left(\gamma >_{\mathbb{R}} 0 \wedge y \in Ax \to x \in \mathrm{dom}(J^A_\gamma)\right).$$

So, using the previous intensional expression

$$x \in \mathrm{dom}(J^A_\gamma) := \exists y^X \left(\frac{1}{\gamma}(x -_X y) \in Ay\right),$$

we are lead to the following formula:

$$\forall \gamma^1, x^X, y^X \left(\gamma >_{\mathbb{R}} 0 \land y \in Ax \to \exists y^X \left(\frac{1}{\gamma}(x -_X y) \in Ay\right)\right).$$

Even further however, by axiom (II′), stating

$$\exists y^X \left(\frac{1}{\gamma}(x -_X y) \in Ay\right)$$

is equivalent to stating

$$\frac{1}{\gamma}(x -_X J^A_\gamma x) \in A(J^A_\gamma x),$$

i.e. that the resolvent at γ is well-defined at x (which is, after all, the meaning of $x \in \mathrm{dom}(J^A_\gamma)$). So, we can immediately simplify the formula from above and consider

$$\forall \gamma^1, x^X, y^X \left(\gamma >_{\mathbb{R}} 0 \land y \in Ax \to \frac{1}{\gamma}(x -_X J^A_\gamma x) \in A(J^A_\gamma x)\right).$$

This axiom expressing the range condition is in particular purely universal and thus can be trivially used in the bound extraction theorems.

3.6 Majorizable operators

Proofs which make essential use of representatives $y \in Ax$ for $x \in \mathrm{dom}A$ (i.e. $Ax \neq \emptyset$) can be treated by providing a suitable witnessing (Skolem) functional for the statement

$$\forall x^X \exists y^X (x \in \mathrm{dom}A \to y \in Ax)$$

which can immediately be treated by adding a further constant a of type $X(X)$ together with a (universal) defining axiom like

$$\forall x^X (x \in \mathrm{dom}A \to ax \in Ax) \qquad (*)$$

where we write $x \in \mathrm{dom}A := \exists y^X (y \in Ax)$. Such a witnessing functional a can take many forms depending on the particular application scenario (which might require additional axioms).

In any way however, such a functional then of course requires majorizing data if used in the bound extraction theorems and we want to shortly discuss this special instance of the majorizability notion here already: a function f of type 1 is called a *majorant* for a, written $f \gtrsim a$, if it is non-decreasing, i.e. $n \geqslant m$ implies $fn \geqslant fm$, and it satisfies

$$n \geqslant \|x\| \to fn \geqslant \|ax\| \text{ for all } n^0, x^X.$$

Thus, any witnessing functional a for an operator A can only be treated in the context of the bound extraction theorems if there is *at least one* choice which is majorizable. The following notion capturing this minimal assumption was then introduced in [165]:

Definition 3.6.1 ([165])**.** An operator A is called *majorizable* if there exists a choice for a satisfying $(*)$ which is majorizable.

A common assumption from the literature is that an operator A is *bounded on bounded sets*, i.e. that $A(\overline{B}_n(0)) = \bigcup_{x \in \overline{B}_n(0)} Ax$ is bounded for any n. This assumption can be seen to impose a uniform majorizability assumption on all selection functionals:

Proposition 3.6.2 ([165])**.** *A is bounded on bounded sets if, and only if*

$$\exists a^{*0(0)} \forall a^{X(X)} \Big(\forall x^X (x \in \mathrm{dom}A \to ax \in Ax) \\ \wedge \forall x^X (x \notin \mathrm{dom}A \to \|ax\|_X =_{\mathbb{R}} 0) \to a^* \gtrsim a \Big).$$

3.7 Bound extraction theorems

We now present the proof mining metatheorems from [165] for the theories $\mathcal{V}^\omega/\mathcal{T}^\omega$ and their partial variants. The outline of the proofs we give in this chapter is rather detailed by which we will for one, provide additional details on the rather informal discussions on the structure and proof of such metatheorems from Chapter 2 as well as, for another, we will be able to shorten later proofs of metatheorems throughout the thesis as these follow a similar outline as the proofs given in this section (which is similar to that of [96]).

In [165], the focus was on metatheorems which allow one to treat classical logic as discussed in Chapter 2. However, in this thesis we will also place an emphasis on systems tailored to semi-constructive proofs which correspondingly, according to the absence of classical logic, allow for certain strengthenings of the conclusions of the metatheorems. The approach to these semi-constructive metatheorems taken here follows the general approach from [70] (and results of that kind were already discussed in [119] for intuitionistic variants of the system(s) $\mathcal{T}^\omega_{(p)}$).

3.7.1 Classical metatheorems

As mentioned in Chapters 1 and 2, the basis for the classical metatheorems is the utilization of *Gödel's functional interpretation* (going back to Gödel's work [74], but we mainly use the presentations from [96, 205]) in combination with a negative translation (which also goes back to Gödel [73] but we rely on a version by Kuroda [130]). We recall the definitions of those interpretations here.

Definition 3.7.1 ([74, 205]). The *Dialectica interpretation* $F^D = \exists \underline{x} \forall \underline{y} F_D(\underline{x}, \underline{y})$ of a formula F in the language of $\mathcal{A}^\omega[X, \|\cdot\|]$ (or any suitable extension thereof) is defined via the following recursion on the structure of the formula:

1. $F^D := F_D := F$ for F being a prime formula.

If $F^D = \exists \underline{x} \forall \underline{y} F_D(\underline{x}, \underline{y})$ and $G^D = \exists \underline{u} \forall \underline{v} G_D(\underline{u}, \underline{v})$, we set

1. $(F \wedge G)^D := \exists \underline{x}, \underline{u} \forall \underline{y}, \underline{v} (F \wedge G)_D$
   where $(F \wedge G)_D(\underline{x}, \underline{u}, \underline{y}, \underline{v}) := F_D(\underline{x}, \underline{y}) \wedge G_D(\underline{u}, \underline{v})$,

2. $(F \vee G)^D := \exists z^0, \underline{x}, \underline{u} \forall \underline{y}, \underline{v} (F \vee G)_D$
   where $(F \vee G)_D(z^0, \underline{x}, \underline{u}, \underline{y}, \underline{v}) := (z = 0 \to F_D(\underline{x}, \underline{y})) \wedge (z \neq 0 \to G_D(\underline{u}, \underline{v}))$,

3. $(F \to G)^D := \exists \underline{U}, \underline{Y} \forall \underline{x}, \underline{v} (F \to G)_D$
   where $(F \to G)_D(\underline{U}, \underline{Y}, \underline{x}, \underline{v}) := F_D(\underline{x}, \underline{Yxv}) \to G_D(\underline{Ux}, \underline{v})$,

4. $(\exists z^\tau F(z))^D := \exists z, \underline{x} \forall \underline{y} (\exists z^\tau F(z))_D$
   where $(\exists z^\tau F(z))_D(z, \underline{x}, \underline{y}) := F_D(\underline{x}, \underline{y}, z)$,

5. $(\forall z^\tau F(z))^D := \exists \underline{X} \forall z, \underline{y} (\forall z^\tau F(z))_D$
   where $(\forall z^\tau F(z))_D(\underline{X}, z, \underline{y}) := F_D(\underline{Xz}, \underline{y}, z)$.

Definition 3.7.2 ([130]). The *negative translation* of F is defined by $F' := \neg\neg F^*$ where F^* is defined by the following recursion on the structure of F:

1. $F^* := F$ for prime F;

2. $(F \circ G)^* := F^* \circ G^*$ for $\circ \in \{\wedge, \vee, \to\}$;

3. $(\exists x^\tau F)^* := \exists x^\tau F^*$;

4. $(\forall x^\tau F)^* := \forall x^\tau \neg\neg F^*$.

Following [71, 95] (see also [96]), we introduce some specific classes of types from T^X (providing precise definitions for the notions already vaguely discussed in Chapter 2). We call a type ξ *of degree* n if $\xi \in T$ and it has degree $\leqslant n$ in the usual sense (recall Chapter 2). Further we call ξ *small* if it is of the form $\xi = \xi_0(0)\dots(0)$ (including $0, X$) for $\xi_0 \in \{0, X\}$ and call it *admissible* if it is of the form $\xi = \xi_0(\tau_k)\dots(\tau_1)$ (including $0, X$) where each τ_i is small and $\xi_0 \in \{0, X\}$ as before.

Further, we define certain subclasses of existential/universal formulas satisfying certain type restrictions: A formula F is called a $\forall$*-formula* if $F = \forall \underline{a}^{\underline{\xi}} F_{qf}(\underline{a})$ with F_{qf} quantifier-free and if all types ξ_i in $\underline{\xi} = (\xi_1, \dots, \xi_k)$ are admissible. A formula F is called an $\exists$*-formula* if $F = \exists \underline{a}^{\underline{\xi}} F_{qf}(\underline{a})$ with similar $\underline{\xi}$.

Following [76, 86, 87], we introduce another certain class of formulas: by Δ we in the following denote a set of formulas of the form

$$\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c})$$

where F_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and $\underline{r}$ is a tuple of closed terms of appropriate types. Here, $\leqslant$ is defined by recursion on the type via

1. $x \leqslant_0 y := x \leqslant_0 y$,

2. $x \leqslant_X y := \|x\|_X \leqslant_{\mathbb{R}} \|y\|_X$,

3. $x \leqslant_{\tau(\xi)} y := \forall z^\xi (xz \leqslant_\tau yz)$,

and we write $\underline{x} \leqslant_{\underline{\sigma}} \underline{y}$ for $x_1 \leqslant_{\sigma_1} y_1 \wedge \dots \wedge x_k \leqslant_{\sigma_k} y_k$ where $\underline{x} = (x_1, \dots, x_k)$ and $\underline{y} = (y_1, \dots, y_k)$ are tuples with x_i, y_i of type σ_i for $\underline{\sigma} = (\sigma_1, \dots, \sigma_k)$.

Given such a set Δ, we write $\widetilde{\Delta}$ for the set of all Skolem normal forms

$$\exists \underline{B} \leqslant_{\underline{\sigma}(\underline{\delta})} \underline{r} \forall \underline{a}^{\underline{\delta}} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{B}\underline{a}, \underline{c})$$

for any $\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c})$ in Δ.

We now write $\mathcal{A}^\omega[X, \|\cdot\|]^-$ for $\mathcal{A}^\omega[X, \|\cdot\|]$ *without* the axioms QF-AC and DC. Further, by (BR), we denote the schema of *simultaneous bar-recursion* for the extended types T^X (see e.g. [96]), extending the notion from the seminal work of Spector [198]. Similarly we introduce $\mathcal{V}^{\omega-}_{(p)}$ and $\mathcal{T}^{\omega-}_{(p)}$ where we write $\mathcal{V}^\omega_{(p)}$ for $\mathcal{V}^\omega$ *or* $\mathcal{V}^\omega_p$ and similar for $\mathcal{T}^\omega_{(p)}$.

Lemma 3.7.3 ([95]). *Let $\mathcal{P}$ be a set of universal sentences and let $F(\underline{a})$ be an arbitrary formula in the language of $\mathcal{A}^\omega[X,\|\cdot\|]$, the latter with only the variables $\underline{a}$ free. Then the rule*

$$\begin{cases}\mathcal{A}^\omega[X,\|\cdot\|]+\mathcal{P}\vdash F(\underline{a})\Rightarrow\\ \mathcal{A}^\omega[X,\|\cdot\|]^-+\mathcal{P}+(\mathrm{BR})\vdash\forall\underline{a},\underline{y}(F')_D(\underline{t}\underline{a},\underline{y},\underline{a})\end{cases}$$

holds where $\underline{t}$ is a tuple of closed terms of the language of $\mathcal{A}^\omega[X,\|\cdot\|]^-+(\mathrm{BR})$ which can be extracted from the respective proof.

This result extends to any suitable extension of the language of $\mathcal{A}^\omega[X,\|\cdot\|]$ (e.g. by any kind of new types and constants) together with any number of additional universal axioms in that language.

In particular, note that the above lemma also holds for $\mathcal{V}^\omega_{(p)}$ and $\mathcal{T}^\omega_{(p)}$.

As discussed in Chapter 2, the central concept for formulating the quantitative bounds obtained by the metatheorems is that of majorization in the sense of the extension to the types in T^X due to [71, 95] of strong majorization due to Bezem [16]. In that way, majorants of objects with types from T^X will be objects with types from T related by the following projection:

Definition 3.7.4 ([71]). Define $\widehat{\tau}\in T$, given $\tau\in T^X$, by recursion on the structure via

$$\widehat{0}:=0,\ \widehat{X}:=0,\ \widehat{\tau(\xi)}:=\widehat{\tau}(\widehat{\xi}).$$

The majorizability relation $\gtrsim_\tau$ is then defined by recursion on the type along with the corresponding structure $\mathcal{M}^{\omega,X}$ of all (strongly) majorizable functionals of finite type as defined in [71, 95]:

Definition 3.7.5 ([71, 95]). Let $(X,\|\cdot\|)$ be a non-empty normed space. The structure $\mathcal{M}^{\omega,X}$ and the majorizability relation $\gtrsim_\tau$ are defined by

$$\begin{cases}M_0:=\mathbb{N}, n\gtrsim_0 m:=n\geqslant m\wedge n,m\in\mathbb{N},\\ M_X:=X, n\gtrsim_X x:=n\geqslant\|x\|\wedge n\in M_0, x\in M_X,\\ x^*\gtrsim_{\tau(\xi)}x:=x^*\in M_{\widehat{\tau}}^{M_{\widehat{\xi}}}\wedge x\in M_\tau^{M_\xi}\\ \qquad\qquad\wedge\forall y^*\in M_{\widehat{\xi}}, y\in M_\xi(y^*\gtrsim_\xi y\to x^*y^*\gtrsim_\tau xy)\\ \qquad\qquad\wedge\forall y^*,y\in M_{\widehat{\xi}}(y^*\gtrsim_{\widehat{\xi}}y\to x^*y^*\gtrsim_{\widehat{\tau}}x^*y),\\ M_{\tau(\xi)}:=\left\{x\in M_\tau^{M_\xi}\mid\exists x^*\in M_{\widehat{\tau}}^{M_{\widehat{\xi}}}:x^*\gtrsim_{\tau(\xi)}x\right\}.\end{cases}$$

Correspondingly, the full set-theoretic type structure $\mathcal{S}^{\omega,X}$ is defined via $S_0 := \mathbb{N}$, $S_X := X$ and

$$S_{\tau(\xi)} := S_\tau^{S_\xi}.$$

For an inner product space, the structures $\mathcal{S}^{\omega,X}$ and $\mathcal{M}^{\omega,X}$ are defined via the norm induced by the inner product.

Now, majorization behaves as expected for functionals with multiple arguments (represented by their "curryied" variants) as the following lemma shows:

Lemma 3.7.6 ([71, 95], see also Kohlenbach [96], Lemma 17.80)**.** *Let* $\xi = \tau(\xi_k)\ldots(\xi_1)$*. For* $x^* : M_{\widehat{\xi}_1} \to (M_{\widehat{\xi}_2} \to \cdots \to M_{\widehat{\tau}})\ldots)$ *and* $x : M_{\xi_1} \to (M_{\xi_2} \to \cdots \to M_\tau)\ldots)$*, we have* $x^* \gtrsim_\xi x$ *iff*

(a) $\forall y_1^*, y_1, \ldots, y_k^*, y_k \left(\bigwedge_{i=1}^k (y_i^* \gtrsim_{\xi_i} y_i) \to x^*y_1^*\ldots y_k^* \gtrsim_\tau xy_1\ldots y_k\right)$ *and*

(b) $\forall y_1^*, y_1, \ldots, y_k^*, y_k \left(\bigwedge_{i=1}^k (y_i^* \gtrsim_{\widehat{\xi}_i} y_i) \to x^*y_1^*\ldots y_k^* \gtrsim_{\widehat{\tau}} x^*y_1\ldots y_k\right)$.

The proof of the main bound extraction result now relies on a combination of functional interpretation and negative translation together with subsequent majorization as outlined in Chapter 2. The following lemma gives the main result for the latter ingredient (akin to, e.g., Lemma 9.11 in [71]).

Lemma 3.7.7 ([165])**.** *Let* $(X, \|\cdot\|)$ *be a (nontrivial) normed space,* A *an m-accretive operator and* J^A_γ *its resolvent with parameter* $\gamma > 0$*. Then* $\mathcal{M}^{\omega,X}$ *is a model of* $\mathcal{V}^{\omega-} +$ (BR) *(for a suitable interpretation of the additional constants). Moreover, for any closed term* t *of* $\mathcal{V}^{\omega-} +$ (BR)*, one can construct a closed term* t^* *of* $\mathcal{A}^\omega +$ (BR) *such that*

$$\mathcal{M}^{\omega,X} \models \forall n^0 \left(n \geqslant_{\mathbb{R}} \left\| c_X -_X J^A_{\widetilde{\gamma}} c_X \right\|_X, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|_X \to t^*(n) \gtrsim t \right).$$

Further, the same claim holds for $\mathcal{V}^\omega$ *replaced with*

1. $\mathcal{T}^\omega$ *where the conclusion is then drawn over inner product spaces using a monotone* A *with total resolvents,*

2. *the partial systems* $\mathcal{V}^\omega_p$ *and* $\mathcal{T}^\omega_p$ *where the conclusion is drawn over the appropriate spaces and operators, assuming that* $\bigcap_{\gamma>0} \mathrm{dom}(J^A_\gamma) \neq \varnothing$*.*

Proof. We sketch the interpretations and the majorization for the new constants and refer for any other details to [71] (see also [165]).[2] The designated interpretation of the constant χ_A in the model $\mathcal{M}^{\omega,X}$ is given by

$$[\chi_A]_{\mathcal{M}} := \lambda x, y \in X. \begin{cases} 0^0 & \text{if } y \in Ax, \\ 1^0 & \text{if } y \notin Ax, \end{cases}$$

where we write $\mathcal{M}$ as an abbreviation for $\mathcal{M}^{\omega,X}$. In the case of total resolvents, we set

$$[J^{\chi_A}]_{\mathcal{M}} := \lambda \alpha \in \mathbb{N}^{\mathbb{N}}, x \in X. \begin{cases} J^A_{r_\alpha} x & \text{if } r_\alpha > 0, \\ 0 & \text{otherwise}, \end{cases}$$

where r_α is the real represented by $\widehat{\alpha}$ as before. We set $[\widetilde{\gamma}]_{\mathcal{M}} := (\lambda)_\circ$ and $[m_{\widetilde{\gamma}}]_{\mathcal{M}} := m_\lambda$ for some real λ and natural m_λ with $\lambda \geqslant 2^{-m_\lambda}$. Lastly, in the case of the total systems, we define $[c_X]_{\mathcal{M}} := c$ for some arbitrary $c \in X$.

These constants are then majorizable (and their interpretations thus belong to $\mathcal{M}^{\omega,X}$): For χ_A, the majorant

$$\lambda x^0, y^0.1 \gtrsim \chi_A$$

is immediate by the previous Lemma 3.7.6.

For J^{χ_A}, assume that we have an n with $n \geqslant \left\| c_X - J^A_{\widetilde{\gamma}} c_X \right\|, \|c_X\|, m_{\widetilde{\gamma}}$ and γ^1, x^X with $x^* \gtrsim x$, i.e. $x^* \geqslant \|x\|$, as well as $\alpha \gtrsim \gamma$. Then in particular (with similar reasoning as in [96], Lemma 17.85), we obtain

$$\alpha(0) + 1 \geqslant \gamma(0) + 1 \geqslant \gamma.$$

Now, if $r_\gamma > 0$, we then have

$$\begin{aligned} \left\| J^A_\gamma x \right\| &\leqslant \|x - c_X\| + \left\| J^A_\gamma c_X \right\| && \text{(nonexpansivity)} \\ &\leqslant \|x\| + \|c_X\| + \left\| c_X - J^A_\gamma c_X \right\| + \|c_X\| \\ &\leqslant \|x\| + 2\|c_X\| + \left(2 + \frac{\gamma}{\widetilde{\gamma}}\right) \left\| c_X - J^A_{\widetilde{\gamma}} c_X \right\| && \text{(Proposition 3.3.2)} \\ &\leqslant x^* + 2n + (2 + 2^n(\alpha(0) + 1))n. \end{aligned}$$

For $r_\gamma \leqslant 0$, we get that $J^A_\gamma x = 0$. Thus, $\left\| J^A_\gamma x \right\| = 0 \leqslant x^* + 2n + (2 + 2^n(\alpha(0) + 1))n$ in that case as well. This implies

$$\lambda \alpha^1, x^{*0}. (x^* + 2n + (2 + 2^n(\alpha(0) + 1))n) \gtrsim J^{\chi_A}$$

[2] Some details on how to deal with some of the other constants of $\mathcal{A}^\omega[X, \|\cdot\|]$ will be given in the later Chapter 8.

using Lemma 3.7.6. Lastly, $(n)_\circ \gtrsim_{0(0)} \widetilde{\gamma}$, $n \geqslant_0 m_{\widetilde{\gamma}}$ and $n \gtrsim_X c_X$ are immediate by the assumptions on n (and using Lemma 2.1.2).

In the partial case, let $c \in \mathrm{dom}(J^A_\gamma)$ for any $\gamma > 0$ and define $[c_X]_\mathcal{M} := c$. Now, the resolvent is interpreted by

$$[J^{\chi_A}]_\mathcal{M} := \lambda\alpha \in \mathbb{N}^\mathbb{N}, x \in X. \begin{cases} J^A_{r_\alpha} x & \text{if } r_\alpha > 0 \text{ and } x \in \mathrm{dom}(J^A_{r_\alpha}), \\ 0 & \text{otherwise.} \end{cases}$$

The argument for majorizability of J^A_γ is the same as before, just restricting to $x \in \mathrm{dom}(J^A_{r_\gamma})$ and using nonexpansivity on the domain and Proposition 3.3.3. The other constants are interpreted and majorized as before.

Note that the corresponding extensions of $\mathcal{M}^{\omega,X}$ to the new constants are indeed models of the respective theories as none of the axioms for J^{χ_A} prescribe any behavior of the resolvent for $\gamma \leqslant 0$. □

The intended interpretation of the language of $\mathcal{V}^\omega/\mathcal{T}^\omega$ in the structure $\mathcal{S}^{\omega,X}$, turning $\mathcal{S}^{\omega,X}$ into a model of the respective theories, is defined exactly as the interpretation of these languages in $\mathcal{M}^{\omega,X}$ given in the above lemma.

We now formulate the bound extraction theorem in which we allow for potential additional axioms Δ of the form discussed before which are treated in spirit of the monotone functional interpretation due to [91] (and conceptually already due to [86, 87] as mentioned before).

We say "in spirit of the monotone functional interpretation" as we actually do not use a monotone variant of the functional interpretation but treat the functional interpretation part and the subsequent majorization separately. This nevertheless allows one to treat the axioms of type Δ for which we follow the presentation given for Corollary 5.14 as it is obtained from Theorem 5.13 in [76]. For that, we need the following lemma:

Lemma 3.7.8 ([76], Lemma 5.11)**.** *Let Δ be a set of formulas of the form considered before. Then $\mathcal{S}^{\omega,X} \models \Delta$ implies $\mathcal{M}^{\omega,X} \models \widetilde{\Delta}$.*

Proof. The proof given in [76] for Lemma 5.11 carries over. See Lemma 8.6.3 later for more details on this lemma and its proof. □

We now get to the main theorem on extractions of bounds from classical proofs, the proof of which we also give here in a more detailed way.

Theorem 3.7.9 ([165]). *Let τ be admissible, δ be of degree 1 and s be a closed term of $\mathcal{V}^\omega$ of type $\sigma(\delta)$ for admissible σ. Let Δ be a set of formulas of the form $\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c})$ where F_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and where $\underline{r}$ is a tuple of closed terms of appropriate type. Let $B_\forall(x,y,z,u)/C_\exists(x,y,z,v)$ be $\forall$-/$\exists$-formulas of $\mathcal{V}^\omega$ with only $x,y,z,u/x,y,z,v$ free. If*

$$\mathcal{V}^\omega + \Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x,y,z,u) \to \exists v^0 C_\exists(x,y,z,v)\right),$$

then one can extract a partial functional $\Phi : S_\delta \times S_{\widehat{\tau}} \times \mathbb{N} \rightharpoonup \mathbb{N}$ which is total and (bar-recursively) computable on $M_\delta \times M_{\widehat{\tau}} \times \mathbb{N}$ and such that for all $x \in S_\delta$, $z \in S_\tau$, $z^* \in S_{\widehat{\tau}}$ and all $n \in \mathbb{N}$, if $z^* \gtrsim z$ and $n \geqslant_{\mathbb{R}} \left\|c_X -_X J^A_{\widetilde{\gamma}}(c_X)\right\|_X, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|_X$, then

$$\mathcal{S}^{\omega,X} \models \forall y \leqslant_\sigma s(x)\big(\forall u \leqslant_0 \Phi(x,z^*,n) B_\forall(x,y,z,u) \to \exists v \leqslant_0 \Phi(x,z^*,n) C_\exists(x,y,z,v)\big)$$

holds whenever $\mathcal{S}^{\omega,X} \models \Delta$ for $\mathcal{S}^{\omega,X}$ defined via any (nontrivial) normed space $(X, \|\cdot\|)$ with χ_A interpreted by the characteristic function of an m-accretive A and J^{χ_A} by the corresponding resolvents J^A_γ for $\gamma > 0$ (and with suitable interpretations of the other constants so that the corresponding axioms hold).

In particular:

1. *If $\widehat{\tau}$ is of degree 1, then Φ is a total computable functional.*
2. *We may have tuples instead of single variables x, y, z, u, v and a finite conjunction instead of a single premise $\forall u^0 B_\forall(x,y,z,u)$.*
3. *If the claim is proved without DC, then τ may be arbitrary and Φ will be a total functional on $S_\delta \times S_{\widehat{\tau}} \times \mathbb{N}$ which is primitive recursive in the sense of Gödel. In that case, also plain majorization[3] can be used instead of strong majorization.*
4. *The claim of the above theorem as well as the items (1) - (3) from above hold similarly for $\mathcal{T}^\omega$ where the conclusion is then drawn over inner product spaces using monotone operators with total resolvents[4] and also for the partial systems $\mathcal{V}^\omega_p$ and $\mathcal{T}^\omega_p$ where the conclusion is drawn over the appropriate spaces and operators, assuming that $\bigcap_{\gamma>0} \mathrm{dom}(J^A_\gamma) \neq \emptyset$.*

Proof. We only treat the case of $\mathcal{V}^\omega$. The set Δ can be treated as in the proof of Theorem 5.13 in [76]: Add the Skolem functionals $\underline{B}$ from $\widetilde{\Delta}$ to the language. Then, $\widetilde{\Delta}$ can

[3]For a precise definition, see in particular the following section.

[4]By Minty's theorem, all conclusions are thus in particular valid over Hilbert spaces using maximally monotone operators.

be seen as another set of universal axioms and all the new constants are majorizable since $\underline{B} \leqslant_{\sigma(\delta)} \underline{r}$ and since $\underline{r}$ is a tuple of closed terms which is majorizable by Lemma 3.7.7 (which extends to this new language by Lemma 3.7.8). Then, the following proof goes through for this extended system instead of $\mathcal{V}^\omega$.

Now, first assume that

$$\mathcal{V}^\omega \vdash \forall z^\tau \left(\forall u^0 B_\forall(z,u) \to \exists v^0 C_\exists(z,v)\right).$$

By assumption, $B_\forall(z,u) = \forall \underline{a} B_{qf}(z,u,\underline{a})$ and $C_\exists(z,v) = \exists \underline{b} C_{qf}(z,v,\underline{b})$ for quantifier-free B_{qf} and C_{qf}. Thus, prenexing the above theorem of $\mathcal{V}^\omega$, we get

$$\mathcal{V}^\omega \vdash \forall z^\tau \exists u,v,\underline{a},\underline{b}(B_{qf}(z,u,\underline{a}) \to C_{qf}(z,v,\underline{b})).$$

Using Lemma 3.7.3, disregarding the realizers for $\underline{a},\underline{b}$ and reintroducing the quantifiers, we get closed terms t_u, t_v of $\mathcal{V}^{\omega-} + (\mathrm{BR})$ such that

$$\mathcal{V}^{\omega-} + (\mathrm{BR}) \vdash \forall z^\tau (B_\forall(z,t_u(z)) \to C_\exists(z,t_v(z))).$$

By Lemma 3.7.7 there are closed terms t^*_u, t^*_v of $\mathcal{A}^\omega + (BR)$ such that for all $n \geqslant \left\|c_X - J^A_{\widetilde{\gamma}}(c_X)\right\|, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|$, we get

$$\mathcal{M}^{\omega,X} \models t^*_u(n) \gtrsim t_u \wedge t^*_v(n) \gtrsim t_v \wedge \forall z^\tau (B_\forall(z,t_u(z)) \to C_\exists(z,t_v(z)))$$

for all (nontrivial) normed spaces $(X, \|\cdot\|)$ and all m-accretive operators A with resolvents J^A_γ defining $\mathcal{M}^{\omega,X}$ as in Lemma 3.7.7. Define

$$\Phi(z^*,n) := \max\{t^*_u(n)(z^*), t^*_v(n)(z^*)\}$$

for z^* of type $\widehat{\tau}$. Then

$$\mathcal{M}^{\omega,X} \models \forall u \leqslant_0 \Phi(z^*,n) B_\forall(z,u) \to \exists v \leqslant_0 \Phi(z^*,n) C_\exists(z,v)$$

holds for all $n \geqslant \left\|c_X - J^A_{\widetilde{\gamma}}(c_X)\right\|, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|$ as well as all $z \in M_\tau$ and $z^* \in M_{\widehat{\tau}}$ with $z^* \gtrsim z$. The conclusion that $\mathcal{S}^{\omega,X}$ satisfies the same sentence can be achieved as in the proof of Theorem 17.52 in [96] which we sketch here: Note that in the conclusion, we restrict ourselves to those z which have majorants z^*. As the type of z is admissible, it takes arguments of small type for which $\mathcal{M}^{\omega,X}$ and $\mathcal{S}^{\omega,X}$ coincide (see [96] and see also the later proof of Lemma 8.6.3 for a discussion of this). Therefore, any such z, z^* from $\mathcal{S}^{\omega,X}$ also live in $\mathcal{M}^{\omega,X}$ so that $\Phi(z^*)$ is well-defined in $\mathcal{S}^{\omega,X}$ for z, z^* with $z^* \gtrsim z$. In

$B_\forall$, all types are admissible to that truth in $\mathcal{S}^{\omega,X}$ implies truth in $\mathcal{M}^{\omega,X}$ and similarly for $C_\exists$ where thus truth in $\mathcal{M}^{\omega,X}$ implies truth in $\mathcal{S}^{\omega,X}$. Lastly, as in Lemma 17.84 in [96], we can show that as Φ is of type $0(\widehat{\tau})$, the interpretations of Φ in $\mathcal{S}^{\omega,X}$ and $\mathcal{M}^{\omega,X}$ coincide on majorizable elements. All in all we have that

$$\mathcal{S}^{\omega,X} \models \forall u \leqslant \Phi(z^*) B_\forall(z,u) \to \exists v \leqslant \Phi(z^*) C_\exists(z,v)$$

holds for all $z \in S_\tau$ and $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$.

For the additional prefix $\forall x^\delta \forall y \leqslant_\sigma s(x)$, let $\delta = 1$ for simplicity. For x of type δ, we then define $x^M(y^0) = \max_{\mathbb{N}}\{x(i) \mid 1 \leqslant i \leqslant y\}$. We get $x^M \gtrsim x$ and if $s(x) \geqslant_\sigma y$, then $s^*(n)(x^M) \gtrsim y$ where $s^*(n)$ is a majorant of s as in Lemma 3.7.7. Note now that the above result immediately extends to tuples $\underline{z}$ instead of a single z. Then by the above result for tuples instead of a single z, there now is a functional $\Phi'(x^*, y^*, z^*, n)$ such that

$$\mathcal{S}^{\omega,X} \models \forall u \leqslant \Phi'(x^*, y^*, z^*, n) B_\forall(x,y,z,u) \to \exists v \leqslant \Phi'(x^*, y^*, z^*, n) C_\exists(x,y,z,v)$$

for all $x \in S_\delta, y \in S_\sigma, z \in S_\tau$ with $x^* \gtrsim x$, $y^* \gtrsim y$, $z^* \gtrsim z$ and n as before. In particular, we have

$$\mathcal{S}^{\omega,X} \models \forall u \leqslant \Phi'(x^M, y^*, z^*, n) B_\forall(x,y,z,u) \to \exists v \leqslant \Phi'(x^M, y^*, z^*, n) C_\exists(x,y,z,v)$$

for any such x, y, z and y^*, z^* and thus, as $y \leqslant_\sigma s(x)$ yields $s^*(n)(x^M) \gtrsim y$, we get

$$\begin{aligned}\mathcal{S}^{\omega,X} \models \forall u \leqslant \Phi'(x^M, s^*(n)(x^M), z^*, n) B_\forall(x,y,u) \\ \to \exists v \leqslant \Phi'(x^M, s^*(n)(x^M), z^*, n) C_\exists(x,y,v)\end{aligned}$$

in that case. Then define $\Phi(x, z^*, n) = \Phi'(x^M, s^*(n)(x^M), z^*, n)$.

Item (1) can be shown as in the proof of Theorem 17.52 from [96] (see page 428 therein). Further, (2) is immediate and (3) follows from the fact that without DC, bar recursion and thus the use of $\mathcal{M}^{\omega,X}$ both become superfluous. □

3.7.2 Semi-constructive metatheorems

As mentioned before in Chapter 1, the basis for the semi-constructive metatheorems is the utilization of *Kreisel's modified realizability interpretation* (going back to Kreisel's work [127, 128], but we again mainly use the presentations from [96, 205]).

For this, we of course also have to rely on suitable semi-constructive systems which are here defined over the common base $\mathcal{A}^\omega_i := \text{E-HA}^\omega + \text{AC}$ where E-HA$^\omega$ is Heyting arithmetic in all finite types with the full axiom of extensionality as in Chapter 2 and $\text{AC} = \bigcup_{\rho,\tau\in T}\{\text{AC}^{\rho,\tau}\}$ is the full axiom of choice with

$$\forall x^\rho \exists y^\tau F(x,y) \to \exists Y^{\tau(\rho)} \forall x^\rho F(x, Yx) \tag{AC$^{\rho,\tau}$}$$

where F is arbitrary. The base $\mathcal{A}^\omega_i$ is then extended with the machinery for abstract types and the constants and axioms for normed linear spaces similar to $\mathcal{A}^\omega[X,\|\cdot\|]$ which results in a system that we denote by $\mathcal{A}^\omega_i[X,\|\cdot\|]$, following [70] (and we in general refer to [70] for any further details surrounding this system). Then we can define $\mathcal{V}^\omega_{i,(p)}$ and $\mathcal{T}^\omega_{i,(p)}$ similar to $\mathcal{V}^\omega_{(p)}$ and $\mathcal{T}^\omega_{(p)}$, but over $\mathcal{A}^\omega_i[X,\|\cdot\|]$ instead of $\mathcal{A}^\omega[X,\|\cdot\|]$.

We now give the definition of the main proof interpretation employed in the context of these semi-intuitionistic systems, the modified realizability interpretation due to Kreisel:

Definition 3.7.10 (Kreisel [127, 128])**.** For any formula F in the language of $\mathcal{A}^\omega_i[X,\|\cdot\|]$ (or any suitable extension thereof), we define its modified realizability interpretation $\underline{x}\, mr\, F$ by recursion on the structure of F:

1. $\langle\rangle\, mr\, F := F$ for a prime formula F where $\langle\rangle$ is the empty tuple.

Further, if $\underline{x}\, mr\, F$ and $\underline{y}\, mr\, G$ are the modified realizability interpretations of F and G, respectively, then:

2. $\underline{x}, \underline{y}\, mr\, (F \wedge G) := \underline{x}\, mr\, F \wedge \underline{y}\, mr\, G$,
3. $z^0, \underline{x}, \underline{y}\, mr\, (F \vee G) := (z =_0 0 \to \underline{x}\, mr\, F) \wedge (z \neq_0 0 \to \underline{y}\, mr\, G)$,
4. $\underline{Y}\, mr\, F \to G := \forall \underline{x}(\underline{x}\, mr\, F \to \underline{Y}\,\underline{x}\, mr\, G)$,
5. $\underline{X}\, mr\, \forall w^\rho F(w) := \forall w^\rho(\underline{X} w\, mr\, F(w))$,
6. $z^\rho, \underline{x}\, mr\, \exists w^\rho F(w) := \underline{x}\, mr\, F(z)$.

Note that in $\underline{x}\, mr\, F$, both the length and the types of $\underline{x}$ depend on the structure of F.

We define $\mathcal{A}^\omega_i[X,\|\cdot\|]^-$ as $\mathcal{A}^\omega_i[X,\|\cdot\|]$ with AC removed and similarly we define $\mathcal{V}^{\omega-}_{i,(p)}$ and $\mathcal{T}^{\omega-}_{i,(p)}$.

Further, we call a formula F to be $\exists$-free if it is built up from prime formulas only via $\wedge, \rightarrow, \neg$ and $\forall$. It is immediate that for an $\exists$-free formula F_{ef} with modified realizability interpretation $\underline{x}\, mr\, F_{ef}$, we have

$$\underline{x}\, mr\, F_{ef} \equiv F_{ef}$$

and further it is clear that the modified realizability interpretation $\underline{x}\, mr\, F$ for any F is $\exists$-free itself.

For the following result, we also define the independence of premise schema for $\exists$-free formulas $\mathrm{IP}_{ef} = \bigcup_{\sigma \in T^X} \{\mathrm{IP}^\sigma_{ef}\}$ where IP^σ_{ef} is defined as

$$(F_{ef} \to \exists x^\sigma G(x)) \to \exists x^\sigma (F_{ef} \to G(x)) \tag{IP^σ_{ef}}$$

where F_{ef} is $\exists$-free and does not contain x freely (and G is arbitrary).

We similarly define $\mathrm{IP}_\neg$ by using negated formulas $\neg F$ instead of formulas F_{ef} with are $\exists$-free.

Lastly, we say that a formula is of type $\Gamma_\neg$ if it is of the form

$$\forall \underline{x}^{\underline{\alpha}} \left(F \to \exists \underline{v} \leqslant_{\underline{\sigma}} \underline{rx} \neg G\right)$$

for some formulas F, G and where the types in $\underline{\sigma}$ are arbitrary and the terms in $\underline{r}$ are closed. Here, $\leqslant$ is defined as before by recursion on the type.

We then get the following soundness result for the modified realizability interpretation.

Theorem 3.7.11 (essentially Troelstra [205], see also Gerhardy and Kohlenbach [70])**.** *Let Δ_{ef} be a set of $\exists$-free sentences. For any formula F in (possibly an extension of) the language of $\mathcal{A}^\omega_i[X, \|\cdot\|]$ with modified realizability interpretation $\underline{x}\, mr\, F$, if $\mathcal{A}^\omega_i[X, \|\cdot\|] + \mathrm{IP}_{\mathrm{ef}} + \Delta_{ef} \vdash F$, then $\mathcal{A}^\omega_i[X, \|\cdot\|]^- + \Delta_{ef} \vdash \underline{t}\, mr\, F$ where the terms $\underline{t}$ satisfy* $\mathrm{free}(\underline{t}) \subseteq \mathrm{free}(F)$ *and can be extracted from the proof of F.*

In particular, the above result also holds for the systems $\mathcal{V}^\omega_{i,(p)}$ and $\mathcal{T}^\omega_{i,(p)}$ as these extend $\mathcal{A}^\omega_i[X, \|\cdot\|]$ only by new constants and further universal (and hence $\exists$-free) axioms.

The following result now provides the principles which characterize the modified realizability interpretation, i.e. which suffice to recover from the modified realizability interpretation to the original formula.

Theorem 3.7.12 (essentially Troelstra [205], see also Gerhardy and Kohlenbach [70])**.** *For any formula F with modified realizability interpretation $\underline{x}\, mr\, F$, we have*

$$\mathcal{A}_i^\omega[X, \|\cdot\|] + \mathrm{IP}_{ef} \vdash F \leftrightarrow \exists \underline{x}(\underline{x}\, mr\, F).$$

Crucially it follows from the characterization result that in strong enough systems, being $\exists$-free is essentially the same as being negated. In this context, we also already consider the scheme of comprehension for $\exists$-free formulas

$$\exists \Phi^{0(\underline{\sigma})} \forall \underline{x}^{\underline{\sigma}} \left(\Phi(\underline{x}) =_0 0 \leftrightarrow F_{ef}(\underline{x})\right) \tag{CA_{ef}}$$

where $\underline{\sigma}$ is an arbitrary tuple of types and Φ is not free in the $\exists$-free formula F_{ef}. Likewise, we can define $\mathrm{CA}_\neg$ for comprehension for all negated formulas $\neg F$.

Lemma 3.7.13 (folklore, see Gerhardy and Kohlenbach [70])**.**

1. *For any formula F in the language of $\mathcal{A}_i^\omega[X, \|\cdot\|]$ (or for suitable extensions), there exists an $\exists$-free formula G_{ef} such that*

$$\mathcal{A}_i^\omega[X, \|\cdot\|] + \mathrm{IP}_{ef} \vdash \neg F \leftrightarrow G_{ef}.$$

2. *For any $\exists$-free formula F_{ef}:*

$$\mathcal{A}_i^\omega[X, \|\cdot\|] + \mathrm{IP}_{ef} \vdash F_{ef} \leftrightarrow \neg\neg F_{ef}.$$

3. *Over $\mathcal{A}_i^\omega[X, \|\cdot\|]$, the following equivalences hold:*

$$\mathrm{IP}_{ef} \leftrightarrow \mathrm{IP}_\neg \text{ and } \mathrm{CA}_{ef} \leftrightarrow \mathrm{CA}_\neg.$$

In the context of the semi-intuitionistic systems, there is no need anymore to rely on strong majorizability since bar-recursion is not needed in that context as choice principles are intuitionistically weak. In that way, we here rely on the "plain" notion of majorizability of Howard [79], again extended to the abstract types similar to the extensions from [71].

Definition 3.7.14. The "plain" majorizability relation $\gtrsim_\tau$ is defined recursively on the type via[5]

$$\begin{cases} n \gtrsim_0 m := n \geqslant m, \\ n \gtrsim_X x := n \geqslant \|x\|, \\ x^* \gtrsim_{\tau(\xi)} x := \forall y^{*\widehat{\xi}}, y^\xi (y^* \gtrsim_\xi y \to x^*y^* \gtrsim_\tau xy). \end{cases}$$

[5] We here use $\gtrsim$ for both the strong and "plain" majorizability relation but the context will make it clear which relation is meant.

The standard structure $\mathcal{S}^{\omega,X}$ is defined as before. Analogous to the previous Lemma 3.7.7, we also get the following majorizability result for the ordinary notion of majorizability (where the interpretations of the additional constants of the respective systems, that turn $\mathcal{S}^{\omega,X}$ into a model of said systems, are defined as before):

Lemma 3.7.15. *Let $(X, \|\cdot\|)$ be a (nontrivial) normed space, A an m-accretive operator and J^A_γ its resolvent with parameter $\gamma > 0$. Then $\mathcal{S}^{\omega,X}$ is a model of $\mathcal{V}^\omega_i + \mathrm{IP}_{ef}$ (for a suitable interpretation of the additional constants). Moreover, for any closed term t of $\mathcal{V}^\omega_i$, one can construct a closed term t^* of $\mathcal{A}^\omega_i$ such that*

$$\mathcal{S}^{\omega,X} \models \forall n^0 \Big(n \geqslant \left\| c_X -_X J^A_{\widetilde{\gamma}} c_X \right\|_X, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|_X \to t^*(n) \gtrsim t \Big).$$

The result holds with suitable modifications (see Lemma 3.7.7) also for $\mathcal{V}^\omega_{i,p}$ and $\mathcal{T}^\omega_{i,(p)}$.

Combining the soundness of the modified realizability interpretation with the majorizability notion (which essentially amounts to applying the monotone modified realizability interpretation, as first considered in [94]), we get the following result on bound extraction for the semi-constructive systems for set-valued accretive and monotone operators. This result (which, as mentioned before, was stated for $\mathcal{T}^\omega_{i,(p)}$ already in [119]) is a natural extension of the results given in [70] (which are in turn based on [94]).

Theorem 3.7.16. *Let δ be of degree 1 and σ, τ be arbitrary, s be a closed term of suitable type. Let $\Gamma_\neg$ be a set of formulas of the form $\forall \underline{x}^{\underline{\alpha}} \Big(C(\underline{x}) \to \exists \underline{v} \leqslant_{\underline{\beta}} \underline{r}\underline{x} \neg D(\underline{x}, \underline{v}) \Big)$ with $\underline{\alpha}, \underline{\beta}$ and $\underline{r}$ arbitrary. Let B, C be arbitrary formulas with only x, y, z, u or x, y, z free, respectively. If*

$$\mathcal{V}^\omega_i + \mathrm{IP}_\neg + \mathrm{CA}_\neg + \Gamma_\neg \vdash \forall x^\delta \forall y \leqslant_\sigma (x) \forall z^\tau (\neg B(x,y,z) \to \exists u^0 C(x,y,z,u)),$$

one can extract a $\Phi : S_\delta \times S_{\widehat{\tau}} \times \mathbb{N} \to \mathbb{N}$ which is primitive recursive in the sense of Gödel and for any $x \in S_\delta$, any $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$, any $z \in S_\tau$ and $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$ and any $n \in \mathbb{N}$ with $n \geqslant \left\| c_X -_X J^A_{\widetilde{\gamma}} c_X \right\|_X, m_{\widetilde{\gamma}}, \widetilde{\gamma}, \|c_X\|_X$, we have

$$\mathcal{S}^{\omega,X} \models \exists u \leqslant_0 \Phi(x, z^*, n)(\neg B(x,y,z) \to C(x,y,z,u))$$

whenever $\mathcal{S}^{\omega,X} \models \Gamma_\neg$ where $\mathcal{S}^{\omega,X}$ is defined by suitably interpreting the constants via a (nontrivial) normed space $(X, \|\cdot\|)$, an m-accretive operator A and its resolvent J^A_γ with parameter $\gamma > 0$.

This result hold similarly for $\mathcal{T}^\omega_i$ where the conclusion is then drawn over inner product spaces using monotone operators with total resolvents and also for the partial *systems $\mathcal{V}^\omega_p$ and $\mathcal{T}^\omega_p$ where the conclusion is drawn over the appropriate spaces and operators, assuming that $\bigcap_{\gamma>0}\mathrm{dom}(J^A_\gamma)\neq\emptyset$.*

Proof. At first, let $\forall \underline{x}^{\underline{\alpha}}\left(C(\underline{x})\to\exists\underline{v}\leqslant_{\underline{\beta}}\underline{rx}\neg D(\underline{x},\underline{v})\right)$ be a formula from $\Gamma_\neg$ and note that using intuitionistic logic, we have

$$\exists\underline{V}\leqslant\underline{r}\neg\neg\forall\underline{x}\left(C(\underline{x})\to\neg D(\underline{x},\underline{Vx})\right)\equiv\exists\underline{V}\leqslant\underline{r}\forall\underline{x}\left(C(\underline{x})\to\neg D(\underline{x},\underline{Vx})\right).$$

Further, we clearly have

$$\exists\underline{V}\leqslant\underline{r}\forall\underline{x}\left(C(\underline{x})\to\neg D(\underline{x},\underline{Vx})\right)\to\forall\underline{x}\left(C(\underline{x})\to\exists\underline{v}\leqslant_{\underline{\beta}}\underline{rx}\neg D(\underline{x},\underline{v})\right).$$

So using $\Gamma_\neg$ can be reduced to considering formulas of the form

$$\exists\underline{V}\leqslant_{\underline{\beta}(\underline{\alpha})}\underline{r}\neg E(\underline{V}).$$

Now using Lemma 3.7.13, we can replace such principles by

$$\exists\underline{V}\leqslant_{\underline{\beta}(\underline{\alpha})}\underline{r}E'_{ef}(\underline{V})$$

where E'_{ef} arises from $\neg E$ by Lemma 3.7.13. We denote the set of all such sentences arising from $\Gamma_\neg$ in that manner by Γ'_{ef}. Similar we can replace $\mathrm{CA}_\neg$ by CA_{ef} and $\mathrm{IP}_\neg$ by IP_{ef} and consequently reason over the modified system $\mathcal{V}^\omega_i+\mathrm{IP}_{ef}+\mathrm{CA}_{ef}+\Gamma'_{ef}$.

At first, regarding the handling of the axioms Γ'_{ef}: For any axiom $\exists\underline{V}\leqslant_{\underline{\beta}(\underline{\alpha})}\underline{r}E'_{ef}(\underline{V})\in\Gamma'_{ef}$, we add new constants $\underline{V}$ with the additional axiom

$$\underline{V}\leqslant_{\underline{\beta}(\underline{\alpha})}\underline{r}\wedge E'_{ef}(\underline{V})$$

to the system. These axioms are $\exists$-free as E'_{ef} is $\exists$-free and thus the soundness result from Theorem 3.7.11 applies for this extension. By considering majorants for the terms $\underline{r}$, we see that these $\underline{V}$ are majorizable and thus the majorizability result from Lemma 3.7.15 extends to this system. Then, the following proof goes through with this modified system if $\mathcal{S}^{\omega,X}\models\Gamma'_{ef}$ holds (which is the case if $\mathcal{S}^{\omega,X}\models\Gamma_\neg$).

Further, we can treat CA_{ef} by reducing it to a formula of a similar form as the formulas from Γ'_{ef}. Note that in the principle

$$\exists\Phi^{0(\underline{\sigma})}\forall\underline{x}^{\underline{\sigma}}\left(\Phi(\underline{x})=_0 0\leftrightarrow F_{ef}(\underline{x})\right),$$

the functional Φ is w.l.o.g. bounded by the constant 1 function, i.e. the principle can be equivalently rewritten as

$$\exists\Phi^{0(\underline{\sigma})}\leqslant_{0(\underline{\sigma})}\lambda x^{\underline{\sigma}}.1\forall\underline{x}^{\underline{\sigma}}\left(\Phi(\underline{x})=_0 0\leftrightarrow F_{ef}(\underline{x})\right).$$

This is now of the same form as the formulas from Γ'_{ef} as the inner matrix is $\exists$-free and thus can be treated in the same manner.

We now therefore only consider the case of $\mathcal{V}^\omega_i$. The other cases can be proved similarly. So, assume that

$$\mathcal{V}^\omega_i + \mathrm{IP}_{ef} \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\neg B(x,y,z) \to \exists u^0 C(x,y,z,u)).$$

Let $\underline{w}\, mr\, B$, $\underline{v}\, mr\, C$ be the modified realizability interpretations of B, C, respectively. The modified realizability interpretation of the above sentence is then given by

$$\begin{aligned}\underline{W}, U\, mr\, &\forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\neg B(x,y,z) \to \exists u^0 C(x,y,z,u))\\ &\equiv \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\forall \underline{v} \neg \underline{v}\, mr\, B(x,y,z) \to \underline{W}xyz\, mr\, C(x,y,z,Uxyz)).\end{aligned}$$

As all the axioms forming $\mathcal{V}^\omega_i$ from $\mathcal{A}^\omega_i[X, \|\cdot\|]$ are $\exists$-free, soundness of the modified realizability interpretation (Theorem 3.7.11) now implies that there are terms $\underline{t_W}, t_U$ such that

$$\mathcal{V}^{\omega-}_i \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\forall \underline{v} \neg \underline{v}\, mr\, B(x,y,z) \to \underline{t_W}xyz\, mr\, C(x,y,z,t_U xyz)).$$

Using the characterization result (Theorem 3.7.12), we get

$$\mathcal{V}^\omega_i + \mathrm{IP}_{ef} \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\neg B(x,y,z) \to C(x,y,z,t_U xyz)).$$

Using the majorizability result (Lemma 3.7.15), we get that there exist terms $t^*_U(n), s^*(n)$ such that for any n with $n \geqslant \left\| c_X - J^A_{\widetilde{\gamma}} c_X \right\|, m_{\widetilde{\gamma}}, |\widetilde{\gamma}|, \|c_X\|$:

$$\mathcal{S}^{\omega,X} \models t^*_U(n) \gtrsim t_U \wedge s^*(n) \gtrsim s \wedge \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau (\neg B(x,y,z) \to C(x,y,z,t_U xyz)).$$

Now, given $x \in S_\delta$ and $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$, we get $\mathcal{S}^{\omega,X} \models s^*(n)(x^M) \gtrsim s(x)$ as in the proof of Theorem 3.7.9 and thus $\mathcal{S}^{\omega,X} \models s^*(n)(x^M) \gtrsim y$. Thus, for any $z \in S_\tau$ and any $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$:

$$\mathcal{S}^{\omega,X} \models t^*_U(n)(x^M, s^*(n)(x^M), z^*) \geqslant t_U xyz.$$

With $\Phi(x, z^*, n) := t^*_U(n)(x^M, s^*(n)(x^M), z^*)$, this gives that

$$\mathcal{S}^{\omega,X} \models \exists u \leqslant_0 \Phi(x, z^*, n)(\neg B(x,y,z) \to C(x,y,z,u))$$

for any $x \in S_\delta$ and $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$ as well as any $z \in S_\tau$ and any $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$. $\square$

4 A proof-theoretic metatheorem for nonlinear semigroups generated by an accretive operator

4.1 Introduction

In this chapter, we now establish the logical tools necessary to treat nonlinear semigroups generated by accretive operators (with applications of these systems presented later on). Already since the pioneering studies of Browder [27], Kato [84] and Komura [123], a major tool in the study of nonlinear evolution equations has been the theory of nonlinear semigroups and through the notion of the generator, these are in particular connected to the theory of accretive operators with a range of correspondences via analogs of the Hille-Yosida theorem.

One of the most important basic results in that context is the representation theorem due to Crandall and Liggett [50] of the solution semigroup associated with the Cauchy problem

$$\begin{cases} u'(t) \in -Au(t),\ 0 < t < \infty \\ u(0) = x \end{cases} \tag{†}$$

over a Banach space X for a given set-valued accretive operator $A : X \to 2^X$. It is straightforward to show that any solution[1] is unique as A is accretive and if the system is solvable[2], then one can consider the family of operators $S(t)x = u_x(t)$ on $\mathrm{dom}A$ induced by the solutions $u_x(t)$ to (†) with initial values $x \in \mathrm{dom}A$ and for $t \geqslant 0$. As

[1] A function $u : [0,\infty) \to X$ is a solution of (†) if $u(0) = x$, $u(t)$ is absolutely continuous, differentiable almost everywhere in $(0,\infty)$ and satisfies (†) almost everywhere. Note that this is often called a strong solution but we omit the prefix strong in the following.

[2] As shown by Crandall and Liggett [50], this is (for strong solutions) in general not the case even for A m-accretive and $\mathrm{dom}A = X$.

these operators are continuous in x, one can consider the resulting extensions to $\overline{\mathrm{dom}A}$ which in that way generate the semigroup $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ on $\overline{\mathrm{dom}A}$ associated with (†). As shown by Brezis and Pazy [25], this solution semigroup, if existent, has a particular fundamental representation in terms of a so-called exponential formula:

$$u_x(t) = \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}A\right)^{-n} x.$$

As shown subsequently by Crandall and Liggett [50], this formula actually always generates a nonexpansive semigroup on $\overline{\mathrm{dom}A}$ and thus facilitates a general study of equations like (†) even in the absence of solutions.

Since the 1970s, an extensive range of results has been established in the theory of these semigroups and the initial value problems in the sense of (†) associated with them, in particular in regard to the asymptotic behavior of the solutions of these differential equations, their connection and use in the study of partial differential equations and their use in the study of zeros of accretive operators (see [4, 5, 11, 149, 158], among many more).

In this chapter, we extend the state-of-the-art of the underlying logical approach to proof mining to be applicable to proofs which make use of nonlinear semigroups generated by an accretive operator via the exponential formula. In particular, we establish logical metatheorems in the vein of the previously discussed results that guarantee, quantify and allow for the extraction of the computational content of theorems pertaining to these nonlinear semigroups. For that, we introduce new underlying logical systems that extend those developed for the treatment of accretive operators on normed spaces as discussed in Chapter 3 by carefully selected additional constants and corresponding axioms such that proofs from the mainstream literature become formalizable. To that end, we show that the initial key properties of these semigroups can be formally proved in these systems.

These logical results provide a formal basis for the previous proof mining application [108] carried out in the context of systems like (†) induced by a certain class of accretive operators and thus remove the ad-hoc nature surrounding it. Even further however, these results are expected to lead to many new case studies for proof mining in the context of that theory and we will see four particular examples of such case studies in the upcoming Chapters 5, 6 and 7.

4.2 Nonlinear semigroups and the Crandall-Liggett formula

The main objects of concern in this chapter are the aforementioned nonlinear (and in this thesis in particular nonexpansive) semigroups:

Definition 4.2.1. Let C be a closed subset of X. A function $S : [0, \infty) \times C \to C$ is a *(nonexpansive) semigroup on* C if

1. $S(t + s)x = S(t)S(s)x$ for all $x \in C$ and all $t, s \geqslant 0$,
2. $S(0)x = x$ for all $x \in C$,
3. $S(t)x$ is continuous in $t \geqslant 0$ for every $x \in C$,
4. $\|S(t)x - S(t)y\| \leqslant \|x - y\|$ for all $t \geqslant 0$ and all $x, y \in C$.

As discussed in the introduction already, these semigroups frequently arise in the study of differential and evolution equations as is e.g. exemplified by the initial value problem (†). In particular, by the results of Crandall and Liggett [50], the exponential formula discussed before always generates such a semigroup on $\overline{\mathrm{dom}A}$ which will be the main object of study of this chapter. Concretely, the following result was established in [50]:

Theorem 4.2.2 (Crandall and Liggett [50])**.** *Let X be a Banach space and A an accretive operator on X such that there exists a $\lambda_0 > 0$ with*

$$\overline{\mathrm{dom}A} \subseteq \mathrm{ran}(Id + \lambda A) \textit{ for all } \lambda \in (0, \lambda_0].$$

Then

$$S(t)x := \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}A\right)^{-n} x$$

exists for all $x \in \overline{\mathrm{dom}A}$ and $t \geqslant 0$ and $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ is a nonlinear semigroup on $\overline{\mathrm{dom}A}$.

We call $\mathcal{S}$ as defined above the semigroup generated by A (via the exponential or Crandall-Liggett formula).[3]

[3]In fact, a large part of the literature calls $-A$ the generator of $\mathcal{S}$ (see e.g. [4] and the references therein) to emphasize that the generator is dissipative. As we want to emphasize the accretiveness of the operator, we here deviated slightly from this convention.

In terms of a logical treatment of these semigroups generated by an accretive operator, all of the later logical considerations naturally depend on the underlying theory of accretive operators over Banach spaces. In that vein, we crucially rely on the basic systems introduced in Chapter 3 for the treatment of those accretive operators. However, these systems need to be extended in order to adequately deal with Theorem 4.2.2 and the associated notions. In particular, we need to provide logical treatments of an alternative notion of accretivity, an extended range condition and the quantification over elements from the closure of the domain of A. We begin with the first of these in the following section.

4.3 The normalized duality map and the alternative notion of accretivity

4.3.1 The duality map and selection functionals

Recall that for a Banach space X with its dual space

$$X^* := \{x^* : X \to \mathbb{R} \mid x^* \text{ linear and continuous}\},$$

its normalized duality mapping

$$J : X \to 2^{X^*},\ x \mapsto \left\{x^* \in X^* \mid \langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2\right\}$$

is non-empty for any $x \in X$ (which follows from the Hahn-Banach theorem). Many works in the context of the theory of accretive operators in general, and the treatment of semigroups generated by those operators in particular, rely on the use of this mapping and in that way, this section is concerned with a proof-theoretic treatment thereof.

As we for now want to refrain from providing a treatment for both the operator norm on the dual space as well as for the full duality map as a set-valued mapping, we follow the approach initiated by Kohlenbach and Leuştean in [111] where the authors handle uses of J by only treating certain selection functionals for J (depending on the situation at hand).

Concretely, a selection functional for the duality map J is just a map $j : X \to X^*$ such that $j(x) \in J(x)$ for any $x \in X$. This general property of being a selection map can then be expressed by corresponding axioms formalizing that

1. $jx : X \to \mathbb{R}$ is a linear operator for any $x \in X$;

2. $\|jx\| \leqslant \|x\|$ where $\|jx\|$ means the operator norm;

3. $jxx = \|x\|^2$ (which, as discussed in [111] already, yields $\|jx\| = \|x\|$).

Given a constant j of type $1(X)(X)$, this can be formally encapsulated by the following universal axiom introduced in [111]:

$$\forall x^X, y^X \Big(jxx =_{\mathbb{R}} \|x\|_X^2 \wedge |jxy| \leqslant_{\mathbb{R}} \|x\|_X \|y\|_X \\ \wedge \forall \alpha^1, \beta^1, u^X, v^X \left(jx(\alpha u +_X \beta v) =_{\mathbb{R}} \alpha jxu +_{\mathbb{R}} \beta jxv \right) \Big),$$

Notice that the operator norm is here avoided by expressing $\|jx\| \leqslant \|x\|$ via stipulating $|jxy|_{\mathbb{R}} \leqslant_{\mathbb{R}} \|x\|_X \|y\|_X$.

Remark 4.3.1. As discussed in [111], the functional j is not provably extensional from the above axiom alone. As indicated by the use of the Dialectica interpretation, if extensionality is to be treated then one has to stipulate an associated modulus of uniform continuity which has been considered in [111]. As not all applications discussed later do require an extensional or continuous selection map, we do not explicitly discuss this issue in this chapter and instead refer to Chapter 5 for a further discussion.

4.3.2 The alternative notion of accretivity

Besides the purely metric notion of accretivity discussed in the preceding Chapter 3, which also forms the basis of the systems $\mathcal{V}^\omega_p$ and its intuitionistic variant $\mathcal{V}^\omega_{i,p}$, the more common notion of accretivity, especially in the context of nonlinear semigroups generated by such operators, is the notion introduced by Kato in [84] where one stipulates that A is accretive if

$$\forall (x,u),(y,v) \in A \exists j \in J(x-y) \left(\langle u-v, j \rangle \geqslant 0 \right).$$

In the language of the preceding subsection, this can be recognized as stipulating the existence of a family of selection functionals $j_{u,v}$ such that, as before, $j_{u,v}x \in J(x)$ and where now further $\langle u-v, j_{u,v}(x-y)\rangle \geqslant 0$ for any $u \in Ax$ and $v \in Ay$.

Formally, this leads us to the following modification of the previous system: we define $\widehat{\mathcal{V}}^\omega_p$ as the extension of $\mathcal{A}^\omega[X, \|\cdot\|]$ with the axiom schemes (I), (II), (IV) and

(V) as defined in Chapter 3, now over the language extended with a constant j of type $1(X)(X)(X)(X)$ together with the axioms

$$\forall x^X, y^X, u^X, v^X \Big(\langle x, j_{u,v}x \rangle =_{\mathbb{R}} \|x\|_X^2 \wedge |\langle y, j_{u,v}x\rangle| \leqslant_{\mathbb{R}} \|x\|_X \, \|y\|_X \\ \wedge \forall \alpha^1, \beta^1, z^X, w^X \left(\langle \alpha z +_X \beta w, j_{u,v}x\rangle =_{\mathbb{R}} \alpha \langle z, j_{u,v}x\rangle + \beta \langle w, j_{u,v}x\rangle \right) \Big) \tag{J}$$

as well as

$$\forall x^X, y^X, u^X, v^X \left(u \in Ax \wedge v \in Ay \to \langle u -_X v, j_{u,v}(x -_X y)\rangle \geqslant_{\mathbb{R}} 0 \right) \tag{A}$$

where we write $j_{u,v}$ for juv as well as $\langle y, j_{u,v}x\rangle$ for $juvxy$.

It is rather immediately clear through the considerations made in [111] that the bound extraction theorems contained in Theorem 3.7.9 and 3.7.16 extend to the system $\widehat{\mathcal{V}}^\omega_p$ as we will discuss now. For this, we first have to give a suitable interpretation to the constant j in the model $\mathcal{M}^{\omega,X}$ associated with an accretive operator A as discussed in Chapter 3. For that, note that the function j is defined by contracting the two parameters besides u, v, namely x and y, into the one argument of j (which is feasible as the witnessing functionals required by the notion of accretivity only have to satisfy $j \in J(x - y)$). The interpretation of this constant in the model now has to "unwind" this contraction (which essentially relies on a choice principle). Concretely, we are lead to the following interpretation of j (writing $\mathcal{M}$ concisely for $\mathcal{M}^{\omega,X}$): given an accretive operator $A \subseteq X \times X$, define $[j]_{\mathcal{M}}$ by

$$[j]_{\mathcal{M}}(u, v, z, w) = \begin{cases} (\langle w, j^A_{u,v}(z)\rangle)_\circ & \text{if } \exists x, y \in X \, (u \in Ax \wedge v \in Ay \wedge z =_X x -_X y), \\ (\langle w, \widetilde{j}(z)\rangle)_\circ & \text{otherwise,} \end{cases}$$

where $\langle \cdot, \cdot \rangle$ is application in the space X^*, the functionals $j^A_{u,v}(z) \in J(z)$ are those guaranteed to exist by the definition of accretivity (if such x, y exist), $\widetilde{j}(z)$ is a generic element of $J(z)$ (which always exists as $J(z) \neq \varnothing$ by the Hahn-Banach theorem) and $(\cdot)_\circ$ is defined as in Chapter 2 on all of $\mathbb{R}$. With this interpretation, the previous axioms are naturally satisfied in the model $\mathcal{M}^{\omega,X}$ associated with an accretive operator A.

The Theorems 3.7.9 and 3.7.16 now extend to this setting as all the additional axioms (J) and (A) are purely universal and since the additional constant j with its interpretation in the model $\mathcal{M}^{\omega,X}$ can be majorized by following the ideas presented in the proof of Theorem 2.2 in [111]: from $|\langle y, j_{u,v}x\rangle| \leqslant \|x\| \, \|y\|$, one obtains that $nm \geqslant |\langle y, j_{u,v}x\rangle|$ for $n \geqslant \|x\|$ and $m \geqslant \|y\|$ which immediately yields that the function

$$(n, m, l, k) \mapsto (mn)_\circ$$

defined for $n, m, k, l \in \mathbb{N}$ with $\|u\| \leqslant k$, $\|v\| \leqslant l$, $\|z\| \leqslant m$, $\|w\| \leqslant n$ is a majorant for j. Note that as discussed in Chapter 2, $\circ$ if restricted to $\mathbb{N}$ can be explicitly given by a term. This majorant is in particular actually independent on the arguments induced by the upper bounds on $\|u\|$ and $\|v\|$, i.e. k and l.

The question of how this notion of accretivity relates to the previously used notion immediately arises. By formalizing one direction of the proof on the equivalence of the two notions of accretivity (essentially due to Kato [84], see also Lemma 3.1 in Chapter II of [4]), we obtain the following:

Proposition 4.3.2. *The system* $\widehat{\mathcal{V}}^{\omega}_p$ *proves:*

1.
$$\forall x^X, y^X, u^X, v^X \left(\langle y, j_{u,v}x\rangle \geqslant_{\mathbb{R}} 0 \to \forall \lambda^1 \left(\|x\|_X \leqslant_{\mathbb{R}} \|x +_X |\lambda| y\|_X\right)\right).$$

2.
$$\forall x^X, y^X, u^X, v^X, \lambda^1 \big((x,u),(y,v) \in A \to \|x -_X y +_X |\lambda|(u -_X v)\|_X \geqslant_{\mathbb{R}} \|x -_X y\|_X\big).$$

Proof. 1. The conclusion is vacuously true for $x = 0$. Thus assume $x \neq 0$ and let $\langle y, j_{u,v}x\rangle \geqslant 0$. Then we get

$$\begin{aligned}
\|x\|^2 &= \langle x, j_{u,v}x\rangle && \text{by } (J)\\
&= \langle x + |\lambda|y - |\lambda|y, j_{u,v}x\rangle && \text{by (qf} - \text{ER)}\\
&= \langle x + |\lambda|y, j_{u,v}x\rangle - |\lambda|\langle y, j_{u,v}x\rangle && \text{by } (J)\\
&\leqslant \langle x + |\lambda|y, j_{u,v}x\rangle \leqslant \|x + |\lambda|y\| \, \|x\| && \text{by } (J).
\end{aligned}$$

Then $\|x\| \leqslant \|x + |\lambda|y\|$ after dividing by $\|x\|$.

2. By using (A), we have $\langle u - v, j_{u,v}(x - y)\rangle \geqslant 0$ for $u \in Ax$ and $v \in Ay$. Then, we get $\|x - y\| \leqslant \|x - y + |\lambda|(u - v)\|$ by (1).

□

Therefore, the system $\widehat{\mathcal{V}}^{\omega}_p$ is an extension of $\mathcal{V}^{\omega}_p$ as all the axioms of $\mathcal{V}^{\omega}_p$ are provable in $\widehat{\mathcal{V}}^{\omega}_p$. In particular, all properties of A and its resolvent exhibited in Proposition 3.3.3 are provable in $\widehat{\mathcal{V}}^{\omega}_p$. Further, the system proves most of the basic facts about such duality selection mappings. One such fact that will be particularly useful later on is the following (proved – in passing – e.g. in the proof of Proposition 1.1 in Chapter I of [4]):

Proposition 4.3.3. *The system* $\widehat{\mathcal{V}}^{\omega}_p$ *proves:*

$$\forall x^X, y^X, u^X, v^X, t^1 \left(t >_{\mathbb{R}} 0 \to \langle y, j_{u,v}x \rangle \leqslant_{\mathbb{R}} \|x\|_X \frac{\|x +_X ty\|_X - \|x\|_X}{t} \right).$$

Proof. We have

$$\|x\|^2 + t\langle y, j_{u,v}x\rangle = \langle x + ty, j_{u,v}x\rangle \leqslant \|x\| \, \|x + ty\|$$

by axiom (J). This implies

$$\langle y, j_{u,v}x\rangle \leqslant \|x\| \frac{\|x + ty\| - \|x\|}{t}.$$

□

4.3.3 The mapping $\langle \cdot, \cdot \rangle_s$

Of crucial importance in the context of many proofs from the theory of nonlinear semigroups, and in particular in the context of the exemplary applications considered later in Chapter 7, is the use of a function $\langle \cdot, \cdot \rangle_s : X \times X \to \overline{\mathbb{R}}$ defined by

$$\langle y, x\rangle_s := \sup \{\langle y, j\rangle \mid j \in J(x)\}.$$

As already observed in the early papers [23, 50], it is easy to see that $\langle y, x\rangle_s < +\infty$ for all $x, y \in X$ and in fact, since $J(x)$ is weak-star compact in X^*, the supremum is actually attained.

While $\langle \cdot, \cdot \rangle_s$ is by virtue of its definition via the supremum and the duality map J a complex object, many proofs only rely on the existence of a mapping which shares some essential properties with $\langle \cdot, \cdot \rangle_s$ and in that case, such a mapping can indeed be treated in the context of the systems discussed above and this is what we want to briefly discuss in the following.

Concretely, under the "essential properties" mentioned above we will understand the following:

1. $\langle \alpha y, \beta x\rangle_s = \alpha\beta\langle y, x\rangle_s$ for $x, y \in X$ and $\alpha, \beta \geqslant 0$;

2. $\langle \alpha x + y, x\rangle_s = \alpha \|x\|^2 + \langle y, x\rangle_s$ for $x, y \in X$ and $\alpha \in \mathbb{R}$;

3. $|\langle y, x\rangle_s| \leqslant \|y\| \, \|x\|$ for $x, y \in X$;

4. $\langle y, j_{u,v}x\rangle \leqslant \langle y, x\rangle_s$ for $x, y \in X$ and $u, v \in X$ where the $j_{u,v}$ are the selection functionals for J guaranteed by accretivity;

5. $\langle\cdot,\cdot\rangle_s$ is upper-semicontinuous (in its right argument).

For a proof for the items (1), (2) and (5), see Proposition 1.2 in Chapter I of [4]. The other items are immediate.

If all that is required of $\langle\cdot,\cdot\rangle_s$ in a proof is that it fulfills these properties, then this proof can, under suitable uniformization of these assumptions, be treated in the context of the above systems by adding a further constant $\langle\cdot,\cdot\rangle_s$ of type $1(X)(X)$ together with the following axioms: the items (1) - (4) are readily formulated as

$(+)_1$ $\forall x^X, y^X, \alpha^1, \beta^1\,(\langle|\alpha|y, |\beta|x\rangle_s =_{\mathbb{R}} |\alpha||\beta|\langle y, x\rangle_s)$,

$(+)_2$ $\forall x^X, y^X, \alpha^1\,\left(\langle \alpha x +_X y, x\rangle_s =_{\mathbb{R}} \alpha\,\|x\|_X^2 + \langle y, x\rangle_s\right)$,

$(+)_3$ $\forall x^X, y^X\,(|\langle y, x\rangle_s| \leqslant_{\mathbb{R}} \|y\|_X\,\|x\|_X)$,

$(+)_4$ $\forall x^X, y^X, u^X, v^X\,(\langle y, j_{u,v}x\rangle \leqslant_{\mathbb{R}} \langle y, x\rangle_s)$,

in the underlying language. For a suitable formulation of item (5), note that the logical methodology based on the monotone Dialectica interpretation suggest that the assumption is upgraded to the existence of a modulus ω^+ of uniform upper-semicontinuity. Concretely, we will consider an additional constant ω^+ of type $0(0)(0)$ together with the axiom

$(+)_5$

$$\forall x^X, y^X, z^X, b^0, k^0\Big(\,\|x\|_X\,,\|z\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\omega^+(b,k)} \\ \to \langle z, y\rangle_s \leqslant_{\mathbb{R}} \langle z, x\rangle_s + 2^{-k}\Big).$$

Note that by the uniformity on x where the rate only depends on the upper bound b, this is actually a full modulus of uniform continuity.

The assumption that $\langle\cdot,\cdot\rangle_s$ is uniformly continuous is in particular true if the space is uniformly smooth and will be in particular also be necessary if the proof to be treated in some form uses the extensionality of the functional $\langle\cdot,\cdot\rangle_s$ (in its right argument) as suggested by the logical methodology. However, if that is not the case and the proof can be formalized just using the axioms $(+)_1, \dots, (+)_4$, then the bound extraction theorem established later in particular guarantees a bound which is valid in all Banach spaces.

Note also that accretivity is sometimes defined by explicitly using the functional $\langle \cdot, \cdot \rangle_s$ through stating that

$$\forall (x, u), (y, v) \in A \left(\langle u -_X v, x -_X y \rangle_s \geqslant_{\mathbb{R}} 0 \right).$$

This version of accretivity is immediately provable in the system $\widehat{\mathcal{V}}^\omega_p + (+)_4$ as, using axioms (A) and $(+)_4$, we have

$$\langle u - v, x - y \rangle_s \geqslant \langle u - v, j_{u,v}(x - y) \rangle \geqslant 0.$$

We later denote the collection of these five axioms $(+)_1$ - $(+)_5$ by $(+)$. Now, the bound extraction results contained in Theorem 3.7.9 and 3.7.16 also extend to the associated extended system(s) $\widehat{\mathcal{V}}^\omega_p + (+)_1 + \cdots + (+)_4 + ((+)_5)$ with the conclusion drawn over any space (or where $\langle \cdot, \cdot \rangle_s$ is additionally uniformly continuous on bounded subsets as above if $(+)_5$ is included). Concretely, this follows as before since, for one, all the axiom schemes are purely universal and, for another, the constant $\langle \cdot, \cdot \rangle_s$ can be immediately majorized: from $|\langle y, x \rangle_s| \leqslant \|y\| \, \|x\|$, we as before infer $mn \geqslant |\langle y, x \rangle_s|$ for $m \geqslant \|y\|$ and $n \geqslant \|x\|$. From this, a majorant for the accompanying interpretation using $(\cdot)_\circ$ in the model $\mathcal{M}^{\omega,X}$ follows by Lemma 2.1.2. Further, the additional constant ω^+ is immediately majorized (essentially by itself) as it is of type $0(0)(0)$ and so, similar to Lemma 17.82 of [96], we have that $\omega^{+,M}$ defined by

$$\omega^{+,M}(b, k) = \max\{\omega^+(a, j) \mid a \leqslant b, j \leqslant k\}$$

is a majorant for ω^+.

4.4 Systems for nonlinear semigroups and bound extraction theorems

In this section, we now are concerned with a formal treatment of the semigroup $\mathcal{S}$ generated by the exponential formula as guaranteed from the result of Crandall and Liggett [50] previously discussed in Theorem 4.2.2. Before diving into the formal treatment of these semigroups, we however need to consider some preliminary formal results for the treatment of $\overline{\mathrm{dom} A}$ (which features in the premise of the range condition in Theorem 4.2.2) as well as how J^A_0 is to be understood.

4.4.1 The treatment of $\overline{\mathrm{dom} A}$

Crucial both for the definition of the semigroup and for the central assumption of Theorem 4.2.2, i.e. the range condition, is the use of the closure of the domain of A and in the following formal investigations, quantification over elements from $\overline{\mathrm{dom} A}$ will therefore be necessary. All the previous systems essentially only considered normed spaces and in that context, we now first have to lift the previous treatment to take the completeness of the underlying Banach space into account. For that, we are following the approach laid out in [96] by which complete spaces are treated by adding another operator C of type $X(X(0))$ which is meant to assign to a Cauchy sequence $x^{X(0)}$ a limit $C(x)$. To discard of the complex premise of Cauchyness in an axiom stating that property, one then restricts oneself to Cauchy sequences with a fixed Cauchy rate (similar to the representation of real numbers in finite type arithmetic discussed in Chapter 2, see again [96]). To implicitly quantify only over all such sequences, a term construction $\widehat{x}$ is used on the objects $x^{X(0)}$. Precisely, $\widehat{x}$ is defined on the level of the representation of the real value of the norm via sequences of rational numbers with fixed Cauchy rate via[4]

$$\widehat{x}_n =_X \begin{cases} x_n & \text{if } \forall k <_0 n \left([\|x_k -_X x_{k+1}\|_X](k+1) <_{\mathbb{Q}} 6 \cdot 2^{-k-1}\right), \\ x_k & \text{for } \min k <_0 n : [\|x_k -_X x_{k+1}\|_X](k+1) \geqslant_{\mathbb{Q}} 6 \cdot 2^{-k-1}, \text{ otherwise.} \end{cases}$$

Then, completeness of the space can be formulated via the universal axiom[5]

$$\forall x^{X(0)}, k^0 \left(\|C(x) -_X \widehat{x}_k\|_X \leqslant_{\mathbb{R}} 2^{-k+3}\right) \tag{$\mathcal{C}$}$$

which indeed implies completeness of the space in the form that from

$$\forall k^0 \exists n^0 \forall m, \widetilde{m} \geqslant_0 n \left(\|x_m -_X x_{\widetilde{m}}\|_X <_{\mathbb{R}} 2^{-k}\right)$$

it follows provably in $\mathcal{A}^\omega[X, \|\cdot\|] + (\mathcal{C})$ that

$$\forall k^0 \exists m^0 \forall l \geqslant_0 m \left(\|C(x) -_X x_l\|_X <_{\mathbb{R}} 2^{-k+1}\right).$$

As further shown in [96], the constant C is majorizable and therefore we find that the bound extraction theorems discussed above immediately extend to $\widehat{\mathcal{V}}^\omega_p + (\mathcal{C})$ or any suitable extension (e.g. by (+)).

---

[4] As discussed already in Chapter 2, we here follow the notion of [96] and denote by $[a](k)$ the k-th element of the Cauchy sequence representation of the real number a.

[5] See the discussion in [96] for the necessity of the additional +3 in the formulation.

Now a statement where one is quantifying over the closure of the domain, i.e. a statement of the form

$$\forall x \in \overline{\mathrm{dom}A}\; B(x) \tag{$*$}$$

can, through the use of C, be (naively) expressed as

$$\forall x^{X(0)} \left(\forall n^0 \exists y^X (y \in A\widehat{x}_n) \to B(C(x))\right).$$

The premise that $x^{X(0)}$ is a Cauchy sequence was removed through the use of $\widehat{x}$ and C but the inclusion of the sequence in the domain, in the form of $\forall n^0 \exists y^X (y \in A\widehat{x}_n)$, remains.

The approach is now to also remove this assumption in a similar style as the $\widehat{\cdot}$-operation by universally quantifying over the potential witnessing sequence y_n and defining a subsequent operation similar to $\widehat{\cdot}$ which potentially alters the sequence such that $x_n \in \mathrm{dom}A$ will always be guaranteed for any n. Concretely, for two objects x, y of type $X(0)$, we define

$$(x \upharpoonright y)_n =_X \begin{cases} x_n & \text{if } \forall k \leqslant_0 n \, (y_k \in Ax_k)\,, \\ x_{k \dot{-} 1} & \text{for } \min k \leqslant_0 n : y_k \notin Ax_k, \text{ otherwise.} \end{cases}$$

Note that since inclusions in the graph of A are quantifier-free, the above indeed can be defined by a closed term in the underlying language.

Now, using the operation $\upharpoonright$ in tandem with $\widehat{\cdot}$, we can implicitly quantify over elements from $\overline{\mathrm{dom}A}$ by quantifying over elements of type $X(0)$ and thus we can express the statement $(*)$ equivalently by

$$\forall x^{X(0)}, y^{X(0)} \, (y_0 \in Ax_0 \to B(C(x \upharpoonright y))).$$

As a feasibility check for using $x \upharpoonright y$, note first that

$$\widehat{x \upharpoonright y} =_{X(0)} \widehat{x} \upharpoonright y.$$

To see this, one can consider a case distinction on whether $\widehat{x} = x$ holds or not and simultaneously on whether $x \upharpoonright y = x$ holds or not. We only consider the one case out of the four where $\widehat{x} \neq x$ and $x \upharpoonright y \neq x$. By definition, we then have a least k such that $[\|x_k - x_{k+1}\|](k+1) \geqslant_{\mathbb{Q}} 6 \cdot 2^{-k-1}$ as well as a least j such that $y_j \notin Ax_j$. Then, it immediately follows by definition of the operations as well as the minimality of k and

j that

$$\begin{aligned}\widehat{x} \upharpoonright y &= (x_0, \dots, x_k, x_k, \dots) \upharpoonright y \\ &= (x_0, \dots, x_{\min\{k, j \dot{-} 1\}}, x_{\min\{k, j \dot{-} 1\}}, \dots) \\ &= (x_0, \dots, x_{j \dot{-} 1}, x_{j \dot{-} 1}, \dots)\widehat{\ } \\ &= \widehat{x \upharpoonright y}\end{aligned}$$

where, in the third line, we wrote $(x_0, \dots, x_{j \dot{-} 1}, x_{j \dot{-} 1}, \dots)\widehat{\ }$ for the operation $\widehat{\cdot}$ applied to the sequence $(x_0, \dots, x_{j \dot{-} 1}, x_{j \dot{-} 1}, \dots)$.

Further, note that the premise $y_0 \in Ax_0$ actually guarantees that $(\widehat{x} \upharpoonright y)_n \in \mathrm{dom}A$ for all n. For this, define

$$(x \upharpoonleft y)_n =_X \begin{cases} y_n & \text{if } \forall k \leqslant_0 n \, (y_k \in Ax_k) \, , \\ y_{k \dot{-} 1} & \text{for } \min k \leqslant_0 n : y_k \notin Ax_k, \text{ otherwise.} \end{cases}$$

Then clearly $y_0 \in Ax_0$ implies $(\widehat{x} \upharpoonleft y)_n \in A((\widehat{x} \upharpoonright y)_n)$ for any n.

4.4.2 Range conditions

A treatment for the canonical variant of a range condition

$$\mathrm{dom}A \subseteq \bigcap_{\lambda > 0} \mathrm{ran}(Id + \lambda A)$$

was already briefly discussed in Chapter 3 where a formal version admissible in the systems for bound extractions was presented with

$$\forall x^X, \lambda^1 \left(x \in \mathrm{dom}A \wedge \lambda >_{\mathbb{R}} 0 \rightarrow \lambda^{-1}(x -_X J^A_\lambda x) \in A(J^A_\lambda x) \right).$$

Recall that this correctly expresses the range condition since stating that $x \in \mathrm{ran}(Id + \lambda A)$ is equivalent to stating that $x \in \mathrm{dom}(J^A_\lambda)$ just via the definition of the resolvent. This latter statement is now equivalently formally encapsulated in our systems by stating the inclusion $\lambda^{-1}(x -_X J^A_\lambda x) \in A(J^A_\lambda x)$. Note also that this axiom is in particular purely universal and thus can be used in the bound extraction theorems.

In the following, we want to consider two modifications: (1) we want to specify that the inclusion is valid even for the closure of the domain; (2) we want to restrict the intersection to $\lambda < \lambda_0$ for some real parameter $\lambda_0 > 0$. The use of such a λ_0 can be facilitated by adding two further constants and an axiom: λ_0 of type 1 and m_{λ_0} of type

0 together with the accompanying axiom $\lambda_0 \geqslant_{\mathbb{R}} 2^{-m_{\lambda_0}}$ providing a verifier to $\lambda_0 > 0$. Note that the bound extraction results stay valid in the context of such an extension if one additionally requires the parameter n from Theorem 3.7.9 to satisfy $n \geqslant |\lambda_0|, m_{\lambda_0}$.

In the context of such additional constants, the above range condition can be immediately modified to represent the restricted range condition

$$\mathrm{dom}A \subseteq \bigcap_{\lambda_0 > \lambda > 0} \mathrm{ran}(Id + \gamma A)$$

by considering

$$\forall x^X, \lambda^1 \left(x \in \mathrm{dom}A \wedge \lambda_0 >_{\mathbb{R}} \lambda >_{\mathbb{R}} 0 \to \lambda^{-1}(x -_X J^A_\lambda x) \in A(J^A_\lambda x) \right).$$

Further, in both cases we can now consider the other main modification of stipulating the range condition also for the closure of the domain, i.e.

$$\overline{\mathrm{dom}A} \subseteq \bigcap_{\lambda_0 > \lambda > 0} \mathrm{ran}(Id + \gamma A),$$

by using the above treatment of quantification over elements in the closure of the domain by quantification over sequences in X together with the operators C and $(\cdot \upharpoonright \cdot)$. Concretely, one rather immediately obtains the following natural extension to the closure of the domain:

$$\forall x^{X(0)}, v^{X(0)}, \lambda^1 \Big(v_0 \in Ax_0 \wedge \lambda_0 >_{\mathbb{R}} \lambda >_{\mathbb{R}} 0 \to \lambda^{-1}(C(x \upharpoonright v) -_X J^A_\lambda(C(x \upharpoonright v))) \in A(J^A_\lambda(C(x \upharpoonright v))) \Big). \qquad (RC)_{\lambda_0}$$

Similarly, we could here lift the restriction via λ_0 again and get a full range condition for the closure of the domain. We denoted this full range condition for the closure of the domain by (RC), but at the same time refrain from spelling this out in any more detail here. Note however that all the other range conditions introduced here are still purely universal and thus are admissible in the context of the bound extraction theorems.

Further, note that e.g. from $(RC)_{\lambda_0}$, the statement

$$\forall x^X, \lambda^1 \left(x \in \mathrm{dom}A \wedge \lambda_0 >_{\mathbb{R}} \lambda >_{\mathbb{R}} 0 \to \lambda^{-1}(x -_X J^A_\lambda x) \in A(J^A_\lambda x) \right)$$

is provable: if $x \in \mathrm{dom}A$ with $v \in Ax$, consider the constant-x and constant-v sequences $\overline{x}$ and $\overline{v}$, respectively. Then clearly $(\overline{x} \upharpoonright \overline{v})_n =_X x$ for any n and thus provably

$C(\overline{x} \upharpoonright \overline{v}) =_X x$ by $(\mathcal{C})$. The statement $(RC)_{\lambda_0}$ yields

$$\lambda^{-1}(C(\overline{x} \upharpoonright \overline{v}) -_X J^A_\lambda(C(\overline{x} \upharpoonright \overline{v}))) \in A(J^A_\lambda(C(\overline{x} \upharpoonright \overline{v})))$$

for $\lambda_0 > \lambda > 0$ and the quantifier-free extensionality rule (as $v \in Ax$ is quantifier-free) yields $\lambda^{-1}(x -_X J^A_\lambda x) \in A(J^A_\lambda x)$.

In the following remark, we lastly collect some subtleties regarding the extension of the metatheorems to systems with these types of axioms.

Remark 4.4.1. The metatheorems exhibited in Theorems 3.7.9 and 3.7.16 require as an assumption that $\bigcap_{\lambda>0} \mathrm{dom} J^A_\lambda \neq \varnothing$, a requirement which would be substantiated via a full range condition together with a witness for $\mathrm{dom} A \neq \varnothing$ (which was previously – in some sense but not precisely – represented by c_X). In the context of the above restricted range conditions, it is however feasible that $\bigcap_{\lambda>0} \mathrm{dom} J^A_\lambda$ is actually empty while only $\bigcap_{\lambda_0>\lambda>0} \mathrm{dom} J^A_\lambda \neq \varnothing$ holds. It should be noted that in this case, Theorems 3.7.9 and 3.7.16 can be modified to stay valid if c_X is interpreted by a point in this restricted intersection. Therefore, if we in the following write $\widehat{\mathcal{V}}^\omega_p + (\mathcal{C}) + (RC)_{\lambda_0}$ or consider any extension, we consider the axioms (IV) and (V) to be replaced by

(IV)′ $\lambda_0 - 2^{m'_{\widetilde{\gamma}}} \geqslant_{\mathbb{R}} \widetilde{\gamma} \geqslant_{\mathbb{R}} 2^{-m_{\widetilde{\gamma}}}$,

(V)′ $d_X \in Ac_X$,

where d_X is a new constant of type X and $m'_{\widetilde{\gamma}}$ is a new constant of type 0, the latter witnessing that $\lambda_0 > \widetilde{\gamma}$. The majorization of all resolvents J^A_γ for $\gamma \in (0, \lambda_0)$ is then achieved similar to before via

$$\begin{aligned} \left\|J^A_\gamma x\right\| &\leqslant \|x\| + 2\|c_X\| + \left(2 + \frac{\gamma}{\widetilde{\gamma}}\right)\left\|c_X - J^A_{\widetilde{\gamma}} c_X\right\| \\ &\leqslant \|x\| + 2\|c_X\| + (2\widetilde{\gamma} + \gamma)\|d_X\|. \end{aligned}$$

In that case however, the interpretation of the resolvent constant J^{χ_A} in the models $\mathcal{M}^{\omega,X}$ and $\mathcal{S}^{\omega,X}$ has to be modified to set $[J^{\chi_A}]_{\mathcal{M}}(x,\gamma) = 0$ for all x if $\gamma \geqslant_{\mathbb{R}} \lambda_0$ (and similar for $\mathcal{S}^{\omega,X}$). Therefore, the extracted bounds only remain meaningful if the theorem does not utilize these resolvents. If it does, further modifications are necessary but we refrain from discussing this here any further as this situation does not arise in this chapter or even in this thesis for that matter.

4.4.3 The resolvent at zero

Something left open by the axioms characterizing the resolvent discussed in Chapter 3 is the behavior of J^A_0. This, however, takes a special role in the context of the treatment of nonlinear semigroups S generated by the associated operator A due to the prominent use often made of $S(0)$.

The reason for this previous ambiguity in the treatment of the resolvent at 0 was the fact that the resolvent does not always behave continuously at 0 if it is naively defined: while the definition of the resolvent via

$$J^A_\gamma = (\mathrm{Id} + \gamma A)^{-1}$$

suggests $J^A_0 x = x$, it is well known (see [11]) that already in Hilbert spaces with a maximally monotone operator A, one has $J^A_t x \to P_{\overline{\mathrm{dom}A}} x$ for $t \to 0$ and all $x \in \mathrm{dom}(J^A_t)$. Therefore, extensionality for the constant J^{χ_A} in its first argument t at 0 can in general not be expected if J^A_0 is defined in this way and the previous axiomatization left the definition of J^A_0 open.

In the following, we nevertheless consider the set of axioms discussed previously forming $\widehat{\mathcal{V}}^\omega_p$ to actually be extended with the sixth axiom

(VI) $\forall x^X \left(J^A_0 x =_X x\right)$,

stating the defining equality $J^A_0 = (\mathrm{Id} + 0A)^{-1} = \mathrm{Id}$.

Now, the above result that $J^A_t x \to P_{\overline{\mathrm{dom}A}} x$ for $t \to 0$ extends to Banach spaces at least partially in the sense that one can show (see Proposition 3.2 of Chapter II in [4]) that $J^A_t x \to x$ for $\lambda_0 > t \to 0$ and

$$x \in \overline{\mathrm{dom}A} \cap \bigcap_{\lambda_0 > \lambda > 0} \mathrm{dom} J^A_\lambda.$$

Therefore, in the presence of a range condition, we should at least have a continuous and thus extensional behavior of the resolvent defined in this manner at $t = 0$ for all $x \in \overline{\mathrm{dom}A}$ and this can indeed be formally verified in the accompanying system.

Lemma 4.4.2. $\widehat{\mathcal{V}}^\omega_p + (\mathcal{C}) + (RC)_{\lambda_0}$ *proves:*

$$\forall x^{X(0)}, v^{X(0)}, \lambda^1, k^0 \Bigg(v_0 \in Ax_0 \wedge 0 <_{\mathbb{R}} \lambda <_{\mathbb{R}} \min\left\{ \frac{2^{-(k+1)}}{\max\{1, \|(\widehat{x} \upharpoonright v)_{k+5}\|_X\}}, \lambda_0 \right\} \to \left\|C(x \upharpoonright v) -_X J^A_\lambda C(x \upharpoonright v)\right\|_X \leqslant_{\mathbb{R}} 2^{-k} \Bigg).$$

Proof. First, by Proposition 3.3.3, we have

$$\forall x^X, v^X, \lambda^1 \left(0 <_{\mathbb{R}} \lambda <_{\mathbb{R}} \lambda_0 \wedge v \in Ax \to \left\|x - J^A_\lambda x\right\| \leqslant \lambda \|v\|\right)$$

as using $(RC)_{\lambda_0}$ and the quantifier-free extensionality rule, we obtain $x \in \mathrm{dom}(J^A_\lambda)$ for all $\lambda \in (0, \lambda_0)$ as discussed before. So, for $x^{X(0)}$ and $v^{X(0)}$ such that $v_0 \in Ax_0$ for all n, we obtain $C(x \upharpoonright v) \in \mathrm{dom}(J^A_\lambda)$ for all $\lambda \in (0, \lambda_0)$ again by $(RC)_{\lambda_0}$. Therefore, using $(\mathcal{C})$ and the nonexpansivity of J^A_λ on its domain:

$$\begin{aligned}\left\|C(x \upharpoonright v) - J^A_\lambda C(x \upharpoonright v)\right\| &\leqslant \|C(x \upharpoonright v) - (\widehat{x} \upharpoonright v)_n\| + \left\|(\widehat{x} \upharpoonright v)_n - J^A_\lambda(\widehat{x} \upharpoonright v)_n\right\| \\ &\quad + \left\|J^A_\lambda(\widehat{x} \upharpoonright v)_n - J^A_\lambda C(x \upharpoonright v)\right\| \\ &\leqslant 2\|C(x \upharpoonright v) - (\widehat{x} \upharpoonright v)_n\| + \left\|(\widehat{x} \upharpoonright v)_n - J^A_\lambda(\widehat{x} \upharpoonright v)_n\right\| \\ &\leqslant 2 \cdot 2^{-n+3} + \lambda \|(\widehat{x} \upharpoonright v)_n\|.\end{aligned}$$

Choosing $n = k + 5$, we get that for $\lambda \leqslant 2^{-(k+1)}/\max\{1, \|(\widehat{x} \upharpoonright v)_{k+5}\|\}$:

$$\left\|C(x \upharpoonright v) - J^A_\lambda C(x \upharpoonright v)\right\| \leqslant 2^{-k}.$$

□

This property will be sufficient in the following as the semigroup operates only on $\overline{\mathrm{dom}A}$.

4.4.4 The semigroup

For treating the semigroup on $\overline{\mathrm{dom}A}$ from Theorem 4.2.2, it is very instructive to first consider the operator S solely on $\mathrm{dom}A$. In that case, we can facilitate a treatment by directly adding a further constant S of type $X(X)(1)$ to the underlying language together with an axiom stating that S on $\mathrm{dom}A$ arises from the Crandall-Liggett formula, i.e. that

$$S(t)x = \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}A\right)^{-n} x$$

for any $x \in \mathrm{dom}A$. This can be achieved by further adding a constant ω^S of type $0(0)(0)(0)$ together with the axiom

$$\begin{aligned}\forall k^0, b^0, T^0, x^X, v^X, t^1 \Bigg(& v \in Ax \wedge \|x\|_X, \|v\|_X <_{\mathbb{R}} b \wedge |t| <_{\mathbb{R}} T \\ & \to \forall n \geqslant_0 \omega^S(k, b, T) \left(|t|/n <_{\mathbb{R}} \lambda_0 \to \left\|S(|t|)x -_X (J^A_{|t|/n})^n x\right\|_X \leqslant_{\mathbb{R}} 2^{-k}\right)\Bigg),\end{aligned} \tag{S1}$$

expressing that ω^S represents a rate of convergence uniform for elements x from bounded subsets $B_b(0) \cap \mathrm{dom}A$ and uniform in t for bounded intervals $[0, T]$ (where we use the absolute value to disperse of the universal premise $t \geqslant 0$). The term $(J^A_{|t|/n})^n$ used here is a shorthand for a term $I(t)(n)(n)$ where $I(t)(m)$ is a closed term of type $X(X)(0)$ defined using the recursors of the underlying language of $\mathcal{A}^\omega[X, \|\cdot\|]$ (recall Chapter 2) via $I(t)(m)(0) = \lambda x.x$ and $I(t)(m)(n+1) = \lambda x.(J^A_{t/m}(I(t)(m)(n)(x)))$.[6] Note also that we in particular treat $S(0)x$ via $J^A_0 x$ by using the absolute value $|t|$ in the above formula to implicitly quantify over non-negative real numbers.

Such a use of a rate of convergence is in particular justified by the fact that the proof given in [50] of the Cauchy-property of the sequence $(J^A_{t/n})^n x$ for given $t > 0$ and $x \in \mathrm{dom}A$ can be immediately recognized to be provable in the system $\widehat{\mathcal{V}}^\omega_{i,p} + (\mathcal{C}) + (RC)_{\lambda_0}$ (naturally defined as $\widehat{\mathcal{V}}^\omega_p + (\mathcal{C}) + (RC)_{\lambda_0}$ just over $\mathcal{A}^\omega_i[X, \|\cdot\|]$ instead of $\mathcal{A}^\omega[X, \|\cdot\|]$). Therefore, the extension of the semi-constructive metatheorem (Theorem 3.7.16) to this system guarantees the existence of a rate of Cauchyness for $(J^A_{t/n})^n x$ and consequently the existence of a modulus ω^S as characterized by the above axiom which can moreover be extracted from the proof given in [50] (which is in fact rather immediate and was essentially already observed in [50]): one can (formally) show that given $x \in \mathrm{dom}A$ with witness $v \in Ax$ and $t \geqslant 0$, we have

$$\left\|(J^A_{t/n})^n x - (J^A_{t/m})^m x\right\| \leqslant 2t \left|\frac{1}{m} - \frac{1}{n}\right|^{1/2} \|v\| .$$

Thus for $T \geqslant t$ and $b \geqslant \|v\|$, we have for a given $\varepsilon > 0$ that for any $m \geqslant n \geqslant \left\lceil \frac{4T^2b^2}{\varepsilon^2} \right\rceil$:

$$\begin{aligned}\left\|(J^A_{t/n})^n x - (J^A_{t/m})^m x\right\| &\leqslant 2Tb \left|\frac{1}{m} - \frac{1}{n}\right|^{1/2} \\ &\leqslant 2Tb \frac{1}{\sqrt{n}} \\ &\leqslant 2Tb \frac{1}{\sqrt{\left\lceil \frac{4T^2b^2}{\varepsilon^2} \right\rceil}} \\ &\leqslant \varepsilon.\end{aligned}$$

Thus the mapping

$$\omega^S(k, b, T) = 2^{2k+2}T^2b^2$$

[6] We consider $I(t)(m)$ to be trivially defined at $m = 0$

is a possible choice for the rate of convergence[7] in the exponential formula as derived from the proof and the upper bound b is here actually even independent of $\|x\|$.

Now, the treatment of the extension of S to $\overline{\mathrm{dom}A}$ is best motivated by considering how it is usually defined in the literature: $S(t)$ as a mapping $\mathrm{dom}A \to X$ is nonexpansive and thus continuous. The object $S(t)x$ for $x \in \overline{\mathrm{dom}A}$ is then defined by considering that as $x \in \overline{\mathrm{dom}A}$, there exists a sequence $x_n \to x$ with $x_n \in \mathrm{dom}A$. By convergence, the sequence x_n is Cauchy and by continuity of $S(t)$, the sequence $S(t)x_n$ is Cauchy as well and thus converges in a Banach space by completeness. Then $S(t)x$ is identified with the limit of that sequence. This crucial use of the completeness of the space prompts us to work in the context of the formal treatment of complete spaces and $\overline{\mathrm{dom}A}$ as discussed before.

In that vein, we now want to provide an axiom classifying the behavior of $S(t)$ for elements of $\overline{\mathrm{dom}A}$ by essentially stating that for any x and any Cauchy sequence $x_n \to x$ with $x_n \in \mathrm{dom}A$, $S(t)x_n$ converges to $S(t)x$. The quantification over all elements of $\overline{\mathrm{dom}A}$ together with their generating sequences can now be achieved as discussed in Section 4.4.1 and in that way, the axiom stating the resulting behavior for $S(t)x$ then takes the form of the following universal axiom[8]

$$\forall x^{X(0)}, y^{X(0)}, t^1 \big(y_0 \in Ax_0 \\ \to \forall n^0 \left(\|S(|t|)(C(x \restriction y)) -_X S(|t|)((\widehat{x} \restriction y)_n)\|_X \leq_{\mathbb{R}} 2^{-n+3} \right) \big). \tag{S2}$$

Note again that the behavior of $S(0)$ is implicitly characterized by the above axioms through the use of $|t|$. We write (S) for $(S1)+(S2)$ as well as H^ω_p for $\widehat{\mathcal{V}}^\omega_p+(\mathcal{C})+(RC)_{\lambda_0}+(S)$ (noting again the additional axioms from Remark 4.4.1 and Section 4.4.3).

Now, the above axioms forming the theory H^ω_p are suitable for formalizing large portions on the theory of nonlinear semigroups as generated by the Crandall-Liggett formula and as a sort of litmus test, we at least provide here sketches of formal proofs in the resulting system of the other main semigroup properties which arise pretty much directly by formalizing the proofs given in [50]. For that, however, some careful consideration for iterations of the semigroup map are required here. Concretely, to

[7]Note that although the function is exponential in k, this is just due to requiring an error of the form 2^{-k}. Abstracting $\varepsilon = 2^{-k}$, the rate is actually linear in $1/\varepsilon$.

[8]Note again that the additional $+3$ is included here as the axiom $(\mathcal{C})$ requires this modification in order to have a model as discussed before and the same rate applies to the semigroup-images here as the semigroup is nonexpansive.

make expressions like $S(t)S(s)x$ meaningful, we have to consider how $S(s)x \in \overline{\mathrm{dom}A}$ is reflected in the system. Based on the representation of $\overline{\mathrm{dom}A}$ chosen above (which also features in how the extension of S is formally defined by means of the axiom (S2)) we thus first have to see how $S(|t|)C(x)$ with $x_n \in \mathrm{dom}A$ for all n can be expressed as an element of the form $C(u)$ for $u^{X(0)}$ such that $u_n \in \mathrm{dom}A$ for all n. To find such a u, note first that the convergence result encoded by (S1) for elements from $\mathrm{dom}A$ extends by means of (S2) to $\overline{\mathrm{dom}A}$ in the following way: provably in H^ω_p, we have

$$\forall x^{X(0)}, y^{X(0)}, t^1, k^0 \exists N^0 \forall n \geqslant_0 N \Big(y_0 \in Ax_0 \wedge |t|/n <_{\mathbb{R}} \lambda_0 \\ \rightarrow \left\| S(|t|)(C(x \upharpoonright y)) -_X \left(J^A_{|t|/n}\right)^n (C(x \upharpoonright y)) \right\|_X \leqslant_{\mathbb{R}} 2^{-k} \Big)$$

were moreover (although we avoid spelling this out here) the choice functional for N can be explicitly given by closed terms build up from ω^S (and the other constants). To see the provability of the above statement, let k, x, y, t be arbitrary with $y_0 \in Ax_0$. Then using nonexpansivity of the semigroup and the resolvent (see item (4) of the following Lemma 4.4.3[9]), we have

$$\begin{aligned}
&\left\| S(|t|)(C(x \upharpoonright y)) - \left(J^A_{|t|/n}\right)^n (C(\widehat{x} \upharpoonright y)) \right\| \\
&\quad \leqslant \left\| S(|t|)(C(\widehat{x} \upharpoonright y)) - S(|t|)((\widehat{x} \upharpoonright y)_{(k+5)}) \right\| \\
&\qquad + \left\| S(|t|)((\widehat{x} \upharpoonright y)_{(k+5)}) - \left(J^A_{|t|/n}\right)^n ((\widehat{x} \upharpoonright y)_{(k+5)}) \right\| \\
&\qquad + \left\| \left(J^A_{|t|/n}\right)^n ((\widehat{x} \upharpoonright y)_{(k+5)}) - \left(J^A_{|t|/n}\right)^n (C(x \upharpoonright y)) \right\| \\
&\quad \leqslant \left\| C(x \upharpoonright y) - (\widehat{x} \upharpoonright y)_{(k+5)} \right\| \\
&\qquad + \left\| S(|t|)((\widehat{x} \upharpoonright y)_{(k+5)}) - \left(J^A_{|t|/n}\right)^n ((\widehat{x} \upharpoonright y)_{(k+5)}) \right\| \\
&\qquad + \left\| (\widehat{x} \upharpoonright y)_{(k+5)} - C(x \upharpoonright y) \right\| \\
&\quad \leqslant 2^{-k-1} + \left\| S(|t|)((\widehat{x} \upharpoonright y)_{(k+5)}) - \left(J^A_{|t|/n}\right)^n ((\widehat{x} \upharpoonright y)_{(k+5)}) \right\| \\
&\quad \leqslant 2^{-k}
\end{aligned}$$

for any n large enough such that $|t|/n < \lambda_0$ as well as

$$\left\| S(|t|)((\widehat{x} \upharpoonright y)_{(k+5)}) - \left(J^A_{|t|/n}\right)^n ((\widehat{x} \upharpoonright y)_{(k+5)}) \right\| \leqslant 2^{-(k+1)}$$

which can be achieved via (S1). In that way, writing $N_{t,x,y}$ for the choice functionals for the quantifier over N in the above statement, we find that $S(|t|)C(x \upharpoonright y)$ is provably

[9]The first four items of this lemma in particular do not rely on this construction as it will only become necessary in the fifth item. Thus, there is no circularity induced by this construction.

$=_X$-equal to

$$C\left(\left(\left(J^A_{|t|/N_{t,x,y}(k)}\right)^{N_{t,x,y}(k)} C(x \upharpoonright y)\right)_k\right).$$

We write $\overline{S(|t|)C(x \upharpoonright y)}$ in the following for this expression (where one should note again that the N-functionals can be explicitly computed, albeit being somewhat messy). In particular note that

$$\left(J^A_{|t|/N_{t,x,y}(k)}\right)^{N_{t,x,y}(k)} C(x \upharpoonright y) \in \mathrm{dom}A$$

with the witnessing terms defined in terms of the Yosida approximates (which follows provably from $(\mathrm{RC})_{\lambda_0}$ if we w.l.o.g. assume that the functionals N, for a given t as a parameter, are large enough such that $|t|/N_{t,x,y}(k) < \lambda_0$). In that way, $S(|t|)S(|s|)C(x \upharpoonright y)$ can be meaningfully represented by

$$S(|t|)\overline{S(|s|)C(x \upharpoonright y)} =_X \overline{S(|t|)\overline{S(|s|)C(x \upharpoonright y)}}.$$

Note that the system can nevertheless *not* prove that

$$S(|t|)S(|s|)C(x \upharpoonright y) =_X S(|t|)\overline{S(|s|)C(x \upharpoonright y)}$$

and so the latter is, in some sense, the only way to talk about iterations meaningfully.

We now get to the main properties of nonexpansive semigroups:

Lemma 4.4.3. *The following are provable in H^ω_p:*

1. $\forall x^X, y^X, t^1, s^1\, (y \in Ax \to \|S(|t|)x -_X S(|s|)x\|_X \leqslant_{\mathbb{R}} 2||t| - |s||\, \|y\|_X)$.

2. $$\begin{cases} \forall x^X, y^X, t^1 \big(x \in \mathrm{dom}A \wedge y \in \mathrm{dom}A \\ \qquad \to \|S(|t|)x -_X S(|t|)y\|_X \leqslant_{\mathbb{R}} \|x -_X y\|_X\big). \end{cases}$$

3. $$\begin{cases} \forall x^{X(0)}, v^{X(0)}, t^1, s^1 \Big(v_0 \in Ax_0 \wedge ||t| - |s|| \leqslant_{\mathbb{R}} 2^{-(k+2)}/\max\{1, \|(\widehat{x} \upharpoonright v)_{k+5}\|\} \\ \qquad \to \|S(|t|)C(x \upharpoonright v) -_X S(|s|)C(x \upharpoonright v)\|_X \leqslant_{\mathbb{R}} 2^{-k}\Big). \end{cases}$$

4. $$\begin{cases} \forall x^{X(0)}, v^{X(0)}, y^{X(0)}, w^{X(0)}, t^1 \Big(v_0 \in Ax_0 \wedge w_0 \in Ay_0 \\ \qquad \to \|S(|t|)(C(x \upharpoonright v)) -_X S(|t|)(C(y \upharpoonright w))\|_X \\ \qquad\qquad \leqslant_{\mathbb{R}} \|C(x \upharpoonright v) -_X C(y \upharpoonright w)\|_X\Big). \end{cases}$$

5. $$\begin{cases} \forall x^{X(0)}, v^{X(0)}, t^1, s^1 \Big(v_0 \in Ax_0 \\ \qquad \to S(|t| + |s|)(C(x \upharpoonright v)) =_X S(|t|)\overline{S(|s|)(C(x \upharpoonright v))}\Big). \end{cases}$$

Proof. 1. At first, note that provably in H_p^ω, we have

$$\begin{cases} \forall x, y, \mu, \lambda, n, m\Big(\lambda_0 > |\lambda| \geqslant |\mu| \wedge n \geqslant m \geqslant 1 \wedge y \in Ax \\ \to \left\|(J^A_{|\mu|})^n x - (J^A_{|\lambda|})^m x\right\| \leqslant \Big(((n|\mu| - m|\lambda|)^2 + n|\mu|(|\lambda| - |\mu|))^{1/2} \\ \qquad + (m|\lambda|(|\lambda| - |\mu|) + (m|\lambda| - n|\mu|)^2)^{1/2} \Big) \|y\| \Big) \end{cases}$$

which can be shown by formalizing the proof given in [50]. Instantiating this with $m = n$, $\mu = |t|/n$ and $\lambda = |s|/n$ for t, s of type 1, where w.l.o.g. $|s| \geqslant |t|$, and where n is large enough that $|t|/n, |s|/n < \lambda_0$, we obtain

$$\left\|(J^A_{|t|/n})^n x - (J^A_{|s|/n})^n x\right\| \leqslant \Big(\big((|t| - |s|)^2 + |t|(|s|/n - |t|/n)\big)^{1/2} + \big(|s|(|s|/n - |t|/n) + (|s| - |t|)^2\big)^{1/2} \Big) \|y\|$$

for any $n \geqslant 1$ and any x, y with $y \in Ax$. Let k be arbitrary. Using axiom (S), we get

$$\begin{aligned} \|S(|t|)x - S(|s|)x\| &\leqslant \left\|S(|t|)x - (J^A_{|t|/n})^n x\right\| + \left\|(J^A_{|t|/n})^n x - (J^A_{|s|/n})^n x\right\| \\ &\quad + \left\|S(|s|)x - (J^A_{|s|/n})^n\right\| \\ &\leqslant \frac{2}{k+1} + \left\|(J^A_{|t|/n})^n x - (J^A_{|s|/n})^n x\right\| \\ &\leqslant \frac{2}{k+1} + \Big(\big((|t| - |s|)^2 + |t|(|s|/n - |t|/n)\big)^{1/2} \\ &\quad + \big(|s|(|s|/n - |t|/n) + (|s| - |t|)^2\big)^{1/2} \Big) \|y\| \end{aligned}$$

for any n additionally satisfying $n \geqslant \omega^S(k, b, T)$ with $b > \|x\|, \|v\|$ and $T > |t|, |s|$. This implies

$$\|S(|t|)x - S(|s|)x\| \leqslant \frac{2}{k+1} + 2||t| - |s|| \, \|y\|$$

and the claim follows as k was arbitrary

2. By Proposition 3.3.3 (essentially), we have provably that

$$\left\|J^A_{|\lambda|} x - J^A_{|\lambda|} y\right\| \leqslant \|x - y\|$$

for any $\lambda_0 > \lambda$ of type 1 and any x, y of type X. By induction, we get

$$\left\|(J^A_{|t|/n})^n x - (J^A_{|t|/n})^n y\right\| \leqslant \|x - y\|$$

for any t of type 1, any x, y of type X and any n large enough such that $|t|/n < \lambda_0$. Now, let k be arbitrary. Then we get

$$\begin{aligned}\|S(|t|)x - S(|t|)y\| &\leqslant \left\|S(|t|)x - (J^A_{|t|/n})^n x\right\| + \left\|(J^A_{|t|/n})^n x - (J^A_{|t|/n})^n y\right\| \\ &\quad + \left\|S(|t|)y - (J^A_{|t|/n})^n y\right\| \\ &\leqslant \frac{2}{k+1} + \left\|(J^A_{|t|/n})^n x - (J^A_{|t|/n})^n y\right\| \\ &\leqslant \frac{2}{k+1} + \|x - y\|\end{aligned}$$

for any $n \geqslant \omega^S(k, b, T)$ with $b > \|x\|, \|y\|, \|v\|, \|w\|$ with $v \in Ax$ and $w \in Ay$ as well as $T > |t|$ using (S). As k was arbitrary, we get the claim.

3. Using item (1) and axiom (S2), we have

$$\begin{aligned}&\|S(|t|)C(x \upharpoonright v) - S(|s|)C(x \upharpoonright v)\| \\ &\qquad \leqslant \|S(|t|)C(x \upharpoonright v) - S(|t|)(\widehat{x} \upharpoonright v)_n\| \\ &\qquad\quad + \|S(|t|)(\widehat{x} \upharpoonright v)_n - S(|s|)(\widehat{x} \upharpoonright v)_n\| \\ &\qquad\quad + \|S(|s|)(\widehat{x} \upharpoonright v)_n - S(|s|)C(x \upharpoonright v)\| \\ &\qquad \leqslant 2 \cdot 2^{-n+3} + \|S(|t|)(\widehat{x} \upharpoonright v)_n - S(|s|)(\widehat{x} \upharpoonright v)_n\| \\ &\qquad \leqslant 2 \cdot 2^{-n+3} + 2||t| - |s|| \, \|(\widehat{x} \upharpoonright v)_n\| .\end{aligned}$$

Choosing $n = k+5$, we get the claim for $||t|-|s|| \leqslant 2^{-(k+2)}/\max\{1, \|(\widehat{x} \upharpoonright v)_{k+5}\|\}$.

4. Using item (2), axiom (S2) as well as $(\mathcal{C})$, we have

$$\begin{aligned}&\|S(|t|)(C(x \upharpoonright v)) - S(|t|)(C(y \upharpoonright w))\| \\ &\qquad \leqslant \|S(|t|)(C(x \upharpoonright v)) - S(|t|)((\widehat{x} \upharpoonright v)_k)\| \\ &\qquad\qquad + \|S(|t|)((\widehat{x} \upharpoonright v)_k) - S(|t|)((\widehat{y} \upharpoonright w)_k)\| \\ &\qquad\qquad + \|S(|t|)(C(y \upharpoonright w)) - S(|t|)((\widehat{y} \upharpoonright w)_k)\| \\ &\qquad \leqslant 2 \cdot 2^{-k+3} + \|S(|t|)((\widehat{x} \upharpoonright v)_k) - S(|t|)((\widehat{y} \upharpoonright w)_k\| \\ &\qquad \leqslant 2 \cdot 2^{-k+3} + \|(\widehat{x} \upharpoonright v)_k - (\widehat{y} \upharpoonright w)_k\| \\ &\qquad \leqslant 2 \cdot 2^{-k+3} + \|(\widehat{x} \upharpoonright v)_k - C(x \upharpoonright v)\| \\ &\qquad\qquad + \|C(x \upharpoonright v) - C(y \upharpoonright w)\| \\ &\qquad\qquad + \|C(y \upharpoonright w) - (\widehat{y} \upharpoonright w)_k\| \\ &\qquad \leqslant 2 \cdot 2^{-k+3} + 2 \cdot 2^{-k+3} + \|C(x \upharpoonright v) - C(y \upharpoonright w)\| .\end{aligned}$$

As this holds for arbitrary k, we get the claim.

5. Let $x \in \mathrm{dom}A$. Using the previously introduced notation of $\overline{\cdot}$, we write

$$\begin{cases} [S(|t|)]^1 x = \overline{S(|t|)x}, \\ [S(|t|)]^{m+1} x = \overline{S(|t|)\left([S(|t|)]^m x\right)}. \end{cases}$$

Note that provably

$$[S(|t|)]^{m+1} x = S(|t|)\left([S(|t|)]^m x\right)$$

which follows as in the discussion previous to this lemma. We now show by induction on m that provably

$$\forall k \exists N_m \forall n \geqslant N_m \left(|t|/n < \lambda_0 \rightarrow \left\|[S(|t|)]^m x - \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| \leqslant 2^{-k}\right).$$

The induction base follows from (S1) as was already discussed above. For the induction step, let $N_m(k)$ be the choice function of the above statement. Then for arbitrary k, we get (using extensionality, see Remark 4.4.4, and nonexpansivity of $S(|t|)$ on the closure of the domain) that

$$\begin{aligned}
&\left\|[S(|t|)]^{m+1} x - \left(\left(J^A_{|t|/n}\right)^{m+1}\right)^n x\right\| \\
&\quad\leqslant \left\|S(|t|)[S(|t|)]^m x - \left(J^A_{|t|/n}\right)^n \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| \\
&\quad\leqslant \left\|S(|t|)[S(|t|)]^m x - \left(J^A_{|t|/n}\right)^n [S(|t|)]^m x\right\| \\
&\qquad\quad + \left\|\left(J^A_{|t|/n}\right)^n [S(|t|)]^m x - \left(J^A_{|t|/n}\right)^n \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| \\
&\quad\leqslant \left\|S(|t|)[S(|t|)]^m x - \left(J^A_{|t|/n}\right)^n [S(|t|)]^m x\right\| \\
&\qquad\quad + \left\|[S(|t|)]^m x - \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| \\
&\quad\leqslant 2^{-k}
\end{aligned}$$

for all n such that $|t|/n < \lambda_0$, $n \geqslant N_m(k+1)$ and such that n is large enough for

$$\left\|S(|t|)[S(|t|)]^m x - \left(J^A_{|t|/n}\right)^n [S(|t|)]^m x\right\| \leqslant 2^{-(k+1)}$$

which can be constructed as in the discussion previous to this lemma. Therefore, given k, we in particular get

$$\begin{aligned}
&\|[S(|t|)]^m x - S(m|t|)x\| \\
&\quad\leqslant \left\|[S(|t|)]^m x - \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| + \left\|\left(\left(J^A_{|t|/n}\right)^m\right)^n x - S(m|t|)x\right\| \\
&\quad\leqslant \left\|[S(|t|)]^m x - \left(\left(J^A_{|t|/n}\right)^m\right)^n x\right\| + \left\|\left(J^A_{m|t|/mn}\right)^{mn} x - S(m|t|)x\right\| \\
&\quad\leqslant 2^{-(k+1)} + 2^{-(k+1)} \\
&\quad\leqslant 2^{-k}
\end{aligned}$$

for any n such that

$$n \geqslant \max\{N_m(k+1), \omega^S(k+1, b, mT)\}$$

for $b > \|x\|, \|v\|$ and $T > |t|$ using (S1) and the previous result. As k was arbitrary, we get $[S(|t|)]^m x = S(m|t|)x$. Using this, we provably we get

$$\begin{aligned} S\left(\frac{l}{k}+\frac{r}{s}\right)x &= S\left(\frac{ls+rk}{ks}\right)x \\ &= \left[S\left(\frac{1}{ks}\right)\right]^{ls+rk} x \\ &= \left[S\left(\frac{1}{ks}\right)\right]^{ls}\left[S\left(\frac{1}{ks}\right)\right]^{rk} x \\ &= \left[S\left(\frac{1}{ks}\right)\right]^{ls}\overline{S\left(\frac{rk}{ks}\right)x} \\ &= S\left(\frac{ls}{ks}\right)\overline{S\left(\frac{rk}{ks}\right)x} \\ &= S\left(\frac{l}{k}\right)\overline{S\left(\frac{r}{s}\right)x} \end{aligned}$$

where we have used the above items for extensionality of S (see again Remark 4.4.4). A continuity argument using item (3) now yields the claim for arbitrary reals $|t|$ and $|s|$. Further, the claim extends to the closure of the domain via another usual continuity argument. Both we do not spell out here.

□

Remark 4.4.4. The constant $S(t)x$ is provably extensional in $x \in \overline{\mathrm{dom}A}$ for any $t \geqslant 0$ by (4) as well as in $t \geqslant 0$ for any $x \in \overline{\mathrm{dom}A}$ by (3).

Remark 4.4.5. Note that by the proof of the above item (3), we have that if the operator A is majorizable in the sense of Chapter 3, i.e. if there exists a function $A^* : \mathbb{N} \to \mathbb{N}$ such that

$$\forall b \in \mathbb{N} \forall x \in \mathrm{dom}A \cap \overline{B}_b(0) \exists y \in X \left(\|y\| \leqslant A^* b \wedge y \in Ax\right),$$

then the semigroup $\mathcal{S}$ generated by A through the Crandall-Liggett formula is uniformly equicontinuous in the sense of [109], i.e. there exists a function $\omega : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that

$$\begin{aligned} &\forall b \in \mathbb{N} \forall q \in \overline{\mathrm{dom}A} \cap B_b(0) \forall m \in \mathbb{N} \forall K \in \mathbb{N} \forall t, t' \in [0, K] \\ &\qquad \left(|t-t'| < 2^{-\omega_{K,b}(m)} \to \|S(t)q - S(t')q\| < 2^{-m}\right). \end{aligned}$$

Concretely, assuming w.l.o.g. that A^* is nondecreasing, this so-called *modulus of uniform equicontinuity* for $\mathcal{S}$ can be given by

$$\omega_{K,b}(m) = (m+2)\max\{1, A^*(b+1)\}.$$

Note in particular that this modulus is independent of the parameter K.

We now come to the main theoretical result of this chapter which comprises a proof-theoretic bound extraction theorem for the system H_p^ω akin to the usual metatheorems of proof mining. The proof of this metatheorem follows the general outline of the proof of Theorem 3.7.9 discussed in Chapter 3 and since the proof is very much standard in this way, we omit most of the details and in the following mainly just sketch the majorizability of the new constant S.

Theorem 4.4.6. *Let τ be admissible, δ be of degree 1 and s be a closed term of H_p^ω of type $\sigma(\delta)$ for admissible σ. Let $B_\forall(x,y,z,u)/C_\exists(x,y,z,v)$ be $\forall$-/$\exists$-formulas of H_p^ω with only $x,y,z,u/x,y,z,v$ free. Let Δ be a set of formulas of the form $\forall \underline{a}^{\underline{\delta}}\exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a}\forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a},\underline{b},\underline{c})$ where F_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and where $\underline{r}$ is a tuple of closed terms of appropriate type. If*

$$H_p^\omega + \Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x,y,z,u) \to \exists v^0 C_\exists(x,y,z,v)\right),$$

then one can extract a partial functional $\Phi : S_\delta \times S_{\widehat{\tau}} \times \mathbb{N} \times \mathbb{N}^\mathbb{N} \rightharpoonup \mathbb{N}$ which is total and (bar-recursively) computable on $M_\delta \times M_{\widehat{\tau}} \times \mathbb{N} \times \mathbb{N}^\mathbb{N}$ and such that for all $x \in S_\delta$, $z \in S_\tau$, $z^* \in S_{\widehat{\tau}}$ and all $n \in \mathbb{N}$ and $\omega \in \mathbb{N}^\mathbb{N}$, if $z^* \gtrsim z$ and $\omega \gtrsim \omega^S$ as well as $n \geqslant_\mathbb{R} m_{\widetilde{\gamma}}, |\widetilde{\gamma}|, \|c_X\|_X, \|d_X\|_X, |\lambda_0|, m_{\lambda_0}, m'_{\widetilde{\gamma}}$, then

$$\begin{aligned}\mathcal{S}^{\omega,X} \models \forall y \leqslant_\sigma s(x)\big(&\forall u \leqslant_0 \Phi(x,z^*,n,\omega) B_\forall(x,y,z,u)\\ &\to \exists v \leqslant_0 \Phi(x,z^*,n,\omega) C_\exists(x,y,z,v)\big)\end{aligned}$$

holds for $\mathcal{S}^{\omega,X}$ whenever $\mathcal{S}^{\omega,X} \models \Delta$ where $\mathcal{S}^{\omega,X}$ is defined via any (nontrivial) Banach space $(X, \|\cdot\|)$ with

1. *χ_A interpreted by the characteristic function of an accretive operator A satisfying the range condition $\overline{\mathrm{dom} A} \subseteq \bigcap_{\lambda_0 > \gamma > 0} \mathrm{dom} J^A_\gamma$,*
2. *J^{χ_A} interpreted by the corresponding resolvents $J^A_\gamma x$ for $\lambda_0 > \gamma \geqslant 0$ and $x \in \mathrm{dom}(J^A_\gamma)$, and by 0 otherwise,*
3. *j interpreted as discussed in Section 4.3.2,*

4. *S interpreted by the semigroup generated by A via the Crandall-Liggett formula on $[0,\infty) \times \overline{\mathrm{dom}A}$, and 0 otherwise,*

5. *d_X, c_X interpreted by a pair $(c,d) \in A$ witnessing $A \neq \emptyset$,*

6. *ω^S interpreted by a rate of convergence for the limit generating the semigroup on* $\mathrm{dom}A$*,*

and with the other constants naturally interpreted so that the respective axioms are satisfied.

Further: If $\hat{\tau}$ is of degree 1, then Φ is a total computable functional. If the claim is proved without DC*, then τ may be arbitrary and Φ will be a total functional on $S_\delta \times S_{\hat{\tau}} \times \mathbb{N}$ which is primitive recursive in the sense of Gödel. In that latter case, also plain majorization can be used instead of strong majorization.*

Proof. The proof given in Chapter 3 immediately extends to this system, noticing the additional considerations on the model of majorizable functionals discussed in the context of j as well as Remark 4.4.1. In particular, note also that all axioms added to H_p^ω are purely universal and that the new constants other than S can be majorized as discussed throughout the previous sections. For the last constant S, we can argue for the majorizability as follows: In the context of the axiom (V)$'$, stating that $\mathrm{dom}A$ is not empty using the constants c_X and d_X, majorization of the constant S on $t \geqslant 0$ and $x \in \mathrm{dom}A$ follows rather immediately. It is straightforward to obtain that

$$\widehat{\mathcal{V}}_p^\omega \vdash \forall x^X, \lambda^1, n^0 \left(x \in \mathrm{dom}(J^A_{|\lambda|}) \to \left\|(J^A_{|\lambda|})^n x -_X x\right\|_X \leqslant_{\mathbb{R}} n \left\|J^A_{|\lambda|} x -_X x\right\|_X \right).$$

Therefore, we have for $x \in \mathrm{dom}A$ with $v \in Ax$ and $b > \|x\|, \|v\|$ and for $t \geqslant 0$ with $T > t$ that for $n \geqslant (\omega^S(0,b,T) + \lceil T/\lambda_0 \rceil)$:[10]

$$\begin{aligned}
\|S(t)x\| &\leqslant \left\|S(t)x - (J^A_{t/n})^n x\right\| + \left\|(J^A_{t/n})^n x\right\| \\
&\leqslant 1 + \left\|(J^A_{t/n})^n x - (J^A_{t/n})^n c_X\right\| + \left\|(J^A_{t/n})^n c_X\right\| \\
&\leqslant 1 + \|x - c_X\| + \|c_X\| + n \left\|J^A_{t/n} c_X - c_X\right\| \\
&\leqslant 1 + \|x\| + 2\|c_X\| + T\|d_X\|
\end{aligned}$$

which follows from the axioms (S) and $(RC)_{\lambda_0}$. This extends to $\overline{\mathrm{dom}A}$ as follows: for $x \in \mathrm{dom}A$ and $x_n \to x$ with rate of convergence 2^{-n} and where $x_n \in \mathrm{dom}A$, we have

[10] We can choose e.g. $n = \omega^S(0,b,T) + [T/\lambda_0](0) + 1$ which can be represented through a closed term.

$\|x_0 - x\| \leqslant 1$ and $\|S(t)x - S(t)x_0\| \leqslant 1$ and thus

$$\begin{aligned}\|S(t)x\| &\leqslant 1 + \|S(t)x_0\| \\ &\leqslant 2 + \|x_0\| + 2\,\|c_X\| + T\,\|d_X\| \\ &\leqslant 3 + \|x\| + 2\,\|c_X\| + T\,\|d_X\|\,.\end{aligned}$$

□

Also Theorem 3.7.16 extends to an intuitionistic version $H^\omega_{i,p}$ of the system H^ω_p in that fashion. Concretely, let $H^\omega_{i,p}$ be defined as the extension/modification of $\mathcal{V}^\omega_{i,p}$ with the same constants and axioms as were added/modified to/in $\mathcal{V}^\omega_p$ to form H^ω_p. Then the following semi-constructive bound extraction theorem holds:

Theorem 4.4.7. *Let δ be of the form $0(0)\dots(0)$ and σ, τ be arbitrary, s be a closed term of suitable type. Let $\Gamma_\neg$ be a set of sentences of the form $\forall \underline{u}^{\underline{\zeta}}(C(\underline{u}) \to \exists \underline{v} \leqslant_{\underline{\beta}} \underline{tu}\neg D(\underline{u}, \underline{v}))$ with $\underline{\zeta}, \underline{\beta}$ and C, D arbitrary types and formulas respectively and where $\underline{t}$ is a tuple of closed terms. Let $B(x, y, z)/C(x, y, z, u)$ be arbitrary formulas of $H^\omega_{i,p}$ with only $x, y, z/x, y, z, u$ free. If*

$$H^\omega_{i,p} + \mathrm{IP}_\neg + \mathrm{CA}_\neg + \Gamma_\neg \vdash \forall x^\delta\, \forall y \leqslant_\sigma (x)\, \forall z^\tau\, (\neg B(x, y, z) \to \exists u^0 C(x, y, z, u)),$$

one can extract a $\Phi : S_\delta \times S_{\widehat{\tau}} \times \mathbb{N} \times \mathbb{N}^\mathbb{N} \to \mathbb{N}$ which is primitive recursive in the sense of Gödel such that for any $x \in S_\delta$, any $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$, any $z \in S_\tau$ and $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$ and any $n \in \mathbb{N}$ and $\omega \in \mathbb{N}^\mathbb{N}$ with $\omega \gtrsim \omega^S$ as well as $n \geqslant_\mathbb{R} m_{\widetilde{\gamma}}, |\widetilde{\gamma}|, \|c_X\|_X, \|d_X\|_X, |\lambda_0|, m_{\lambda_0}, m'_{\widetilde{\gamma}}$, we have that

$$\mathcal{S}^{\omega,X} \models \exists u \leqslant_0 \Phi(x, z^*, n, \omega)\, (\neg B(x, y, z) \to C(x, y, z, u))$$

holds for $\mathcal{S}^{\omega,X}$ whenever $\mathcal{S}^{\omega,X} \models \Gamma_\neg$ where $\mathcal{S}^{\omega,X}$ is defined via any (nontrivial) Banach space $(X, \|\cdot\|)$ with the constants interpreted as in Theorem 4.4.6.

Using the previous arguments regarding the majorizability of the new constants, the proof is a straightforward adaptation of the proof of Theorem 3.7.16 given in Chapter 3 and we thus omit any further details.

5 Quantitative results on Pazy's convergence condition and first-order Cauchy problems

5.1 Introduction

As discussed in the introduction of Chapter 4, one of the fundamental questions in the theory of differential equations is that of the asymptotic behavior of the solutions to a particular system. Concretely, consider again the initial value problem

$$\begin{cases} u'(t) \in -Au(t),\ 0 < t < \infty \\ u(0) = x \end{cases} \tag{$*$}$$

over a Banach space X generated by an initial value $x \in X$ and an accretive set-valued operator $A : X \to 2^X$. The focus of Chapter 4 was on correctly representing the semigroup generated by an accretive operator A via the Crandall-Liggett formula which, as also discussed in the introduction of Chapter 4, generalizes the solution semigroup of the above system in the sense that if the system is solvable, then the solution semigroup coincides with the semigroup generated by A.

Even further however, Crandall and Liggett in [50] also obtained a characterizing condition for when

$$S(t)x = \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}A\right)^{-n} x$$

actually is a solution to $(*)$. Namely, their result yields in particular that if $0 < T \leqslant \infty$ and A is m-accretive, then u_x is a solution of the initial value problem with $x \in \mathrm{dom}A$ on $[0, T)$ if and only if $u_x(t) = \lim_{n\to\infty} \left(\mathrm{Id} + \frac{t}{n}A\right)^{-n} x$ for $t \in [0, T)$ and u_x is differentiable almost everywhere.

As this function $S(t)x$ is Lipschitz continuous in t (see, e.g., the proof of Theorem 1.3 in [4], Chapter III or see also the previous Chapter 4), the additional differentiability condition is in particular immediately satisfied if any Lipschitz continuous function from the real numbers into X is differentiable almost everywhere. This in turn is true in any reflexive space X by (an extension of) Rademacher's theorem which, as is well-known, in particular includes uniformly convex spaces by the Milman–Pettis theorem.

In this chapter, we are concerned with the asymptotic behavior of $S(t)x$ for $t \to \infty$ in the context of uniformly convex and uniformly smooth spaces. It is well-known that $S(t)x$ does not always converge in that case. Motivated by these circumstances, there has been a search for potential conditions guaranteeing the convergence of the orbits and, in that context, Pazy in [160] introduced the so-called convergence condition for the operator A. Concretely, over a Hilbert space X with inner product $\langle \cdot, \cdot \rangle$ (and assuming $A^{-1}0 \neq \emptyset$), we say (following Pazy) that A satisfies the convergence condition[1] if for all bounded sequences $(x_n, y_n) \subseteq A$ such that

$$\lim_{n\to\infty} \langle y_n, x_n - Px_n \rangle = 0,$$

it holds that $\liminf_{n\to\infty} \|x_n - Px_n\| = 0$ where P is the projection onto the closed and convex set $A^{-1}0$ (if A is maximally monotone). Then Pazy obtained the following result:

Theorem 5.1.1 (Pazy [160]). *Let X be a Hilbert space and A be maximally monotone and let $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ be the semigroup generated by A where $A^{-1}0 \neq \emptyset$. If A satisfies the convergence condition then, for every $x \in \overline{\mathrm{dom} A}$, $S(t)x$ converges strongly to a zero of A as $t \to \infty$.*

This convergence result was subsequently extended to uniformly convex and uniformly smooth Banach spaces by Nevanlinna and Reich in [154] who simultaneously adapted the above convergence condition to a suitable variant in said classes of Banach spaces by modifying the premise to the assumption that

$$\lim_{n\to\infty} \langle y_n, J(x_n - Px_n) \rangle = 0$$

where J is the normalized-duality map (see again Chapter 4) which is single-valued here as the space is smooth. Concretely, the following result was obtained:

Theorem 5.1.2 (Nevanlinna and Reich [154]). *Let X be uniformly convex and uniformly smooth and A be m-accretive with $A^{-1}0 \neq \emptyset$ and such that it satisfies the*

[1] Actually, Pazy also emphasized a particular consequence of the above condition as a separate additional property for the convergence condition, but we refrain from doing so (in line with the presentation in [154]).

convergence condition. If $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ is the semigroup generated by A via the exponential formula then, for any $x \in \overline{\mathrm{dom} A}$, $S(t)x$ converges strongly to a zero of A as $t \to \infty$.

This result was further generalized by Xu in [208] who studied the behavior of almost-orbits associated with the semigroup generated by A as introduced by Miyadera and Kobayasi [150]: an almost-orbit of $\mathcal{S}$ is a continuous function $u : [0, \infty) \to \overline{\mathrm{dom} A}$ such that

$$\lim_{s\to\infty} \sup\{\|u(t+s) - S(t)u(s)\| \mid t \geqslant 0\} = 0.$$

Concretely, Xu obtained the following result:

Theorem 5.1.3 (Xu [208])**.** *Let X be uniformly convex and uniformly smooth and A be m-accretive with $A^{-1}0 \neq \emptyset$ and such that it satisfies the convergence condition. If $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ is the semigroup generated by A via the exponential formula, then every almost-orbit $u(t)$ of $\mathcal{S}$ converges strongly to a zero of A as $t \to \infty$.*

All the above results do not offer any quantitative information on the convergence of the orbits or almost-orbits. We here analyze the proofs of Theorem 5.1.2 as well as Theorem 5.1.3 and extract from these explicit computable transformations which translate a modulus witnessing a quantitative reformulation of the convergence condition into quantitative information on the convergence result. By this latter statement, we mean in particular full rates of convergence for $S(t)x$ for $t \to \infty$ in the context of the result of Nevanlinna and Reich. In the case of the result of Xu, this amounts to two kinds of quantitative "translations" with the first translating a rate of convergence for the almost-orbit into a rate of convergence of the solution of the Cauchy problem towards a zero of the operator A.

Akin to fundamental results of Specker [197] from recursion theory whereas even computable monotone sequences of rational numbers in $[0, 1]$ do not have a computable rate of convergence, one can see that those rates will in general not be computable (see for similar results also the work of Neumann [153]). The second quantitative result on Theorem 5.1.3 then takes the form of a translation converting a rate of metastability of the almost-orbit (which will be discussed later on) into a rate of metastability for the convergence towards a zero of the operator A. For this, note in particular the example presented in [108] for a concrete almost-orbit where such a rate of metastability can be naturally obtained and is moreover computable and highly uniform while any rate of convergence will not even be computable in this case.

In particular, we want to note that the theorem of García-Falset [66] analyzed in the work [108] is strongly related to the results of Pazy, Nevanlinna and Reich as well as Xu presented above. Concretely, García-Falset obtains a similar result on the asymptotic behavior of the almost-orbits of the solution semigroup of the abstract Cauchy problem generated by an operator A under the condition that A is ϕ-accretive at zero as defined in [66]. The generality gained by assuming ϕ-accretivity at zero of A is that the space is allowed to be an arbitrary Banach space.

In that context, our dichotomous situation of the two quantitative versions of the result of Xu is also similar to the results from [108] and, as will be discussed later, the work [108] is where the metastable version of the almost-orbit condition was first introduced.

In contrast to the results by García-Falset in [66] where the notion of ϕ-accretive at zero carries the strength of removing the convergence condition as well as the assumptions on the space X but simultaneously provides a strong restriction on the operator (by, among others, making the zero of the operator unique), the results given by Pazy, Nevanlinna and Reich as well as Xu offer a practically higher generality at the modest price of a uniformly convex and uniformly smooth space, a property which is still fulfilled for most spaces of interest, in particular for all L^p-spaces as is the case for all examples of application given in [66].

5.2 Preliminaries: convexity and smoothness in Banach spaces

Consider a Banach space $(X, \|\cdot\|)$. We assume throughout that X is *uniformly convex*, i.e.

$$\forall \varepsilon \in (0,2] \exists \delta \in (0,1] \forall x, y \in B_1(0) \left(\|x-y\| \geqslant \varepsilon \to \left\| \frac{x+y}{2} \right\| \leqslant 1-\delta \right),$$

and *uniformly smooth*, i.e.

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in X \left(\|x\| = 1 \wedge \|y\| \leqslant \delta \to \|x+y\| + \|x-y\| \leqslant 2 + \varepsilon \|y\| \right).$$

Note that X is uniformly convex if, and only if, its dual X^* is uniformly smooth.

Recall the *normalized duality mapping* $J : X \to 2^{X^*}$ associated with X from Chapter 4, i.e.

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

for $x \in X$. This mapping is single-valued and uniformly continuous if, and only if, X is uniformly smooth (see [43]). As is common in that context, we identify J with this unique selection mapping $X \to X^*$.

As X is uniformly convex, if $C \subseteq X$ is a non-empty, closed, convex subset of X, then the *nearest point projection* $P_C : X \to C$ is single-valued and outputs the unique point satisfying the condition

$$\|x - P_C x\| = \inf\{\|x - y\| \mid y \in C\}.$$

Even further, the projection map P_C is continuous and in fact even uniformly continuous in uniformly convex spaces as will be used later (see [189] for this).

5.3 The convergence condition and quantitative versions

As discussed in the introduction, the central notion for the asymptotic results from [154, 160, 208] is that of the convergence condition for the operator A inducing the differential equation. In the quantitative versions of these results of Pazy, Reich and Nevanlinna as well as Xu, we will rely on a (or rather multiple) particular quantitative version(s) of that condition, which we shall call a convergence condition with a modulus. These quantitative reformulations are motivated by logical considerations on different equivalent variants of the convergence condition as suggested by the classical and constructive metatheorems for accretive operators from Chapter 3. This will be discussed in more detail in Section 5.5 later on. In particular, there we will discuss that these moduli have the following two central properties guaranteed by the general logical metatheorems: For one, the extractability of such moduli for a large class of operators which provably satisfy the convergence condition is guaranteed. For another, the same logical metatheorems guarantee that from any (formalizable) proof using the assumption that an operator satisfies the convergence condition, quantitative results can be extracted which depend on such a modulus.

5.3.1 Variants of the convergence condition

To begin with, as mentioned in the introduction, the original formulation of the convergence condition is due to Pazy [160], but in our setting of uniformly convex and uniformly smooth Banach spaces, we follow the notion of Nevanlinna and Reich [154] and, therefore, say that an A with $A^{-1}0 \neq \varnothing$ satisfies the *convergence condition* if for all bounded sequences $(x_n, y_n) \subseteq A$:

$$\lim\langle y_n, J(x_n - Px_n)\rangle = 0 \to \liminf \|x_n - Px_n\| = 0.$$

Already in the literature, other equivalent variants are sometimes mentioned, e.g. replacing the limit in the premise of the implication by a limit inferior or conversely replacing the limit inferior in the conclusion by a limit (see for example [171]). However, in the following we only focus on the usual formulation of the convergence condition in the form above, together with one particular equivalent version which is of a different spirit entirely:

Lemma 5.3.1. *An operator A satisfies the convergence condition if, and only if, for all natural numbers $k, K \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that*[2]

$$\forall (x,y) \in A \left(\|x\|, \|y\| \leqslant K \wedge |\langle y, J(x - Px)\rangle| \leqslant \frac{1}{n+1} \to \|x - Px\| \leqslant \frac{1}{k+1} \right). \quad (+)$$

Proof. For sufficiency assume (+) and consider arbitrary sequences $(x_n), (y_n)$ such that $y_n \in Ax_n$, and $\|x_n\|, \|y_n\| \leqslant K$ for some $K \in \mathbb{N}$. Assume that $\lim\langle y_n, J(x_n - Px_n)\rangle = 0$ and let $k \in \mathbb{N}$ be given. By (+), there is an $n \in \mathbb{N}$ such that

$$\forall m \in \mathbb{N} \left(|\langle y_m, J(x_m - Px_m)\rangle| \leqslant \frac{1}{n+1} \to \|x_m - Px_m\| \leqslant \frac{1}{k+1} \right). \quad (++)$$

Then, by $\lim\langle y_n, J(x_n - Px_n)\rangle = 0$ there exists $N \in \mathbb{N}$ such that

$$\forall m \geqslant N \left(|\langle y_m, J(x_m - Px_m)\rangle| \leqslant \frac{1}{n+1} \right),$$

which by (++) entails that $\|x_m - Px_m\| \leqslant \frac{1}{k+1}$, for all $m \geqslant N$. This means that $\lim \|x_n - Px_n\| = 0$, and we conclude that A satisfies the convergence condition.

For necessity, suppose that (+) fails. Then for some $k, K \in \mathbb{N}$, we have

$$\forall n \in \mathbb{N} \exists (x_n, y_n) \in A$$
$$\left(\|x_n\|, \|y_n\| \leqslant K \wedge |\langle y_n, J(x_n - Px_n)\rangle| \leqslant \frac{1}{n+1} \wedge \|x_n - Px_n\| > \frac{1}{k+1} \right).$$

[2]The absolute values are actually not necessary in the premise as A is accretive.

Then in particular $|\langle y_n, J(x_n - Px_n)\rangle| \leqslant \frac{1}{n+1}$ for all $n \in \mathbb{N}$ which entails that

$$\lim\langle y_n, J(x_n - Px_n)\rangle = 0.$$

However $(\|x_n - Px_n\|)$ is bounded away from zero by $\frac{1}{k+1}$, and so A can not satisfy the convergence condition. □

The above equivalent version does not feature sequences at all and, in this way, is of a much more local nature than the original formulation. By applying the underlying logical considerations of proof mining to these two formulations, we will now derive the previously mentioned quantitative versions of the convergence condition in the form of two different moduli (where this difference of the moduli can actually be recognized in terms of logical properties of their equivalence proof as will be discussed in Section 5.5 later on). We want to note that both the above equivalence and the following quantitative versions are similar in character to the alternative characterization of strongly nonexpansive mappings introduced in [99] as well as the moduli introduced there.

5.3.2 Quantitative versions of the convergence condition

Note that the convergence condition is essentially (modulo the boundedness condition) of the general form

$$\lim a_n = 0 \rightarrow \liminf b_n = 0$$

with $a_n = \langle y_n, J(x_n - Px_n)\rangle$ and $b_n = \|x_n - Px_n\|$. In that conceptual vein, two of our quantitative versions of the convergence condition will be certain *moduli* translating a quantitative witness for the convergence $\lim a_n = 0$ in the premise into a quantitative witness for $\liminf b_n = 0$ in the conclusion (or even for a weakening of that).

In that way, two of these moduli arise by considering combinations of a quantitative witness for the "convergences" in the premise or conclusion and for that, we rely on the following notions providing such a quantitative account in various ways:

Definition 5.3.2. Let (a_n) be a sequence of non-negative real numbers.

1. We say that a functional $\varphi : \mathbb{N} \to \mathbb{N}$ is a rate of convergence for (a_n) (towards zero) if

$$\forall k \in \mathbb{N} \forall n \geqslant \varphi(k) \left(a_n \leqslant \frac{1}{k+1} \right).$$

2. We say that a functional $\varphi : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is a lim inf-rate for (a_n) (towards zero) if
$$\forall k, m \in \mathbb{N} \exists n \in [m; \varphi(k,m)] \left(a_n \leqslant \frac{1}{k+1} \right).$$

3. We say that a functional $\varphi : \mathbb{N} \to \mathbb{N}$ is a rate of approximate zeros for (a_n) if
$$\forall k \in \mathbb{N} \exists n \leqslant \varphi(k) \left(a_n \leqslant \frac{1}{k+1} \right).$$

Combinations of these quantitative versions of $\lim / \liminf = 0$ (or the even weaker property of approximate zeros) now yield the previously mentioned different quantitative versions of the convergence condition. We begin with the most immediate version which translates a rate of convergence for the premise together with the upper bound on the sequence into a lim inf-rate for the conclusion.

Definition 5.3.3. A *modulus for the convergence condition* is a functional $\Omega : \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N} \times \mathbb{N}}$ such that for any $(x_n), (y_n) \subseteq X$ and any $K \in \mathbb{N}$ and $\varphi : \mathbb{N} \to \mathbb{N}$:

if $\forall n \in \mathbb{N}\, (y_n \in Ax_n \wedge \|x_n\|, \|y_n\| \leqslant K)$
and φ is a rate of convergence for $|\langle y_n, J(x_n - Px_n)\rangle|$,
then $\Omega(K, \varphi)$ is a lim inf-rate for $\|x_n - Px_n\|$.

While conceptually appealing due to its naturality, the logical considerations underlying the approach of proof mining actually in general suggest a stronger type of modulus, named a *full* modulus here, to be necessary in the context of general quantitative analyses of results relying on the convergence condition as well as *classical logic*. Actually, in Section 5.5, we will present instances of the general logical metatheorems for the systems for semigroups from Chapter 4 that guarantee both

1. the extractability of a computable full modulus (and thus of a "plain" modulus above) for the convergence condition from a wide range of (noneffective) proofs of the convergence condition for definable classes of operators, as well as,

2. that from a proof using the convergence condition as a premise, a transformation can be extracted that maps
   (a) a full modulus into quantitative information on the conclusion if the underlying proof is nonconstructive,
   (b) a "plain" modulus into quantitative information on the conclusion if the underlying proof is "essentially" constructive,

where, moreover, the complexity of the principles used in the proof is reflected in the complexity of the extracted transformation.

In that way, while the above modulus is derived from a "constructive" perspective on the convergence condition, the following full modulus is attained from a "classical" perspective on it. We however postpone a detailed discussion of these logical aspects to the end of the chapter (see Section 5.5) where we in particular will give formal justifications for the above statements. We now give the definition of a full modulus:

Definition 5.3.4. A *full modulus for the convergence condition* is a functional Ω^f : $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ satisfying that for any $k, K \in \mathbb{N}$: if $y \in Ax$ are such that $\|x\|, \|y\| \leqslant K$, then

$$|\langle y, J(x - Px)\rangle| \leqslant \frac{1}{\Omega^f(K,k)+1} \Rightarrow \|x - Px\| \leqslant \frac{1}{k+1}.$$

In a way, the above is a true finitization of the convergence condition in the sense that the above notion only refers to finitely many objects together with the fact that by the result given in Lemma 5.3.1, we have effectively shown the following:

Proposition 5.3.5. *An operator A satisfies the convergence condition if, and only if, it has a full modulus for the convergence condition Ω^f.*

Remark 5.3.6. Note by Lemma 5.3.1 that the convergence condition is nothing else but a uniform version of the property

$$\forall (x,y) \in A \forall k \in \mathbb{N} \exists n \in \mathbb{N} \left(|\langle y, J(x - Px)\rangle| \leqslant \frac{1}{n+1} \to \|x - Px\| < \frac{1}{k+1} \right),$$

which can easily be seen to be equivalent to

$$\forall (x,y) \in A \ (\langle y, J(x - Px)\rangle = 0 \to \|x - Px\| = 0).$$

This property was already singled out as an important special case of the convergence condition in Pazy's original paper [160] (as mentioned already in a footnote in the introduction to this chapter). In particular, based on the logical form of the above statement, suitable extensions of the logical metatheorems for nonlinear semigroups mentioned above actually guarantee a strengthened form of item (1) discussed above in the sense that already from a (possibly noneffective) proof of the above property for a class of operators, one can extract a computable full modulus (and thus a "plain" modulus) for the convergence condition, provided the proof is as before confined by the logical conditions of the metatheorem. Also this situation is conceptually similar to the results on strongly nonexpansive mappings from [99], in particular to the fact that the SNE-modulus introduced there arises as the uniform version of the notion of strict nonexpansivity.

In any way, even in the case of a (semi-)constructive proof and in the context of a "plain" modulus, the required modulus can often further be weakened. While our quantitative versions of the convergence results of Nevanlinna and Reich as well as Xu can, for one, be stated already in terms of a "plain" modulus for the convergence condition, the only sequences to which the convergence condition is ever applied (in the context of this thesis) are such that $\|x_n - Px_n\|$ is nonincreasing. In that case, it is clear that it already suffices to require a modulus which translates a rate of convergence φ for the sequence $|\langle y_n, J(x_n - Px_n)\rangle|$ together with the bound K into a rate of approximate zeros $\Omega(K, \varphi)$ for the sequence $\|x_n - Px_n\|$. As this circumstance seems to occur rather frequently,[3] we introduce this special case as a particular other notion for a quantitative form of the convergence condition:

Definition 5.3.7. A *weak modulus for the convergence condition* is a functional $\Omega^w : \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$ such that for any $(x_n), (y_n) \subseteq X$ and any $K \in \mathbb{N}$ and $\varphi : \mathbb{N} \to \mathbb{N}$:

$$\begin{aligned}&\textbf{if } \forall n \in \mathbb{N}\, (y_n \in Ax_n \wedge \|x_n\|, \|y_n\| \leqslant K)\\ &\text{and } \varphi \text{ is a rate of convergence for } |\langle y_n, J(x_n - Px_n)\rangle|,\\ &\textbf{then } \Omega^w(K, \varphi) \text{ is a rate of approximate zeros for } \|x_n - Px_n\|.\end{aligned}$$

In that way, while both the full and "plain" moduli represent the correct quantitative content of the convergence condition (from a classical and a constructive perspective, i.e. complying with the properties (1) and (2) mentioned above, respectively), the extractions formulated here will be phrased in terms of the weaker quantitative assumption of a weak modulus for the convergence condition. Note for this that there is of course no loss of generality as given a full modulus Ω^f, a "plain" modulus Ω can be defined via $\Omega(K, \varphi)(k, m) = \max\{m, \varphi(\Omega^f(K, k))\}$ and in turn, given a "plain" modulus Ω, a weak modulus Ω^w can be defined just via $\Omega^w(K, \varphi)(k) = \Omega(K, \varphi)(k, 0)$.

5.3.3 Examples for operators and their moduli

In the following, we survey various examples given in the works [154, 160] and beyond for classes of operators which naturally satisfy the convergence condition. Based on

[3] In fact, in e.g. the related work [66] on quantitative behavior of semigroups generated by ϕ-accretive operators, the requirements in the condition of ϕ-accretivity (essentially replacing the convergence condition) are such that they restrict the conclusion essentially to sequences x_n such that $\|x_n - Px_n\|$ is decreasing. A similar restriction could have been made in the case of the convergence condition since, as said above, the applications given in [154, 160, 208] satisfy the requirement but it seems that the authors have refrained from doing so to make the condition less technical.

the corresponding proofs, we extract respective full moduli in the sense of the previous section.

Strongly accretive operators

The following is an immediate generalization of Example 4.3 in [160].

Lemma 5.3.8. *If A is* strongly accretive*, by which we mean there exists an $\alpha > 0$ such that*

$$\langle u - v, J(x-y)\rangle \geqslant \alpha \left\|x-y\right\|^2$$

for any $(x,u),(y,v) \in A$ and additionally $A^{-1}0 \neq \varnothing$, then A satisfies the convergence condition with a full modulus for the convergence condition $\Omega^f_a(K,k) = a(k+1)^2 \dot{-} 1$ for any $a \in \mathbb{N}^*$ such that $\alpha \geqslant a^{-1}$.

Proof. Let $(x,y) \subseteq A$ with $\|x\|, \|y\| \leqslant K$ and where

$$\langle y, J(x-Px)\rangle \leqslant \frac{1}{\Omega^f_a(K,k)+1}.$$

Then as $\langle y, J(x-Px)\rangle \geqslant \alpha \left\|x-Px\right\|^2$ we get

$$\alpha \left\|x-Px\right\|^2 \leqslant \frac{1}{\Omega^f_a(K,k)+1} \leqslant \frac{1}{a(k+1)^2} \leqslant \frac{\alpha}{(k+1)^2}$$

which yields $\|x-Px\| \leqslant 1/(k+1)$. □

As already mentioned in [160], a particular example of a strongly monotone operator is the negative Laplacian: Let Ω be a bounded domain in $\mathbb{R}^n$ with smooth boundary. $\mathrm{L}^2(\Omega)$ is the space of square-integrable functions as usual and $\mathrm{W}^{1,2}_0(\Omega)$ the associated subspace of the Sobolev-space $\mathrm{W}^{1,2}(\Omega)$ containing functions of zero-trace. Then using Poncairé's inequality (see e.g. [131]), we get that

$$-\int_\Omega \Delta u \cdot u \,\mathrm{dx} = \int_\Omega |\nabla u|^2 \,\mathrm{dx} \geqslant \lambda_1 \int_\Omega |u|^2 \,\mathrm{dx}$$

where Δ is the usual Laplacian operator and $\lambda_1 > 0$ is the minimal eigenvalue of $-\Delta$. Therefore, $A = -\Delta$ is strongly monotone and by the above lemma satisfies the convergence condition with a full modulus for the convergence condition

$$\Omega^f_\Lambda(K,k) = \Lambda(k+1)^2 \dot{-} 1,$$

where $\Lambda \in \mathbb{N}^*$ is such that Λ^{-1} is a lower bound on the eigenvalues of $-\Delta$.

Operators that are ϕ-accretive at zero or uniformly accretive at zero

The above case of strongly monotone operators is a special case of the notion of operators which are ϕ-accretive at zero introduced in [66] over general Banach spaces.

Definition 5.3.9 ([66])**.** An operator A with $0 \in Az$ is ϕ-accretive at zero in the sense of [66] if $\phi : X \to [0, \infty)$ is a continuous function with $\phi(0) = 0$, $\phi(x) > 0$ for $x \neq 0$ and

$$\phi(x_n) \to 0 \Rightarrow \|x_n\| \to 0$$

for every sequence $(x_n) \subseteq X$ such that $\|x_n\|$ is nonincreasing and we have that

$$\langle y, J(x - z)\rangle \geqslant \phi(x - z)$$

for all $(x, y) \in A$.

As already mentioned in [66], it is a straightforward consequence of [68, Theorem 8] that if A is m-ψ-strongly accretive in the sense of [66], then A is $(\psi \circ \|\cdot\|)$-accretive at zero.

In the course of their proof-theoretic analysis of the main result of [66], which is similar in kind to the results analyzed here, Kohlenbach and Koutsoukou-Argyraki in [108] introduced (similarly motivated by proof-theoretic considerations) a generalized uniform version of the above property (without any reference to a function ϕ) under the name of uniform accretivity at zero:

Definition 5.3.10 ([108])**.** An accretive operator A with $0 \in Az$ is called *uniformly accretive at zero* if for all $k \in \mathbb{N}$ and all $K \in \mathbb{N}^*$, there exists an $m \in \mathbb{N}$ such that

$$\forall (x, u) \in A \ \left(\|x - z\| \in [2^{-k}, K] \to \langle u, x - z\rangle_s \geqslant 2^{-m}\right)$$

with $\langle \cdot, \cdot \rangle_s$ defined as in Chapter 4, i.e.

$$\langle y, x\rangle_s := \max\{\langle y, j\rangle \mid j \in J(x)\}.$$

This notion was accompanied in [108] with a corresponding uniform quantitative modulus of being uniformly accretive at zero which is defined in the following sense:

Definition 5.3.11 ([108])**.** A function $\Theta : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$ is a *modulus of accretivity at zero for* A if $m := \Theta_K(k)$ satisfies the condition in Definition 5.3.10.

Note that this notion in particular encompasses the moduli of uniform ϕ-accretivity at zero also introduced in [108] which provide a quantitative perspective on the above notion of ϕ-accretivity at zero.

Now, while our setting is more restrictive in terms of the space, we can nevertheless recognize the above notion as essentially stating the existence a full modulus for the convergence condition for A, at least in our context of uniformly convex and uniformly smooth spaces: At first, the expression $\langle u, x-z\rangle_s$ reduces to $\langle u, J(x-z)\rangle$ in a uniformly smooth space while in the context of uniformly convex spaces, through the presence of the projection P and as the zero z is unique, the point z can be replaced by the projection Px for any point x. Reading the resulting condition as its contraposition, we obtain that a modulus of accretivity at zero for A satisfies that for any k and K, if $\|x - Px\| \leqslant K$, then

$$\forall (x, u) \in A \left(|\langle u, J(x - Px)\rangle| < 2^{-\Theta_K(k)} \to \|x - Px\| < 2^{-k}\right).$$

Since we can bound $\|x - Px\|$ by

$$\|x - Px\| \leqslant \|x\| + \|z\|$$

using the single witness $z \in \mathrm{zer}A$ for $\mathrm{zer}A \neq \varnothing$ (as required in the context of the convergence condition), we get that therefore Ω^f defined by

$$\Omega^f(K, k) = 2^{\Theta_{K+Z}(k)},$$

where $Z \geqslant \|z\|$, is a full modulus for the convergence condition of A which is even independent of an upper bound for $u \in Ax$. In that way, we find that the notion of being uniformly accretive at zero is essentially an equivalent formulation of the convergence condition in that context.

Thus, if restricted to the class of spaces considered here, we find that the quantitative results on the behavior of the semigroups generated by A as derived in [108] can also be recognized as applications of our general quantitative results, using the notion of a full modulus for the convergence condition Ω^f.

Operators without unique zeros

All operators discussed so far are ϕ-accretive in the sense of [66]. The convergence condition however encompasses a far larger class of operators and the difference set of those two notions is already populated with fairly simple examples of which we exhibit one in the following. For this, we recall the following result due to Pazy:

Proposition 5.3.12 (Pazy [160]). *Let $\varphi : X \to \overline{\mathbb{R}}$ be proper, convex and l.s.c. on a Hilbert space X and assume that $\varphi(x) \geqslant 0$ for all $x \in X$ as well as $\min_{x \in X} \varphi(x) = 0$. If the level-sets*

$$K_R = \{x \mid \|x\| \leqslant R, \varphi(x) \leqslant R\}$$

are totally bounded, then the maximally monotone operator $\partial\varphi(x) = \{u \in X \mid f(y) \geqslant f(x) + \langle y - x, u\rangle$ for all $y \in X\}$ satisfies the convergence condition.

Now, for an example of an operator which satisfies the convergence condition but is not ϕ-accretive at 0 for any ϕ, consider the following function $f : \mathbb{R} \to \mathbb{R}$:

$$f(x) = \begin{cases} (x+1)^4 & \text{if } x \in (-\infty, -1], \\ 0 & \text{if } x \in [-1, 1], \\ (x-1)^4 & \text{if } x \in [1, \infty). \end{cases}$$

This function is continuously differentiable with first derivative

$$f'(x) = \begin{cases} 4(x+1)^3 & \text{if } x \in (-\infty, -1], \\ 0 & \text{if } x \in [-1, 1], \\ 4(x-1)^3 & \text{if } x \in [1, \infty). \end{cases}$$

Therefore $\partial f(x) = \{f'(x)\}$ for any $x \in \mathbb{R}$ (see e.g. Proposition 17.31 in [11]) and it is easy to see that f is convex and that the level sets K_R are compact. Thus ∂f satisfies the convergence condition. However, we have $\mathrm{zer}\partial f = [-1, 1]$ and thus ∂f does not have an unique zero. The uniqueness of the zero is, however, a property of every operator that is ϕ-accretive at zero (see [66]) or even of every operator that is uniformly accretive at zero (see [108]).

Nevertheless, by a quantitative analysis of the application of Proposition 5.3.12 to the function f, we can immediately extract a full modulus for the convergence condition $\Omega^f(K, k) = (k+1)^4 - 1$ for the convergence condition of ∂f: Let consider $x \in \mathbb{R}$ and assume $|x|, |f'(x)| \leqslant K$ as well as

$$\langle y, x - Px\rangle \leqslant \frac{1}{((k+1)^4 - 1) + 1}.$$

As in [160], i.e. using the subgradient inequality, one can show $\langle y, x - Px\rangle \geqslant f(x)$. Thus in particular

$$f(x) \leqslant \frac{1}{((k+1)^4 - 1) + 1}.$$

One can immediately show that if $f(x) \leqslant \varepsilon$ for $\varepsilon > 0$, then $x \in [-1 - \sqrt[4]{\varepsilon}, 1 + \sqrt[4]{\varepsilon}]$ and thus $\|x - Px\| \leqslant \sqrt[4]{\varepsilon}$. Therefore the above implies

$$\|x - Px\| \leqslant \frac{1}{k+1}$$

as desired.

5.4 Quantitative results on the asymptotic behavior of semigroups and their almost-orbits

In this section, we employ the previous quantitative considerations on the convergence condition for establishing quantitative versions of the theorems of Nevanlinna and Reich as well as of Xu outlined in the introduction. Note that since the proofs of the respective results are essentially constructive, a dependence on a "plain" (or even weak) modulus for the convergence condition can be guaranteed a priori for the extracted results (see the logical remarks in Section 5.5) which is also the case for the concrete rates presented below. In that vein, we in the following denote all moduli just by an Ω without the previous superscripts. We begin with the result of Nevanlinna and Reich.

5.4.1 The asymptotic behavior of nonlinear semigroups

Consider again the setup from Theorem 5.1.2 and write $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ for the semigroup generated by A via the exponential formula. In the following, if not stated otherwise, let $x \in \mathrm{dom} A$. We write $w_x(t)$ for $S(t)x$ (in the spirit of Xu [208]), $v_x(t)$ for $-w_x'(t)$ and $j_x(t)$ for $J(w_x(t) - Pw_x(t))$. Note that $w_x'(t)$ is defined almost-everywhere and $(w_x(t), -w_x'(t)) \in A$ is satisfied almost-everywhere (see [4]), say both on $[0,\infty)\backslash N_1$ where N_1 is a Lebesgue null set.

The first step in the proof is to establish $\langle v_x(t), j_x(t)\rangle \geqslant 0$ and subsequently that $\liminf_{t\to\infty}\langle v_x(t), j_x(t)\rangle = 0$. The following results extract from their proof a rate for the lim inf expression.

Lemma 5.4.1. *If $f : [0,\infty) \to [0,\infty)$ is Lebesgue integrable with*

$$\int_0^\infty f(t)\mathrm{dt} \leqslant L$$

for some $L \in [0,\infty)$, then for any Lebesgue null set $N \subseteq [0,\infty)$ and any k, n:

$$\exists t \in [n, \lceil L+1\rceil(k+1) + n]\backslash N \left(f(t) \leqslant \frac{1}{k+1} \right).$$

Proof. Suppose not. Then there are a Lebesgue null set N and k, n such that for any $t \in [n, \lceil L+1 \rceil (k+1) + n] \backslash N$ it holds that $f(t) > 1/(k+1)$. As f is nonnegative, we get that

$$\int_0^\infty f(t)\mathrm{dt} \geqslant \int_{[n,(L+1)(k+1)+n]\backslash N} f(t)\mathrm{dt} \geqslant \frac{((L+1)(k+1)+n-n)}{k+1} = (L+1)$$

which is a contradiction. □

Now, $\|w_x(t) - Pw_x(t)\|$ is Lipschitz-continuous as $\|x - Px\|$ is nonexpansive and $w_x(t)$ is Lipschitz with $\|w_x(t) - w_x(s)\| \leqslant 2\|v\| |t-s|$ where $v \in Ax$ which exists as $x \in \mathrm{dom}A$ (see the proof of Theorem 1.3 in Chapter III of [4]). Thus $\|w_x(t) - Pw_x(t)\|$ is absolutely continuous on every $[0, T]$ which implies that the derivative $\frac{\mathrm{d}}{\mathrm{dt}} \|w_x(t) - Pw_x(t)\|^2$ exists almost everywhere, say on $[0, \infty)\backslash N_2$, and that this derivative is Lebesgue-integrable such that the fundamental theorem of calculus is valid. Further, as shown in [154], we have that

$$\langle v_x(t), j_x(t)\rangle \leqslant -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{dt}} \|w_x(t) - Pw_x(t)\|^2$$

holds almost everywhere, say w.l.o.g. also on $[0, \infty)\backslash N_2$ where we assume, also without loss of generality, that $N_2 \supseteq N_1$. Using these properties, we get the following lemma:

Lemma 5.4.2. *Let $b \geqslant \|x - Px\|$. For any Lebesgue null set $N \supseteq N_2$ and any k, n:*

$$\exists t \in \left[n, \left\lceil \frac{1}{2}b^2 + 1 \right\rceil (k+1) + n\right] \backslash N \left(\langle v_x(t), j_x(t)\rangle \leqslant \frac{1}{k+1}\right).$$

Proof. We have $\langle v_x(t), j_x(t)\rangle \geqslant 0$ for any $t \in [0, \infty)\backslash N_1$ by accretivity of A. As

$$\langle v_x(t), j_x(t)\rangle \leqslant -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{dt}} \|w_x(t) - Pw_x(t)\|^2$$

holds almost everywhere, we get

$$\begin{aligned}
\int_0^\infty \langle v_x(t), j_x(t)\rangle \mathrm{dt} &\leqslant -\frac{1}{2}\int_0^\infty \frac{\mathrm{d}}{\mathrm{dt}} \|w_x(t) - Pw_x(t)\|^2 \mathrm{dt} \\
&= -\frac{1}{2}\lim_{T\to\infty} (\|w_x(T) - Pw_x(T)\|^2 - \|w_x(0) - Pw_x(0)\|^2) \\
&= \frac{1}{2}\lim_{T\to\infty} (\|w_x(0) - Pw_x(0)\|^2 - \|w_x(T) - Pw_x(T)\|^2) \\
&\leqslant \frac{1}{2}\|w_x(0) - Pw_x(0)\|^2 \\
&\leqslant \frac{1}{2}b^2.
\end{aligned}$$

By Lemma 5.4.1, we get that for any $N \supseteq N_2$ and any k, n:

$$\exists t \in \left[n, \left\lceil \frac{1}{2}b^2 + 1 \right\rceil (k+1) + n\right] \setminus N \left(\langle v_x(t), j_x(t) \rangle \leqslant \frac{1}{k+1} \right)$$

which is the claim. □

The next step in the proof of Nevanlinna and Reich infers the respective lim inf result for the function $\|w_x(t) - Pw_x(t)\|$ via the convergence condition together with Lemma 5.4.2 and then, using the fact that $\|w_x(t) - Pw_x(t)\|$ is nonincreasing, infers the convergence of $w_x(t)$. An analysis of this proof yields, in combination with the above, the following quantitative version of Theorem 5.1.2. For this, we first focus on the special case when $x \in \mathrm{dom}A$. Note that the following theorem does not use the full lim inf-rate of the previous lemma but only requires an instantiation of the above for $n = 0$.

Theorem 5.4.3. *Let X be uniformly convex and uniformly smooth and A be m-accretive such that there exists a weak modulus for the convergence condition Ω. Let $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ be the semigroup generated by A via the exponential formula. Let $A^{-1}0 \neq \varnothing$ with $p \in A^{-1}0$. For any $x \in \mathrm{dom}A$ with $v \in Ax$, we have*

$$\forall k \in \mathbb{N} \forall s, s' \geqslant \chi(\Omega(K, \mathrm{id})(2k+1)) \left(\|S(s)x - S(s')x\| \leqslant \frac{1}{k+1} \right)$$

where

$$\chi(k) = \left\lceil \frac{1}{2}b^2 + 1 \right\rceil (k+1)$$

and where $b \geqslant \|x - Px\|$ as well as $K \geqslant \max\{\|v\|, \|x-p\| + \|p\|\}$.

Proof. First, note that we have

$$\|w_x(t) - p\| = \left\| \lim_{n\to\infty} J^n_{t/n}x - p \right\| \leqslant \|x - p\|$$

as $p \in A^{-1}0$ and thus p is a fixed point for any resolvent. Therefore

$$\|w_x(t)\| \leqslant \|x - p\| + \|p\|$$

for any $t \in [0, \infty)$. Further, Proposition 1.2 in [4] implies

$$\|w'_x(t)\| \leqslant \|v\|$$

almost everywhere as $v \in Ax$, say for $t \in [0, \infty) \setminus N_3$. W.l.o.g. we assume that $N_3 \supseteq N_2 \supseteq N_1$.

Now, Lemma 5.4.2 yields that for any k:

$$\exists t \in [0, \chi(k)] \backslash N_3 \left(\langle v_x(t), j_x(t) \rangle \leqslant \frac{1}{k+1} \right). \tag{†}$$

Now we choose a sequence $(t_n) \subseteq [0, \infty) \backslash N_3$ using the previous (†) such that $\langle v_x(t_n), j_x(t_n) \rangle \leqslant \frac{1}{n+1}$ and $t_n \leqslant \chi(n)$.

This is well-defined as $N_3 \supseteq N_2$ and by the above, we have $\|w_x(t_n)\|, \|w_x'(t_n)\| \leqslant K$ for all n where also the latter is well-defined. Now, $\mathrm{id} : \mathbb{N} \to \mathbb{N}$ is a rate of convergence for $\langle v_x(t_n), j_x(t_n) \rangle \to 0$. Then by assumption on Ω, we get

$$\forall k \exists n \leqslant \Omega(K, \mathrm{id})(2k+1) \left(\|w_x(t_n) - P w_x(t_n)\| \leqslant \frac{1}{2(k+1)} \right)$$

and thus, as $t_n \leqslant \chi(n)$, we get

$$\forall k \exists t \leqslant \chi(\Omega(K, \mathrm{id})(2k+1)) \left(\|w_x(t) - P w_x(t)\| \leqslant \frac{1}{2(k+1)} \right).$$

Similar as in [154], using that

$$0 \leqslant \langle v_x(t), j_x(t) \rangle \leqslant -\frac{1}{2} \frac{\mathrm{d}}{\mathrm{dt}} \|w_x(t) - P w_x(t)\|^2$$

almost everywhere, we have that $\|w_x(t) - P w_x(t)\|$ is nonincreasing and thus

$$\forall k \forall t \geqslant \chi(\Omega(K, \mathrm{id})(2k+1)) \left(\|w_x(t) - P w_x(t)\| \leqslant \frac{1}{2(k+1)} \right).$$

We then get

$$\begin{aligned} \|w_x(t) - w_x(t+h)\| &\leqslant \|w_x(t) - P w_x(t)\| + \|P w_x(t) - w_x(t+h)\| \\ &\leqslant 2 \|w_x(t) - P w_x(t)\| \end{aligned}$$

for all $t, h \geqslant 0$ (as $\|w_x(t) - p\|$ is nonincreasing for any $p \in A^{-1}0$) and therefore

$$\forall k \forall t \geqslant \chi(\Omega(K, \mathrm{id})(2k+1)) \forall h \left(\|w_x(t) - w_x(t+h)\| \leqslant \frac{1}{k+1} \right)$$

which yields the claim. □

The following is then an immediate extension to the case of $x \in \overline{\mathrm{dom} A}$.

Theorem 5.4.4. *Assume the conditions of Theorem 5.4.3. Let* $x \in \overline{\mathrm{dom} A}$ *where* $f : \mathbb{N} \to \mathbb{N}$ *is such that* f *is nondecreasing and*

$$\forall n \in \mathbb{N} \exists u_n, v_n \in X \left(v_n \in A u_n \wedge \|u_n\|, \|v_n\| \leqslant f(n) \wedge \|u_n - x\| \leqslant \frac{1}{n+1} \right).$$

Then

$$\forall k \in \mathbb{N} \forall s, s' \geqslant \chi_k(\Omega(K_k, \mathrm{id})(6k+5)) \left(\|S(s)x - S(s')x\| \leqslant \frac{1}{k+1} \right)$$

where

$$\chi_k(j) = \left\lceil \frac{1}{2} b_k^2 + 1 \right\rceil (j+1)$$

and where $b_k \geqslant \|x - Px\| + \|x\| + f(3k+2)$ *as well as* $K_k \geqslant f(3k+2) + 2\|p\|$.

Proof. By assumption on f, we get that there exists a u, v with $v \in Au$ such that $\|u\|, \|v\| \leqslant f(3k+2)$ and such that $\|x - u\| \leqslant 1/(3(k+1))$. Therefore, as $S(t)$ is nonexpansive for every t, we have

$$\begin{aligned} \|S(s)x - S(s')x\| &\leqslant \|S(s)x - S(s)u\| + \|S(s)u - S(s')u\| + \|S(s')x - S(s')u\| \\ &\leqslant 2\|x - u\| + \|S(s)u - S(s')u\| \\ &\leqslant \frac{2}{3(k+1)} + \|S(s)u - S(s')u\|. \end{aligned}$$

Using the previous Theorem 5.4.3, since $v \in Au$, we get that

$$\forall k \forall s, s' \geqslant \chi_k(\Omega(K_k, \mathrm{id})(6k+5)) \left(\|S(s)u - S(s')u\| \leqslant \frac{1}{3(k+1)} \right)$$

and thus

$$\forall k \forall s, s' \geqslant \chi_k(\Omega(K_k, \mathrm{id})(6k+5)) \left(\|S(s)x - S(s')x\| \leqslant \frac{1}{k+1} \right)$$

since

$$\max\{\|v\|, \|u - p\| + \|p\|\} \leqslant \max\{f(3k+2), f(3k+2) + 2\|p\|\} \leqslant K_k$$

as well as

$$\|u - Pu\| \leqslant \|u - Px\| \leqslant \|u - x\| + \|x - Px\|$$

and thus $\|u - Pu\| \leqslant b_k$. □

Remark 5.4.5. As revealed by the quantitative analysis, the above result as well as Theorem 5.4.3 already hold in general Banach spaces whenever there exist selections of the duality map and of the projection satisfying some simple requirements. See Section 5.5 for further comments on this.

5.4.2 The asymptotic behavior of almost-orbits of nonlinear semigroups

We now turn to an analysis of Xu's result. For that, consider the setup from Theorem 5.1.3 and write $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ for the semigroup generated by A via the exponential formula as before.

As already discussed in the introduction, the (logically speaking) complicated premise of u being an almost-orbit in that context induces two natural quantitative versions of that property which were introduced in [108] and also feature in the finitary variants of Xu's result given here. Concretely, in the following, we will obtain (similar to [108]) two translations converting respectively

1. a rate of metastability Φ of the almost-orbit as introduced in [108], i.e. Φ satisfies

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Phi(k, f) \forall t \in [0, f(n)] \left(\|S_{1/2}(t)u(n) - u(t+n)\| \leqslant \frac{1}{k+1} \right),$$

into a rate of metastability Γ for the Cauchy property of the almost-orbit, i.e. Γ satisfies

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Gamma(k, f) \forall t, t' \in [n, n + f(n)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right),$$

2. a rate of convergence Φ for the almost-orbit, i.e. Φ satisfies

$$\forall k \in \mathbb{N} \forall s \geqslant \Phi(k) \left(\sup_{t \geqslant 0} \left\|u(s+t) - S_{1/2}(t)u(s)\right\| \leqslant \frac{1}{k+1} \right).$$

into a rate of Cauchyness of the almost-orbit of the Cauchy problem in a similar manner as before.

We begin with the former.

Theorem 5.4.6. *Let X be uniformly convex and uniformly smooth and A be m-accretive such that there exists a weak modulus for the convergence condition Ω. Let $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ be the semigroup generated by A via the exponential formula. Let $A^{-1}0 \neq \varnothing$ with $p \in A^{-1}0$ and assume that P, the nearest point projection onto $A^{-1}0$, is uniformly continuous on bounded subsets of X with a modulus $\omega : \mathbb{N}^2 \to \mathbb{N}$, i.e.*

$$\forall r, k \in \mathbb{N} \forall x, y \in \overline{B}_r(p) \left(\|x - y\| \leqslant \frac{1}{\omega(r,k)+1} \to \|Px - Py\| \leqslant \frac{1}{k+1} \right),$$

and, without loss of generality, assume that $\omega(r,k) \geqslant k$ for all $r,k \in \mathbb{N}$. Let u be an almost-orbit of $\mathcal{S}$ with a rate of metastability Φ on the almost-orbit condition, i.e.

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Phi(k,f) \forall t \in [0, f(n)] \left(\|S(t)u(n) - u(t+n)\| \leqslant \frac{1}{k+1} \right).$$

Let $B \in \mathbb{N}^*$ be such that $\|u(t) - p\| \leqslant B$ for all $t \geqslant 0$ and let $f_s : \mathbb{N} \to \mathbb{N}$ for $s \geqslant 0$ be such that f_s is nondecreasing and

$$\forall n \in \mathbb{N} \exists x_{s,n}, y_{s,n} \in X \left(y_{s,n} \in Ax_{s,n} \wedge \|x_{s,n}\|, \|y_{s,n}\| \leqslant f_s(n) \wedge \|x_{s,n} - u(s)\| \leqslant \frac{1}{n+1} \right).$$

Then we have

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Gamma(k,f) \forall t,t' \in [n, n+f(n)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right),$$

where

$$\Gamma(k,f) := \max\{\Gamma'(8k+7, j_{k,f}), \Phi(8k+7, h_{N,f}) \mid N \leqslant \Gamma'(8k+7, j_{k,f})\}$$

with

$$\begin{gathered} h_{N,f}(n) := f(\max\{N,n\}) + \max\{N,n\} - n, \\ j_{k,f}(n) := \max\{n, \Phi(8k+7, h_{n,f})\} - n \\ g_{k,f}(m) := \Omega_m(3k+2) + f(m + \Omega_m(3k+2)), \\ \Gamma'(k,f) := \Phi(\omega(B, 3k+2), g_{k,f}) + \max\{\Omega_m(3k+2) \mid m \leqslant \Phi(\omega(B,3k+2), g_{k,f})\}, \end{gathered}$$

for $\Omega_s(k)$ with $s \geqslant 0$ defined by

$$\Omega_s(k) := \chi(\Omega(K_{s,k}, \mathrm{id})(3k+2)),$$

now with

$$\chi(k) := \left\lceil \frac{1}{2}(B+1)^2 + 1 \right\rceil (k+1)$$

and where $K_{s,k} \geqslant \max\{f_s(\omega(B+1, 3k+2)), B+1+\|p\|\}$.

Proof. For $x \in \mathrm{dom}A$ with $v \in Ax$ consider $S(t)x$. As in the proof of Theorem 5.4.3, we get

$$\forall k \in \mathbb{N} \forall t \geqslant \Omega'_{K,b}(k) \left(\|S(t)x - PS(t)x\| \leqslant \frac{1}{k+1} \right), \tag{$-$}$$

where

$$\Omega'_{K,b}(k) := \left\lceil \frac{1}{2}b^2 + 1 \right\rceil (\Omega(K, \mathrm{id})(k) + 1),$$

with $K \geqslant \max\{\|v\|, \|x-p\|+\|p\|\}$ and $b \geqslant \|x-Px\|$.

Claim 1: For all $s \geqslant 0$,

$$\forall k \in \mathbb{N} \forall t \geqslant \Omega_s(k) \left(\|S(t)u(s) - PS(t)u(s)\| \leqslant \frac{1}{k+1} \right).$$

Proof of claim 1: For given $s \geqslant 0$, note that by assumption on f_s there exist $y_{s,k} \in Ax_{s,k}$ with $\|x_{s,k}\|, \|y_{s,k}\| \leqslant f_s(\omega(B+1, 3k+2))$ such that

$$\|x_{s,k} - u(s)\| \leqslant \frac{1}{\omega(B+1,3k+2)+1} \left(\leqslant \frac{1}{3(k+1)} \right).$$

For χ and $K_{s,k}$ as above, since

$$\|x_{s,k} - Px_{s,k}\| \leqslant \|x_{s,k} - p\| \leqslant \|x_{s,k} - u(s)\| + \|u(s) - p\| \leqslant B+1,$$

we have by $(-)$ that

$$\forall k \in \mathbb{N} \forall t \geqslant \Omega_s(k) \left(\|S(t)x_{s,k} - PS(t)x_{s,k}\| \leqslant \frac{1}{3(k+1)} \right),$$

with $\Omega_s(k)$ defined as above since $\Omega_s(k) = \Omega'_{K_{s,k},(B+1)}(3k+2)$. For $t \geqslant \Omega_s(k)$, we thus also have

$$\begin{aligned}
\|S(t)u(s) - PS(t)u(s)\| &\leqslant \|S(t)u(s) - S(t)x_{s,k}\| + \|S(t)x_{s,k} - PS(t)x_{s,k}\| \\
&\quad + \|PS(t)x_{s,k} - PS(t)u(s)\| \\
&\leqslant \|u(s) - x_{s,k}\| + \|S(t)x_{s,k} - PS(t)x_{s,k}\| \\
&\quad + \|PS(t)x_{s,k} - PS(t)u(s)\| \\
&\leqslant \frac{1}{3(k+1)} + \frac{1}{3(k+1)} + \|PS(t)x_{s,k} - PS(t)u(s)\|.
\end{aligned}$$

Since $\|S(t)x_{s,k} - S(t)u(s)\| \leqslant \|x_{s,k} - u(s)\| \leqslant 1/(\omega(B+1,3k+2)+1)$ (using nonexpansivity of $S(t)$) as well as $\|S(t)x_{s,k} - p\| \leqslant \|x_{s,k} - p\| \leqslant B+1$ and $\|S(t)u(s) - p\| \leqslant B \leqslant B+1$ (using nonexpansivity of $S(t)$ and that p is a common fixed-point of all $S(t)$), we conclude that $\|PS(t)x_{s,k} - PS(t)u(s)\| \leqslant 1/(3(k+1))$. This yields the claim. ■

Claim 2: For all $k \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$:

$$\exists n \leqslant \Gamma'(k,f) \forall t \in [n, n+f(n)] \left(\|u(t) - Pu(t)\| \leqslant \frac{1}{k+1} \right).$$

Proof of claim 2: For given $k \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$, consider the function $g_{k,f}$ as defined above. Using the fact that u is an almost-orbit with rate of metastability Φ, there is some $n_0 \leqslant \Phi(\omega(B, 3k+2), g_{k,f})$ such that

$$\forall t \in [0, g_{k,f}(n_0)] \left(\|S(t)u(n_0) - u(t+n_0)\| \leqslant \frac{1}{\omega(B, 3k+2)+1} \right).$$

Since $\|S(t)u(n_0) - p\|, \|u(t+n_0) - p\| \leqslant B$, we conclude that

$$\forall t \in [0, g_{k,f}(n_0)] \left(\|PS(t)u(n_0) - Pu(t+n_0)\| \leqslant \frac{1}{3(k+1)} \right).$$

Thus, for $t \in [0, g_{k,f}(n_0)]$, we get

$$\begin{aligned} \|u(t+n_0) - Pu(t+n_0)\| &\leqslant \|u(t+n_0) - S(t)u(n_0)\| + \|S(t)u(n_0) - PS(t)u(n_0)\| \\ &\quad + \|PS(t)u(n_0) - Pu(t+n_0)\| \\ &\leqslant \frac{2}{3(k+1)} + \|S(t)u(n_0) - PS(t)u(n_0)\|. \end{aligned}$$

Using Claim 1, we get

$$\forall t \geqslant \Omega_{n_0}(3k+2) \left(\|S(t)u(n_0) - PS(t)u(n_0)\| \leqslant \frac{1}{3(k+1)} \right),$$

from which follows that

$$\forall t \in [\Omega_{n_0}(3k+2), g_{k,f}(n_0)] \left(\|u(t+n_0) - Pu(t+n_0)\| \leqslant \frac{1}{k+1} \right),$$

and thus

$$\forall t \in [n_0 + \Omega_{n_0}(3k+2), n_0 + g_{k,f}(n_0)] \left(\|u(t) - Pu(t)\| \leqslant \frac{1}{k+1} \right).$$

This yields the claim by the definition of $g_{k,f}$. ■

Claim 3: For all $k, N \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$:

$$\exists n \in [N, \max\{N, \Phi(2k+1, h_{N,f})\}] \forall t \leqslant f(n) \left(\|S(t)u(n) - u(t+n)\| \leqslant \frac{1}{k+1} \right).$$

Proof of claim 3: Since Φ is a rate of metastability for the almost-orbit u, there is $n_0 \leqslant \Phi(2k+1, h_{N,f})$ such that

$$\forall t \leqslant h_{N,f}(n_0) \left(\|S(t)u(n_0) - u(t+n_0)\| \leqslant \frac{1}{2(k+1)} \right),$$

with $h_{N,f}$ defined as above. Writing $n := \max\{N, n_0\} \in [N, \max\{N, \Phi(2k+1, h_{N,f})\}]$, we have for $t \leqslant f(n)$ that

$$\begin{aligned}\|S(t)u(n) - u(t+n)\| &\leqslant \|S(t)u(n) - S(t+n-n_0)u(n_0)\| \\ &\qquad + \|S(t+n+n_0)u(n_0) - u(t+n)\| \\ &\leqslant \|u(n) - S(n-n_0)u(n_0)\| \\ &\qquad + \|S(t+n-n_0)u(n_0) - u(t+n)\|.\end{aligned}$$

Since $n - n_0 \leqslant t + n - n_0 \leqslant h_{N,f}(n_0)$, we conclude the claim. ■

Claim 4: For all $k \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$, there is some $n_0 \leqslant \Gamma'(8k+7, j_{k,f})$ such that

$$\exists n_1 \leqslant \max\{n_0, \Phi(8k+7, h_{n_0,f})\} \forall t \leqslant f(n_1) \left(\|u(n_1) - u(t+n_1)\| \leqslant \frac{1}{2(k+1)} \right).$$

Proof of claim 4: Let $k \in \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{N}$ be given. From Claim 2 with the function $j_{k,f}(n)$ defined as above, we may consider $n_0 \leqslant \Gamma'(8k+7, j_{k,f})$ such that

$$\forall t \in [n_0, n_0 + j_{k,f}(n_0)] \left(\|u(t) - Pu(t)\| \leqslant \frac{1}{8(k+1)} \right).$$

By Claim 3, there exists $n_1 \in [n_0, \max\{n_0, \Phi(8k+7, h_{n_0,f})\}]$ satisfying

$$\forall t \leqslant f(n_1) \left(\|S(t)u(n_1) - u(t+n_1)\| \leqslant \frac{1}{4(k+1)} \right).$$

Since $n_1 \in [n_0, \max\{n_0, \Phi(8k+7, h_{n_0,f})\}] = [n_0, n_0 + j_{k,f}(n_0)]$, we also have $\|u(n_1) - Pu(n_1)\| \leqslant 1/(8(k+1))$. Thus, for any $t \leqslant f(n_1)$:

$$\begin{aligned}\|u(n_1) - u(t+n_1)\| &\leqslant \|u(n_1) - Pu(n_1)\| + \|Pu(n_1) - S(t)u(n_1)\| \\ &\qquad + \|S(t)u(n_1) - u(t+n_1)\| \\ &\leqslant 2\|u(n_1) - Pu(n_1)\| + \|S(t)u(n_1) - u(t+n_1)\| \\ &\leqslant \frac{2}{8(k+1)} + \frac{1}{4(k+1)} = \frac{1}{2(k+1)}\end{aligned}$$

which yields the claim. ■

Lastly, using the n_1 from Claim 4, by triangle inequality it follows that

$$\forall t, t' \in [n_1, n_1 + f(n_1)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right)$$

and this yields the claim of the theorem, noticing that $n_1 \leqslant \Gamma(k, f)$. □

Remark 5.4.7. Similar to Remark 5.4.5, as revealed by the quantitative analysis, the above result already holds in general Banach spaces whenever there exist suitable selections of the duality map and projection. We again refer to Section 5.5 for further comments on this.

This theorem is (essentially) now a true finitization of Xu's original convergence result since it trivially (though non-effectively) implies back the original statement but (if instantiated to sequences t_n with $t_n \to \infty$) only talks about finite initial segments.

Remark 5.4.8. As used above, if X is uniformly convex, then P is uniformly continuous on bounded subsets of X and it should be noted that given a modulus of uniform convexity $\eta : (0,2] \to (0,1]$ in the sense that

$$\forall \varepsilon \in (0,2] \forall x, y \in X \left(\|x\|, \|y\| \leqslant 1 \wedge \|x-y\| \geqslant \varepsilon \to \left\| \frac{x+y}{2} \right\| \leqslant 1 - \eta(\varepsilon) \right),$$

one can compute a modulus of uniform continuity ω for P as used above. Concretely, we want to mention the following result given e.g. in [189]: if $\mathrm{dist}(x, A^{-1}0) \leqslant r$ and

$$\|x-y\| \leqslant \frac{1}{2}\alpha\left(\frac{\varepsilon}{1+r}\right)$$

where

$$\alpha(\varepsilon) = \min\left\{1, \frac{\varepsilon}{4}, \frac{\varepsilon\eta(\varepsilon)}{4(1-\eta(\varepsilon))}\right\},$$

then $\|Px - Py\| \leqslant \varepsilon$. From this, a suitable modulus $\omega(r,k)$ can be immediately derived.

Now, similarly to [108] and as discussed before already, the analysis of Xu's result (by being essentially constructive) allows for the extraction of two kinds of quantitative "translations" and we now focus on the other variant compared to the above which translates the stronger quantitative assumption of a rate of convergence for the almost-orbit into a rate of convergence of the solution of the Cauchy problem towards a zero of the operator A.

Theorem 5.4.9. *Let X be uniformly convex and uniformly smooth and A be m-accretive such that there exists a weak modulus for the convergence condition Ω. Let $\mathcal{S} = \{S(t) \mid t \geqslant 0\}$ be the semigroup generated by A via the exponential formula. Let $A^{-1}0 \neq \emptyset$ with $p \in A^{-1}0$ and assume that P, the nearest point projection onto $A^{-1}0$, is uniformly continuous on bounded subsets of X with a modulus $\omega : \mathbb{N}^2 \to \mathbb{N}$, i.e.*

$$\forall r, k \in \mathbb{N} \forall x, y \in B_r(p) \left(\|x-y\| \leqslant \frac{1}{\omega(r,k)+1} \to \|Px - Py\| \leqslant \frac{1}{k+1} \right),$$

and, without loss of generality, assume that $\omega(r,k) \geqslant k$ for all $r, k \in \mathbb{N}$. Let u be an almost-orbit with a rate of convergence $\Phi : \mathbb{N} \to \mathbb{N}$ on the almost-orbit condition, i.e.

$$\forall k \in \mathbb{N} \forall s \geqslant \Phi(k) \left(\sup_{t \geqslant 0} \|u(s+t) - S(t)u(s)\| \leqslant \frac{1}{k+1} \right).$$

Let $B \in \mathbb{N}^*$ be such that $\|u(t) - p\| \leqslant B$ for all $t \geqslant 0$ and let $f_s : \mathbb{N} \to \mathbb{N}$ for $s \geqslant 0$ be such that f_s is nondecreasing and

$$\forall n \in \mathbb{N} \exists x_{s,n}, y_{s,n} \in X \left(y_{s,n} \in Ax_{s,n} \wedge \|x_{s,n}\|, \|y_{s,n}\| \leqslant f_s(n) \wedge \|x_{s,n} - u(s)\| \leqslant \frac{1}{n+1} \right).$$

Then we have

$$\forall k \forall t, t' \geqslant \max\{\Phi(8k+7), s^* + \max\{\Omega_m(24k+23) \mid m \leqslant s^*\}\} \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right)$$

where $s^* = \Phi(\omega(B, 24k+23))$ and where $\Omega_s(k)$ is defined as in Theorem 5.4.6.

Proof. Given a rate of convergence Φ on the almost-orbit condition, it is clear that $\Phi(k,f) := \Phi(k)$ (ignoring the slight abuse of notation) is a rate of metastability for the almost-orbit. Therefore, by Theorem 5.4.6, we get that the previously constructed $\Gamma(k,f)$ is rate of metastability for the conclusion. As shown in Proposition 2.6 of [118], a function[4] $\rho : (0,\infty) \to \mathbb{N}$ is a Cauchy rate of a sequence iff $\varphi(\varepsilon, f) := \rho(\varepsilon)$ is a rate of metastability (which also holds in our adapted context where we consider rates to be functions operating on natural numbers as errors). Now, using that $\Phi(k,f) = \Phi(k)$, we find by inspection of the defining term that also $\Gamma(k,f)$ is independent of the parameter f. Thus, we get that $\Gamma(k) := \Gamma(k,f)$ is a rate of convergence and the given bound in the above theorem just results by simplifying the expressions accordingly. □

Note in particular that the above result is indeed a consequence of the previous metastability result and does not require one to reiterate the proof. In that way, the metastability result already contained the quantitative information regarding rates of convergence. We refer to [97] for further discussions of such phenomena.

Remark 5.4.10. Such a rate of convergence of the almost-orbit condition as required as a premise in the above theorem, i.e. a $\Phi : \mathbb{N} \to \mathbb{N}$ such that

$$\forall k \in \mathbb{N} \forall s \geqslant \Phi(k) \left(\sup_{t \geqslant 0} \|u(s+t) - S(t)u(s)\| \leqslant \frac{1}{k+1} \right)$$

[4]Note the typo in [118] where it instead says $\rho : (0,\infty) \to (0,\infty)$

can actually be derived from the seemingly weaker assumption on the existence of a Φ such that

$$\forall k \in \mathbb{N} \exists s_0 \leqslant \varphi(k) \left(\sup_{t \geqslant 0} \|u(s_0 + t) - S(t)u(s_0)\| \leqslant \frac{1}{k+1} \right).$$

Namely, for given $k \in \mathbb{N}$ and for $s \geqslant s_0$ with $s_0 \leqslant \varphi(2k+1)$ as stipulated above, we can express $s = s_0 + s_1$ for $s_1 \geqslant 0$ and then compute for all $t \geqslant 0$:

$$\begin{aligned}
\|u(s+t) - S(t)u(s)\| &= \|u(s_0 + s_1 + t) - S(t)u(s_0 + s_1)\| \\
&\leqslant \|u(s_0 + s_1 + t) - S(t + s_1)u(s_0)\| \\
&\qquad + \|S(t + s_1)u(s_0) - S(t)u(s_0 + s_1)\| \\
&\leqslant \frac{1}{2(k+1)} + \|S(t)S(s_1)u(s_0) - S(t)u(s_0 + s_1)\| \\
&\leqslant \frac{1}{2(k+1)} + \|S(s_1)u(s_0) - u(s_0 + s_1)\| \\
&\leqslant \frac{1}{k+1}.
\end{aligned}$$

Thus, $\Phi(k) = \varphi(2k+1)$ is actually a full rate of convergence. This remark also applies to the results regarding rates of convergence presented in [108] which thus also essentially depend on a rate of convergence of the almost-orbit condition.

Further, note that in both cases the existence of the assumed bound on $\|u(t) - p\|$ is actually guaranteed by the assumption of u being an almost-orbit and that $A^{-1}0 \neq \varnothing$: the definition implies

$$\exists s^* \sup_{t \geqslant 0} \|u(t + s^*) - S(t)u(s^*)\| \leqslant 1$$

and thus for $p \in A^{-1}0$, we have

$$\begin{aligned}
\|u(t + s^*) - p\| &\leqslant \|u(t + s^*) - S(t)u(s^*)\| + \|S(t)u(s^*) - p\| \\
&\leqslant 1 + \|u(s^*) - p\| < \infty
\end{aligned}$$

for all $t \geqslant 0$. As u is continuous, we get that

$$\sup_{t \in [0, s^*]} \|u(t) - p\| < \infty$$

which implies that $u(t) - p$ is bounded in norm. Note that a concrete bound can therefore be computed using a modulus of continuity for u on bounded sets together with a rate of convergence Φ on the almost-orbit condition and a norm upper bound on p.

5.5 Logical aspects of the above results

The extractions presented in this chapter rest on some logical considerations which we want to discuss in this section by sketching in what ways the previous system H^ω_p needs to be extended to recognize the quantitative results presented in this chapter as instances of the general logical metatheorems established in Chapter 4. For this, we write H^ω for the variant of H^ω_p where the range condition is replaced with axiom (II) from Chapter 3 (or, in other words, if the extensions that make H^ω_p result from $\mathcal{V}^\omega_p$ are applied instead to $\mathcal{V}^\omega$). This H^ω provides the basic system for this section. In this section, we switch from a representation of errors via $1/(k+1)$ to 2^{-k} to be more in line with the presentation from Chapter 4. This has no real practical consequences.

5.5.1 Uniform convexity and projections

As discussed already in some of the earliest papers on the treatment of abstract spaces in proof mining (see [71, 95]), uniformly convex spaces can be treated by adding an additional constant together with a corresponding universal axiom to express that this new constant represents a modulus of uniform convexity (see also the later Chapter 7).

In the works [154, 208], the uniform convexity is only assumed to infer the existence of an (in the case of Xu, uniformly continuous) selection of the projection map onto closed and convex subsets of X. In fact, the only selection map of a projection ever needed is a selection of the projection onto the set $A^{-1}0$ which we as before denote just by P. For that, the set $A^{-1}0$ is assumed to be non-empty which can be hardwired into the language of the systems by adding a designated constant p_0 of type X together with the corresponding axiom

$$0 \in Ap_0. \tag{NE}$$

In the context of the above systems for the treatment of m-accretive operators and their extensions, this kind of projection map can be immediately treated by adding a further constant P of type $X(X)$ together with the axiom scheme

$$\forall x^X, p^X \left(0 \in A(Px) \wedge (0 \in Ap \rightarrow \|x -_X Px\|_X \leq_{\mathbb{R}} \|x -_X p\|_X)\right), \tag{P1}$$

characterizing that P is indeed a selection of the projection onto the set $A^{-1}0$. In particular, note also that these axioms are in particular purely universal as the statement $0 \in Ap$ in the context of the system H^ω is quantifier-free, being an abbreviation for $\chi_A(p, 0) =_0 0$ (see again Chapter 3). Note also again that in that way, as stressed before, the treatment of the projection does not require it to be unique but only to be a suitable selection from the potentially multi-valued nearest point projection.

Further, it is immediate from the axioms that P is provably majorizable in $H^\omega + (P1) + (NE)$ as we can prove

$$\|Px\| \leqslant \|x\| + \|x - Px\| \leqslant \|x\| + \|x - p_0\| \leqslant 2\|x\| + \|p_0\|$$

from the axioms $(P1)$ and (NE).

If extensionality or continuity is needed for the projection P (as is the case in the context of Xu's result), the above system needs to be extended with a modulus of uniform continuity ω^P of type $0(0)(0)$ together with a corresponding axiom like

$$\begin{cases} \forall r^0, k^0, x^X, y^X \Big(\|x -_X p_0\|_X, \|y -_X p_0\|_X <_{\mathbb{R}} r \\ \qquad \wedge \ \|x -_X y\|_X <_{\mathbb{R}} 2^{-\omega^P(r,k)} \to \|Px -_X Py\|_X \leqslant_{\mathbb{R}} 2^{-k} \Big). \end{cases} \tag{P2}$$

In that way, the bound extraction theorems stated in Theorem 4.4.6 and Theorem 4.4.7 immediately extend to the system $H^\omega + (NE) + (P1)$ $(+(P2))$ where one then additionally requires n to satisfy $n \geqslant \|p_0\|$ (and in the case of $(P2)$, Φ additionally depends on ω^P).

5.5.2 Uniform smoothness and the normalized duality map

Regarding uniformly smooth spaces, we focus on the dual characterization of such spaces via the requirement of a single-valued duality map J which is norm-to-norm uniformly continuous on bounded subsets (recall again Section 5.2).

As becomes clear through inspection of the analyses presented above, they actually "only" require a function $j' : X \to X^*$ which selects a specific point from the duality set, namely such that

1. $\langle y - Px, j'(x - Px)\rangle \leqslant 0$ for all $x \in X$ and all $y \in A^{-1}0$;

2. $\langle u - v, j'(x - y)\rangle \geqslant 0$ for all $(x, u), (y, v) \in A$.

Both properties are satisfied for the unique selection if X is uniformly smooth and A is m-accretive with $A^{-1}0 \neq \varnothing$ and where a selection P of the projection onto that set exists as above. But, actually, *any* such selection suffices which is in particular suggested by the proof-theoretic perspective.

In that way, we find that the use of the duality map made in the above extractions can be formalized by using the approach to axiomatizations of duality selection maps developed in [111] and previously used for our treatment of the alternative notion of accretivity in Chapter 4. In that vein, we can thus treat the existence of such a map by extending $H^\omega + (P1) + (NE)$ $(+(P2))$ by a constant j' of type $1(X)(X)$ together with the axiom

$$(j')\quad \begin{cases} \forall x^X, y^X \Big(\langle x, j'x \rangle =_{\mathbb{R}} \|x\|_X^2 \wedge |\langle y, j'x \rangle| \leqslant_{\mathbb{R}} \|x\|_X \|y\|_X \\ \qquad\qquad \wedge\ \forall \alpha^1, \beta^1, u^X, v^X \left(\langle \alpha u +_X \beta v, j'x \rangle =_{\mathbb{R}} \alpha\langle u, j'x\rangle + \beta\langle v, j'x \rangle \right) \Big), \end{cases}$$

which instantiates the axiom from [111], stating that j' is indeed a selection (recall also Chapter 4), together with two additional axioms expressing the above properties (1) and (2)

$$(M1)\ \forall x^X, y^X \left(0 \in Ay \to \langle y -_X Px, j'(x -_X Px)\rangle_X \leqslant_{\mathbb{R}} 0\right),$$

$$(M2)\ \forall x^X, y^X, u^X, v^X \left(u \in Ax \wedge v \in Ay \to \langle u -_X v, j'(x -_X y)\rangle_X \geqslant_{\mathbb{R}} 0\right).$$

The bound extraction theorems stated before also here immediately extend to the system $H^\omega + (P1) + (NE) + (j') + (M1) + (M2)$ $(+(P2))$ as is immediately clear through the discussion in [111] and the fact that all the new axioms are universal.

Remark 5.5.1. Clearly, the system $H^\omega + (P1) + (NE) + (j') + (M1) + (M2)$ is over-specified regarding the accretivity of A as H^ω already contains a family of selection functionals witnessing accretivity but $(M2)$ says that j' is a uniform witness for this property. This over-specification has no impact on the theoretical results, however, so we do not "trim" the system.

Now, by itself, the existence of a selection functional for the duality map in particular does not imply that the latter has to be single-valued. However, as shown by Körnlein [124], the existence of a selection functional which is uniformly norm-to-norm continuous is actually equivalent to uniform smoothness of the space X and thus actually implies that this selection is the unique selection. As discussed in [111], the uniform continuity of the selection is already implied by the logical methodology in the case that the proof relies on the extensionality of it.

However, as the above analysis shows, the proofs of Nevanlinna and Reich as well as Xu do not rely on uniform continuity or even extensionality of j' and, in that way, can be formalized in the system $H^\omega + (P1) + (NE) + (j') + (M1) + (M2)$ (with $(P2)$ in the case of Xu) which also explains the absence of any moduli of uniform smoothness for the space or any moduli of uniform continuity for the selection j' in the analysis.

In particular this additionally shows that the results are already valid in the context of the existence of a selection functional satisfying (1) and (2) which is potentially weaker than uniform smoothness.

This insight that conditions (1) and (2) are sufficient, now here facilitated via a proof-theoretic method, was essentially already observed in the last section of the work of Nevanlinna and Reich [154] although it was not stressed in this abstract nature. Instead, they list additional conditions on the operator in order to guarantee that the conditions (1) and (2) are naturally satisfied. Concretely, they require that the operator then is accretive in the sense of Browder [29] to enable that the condition (2) is satisfied for any possible selection j' of J and they require that $A^{-1}0$ is a so-called proximal sun (see [154]) in order to guarantee that a selection satisfying (1) always exists and they require that the semigroup is differentiable so that the orbit generated by the Crandall-Liggett formula is actually a solution of the corresponding initial valued problem (as shown in [50]) which was previously guaranteed by the uniform convexity and uniform smoothness. In the vein of the previous logical discussion, we thus find that our above quantitative results also apply to these generalizations.

5.5.3 Logical aspects of the convergence condition

Besides the quantitative analyses of the results of Nevanlinna and Reich as well as Xu, the main contribution of this chapter is the introduction of the new notions of "moduli for the convergence condition". Already in Bishop's work [17], arguments for the functional interpretation as the correct numerical interpretation of theorems of the form $\exists\forall \to \exists\forall$ are given and, in modern times, the proof mining program has been very effective in arguing that the monotone functional interpretation (in combination with a negative translation) provides the right numerical information in the search for uniform bounds in analysis (see in particular the detailed discussion in [116]). In the following, we will now see how, through this lens, these moduli actually arise from the underlying logical methodology and thus, in various ways, represent the real finitary core of the convergence condition from both a classical and a constructive perspective.

The convergence condition from a classical perspective

Based on the equivalence laid out in Lemma 5.3.1, any proof that a class of operators satisfies the convergence condition, written in the immediate formal translation

$$\forall (x_n)^{X(0)}, (y_n)^{X(0)}, K^0 \Bigg(\forall i^0 \left(y_i \in Ax_i \wedge \|x_i\|_X , \|y_i\|_X \leqslant_{\mathbb{R}} K\right) \\ \wedge \forall a^0 \exists b^0 \forall c^0 \left(c \geqslant_0 b \to |\langle y_c, j'(x_c -_X Px_c)\rangle| \leqslant_{\mathbb{R}} 2^{-a}\right) \\ \to \forall k^0, N^0 \exists n^0 \left(n \geqslant_0 N \wedge \|x_n -_X Px_n\|_X \leqslant_{\mathbb{R}} 2^{-k}\right) \Bigg), \tag{1}$$

can be transformed into a proof for satisfying the equivalent statement

$$\forall x^X, y^X, K^0, k^0 \exists n^0 \Bigg(y \in Ax \wedge \|x\|_X , \|y\|_X \leqslant_{\mathbb{R}} K \\ \wedge\ |\langle y, j'(x -_X Px)\rangle| \leqslant_{\mathbb{R}} 2^{-n} \to \|x -_X Px\|_X \leqslant_{\mathbb{R}} 2^{-k} \Bigg), \tag{2}$$

however at the expense of using classical logic as well as countable choice. However, this use of countable choice is in essence only applied to a quantifier-free formula and thus is an instance of QF-AC. It is clear that (after equivalently writing (2) with $<_{\mathbb{R}}$ in the conclusion to make the inner matrix existential) the negative translation of (2) is equivalent to its original version by the use of Markov's principle and thus that the negative translation followed by the monotone functional interpretation, applied to (2), immediately produces a full modulus (as defined in Definition 5.3.4) as the suggested finitization of this variant of the convergence condition.

Thus, a priori, through the application of the classical metatheorem given in Theorem 4.4.6, we have the following:

Proposition 5.5.2. *There are primitive-recursive (in the sense of Gödel) translations which transform any full modulus for the convergence condition into a solution of the negative translation followed by the monotone functional interpretation of* (1)*, and vice versa.*

Therefore, the two variants of the convergence condition and the accompanying moduli can be extracted from proofs and used interchangeably without yielding a far increase of complexity beyond the principles used in the proof. Thus the bound extraction result discussed in Theorem 4.4.6 guarantees the extractability of such moduli even from classical proofs that A satisfies the convergence condition, provided that the proof can be formalized in $H^\omega + (P1) + (NE) + (j') + (M1) + (M2) + \Delta$ (which we abbreviate in the following by $\mathcal{C}^\omega$) for suitable Δ or any extension/fragment thereof pertaining to the bound extraction theorems. Even further, as already hinted on in Remark 5.3.6, this extraction is already possible from suitable proofs of the much weaker requirement

$$\forall (x,y) \in A\ (\langle y, j'(x - Px)\rangle = 0 \to \|x - Px\| = 0)\,.$$

In that way, a formalized version of the argument in Remark 5.3.6 in fact shows the following:

Proposition 5.5.3. *If $\mathcal{C}^\omega$ (or any suitable extension or fragment thereof) proves that A satisfies*

$$\forall (x,y) \in A\ (\langle y, j'(x -_X Px)\rangle =_{\mathbb{R}} 0 \to \|x -_X Px\|_X =_{\mathbb{R}} 0)\,,$$

then from the proof one can extract a (potentially bar-recursively) computable full modulus for the convergence condition. If the proof does not use DC*, then the modulus is even primitive recursive in the sense of Gödel.*

This in particular also holds if there exists a suitable proof of the convergence condition itself as this proof can be transformed into a proof of the above property (without any additional use of classical logic or choice).

However, the modularity of the approach to quantitative information via the monotone functional interpretation further yields that from any proof using the convergence condition as a premise (formulated in any variant (1) or (2) as discussed above) and formalizable in the respective systems, quantitative information on the conclusion can be extracted which depends then additionally on such a modulus solving the monotone functional interpretation of the convergence condition. This is collected in the following derived metatheorem:

Theorem 5.5.4. *Under the assumptions of Theorem 4.4.6, we have the following: If*

$$\mathcal{C}^\omega \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(A \text{ satisfies the convergence condition } \to \exists v^0 C_\exists(x,y,z,v)\right),$$

then one can extract a bar-recursively computable partial function

$$\Phi : S_\delta \times S_{\hat{\tau}} \times (S_{0(0)(0)})^2 \times \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$$

such that for all $x \in S_\delta$, $z \in S_\tau$, $z^* \in S_{\hat{\tau}}$, $\Omega^f, \omega' \in S_{0(0)(0)}$ and all $n \in \mathbb{N}$ and $\omega \in \mathbb{N}^{\mathbb{N}}$, if

$z^* \gtrsim z$, $\omega \gtrsim \omega^S$ and $\omega' \gtrsim \omega^P$ as well as $n \geqslant_{\mathbb{R}} \left\|J^A_1(0)\right\|_X, \|p_0\|_X$,[5]*, then*

$$\mathcal{S}^{\omega,X} \models \forall y \leqslant_\sigma s(x)\Big(\Omega^f \text{ is a full modulus for the convergence condition for } A \\ \to \exists v \leqslant_0 \Phi(x,z^*,\Omega^f,\omega',n,\omega)\, C_\exists(x,y,z,v)\Big)$$

holds whenever $\mathcal{S}^{\omega,X} \models \Delta$ for $\mathcal{S}^{\omega,X}$ defined as in Theorem 4.4.6 for the constants of H^ω and with the other new constants of $\mathcal{C}^\omega$ naturally interpreted so that the respective axioms are satisfied.

Moreover: if the proof does not use DC*, then the modulus is even primitive recursive in the sense of Gödel. The result remains true for any suitable extension or fragment of $\mathcal{C}^\omega$.*

In that way, by Proposition 5.5.3 and Theorem 5.5.4, we find that a full modulus is indeed the right quantitative notion for the convergence condition in the sense that both items (1) and (2), discussed before as the central properties on page 85, are fulfilled.

The convergence condition from a constructive perspective

From the semi-constructive perspective of the monotone modified realizability interpretation and the associated system $\mathcal{V}^\omega_i$ and its extensions, the quantitative version of the convergence condition is exactly what is captured by the notion of the "plain" modulus introduced in Definition 5.3.3.

For this, we now work over the semi-constructive variant of the previous theories. Concretely, we abbreviate with $\mathcal{C}^\omega_i$ in the following the system $H^\omega_i + \mathrm{IP}_\neg + \mathrm{CA}_\neg + (P1) + (NE) + (j') + (M1) + (M2) + \Gamma_\neg$ for suitable $\Gamma_\neg$ where H^ω_i results from $\mathcal{V}^\omega_i$ by extending it in the same manner as $\mathcal{V}^\omega$ is extended to form H^ω.

Concretely, applying the monotone modified realizability interpretation to the formal statement (1) considered previously, we get that it asks for a functional Ω which transforms K and majorants for $(x_n),(y_n)$ (which w.l.o.g. are assumed to coincide with the constant K-function and are in that way represented by the input K) and a majorant of a realizer for the premise

$$\forall a \exists b \forall c \left(c \geqslant b \to |\langle y_c, j'(x_c - Px_c)\rangle| \leqslant 2^{-a}\right),$$

[5]We can here simplify the assumptions on n compared to Theorem 4.4.6 as, in the context of H^ω, the resolvents are all total so that we can pick $m_{\widetilde{\gamma}}, m_{\lambda_0}, m'_{\widetilde{\gamma}} = 0$ with $\widetilde{\gamma} = 1$, $\lambda_0 = 2$ as well as $c_X = 0$ and $d_X = 0 -_X J^A_1(0)$.

i.e. of a φ of type 1 such that

$$\forall a, c \left(c \geqslant \varphi(a) \to |\langle y_c, j'(x_c - Px_c)\rangle| \leqslant 2^{-a}\right)$$

into a majorant of a realizer for the conclusion $\forall k, N \exists n \left(n \geqslant N \wedge \|x_n - Px_n\| \leqslant 2^{-k}\right)$, i.e. into an $\Omega(K, \varphi)$ of type $0(0)(0)$ such that

$$\forall k, N \exists n \leqslant \Omega(K, \varphi)(k, N) \left(n \geqslant N \wedge \|x_n - Px_n\| \leqslant 2^{-k}\right).$$

Thus, this is exactly what is represented by a "plain" modulus for the convergence condition.

An immediate application of the bound extraction result contained in Theorem 4.4.7 yields the following result, similarly to the previous Proposition 5.5.3.

Proposition 5.5.5. *If $\mathcal{C}_i^\omega$ (or any suitable extension or fragment thereof) proves*

$$\forall (x, y) \in A \ (\langle y, j'(x -_X Px)\rangle =_{\mathbb{R}} 0 \to \|x -_X Px\|_X =_{\mathbb{R}} 0),$$

then from the proof one can extract a primitive-recursive full modulus for the convergence condition.

As discussed before, this in particular also holds if there exists a suitable proof of the convergence condition.

Note that in the presence of the previous Proposition 5.5.3, the above result is nevertheless not void. While an intuitionistic proof is especially a classical proof, Proposition 5.5.3 of course guarantees already the extractability of a full modulus. However, this only applies in the case that the additional axioms $\Gamma_\neg$ potentially contained in the above system $\mathcal{C}_i^\omega$ have a monotone functional interpretation as required by Proposition 5.5.3. So if the real strength of $\Gamma_\neg$ is used while restricting to intuitionistic logic, then the above result nevertheless guarantees the existence and extractability of a primitive recursive full modulus.

Now, in similarity to Theorem 5.5.4, we obtain a macro for the logical metatheorem contained in Theorem 4.4.7 which guarantees that now from a semi-constructive proof of a result using the convergence condition as a premise, one can extract a transformation which transforms any modulus for the convergence condition into information on the conclusion, even in the presence of the axioms $\Gamma_\neg$.

Theorem 5.5.6. *Under the assumptions of Theorem 4.4.7 we have the following: If*

$$\mathcal{C}_i^\omega \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(A \text{ satisfies the convergence condition } \to \exists u^0 C(x,y,z,u)\right),$$

one can extract a

$$\Phi : S_\delta \times S_{\widehat{\tau}} \times S_{1(0)(1)(0)} \times S_{0(0)(0)} \times \mathbb{N} \times \mathbb{N}^\mathbb{N} \to \mathbb{N}$$

with is primitive recursive in the sense of Gödel such that for any $x \in S_\delta$, any $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$, any $z \in S_\tau$ and $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$[6] and any $n \in \mathbb{N}$, $\Omega \in S_{1(0)(1)(0)}$, $\omega' \in S_{0(0)(0)}$, $\omega \in \mathbb{N}^\mathbb{N}$ with $n \geqslant_\mathbb{R} \left\|J_1^A(0)\right\|_X, \|p_0\|_X$ as well as $\omega \gtrsim \omega^S$ and $\omega' \gtrsim \omega^P$ we have that

$$\mathcal{S}^{\omega,X} \models \exists u \leqslant_0 \Phi(x, z^*, \Omega, \omega', n, \omega)\Big(\Omega \text{ is a modulus for the convergence condition for } A \to C(x,y,z,u)\Big)$$

holds whenever $\mathcal{S}^{\omega,X} \models \Gamma_\neg$ for $\mathcal{S}^{\omega,X}$ defined as in Theorem 4.4.7 for the constants of H^ω and with the other new constants of $\mathcal{C}_i^\omega$ naturally interpreted so that the respective axioms are satisfied.

Note lastly that it is also this result which a priori guaranteed the dependence of the quantitative versions of the result of Nevanlinna and Reich as well as Xu on our "plain" modulus instead of on the full modulus and which in that way lies behind the extraction.

[6] Here, $\gtrsim$ denotes (not necessarily strong) majorization interpreted in the model $\mathcal{S}^{\omega,X}$, as before.

6 Rates of convergence for the asymptotic behavior of second-order Cauchy problems

6.1 Introduction

While the previous Chapter 5 was concerned with the theorems of Nevanlinna and Reich as well as Xu in the context of first-order systems, we are in this short chapter now concerned with a result due to Poffald and Reich [171] which extends the work of Nevanlinna and Reich to incomplete second-order Cauchy problems. Namely, for the second-order system

$$\begin{cases} u''(t) \in Au(t),\ 0 < t < \infty, \\ u(0) = x, \\ \sup\{\|u(t)\| \mid t \geqslant 0\} < \infty, \end{cases} \tag{$\dagger$}$$

over a uniformly smooth and uniformly convex Banach space X with a strongly monotone duality map[1] J, i.e. for a constant $M > 0$ it holds that $\langle x - y, Jx - Jy\rangle \geqslant M\|x - y\|^2$ for all $x, y \in X$, and A m-accretive as before, the solution set

$$\mathcal{S} = \{u(t) \mid u \text{ is a solution to } (\dagger) \text{ for some } x \text{ in the sense of [171, Theorem 2.8]}\}$$

is a nonlinear semigroup for $x \in \mathrm{dom}A$ as shown in [171]. Thus, by the results from [172], this semigroup is generated by some unique m-accretive operator which is denoted by $A_{1/2}$ and called the square root of A. Similarly, we write $\mathcal{S}_{1/2}$ for this semigroup. Various properties of this semigroup and the accompanying system were exhibited in [171], generalizing previous work in the context of Hilbert spaces by Barbu [3] as well as Brezis [24]. In particular, Poffald and Reich obtained the following result on the asymptotic behavior of the semigroup:

[1] As shown in [171, Proposition 2.11], a smooth Banach space has a strongly monotone duality map if, and only if, it is uniformly convex with a modulus of convexity of power type 2.

Theorem 6.1.1 (Poffald and Reich [171]). *Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J, i.e. for a constant $M > 0$ it holds that $\langle x - y, Jx - Jy\rangle \geqslant M \|x - y\|^2$ for all $x, y \in X$, and A be m-accretive with $A^{-1}0 \neq \varnothing$ and such that it satisfies the convergence condition. If $\mathcal{S}_{1/2} = \{S_{1/2}(t) \mid t \geqslant 0\}$ is the semigroup generated by $A_{1/2}$ via the exponential formula as above, then $S_{1/2}(t)x$ converges strongly to a zero of A for $t \to \infty$ for any $x \in \overline{\mathrm{dom} A}$.*

In this chapter, we exhibit the quantitative content of this result by extracting an explicit and computable transformation from the proof of Theorem 6.1.1 which translates the previously introduced *modulus of the convergence condition*, together with some minor quantitative data, into a full rate of convergence for the strong convergence of $S_{1/2}(t)x$ to a zero of A.

For simplicity, we formulate all the results only for the full moduli from Chapter 5 for simplicity. It should however be noted that the results also hold already in the context of a "plain" modulus for the convergence condition as before.

Going beyond the range of proof mining however, we are here further concerned with generalizations of the theorem of Poffald and Reich to new results. As discussed in Chapter 5, Xu [208] studied the behavior of almost-orbits associated with the semigroup generated by A (see Theorem 5.1.3).

Combining the ideas of the quantitative analysis obtained in the previous Chapter 5 of this result of Xu together with the quantitative version of the result of Poffald and Reich established in this chapter, we obtain a similar quantitative version of a result on almost-orbit convergence for the semigroup $\mathcal{S}_{1/2}$. This result, while finitary in nature, in particular also implies back the following basic "infinitary" result for $\mathcal{S}_{1/2}$ which is similar to Xu's result from before:

Theorem 6.1.2. *Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J with value $M > 0$, i.e. $\langle x - y, Jx - Jy\rangle \geqslant M \|x - y\|^2$ for all $x, y \in X$. Let A be m-accretive such that it satisfies the convergence condition and that $A^{-1}0 \neq \varnothing$. Let $\mathcal{S}_{1/2} = \{S_{1/2}(t) \mid t \geqslant 0\}$ be the semigroup generated by $A_{1/2}$ via the exponential formula. Then every almost-orbit $u(t)$ of $\mathcal{S}_{1/2}$ converges strongly to a zero of A as $t \to \infty$.*

This result on the behavior of almost-orbits in the case of $\mathcal{S}_{1/2}$ seems to be new to the literature and the approach taken here to establish it in particular exhibits the strength of quantitative analyses obtained in the proof mining program as these exhibit the real finitary core of a mathematical proof, stripped of any non-essential notions and arguments, which sometimes allows for easy generalizations that lead to new results.

6.2 An analysis of Poffald's and Reich's result

To derive a quantitative version of the convergence result contained in Theorem 6.1.1, by means of applying a modulus for the convergence condition, we first have to extract from the proof given in [171] explicit quantitative bounds on the norms of the orbits and their derivatives involved.

For that, we follow the way a solution for the associated system (†) is constructed in [171] (which differs in comparison to the construction of Barbu [3] (see also [24]) who considered this problem in the context of Hilbert spaces before Poffald and Reich). To solve (†), Poffald and Reich first solve the system

$$\begin{cases} u''(t) \in Au(t) + pu(t),\ 0 < t < \infty, \\ u(0) = x, \\ u \in L^2(0,\infty;X), \end{cases} \tag*{$(\dagger)_p$}$$

in $W^{2,2}(0,\infty;X)$ for $p \to 0^+$ which in turn is solved by solving the approximate system

$$\begin{cases} u''(t) = A_r u(t) + pu(t),\ 0 < t < \infty, \\ u(0) = x, \\ u \in L^2(0,\infty;X), \end{cases} \tag*{$(\dagger)_p^r$}$$

for $r \to 0^+$ where A_r is the Yosida approximate.

In the latter case, they conclude that the unique solution u_p^r of $(\dagger)_p^r$ converges in $L^2(0,\infty;X)$ and $C([0,\infty);X)$ to a (unique) solution u_p of $(\dagger)_p$. For the approximate solutions u_p^r, the following bounds on u_p^r and its derivatives are obtained in [171]:

- $\left\|u_p^r(t)\right\| \leqslant \|x\|$ for all $t \geqslant 0$ (p. 521, (2.7));
- $\int_0^\infty \left\|u_p^{r\prime}(t)\right\|^2 \mathrm{dt} \leqslant 2/M^2(d(0,Ax) + p\|x\|)^{1/2}\|x\|^{3/2}$ (p. 522, (2.17));
- $\int_0^\infty \left\|u_p^{r\prime\prime}(t)\right\|^2 \mathrm{dt} \leqslant 2/M^2(d(0,Ax) + p\|x\|)^{3/2}\|x\|^{1/2}$ (p. 522, (2.14)).

As remarked in [171], these bounds immediately transfer to the solution u_p of $(\dagger)_p$ by applying Lemma 2.6 of [171] to $u_p^r \to u_p$ for $r \to 0^+$.

Following [171], these bounds can then be used to establish bounds on the respective norms of a solution u to (†) by applying Lemma 2.7 of [171] to the convergence $u_p \to u$ for $p \to 0^+$ which immediately yields the following bounds for the solution of (†) corresponding to the initial value x:

- $\|u(t)\| \leqslant \|x\|$ for all $t \geqslant 0$;
- $\int_0^\infty \|u'(t)\|^2 \,\mathrm{dt} \leqslant 2/M^2 d(0,Ax)^{1/2}\|x\|^{3/2}$;
- $\int_0^\infty \|u''(t)\|^2 \,\mathrm{dt} \leqslant 2/M^2 d(0,Ax)^{3/2}\|x\|^{1/2}$.

The quantitative version of Theorem 6.1.1 now takes the following form for the case of $x \in \mathrm{dom}A$.

Theorem 6.2.1. *Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J with value $M > 0$, i.e. $\langle x-y, Jx-Jy\rangle \geqslant M\|x-y\|^2$ for all $x, y \in X$. Let A be m-accretive with $A^{-1}0 \neq \varnothing$ with $p \in A^{-1}0$ and such that it satisfies the convergence condition with a full modulus for the convergence condition Ω. Let $\mathcal{S}_{1/2} = \{S_{1/2}(t) \mid t \geqslant 0\}$ be the semigroup generated by $A_{1/2}$ via the exponential formula. For any $x \in \mathrm{dom}A$, we have*

$$\forall k \in \mathbb{N}\forall t, t' \geqslant \chi((\Omega(2k+1, \max\{1,d\})+1)^2 \dot{-} 1)\left(\left\|S_{1/2}(t)x - S_{1/2}(t')x\right\| \leqslant \frac{1}{k+1}\right)$$

with $\chi(k) = (D+1)(k+1)$ and where

$$D \geqslant (1+b^2)\frac{2}{M^2}d(0,Ax)^{3/2}d^{1/2}$$

as well as $b \geqslant \|x - Px\|$ and $d \geqslant \|x\|$.

Proof. We write $u(t) = S_{1/2}(t)x$. Then u'' exists almost everywhere, say on $[0,\infty)\backslash N$. As outlined in the discussion before, we have $\|u(t)\| \leqslant \|x\|$ for all $t \geqslant 0$ as well as

$$\int_0^\infty \|u''(t)\|^2 \,\mathrm{dt} \leqslant \frac{2}{M^2}d(0,Ax)^{3/2}\|x\|^{1/2}.$$

Now, using the defining property of the projection P and the definition of u, we have

$$\|u(t+h) - Pu(t+h)\| \leqslant \|u(t+h) - Pu(t)\| \leqslant \|u(t) - Pu(t)\|$$

which in particular implies that

$$\begin{aligned}\int_0^\infty \langle u''(t), J(u(t)-Pu(t))\rangle^2 \,\mathrm{dt} &\leqslant \int_0^\infty \|u''(t)\|^2 \|J(u(t)-Pu(t))\|^2 \,\mathrm{dt}\\ &= \int_0^\infty \|u''(t)\|^2 \|u(t)-Pu(t)\|^2 \,\mathrm{dt}\\ &\leqslant \int_0^\infty \|u''(t)\|^2 \|u(0)-Pu(0)\|^2 \,\mathrm{dt}\\ &\leqslant \frac{2}{M^2} d(0,Ax)^{3/2} \|x\|^{1/2} \|x-Px\|^2 .\end{aligned}$$

Therefore also

$$\begin{aligned}&\int_0^\infty \left(\|u''(t)\|^2 + \langle u''(t), J(u(t)-Pu(t))\rangle^2\right) \mathrm{dt}\\ &\qquad \leqslant (1+\|x-Px\|^2)\frac{2}{M^2} d(0,Ax)^{3/2} \|x\|^{1/2}\\ &\qquad \leqslant D.\end{aligned}$$

Lemma 5.4.1 now implies that for any $k \in \mathbb{N}$:

$$\exists t \in [0,\chi(k)]\backslash N \left(\max\left\{\|u''(t)\|^2, \langle u''(t), J(u(t)-Pu(t))\rangle^2\right\} \leqslant \frac{1}{k+1}\right).$$

Thus in particular, we have

$$\exists t \in [0,\chi((k+1)^2 \dot{-} 1)]\backslash N \left(\max\left\{\|u''(t)\|, \langle u''(t), J(u(t)-Pu(t))\rangle\right\} \leqslant \frac{1}{k+1}\right)$$

which yields

$$\begin{aligned}&\exists t \leqslant \chi((\Omega(k,\max\{1,d\})+1)^2 \dot{-} 1)\\ &\qquad \left(\max\left\{\|u''(t)\|, \langle u''(t), J(u(t)-Pu(t))\rangle\right\} \leqslant \frac{1}{\Omega(k,\max\{1,d\})+1}\right)\end{aligned}$$

and thus, as $\|u''(t)\| \leqslant 1$ for such a t, the properties of Ω yield that

$$\exists t \leqslant \chi((\Omega(k,\max\{1,d\})+1)^2 \dot{-} 1) \left(\|u(t)-Pu(t)\| \leqslant \frac{1}{k+1}\right).$$

But as discussed above, $\|u(t)-Pu(t)\|$ is decreasing and thus actually

$$\forall t \geqslant \chi((\Omega(k,\max\{1,d\})+1)^2 \dot{-} 1) \left(\|u(t)-Pu(t)\| \leqslant \frac{1}{k+1}\right).$$

As in [171], we can now show

$$\|u(t+h)-u(t)\| \leqslant 2\|u(t)-Pu(t)\|$$

and thus we obtain

$$\forall t \geqslant \chi((\Omega(2k+1, \max\{1,d\})+1)^2 \dot{-} 1)\forall h\left(\|u(t+h)-u(t)\| \leqslant \frac{1}{k+1}\right)$$

which is the claim. □

By continuity of $S_{1/2}$, the result for $x \in \mathrm{dom}A$ extends to $x \in \overline{\mathrm{dom}A}$ and by an analysis of this proof, we obtain the following quantitative result for the extension.

Theorem 6.2.2. *Assume the conditions of Theorem 6.2.1. Let $x \in \overline{\mathrm{dom}A}$ with $f : \mathbb{N} \to \mathbb{N}$ be such that f is nondecreasing and*

$$\forall k \in \mathbb{N} \exists z, y \in X\left(z \in Ay \wedge \|y\|, \|z\| \leqslant f(k) \wedge \|x-y\| \leqslant \frac{1}{k+1}\right).$$

Then

$$\forall k \in \mathbb{N}\forall t, t' \geqslant \chi_k((\Omega(6k+5, \max\{1, f(3k+2)\})+1)^2 \dot{-} 1)$$
$$\left(\left\|S_{1/2}(t)x - S_{1/2}(t')x\right\| \leqslant \frac{1}{k+1}\right)$$

with $\chi_k(k) = (D_k+1)(k+1)$ and where

$$D_k \geqslant (1+b_k^2)\frac{2}{M^2}f(3k+2)^2$$

as well as $b_k \geqslant \|x-Px\| + \|x\| + f(3k+2)$.

Proof. By the properties of f, we get that there exists $z \in Ay$ such that $\|z\|, \|y\| \leqslant f(3k+2)$ and $\|x-y\| \leqslant 1/(3k+3)$. Therefore

$$\begin{aligned}\left\|S_{1/2}(t)x - S_{1/2}(t')x\right\| &\leqslant \left\|S_{1/2}(t)x - S_{1/2}(t')y\right\| + \left\|S_{1/2}(t)y - S_{1/2}(t')y\right\| \\ &\quad + \left\|S_{1/2}(t')x - S_{1/2}(t')y\right\| \\ &\leqslant 2\|x-y\| + \left\|S_{1/2}(t)y - S_{1/2}(t')y\right\| \\ &\leqslant \frac{2}{3(k+1)} + \left\|S_{1/2}(t)y - S_{1/2}(t')y\right\|.\end{aligned}$$

Using the previous Theorem 6.2.1, we get that

$$\forall k \in \mathbb{N}\forall t, t' \geqslant \chi_k((\Omega(6k+5, \max\{1, f(3k+2)\})+1)^2 \dot{-} 1)$$
$$\left(\left\|S_{1/2}(t)y - S_{1/2}(t')y\right\| \leqslant \frac{1}{3k+3}\right)$$

since

$$\begin{aligned}\|y - Py\| &\leqslant \|x - Px\| + |\|y - Py\| - \|x - Px\|| \\ &\leqslant \|x - Px\| + \|y - x\| \\ &\leqslant \|x - Px\| + \|x\| + f(3k + 2) \\ &\leqslant b_k\end{aligned}$$

as well as $\|y\| \leqslant f(3k + 2)$ and $d(0, Ay) \leqslant \|z\| \leqslant f(3k + 2)$. This gives the claim. □

6.3 A generalization to almost-orbits

We now generalize the result of Xu from the first-order to the second-order case. The new result in Theorem 6.1.2 follows from the following quantitative result which itself arises as a generalization of the quantitative version of Xu's result (Theorem 5.4.6) given in Chapter 5.

Theorem 6.3.1. *Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J with value $M > 0$, i.e. $\langle x - y, Jx - Jy\rangle \geqslant M \|x - y\|^2$ for all $x, y \in X$. Let A be m-accretive such that it satisfies the convergence condition with a full modulus for the convergence condition Ω. Let $\mathcal{S}_{1/2} = \{S_{1/2}(t) \mid t \geqslant 0\}$ be the semigroup generated by $A_{1/2}$ via the exponential formula. Let $A^{-1}0 \neq \varnothing$ with $p \in A^{-1}0$ and assume that P, the nearest point projection onto $A^{-1}0$, is uniformly continuous on bounded subsets of X with a modulus $\omega : \mathbb{N}^2 \to \mathbb{N}$, i.e.*

$$\forall r, k \in \mathbb{N} \forall x, y \in \overline{B}_r(p) \left(\|x - y\| \leqslant \frac{1}{\omega(r, k) + 1} \to \|Px - Py\| \leqslant \frac{1}{k + 1} \right),$$

and, without loss of generality, assume that $\omega(r, k) \geqslant k$ for all $r, k \in \mathbb{N}$. Let u be an almost-orbit of $\mathcal{S}_{1/2}$ with a rate of metastability Φ on the almost-orbit condition, i.e.

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Phi(k, f) \forall t \in [0, f(n)] \\ \left(\|S_{1/2}(t)u(n) - u(t + n)\| \leqslant \frac{1}{k + 1} \right).$$

Let $B \in \mathbb{N}^*$ be such that $\|u(t) - p\| \leqslant B$ for all $t \geqslant 0$ and let $f_s : \mathbb{N} \to \mathbb{N}$ for $s \geqslant 0$ be such that f_s is nondecreasing and

$$\forall n \in \mathbb{N} \exists x_{s,n}, y_{s,n} \in X \left(y_{s,n} \in Ax_{s,n} \wedge \|x_{s,n}\|, \|y_{s,n}\| \leqslant f_s(n) \wedge \|x_{s,n} - u(s)\| \leqslant \frac{1}{n + 1} \right).$$

Then we have

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Gamma(k, f) \forall t, t' \in [n, n + f(n)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right),$$

where

$$\Gamma(k, f) := \max\{\Gamma'(8k+7, j_{k,f}), \Phi(8k+7, h_{N,f}) \mid N \leqslant \Gamma'(8k+7, j_{k,f})\}$$

with

$$\begin{aligned} h_{N,f}(n) &:= f(\max\{N, n\}) + \max\{N, n\} - n, \\ j_{k,f}(n) &:= \max\{n, \Phi(8k+7, h_{n,f})\} - n \\ g_{k,f}(m) &:= \Omega_m(3k+2) + f(m + \Omega_m(3k+2)), \\ \Gamma'(k, f) &:= \Phi(\omega(B, 3k+2), g_{k,f}) + \max\{\Omega_m(3k+2) \mid m \leqslant \Phi(\omega(B, 3k+2), g_{k,f})\}, \end{aligned}$$

for $\Omega_s(k)$ with $s \geqslant 0$ defined by

$$\Omega_s(k) := \chi_{s,k}((\Omega(3k+2, \max\{1, f_s(\omega(B+1, 3k+2))\}) + 1)^2 \dot{-} 1)$$

with $\chi_{s,k}(k) := (D_{s,k}+1)(k+1)$ and where

$$D_{s,k} \geqslant (1 + (B+1)^2) \frac{2}{M^2} f_s(\omega(B+1, 3k+2))^2.$$

We omit the proof as it is, in essence, a careful reimplementation of the proof of Theorem 5.4.6, now using Theorem 6.2.1 in the beginning instead of Theorem 5.4.3 as before.

This finitary result now in particular implies a usual infinitary result on the convergence of almost-orbits of $\mathcal{S}_{1/2}$ as formulated in Theorem 6.1.2 since metastability trivially (though non-effectively) implies back convergence of the respective sequence.

Proof of Theorem 6.1.2. Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J with value $M > 0$ and let A be m-accretive such that it satisfies the convergence condition and that $A^{-1}0 \neq \varnothing$. Let u be an almost-orbit of $\mathcal{S}_{1/2}$. By Proposition 5.3.5, there exists a full modulus for the convergence condition Ω. As in [108], it is rather immediate to see that u has a rate of metastability Φ. Then, for the other minor quantitative data as required in the above theorem (which naturally exist), we get that there exists a function Γ such that

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Gamma(k, f) \forall t, t' \in [n, n + f(n)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right).$$

In particular

$$\forall k \in \mathbb{N} \forall f : \mathbb{N} \to \mathbb{N} \exists n \in \mathbb{N} \forall t, t' \in [n, n + f(n)] \left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right)$$

and this formulation implies the Cauchy property of $u(t)$ as follows: suppose

$$\exists k \in \mathbb{N} \forall n \in \mathbb{N} \exists t, t' \geqslant n \left(\|u(t) - u(t')\| > \frac{1}{k+1} \right)$$

and define $f(n)$ non-effectively such that $f(n) + n \geqslant t, t'$ for these two t, t' guaranteed by this property. Then for that k and f:

$$\forall n \in \mathbb{N} \exists t, t' \in [n, n + f(n)] \left(\|u(t) - u(t')\| > \frac{1}{k+1} \right)$$

which is in contradiction to the metastability of u. □

Lastly, similar to both [108] and to the previous chapter, we can also give the second quantitative version of Theorem 6.1.2, based on the previously discussed strengthened premise of a rate of convergence for the almost-orbit. This then takes the form of the following theorem.

Theorem 6.3.2. *Let X be uniformly convex and uniformly smooth with a strongly monotone duality map J with value $M > 0$, i.e. $\langle x - y, Jx - Jy \rangle \geqslant M \|x - y\|^2$ for all $x, y \in X$. Let A be m-accretive such that there exists a weak modulus for the convergence condition Ω. Let $\mathcal{S}_{1/2} = \{S_{1/2}(t) \mid t \geqslant 0\}$ be the semigroup generated by $A_{1/2}$ via the exponential formula. Let $A^{-1}0 \neq \varnothing$ with $p \in A^{-1}0$ and assume that P, the nearest point projection onto $A^{-1}0$, is uniformly continuous on bounded subsets of X with a modulus $\omega : \mathbb{N}^2 \to \mathbb{N}$, i.e.*

$$\forall r, k \in \mathbb{N} \forall x, y \in \overline{B}_r(p) \left(\|x - y\| \leqslant \frac{1}{\omega(r,k)+1} \to \|Px - Py\| \leqslant \frac{1}{k+1} \right),$$

and, without loss of generality, assume that $\omega(r, k) \geqslant k$ for all $r, k \in \mathbb{N}$. Let u be an almost orbit with a rate of convergence $\Phi : \mathbb{N} \to \mathbb{N}$ on the almost-orbit condition, i.e.

$$\forall k \in \mathbb{N} \forall s \geqslant \Phi(k) \left(\sup_{t \geqslant 0} \left\| u(s+t) - S_{1/2}(t)u(s) \right\| \leqslant \frac{1}{k+1} \right).$$

Let $B \in \mathbb{N}^*$ be such that $\|u(t) - p\| \leqslant B$ for all $t \geqslant 0$ and let $f_s : \mathbb{N} \to \mathbb{N}$ for $s \geqslant 0$ be such that f is nondecreasing and

$$\forall n \in \mathbb{N} \exists x_{s,n}, y_{s,n} \in X \Big(y_{s,n} \in Ax_{s,n} \\ \wedge \|x_{s,n}\|, \|y_{s,n}\| \leqslant f_s(n) \wedge \|x_{s,n} - u(s)\| \leqslant \frac{1}{n+1} \Big).$$

Then we have

$$\forall k \in \mathbb{N} \forall t, t' \geqslant \max\{\Phi(8k+7), s^* + \max\{\Omega_m(24k+23) \mid m \leqslant s^*\}\}$$
$$\left(\|u(t) - u(t')\| \leqslant \frac{1}{k+1} \right)$$

where $s^* = \Phi(\omega(B, 24k+23))$ *and where* $\Omega_s(k)$ *is defined as in Theorem 6.3.1.*

Also here, we omit the proof as it is completely analogous to the proof of Theorem 6.3.2 from Chapter 5.

7 Quantitative asymptotic behavior of nonlinear semigroups

7.1 Introduction

The previous chapters were concerned with conditions under which strong convergence of $S(t)x$ can be guaranteed. In this chapter, we now care for asymptotic results for semigroups generated by accretive operators which are of a more relative flavor, i.e. which provide results that link the asymptotic behavior of the semigroup with that of other objects without guaranteeing convergence outright. Concretely, this chapter provides two case studies on results due to Plant [170] and Reich [174] for the asymptotic behavior of these semigroups and in that context, under suitable quantitative translations of the assumptions used in the respective results, we are able to extract rates of convergence for the limits involved which are moreover polynomial in all data. In particular, we want to note that full rates of convergence are obtained here despite the fact that the sequence in question is not monotone and that the original proof is classical. This is due to a logical particularity that will be discussed after the extractions.

In that way, the current chapter further illustrates the applicability of the formal systems developed in Chapter 4 for the semigroups generated by accretive operators via the exponential formula as it employs the bound extraction theorems introduced before to provide quantitative information on the above mentioned two results. These two theorems of Reich and Plant, respectively, were motivated by the results of Pazy [159] for iterations of nonexpansive mappings and take the following form:[1]

Theorem 7.1.1 (Plant [170]). *Let X be uniformly convex, A be an accretive operator*

[1] To simplify the notation in the following, we drop the superscript of the operator A from the resolvent.

that satisfies the range condition $(RC)_{\lambda_0}$ *and let* $x \in \mathrm{dom}A$*. Then*

$$\lim_{\lambda_0 > t \to 0^+} \frac{\|J_t x - S(t)x\|}{t} = 0.$$

Theorem 7.1.2 (Reich [174])**.** *Let* X *be uniformly convex,* A *be an accretive operator that satisfies the range condition* (RC) *and let* $x \in \mathrm{dom}A$*. Then*

$$\lim_{t \to \infty} \frac{\|J_t x - S(t)x\|}{t} = 0.$$

A usual application of negative translation and monotone functional interpretation as used by the metatheorems suggest the extractability of "metastability-like" rates here (provided that the proof formalizes in the underlying systems). However, as we will see, classical logic features in these proofs only in two ways: at first, it features in some of the basic underlying convergence results in which case the limits are decreasing and a rate of convergence can thus nevertheless be obtained using the metatheorem from Theorem 4.4.6. For both results, the proof then proceeds via a case distinction on real numbers between $= 0$ and > 0. In both results, the proofs for the "$= 0$"-cases are trivial and rates of convergence can be immediately extracted. While the proofs for the "> 0'-cases are nontrivial, they are nevertheless essentially constructive which allows, through the use of the semi-constructive metatheorem of Theorem 4.4.7, for the extraction of full rates of convergence for both limits exhibited above, under the appropriate quantitative reformulations of the "> 0'-assumption, respectively. So, a rate of convergence can be obtained in either case, for both results. Only in the combination of these rates to a rate for the full result, the issues from the use of classical logic could feature but as we will see, in both cases the rates can be smoothed to be combined into a full rate of convergence for the whole result.

The next two sections now present the extractions of the quantitative results and in that context do not explicitly focus on the logical particularities of the extraction which will only be discussed in the last section. In that way, to present these results in a way more amenable to the usual literature of the theory of semigroups, we also move to using ε's for the errors instead of 2^{-k} or similar constructions using natural numbers like $1/(k+1)$.

In general, the main assumption featuring in both results is the uniform convexity of the underlying space which can be treated, as extensively discussed in the proof mining literature starting from the earliest works on the treatment of abstract spaces (see [95]), by a so-called modulus of uniform convexity:

Definition 7.1.3. A modulus of uniform convexity for a space X is a mapping $\eta : (0,2] \to (0,1]$ such that

$$\forall \varepsilon \in (0,2] \forall x, y \in X \left(\|x\|, \|y\| \leqslant 1 \wedge \|x-y\| \geqslant \varepsilon \to \left\| \frac{x+y}{2} \right\| \leqslant 1 - \eta(\varepsilon) \right).$$

Of course, the rates in general will then depend on such a modulus. It should be further noted that this modulus is conceptually related to the common analytic notion of a modulus of convexity $\delta : [0,2] \to [0,1]$ (implicit already used in e.g. [45]) defined as

$$\delta(\varepsilon) = \inf \{1 - \|x+y\|/2 \mid \|x\| = \|y\| = 1, \|x-y\| = \varepsilon\}.$$

In fact, as is well-known, uniformly convex spaces are characterized by the property that $\delta(\varepsilon) > 0$ whenever $\varepsilon > 0$ and the modulus of uniform convexity η effectively provides a witness for this inequality in the form of a lower bound, i.e. that $\delta(\varepsilon) \geqslant \eta(\varepsilon) > 0$ for $\varepsilon \in (0,2]$.

The proofs of the results of both Plant and Reich make an essential use of δ but closer inspection reveals that they only rely on a lower bound on $\delta(\varepsilon)$ greater than 0 which therefore can be substituted by the modulus of uniform convexity η. Note that η can be assumed to be nondecreasing which we will do w.l.o.g. in the following. In that case, one in particular has that $\eta(\varepsilon) < \eta(\delta)$ implies $\varepsilon \leqslant \delta$.

7.2 An analysis of Plant's result

In this section, if not said otherwise, let X be a fixed Banach space, A be a fixed accretive operator that satisfies the range condition $(RC)_{\lambda_0}$ and let S be the semigroup on $\overline{\mathrm{dom} A}$ generated by A using the Crandall-Liggett formula. The proof of Plant's result now proceeds by establishing that the sequence

$$\frac{x - J_t x}{t}, \quad (t \to 0^+)$$

is Cauchy and that we have the limit

$$\lim_{t,s/t \to 0^+} \left\| \frac{x - J_t x}{t} - \frac{x - S(s)x}{s} \right\| = 0.$$

Both results rely crucially on the existence and equality of the limits

$$\lim_{t \to 0^+} \frac{\|x - J_t x\|}{t} \text{ and } \lim_{t \to 0^+} \frac{\|x - S(t)x\|}{t}.$$

The first sequence is nondecreasing for $t \to 0^+$ (see e.g. [52]) and bounded by $\|v\|$ for $v \in Ax$ witnessing $x \in \mathrm{dom} A$ (see e.g. [4]). Following [48], we denote the first limit by $|Ax|$ which naturally satisfies $|Ax| \leqslant \|v\|$. The second limit was shown to coincide with $|Ax|$ in [48].

Now, the proof given in [170] crucially relies on the use of the limit operator $|Ax|$ and some elementary properties thereof. For the following, we denote the expression $(x - J_t x)/t$ (which is just the Yosida approximate) by $A_t x$, in contrast to Plants notation.

As discussed in Chapter 3 (and as will be discussed further later on), one of the main theoretical obstacles in treating accretive and monotone operators is the use of extensionality in proofs as this requires one to provide some sort of quantitative modulus of uniform continuity as dictated by the monotone functional interpretation. While this will be discussed in more detail in the later logical remarks, we also find here that the main convergence principle

$$\|A_t x\| \to |Ax| \text{ for } t \to 0^+ \text{ with } x \in \mathrm{dom}A,$$

on which the proof of Plant relies can be recognized as a particular weak version of such a kind of extensionality statement, namely it can be shown that it is provably equivalent to the lower semi-continuity on $\mathrm{dom}A$ of the operator $|A\cdot|$ associated with A (see Proposition 7.4.1 later on).

As in the case of the functional $\langle \cdot, \cdot \rangle_s$ (recall Chapter 4), the logical methodology based on the monotone Dialectica interpretation now implies the following quantitative version of this statement: under this interpretation, the statement is upgraded to the existence of a "modulus of uniform lower-semicontinuity" $\varphi : \mathbb{R}_{>0} \times \mathbb{N} \to \mathbb{R}_{>0}$, i.e.

$$\forall b \in \mathbb{N}, \varepsilon \in \mathbb{R}_{>0}, (x, u), (y, v) \in A$$
$$(\|x\|, \|u\|, \|y\|, \|v\| \leqslant b \land \|x - y\| \leqslant \varphi(\varepsilon, b) \to |Ax| - |Ay| \leqslant \varepsilon),$$

which, as discussed already in the context of $\langle \cdot, \cdot \rangle_s$, is essentially a modulus of uniform continuity.

Based on the above mentioned equivalence, this modulus can then be used to derive a rate of convergence for the Yosida approximates towards $|Ax|$.

Lemma 7.2.1. *Let φ be a modulus of uniform continuity for $|A\cdot|$ and let n satisfy $n \geqslant \|c\|, \|d\|, \lambda_0, \tilde{\gamma}$ for $(c, d) \in A$ and $0 < \tilde{\gamma} < \lambda_0$. Then for $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$, we have*

$$\forall \varepsilon > 0 \forall t \in (0, \varphi_1(\varepsilon, b, n, \varphi)] \left(|Ax| - \frac{\|x - J_t^A x\|}{t} \leqslant \varepsilon \right)$$

where

$$\varphi_1(\varepsilon, b, n, \varphi) := \min\{\varphi(\varepsilon, b + 2n + 3n^2)/b, \lambda_0/2\}.$$

Proof. Let ε be given and let $t \leqslant \varphi_1(\varepsilon, b, n, \varphi)$. We at first have $\|A_t x\| \leqslant \|v\| \leqslant b$ as well as

$$\begin{aligned}\|J_t x\| &\leqslant \|x\| + 2\|c\| + (2\tilde{\gamma} + t)\|d\| \\ &\leqslant \|x\| + 2\|c\| + (2\tilde{\gamma} + \lambda_0)\|d\| \\ &\leqslant \|x\| + 2n + (2n + n)\, n \\ &\leqslant \|x\| + 2n + 3n^2\end{aligned}$$

using Proposition 3.3.3 (as $t < \lambda_0$). Now as $A_t x \in AJ_t x$, we have $|AJ_t x| \leqslant \|A_t x\|$ and thus

$$|Ax| - \|A_t x\| \leqslant |Ax| - |AJ_t x|.$$

Now, we get[2]

$$\|x - J_t x\| \leqslant t\|v\| \leqslant \varphi_1(\varepsilon, b, n, \varphi) b \leqslant \varphi(\varepsilon, b + 2n + 3n^2)$$

and thus, as $v \in Ax$ and $A_t x \in AJ_t x$ with $\|x\|, \|v\|, \|J_t x\|, \|A_t x\| \leqslant b + 2n + 3n^2$, we have

$$|Ax| - \|A_t x\| \leqslant |Ax| - |AJ_t x| \leqslant \varepsilon$$

which is the claim. □

As mentioned before, the fact that

$$\lim_{t \to 0^+} \frac{\|x - S(t)x\|}{t} = |Ax|$$

was proved by Crandall in [48] and the proof proceeds by establishing that $\|x - S(t)x\|/t \leqslant |Ax|$ for any $t > 0$ as well as

$$\liminf_{t \to 0^+} \frac{\|S(t)x - x\|}{t} \geqslant |Ax|$$

and in that way crucially relies on the limit operator $|A\cdot|$ as well. The latter of these results relies on a result established by Miyadera in [148][3] that

$$\limsup_{t \to 0^+} \left\langle \frac{S(t)x - x}{t}, \zeta^* \right\rangle \leqslant \langle y_0, x_0 - x \rangle_s$$

[2]As $v \in Ax$ and $t < \lambda_0$, we have $x \in \mathrm{dom} J_t$ using $(RC)_{\lambda_0}$. Clearly $x + tv \in (Id + tA)x$. So $J_t(x + tv) = x$ by uniqueness of J_t. Thus by the nonexpansivity of J_t on its domain: $\|x - J_t x\| = \|J_t(x + tv) - J_t x\| \leqslant \|x + tv - x\| = t\|v\|$.

[3]The result goes back to earlier work by Brezis [23] with a special case already contained in [50] and more general results proved in [52].

for $y_0 \in Ax_0$, $x \in \overline{\mathrm{dom}A}$ and $\zeta^* \in J(x - x_0)$.

The proof given by Crandall actually only invokes this result for $x \in \mathrm{dom}A$ and, for the proof of Plant's result, it is further sufficient to obtain it only for *some* $\zeta^* \in J(x - x_0)$. Lastly, the proof relies crucially on the use of the functional $\langle \cdot, \cdot \rangle_s$ and in particular on the upper semicontinuity of this functional. In that way, based on the logical methodology that upgrades this upper semicontinuity to a modulus of uniform continuity, we extract the following quantitative version of the above fragment of Miyadera's result:

Lemma 7.2.2. *Let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y \rangle_s \leqslant \langle z, x \rangle_s + \varepsilon \right).$$

For $\zeta^* \in J(x - x_0)$ as well as $x \in \mathrm{dom}A$ with $v \in Ax$ and $y_0 \in Ax_0$ where $b \in \mathbb{N}^*$ with $\|x\|, \|v\|, \|x_0\|, \|y_0\| \leqslant b$:

$$\forall \varepsilon > 0 \forall t \in (0, \psi(\varepsilon, b, \omega)] \left(\left\langle \frac{S(t)x - x}{t}, \zeta^* \right\rangle \leqslant \langle y_0, x_0 - x \rangle_s + \varepsilon \right)$$

where

$$\psi(\varepsilon, b, \omega) := \frac{\omega(2b, \varepsilon)}{2b}.$$

Proof. At first, given an ε, we get for any $t \in \left(0, \frac{\varepsilon}{2b}\right]$ and for all $v \in Ax$ with $\|x\|, \|v\| \leqslant b$ that

$$\|x - S(t)x\| = \|S(0)x - S(t)x\| \leqslant 2\|v\| t \leqslant 2b \frac{\varepsilon}{2b} \leqslant \varepsilon$$

by Lemma 4.4.3, (1). Now, as in Miyadera's proof from [148], we get

$$\langle S(t)x - x, \zeta^* \rangle \leqslant \int_0^t \langle y_0, x_0 - S(\tau)x \rangle_s \mathrm{d}\tau.$$

Then for $b \geqslant \|x\|, \|v\|, \|x_0\|, \|y_0\|$, we get

$$\langle y_0, x_0 - S(t)x \rangle_s \leqslant \langle y_0, x_0 - x \rangle_s + \varepsilon$$

for any $t \in (0, \psi(\varepsilon, b, \omega)]$ as by the above, we have

$$\|S(t)x - x\| \leqslant \omega(2b, \varepsilon)$$

for all such t by assumption on ω and since we trivially have $\|x - x_0\| \leqslant 2b$. Thus in particular we have

$$\begin{aligned} \langle S(t)x - x, \zeta^* \rangle &\leqslant \int_0^t \langle y_0, x_0 - S(\tau)x \rangle_s \mathrm{d}\tau \\ &\leqslant t \left(\langle y_0, x_0 - x \rangle_s + \varepsilon \right) \end{aligned}$$

which gives the claim. □

Then, by following the proof given in [48], we obtain a quantitative version of the crucial direction

$$\liminf_{t\to 0^+} \frac{\|S(t)x - x\|}{t} \geqslant |Ax|$$

of Crandall's proof. Now, already here, a case distinction on whether $|Ax| = 0$ or $|Ax| > 0$ features in the proof of Crandall and the following result first provides a quantitative result on the latter case.

Lemma 7.2.3. *Let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y\rangle_s \leqslant \langle z, x\rangle_s + \varepsilon\right).$$

Let further φ be a modulus of uniform continuity for $|A\cdot|$ and let n be as in Lemma 7.2.1. Then for $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$ and where $|Ax| \geqslant c$ for $c \in \mathbb{R}_{>0}$, we have

$$\forall \varepsilon > 0 \forall t \in (0, \varphi_2'(\varepsilon, b, c, n, \varphi, \omega)] \left(|Ax| - \frac{\|x - S(t)x\|}{t} \leqslant \varepsilon\right)$$

where

$$\varphi_2'(\varepsilon, b, c, n, \varphi, \omega) := \psi(\varepsilon c \min\{\varphi_1(\min\{\varepsilon/2, c/2\}, b, n, \varphi), \lambda_0/2\}/4, b + 2n + 3n^2, \omega)$$

with ψ as in Lemma 7.2.2 and φ_1 as in Lemma 7.2.1.

Proof. Using Lemma 7.2.2, we get

$$\forall \varepsilon > 0 \forall t \in (0, \psi(\varepsilon, b, \omega)] \left(\left\langle \frac{S(t)x - x}{t}, \zeta^* \right\rangle \leqslant \langle y_0, x_0 - x\rangle_s + \varepsilon\right)$$

for $\|x\|, \|v\|, \|x_0\|, \|y_0\| \leqslant b$ and $\zeta^* \in J(x - x_0)$. Now, for $y_0 = A_\lambda x$ and $x_0 = J_\lambda x$ with $\lambda < \lambda_0$, we have

$$\langle y_0, x_0 - x\rangle_s = -\lambda \|A_\lambda x\|^2$$

as well as

$$\begin{aligned}\left\langle \frac{S(t)x - x}{t}, \zeta^* \right\rangle &\geqslant - \left\|\frac{S(t)x - x}{t}\right\| \|x - J_\lambda x\| \\ &= - \left\|\frac{S(t)x - x}{t}\right\| \lambda \|A_\lambda x\|.\end{aligned}$$

Therefore, we obtain

$$\forall \varepsilon > 0 \forall t \in \big(0, \psi(\varepsilon, b + 2n + 3n^2, \omega)\big] \left(\left\|\frac{S(t)x - x}{t}\right\| \|A_\lambda x\| \geqslant \|A_\lambda x\|^2 - \frac{\varepsilon}{\lambda}\right)$$

for all such λ since $b + 2n + 3n^2 \geqslant \|J_\lambda x\|$ and $b \geqslant \|A_\lambda x\|$ as before. Since $|Ax| \geqslant c$, we have that for $\lambda \leqslant \min\{\varphi_1(c/2, b, n, \varphi), \lambda_0/2\}$ that

$$c/2 = c - c/2 \leqslant |Ax| - c/2 \leqslant \|A_\lambda x\|$$

by Lemma 7.2.1. Therefore, we have that

$$\forall \varepsilon > 0 \forall t \in \big(0, \psi(\varepsilon, b + 2n + 3n^2, \omega)\big] \left(\left\| \frac{S(t)x - x}{t} \right\| \geqslant \|A_\lambda x\| - \frac{\varepsilon}{\lambda c/2} \right)$$

for all $\lambda \leqslant \min\{\varphi_1(c/2, b, n, \varphi), \lambda_0/2\}$ and thus in particular

$$\begin{aligned} |Ax| - \left\| \frac{S(t)x - x}{t} \right\| &\leqslant |Ax| - \|A_\lambda x\| + \frac{\varepsilon}{\lambda c/2} \\ &\leqslant \delta/2 + \frac{\varepsilon}{\lambda c/2} \end{aligned}$$

for all $t \leqslant \psi(\varepsilon, b + 2n + 3n^2, \omega)$ and for all $\lambda \leqslant \min\{\varphi_1(\min\{\delta/2, c/2\}, b, n, \varphi), \lambda_0/2\}$. Thus, lastly, for

$$t \leqslant \psi(\varepsilon c \min\{\varphi_1(\min\{\varepsilon/2, c/2\}, b, n, \varphi), \lambda_0/2\}/4, b + 2n + 3n^2, \omega)$$

we have

$$|Ax| - \left\| \frac{S(t)x - x}{t} \right\| \leqslant \varepsilon.$$

□

For the other case, i.e. where $|Ax| = 0$, it is immediately clear that for $|Ax| \leqslant \varepsilon$, we get

$$|Ax| - \frac{\|x - S(t)x\|}{t} \leqslant |Ax| \leqslant \varepsilon$$

for all t. However, this allows for a smoothening of the above case distinction (see the later logical remarks for further discussions of this) in the form of the following lemma:

Lemma 7.2.4. *Let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y \rangle_s \leqslant \langle z, x \rangle_s + \varepsilon \right).$$

Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Then for $x \in \mathrm{dom} A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$, we have

$$\forall \varepsilon > 0 \forall t \in (0, \varphi_2(\varepsilon, b, n, \omega, \varphi)] \left(|Ax| - \frac{\|x - S(t)x\|}{t} \leqslant \varepsilon \right)$$

where

$$\varphi_2(\varepsilon, b, n, \omega, \varphi) := \psi(\varepsilon^2 \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \lambda_0/2\}/4, b + 2n + 3n^2, \omega).$$

with ψ as in Lemma 7.2.2 and φ_1 as in Lemma 7.2.1.

Proof. Let ε be given. Then either $|Ax| \leqslant \varepsilon$ whereas

$$|Ax| - \frac{\|x - S(t)x\|}{t} \leqslant \varepsilon$$

for any t. Otherwise, we have $|Ax| \geqslant \varepsilon$ and thus from Lemma 7.2.3 with $c = \varepsilon$, it follows that

$$|Ax| - \frac{\|x - S(t)x\|}{t} \leqslant \varepsilon$$

for all

$$t \leqslant \psi(\varepsilon^2 \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \lambda_0/2\}/4, b + 2n + 3n^2, \omega).$$

□

Using those two results, we can then give a quantitative version of the partial results on the way to Plants results discussed above, in the form of a rate of Cauchyness and a rate of convergence, respectively.

In that context, we follow the notation used in [170] and write

$$\alpha(a, b) = \left\| \frac{a}{\|a\|} - \frac{b}{\|b\|} \right\| \leqslant 2$$

where $a, b \neq 0$ for the generalized angle of Clarkson [45]. Similar to the proof given in [170], we rely on two fundamental inequalities of α:

Lemma 7.2.5 (essentially [45])**.** *Let $a, b \neq 0$. Then*

$$|\, \|a\| \,\alpha(a, b) - \|a - b\| \,| \leqslant |\, \|a\| - \|b\| \,|.$$

If further $a + b \neq 0$, then

$$\|a + b\| \leqslant (1 - 2\eta(\alpha(a + b, a)))\, \|a\| + \|b\|$$

where η is a modulus of uniform convexity for the space X.

Lemma 7.2.6. *Let X be a uniformly convex Banach space with a nondecreasing modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$. Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Let further $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Suppose $|Ax| \geqslant c$ for $c \in \mathbb{R}_{>0}$. Then*

$$\forall \varepsilon > 0 \forall t \in (0, \varphi_3'(\varepsilon, b, c, \eta, n, \varphi)]\, \forall s \in (0, t) \left(\left\| \frac{x - J_t x}{t} - \frac{x - J_s x}{s} \right\| \leqslant \varepsilon \right)$$

where

$$\varphi_3'(\varepsilon, b, c, \eta, n, \varphi) := \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \varphi_1(\eta(\min\{\varepsilon/2b, 2\})c/2, b, n, \varphi), \varphi_1(c/2, b, n, \varphi), \lambda_0/2\}$$

with φ_1 *as in Lemma 7.2.1.*

Proof. If $x = J_t x$ or $x = J_s x$, then $0 \in Ax$ and thus $|Ax| = 0$. As we have assumed $|Ax| \geqslant c > 0$, we get $x \neq J_t x$ and $x \neq J_s x$. We write $\alpha_{s,t} = \alpha(x - J_s x, x - J_t x)$ where $s \in (0, t)$ and $t \leqslant \lambda_0/2 < \lambda_0$. Using Lemma 7.2.5 with $a = x - J_s x$ and $b = J_s x - J_t x$, we have

$$\|x - J_t x\| \leqslant (1 - 2\eta(\alpha_{s,t})) \|x - J_s x\| + \|J_t x - J_s x\|.$$

Using Proposition 3.3.3, items (3) and (5), we get

$$\begin{aligned}\|J_t x - J_s x\| &= \left\| J_s\left(\frac{s}{t}x + \frac{t-s}{t}J_t x\right) - J_s x\right\| \\ &\leqslant \left(1 - \frac{s}{t}\right)\|x - J_t x\|\end{aligned}$$

and thus we have

$$s\|A_t x\| \leqslant (1 - 2\eta(\alpha_{s,t}))\|x - J_s x\|,$$

i.e.

$$2\eta(\alpha_{s,t})\|A_s x\| \leqslant \|A_s x\| - \|A_t x\|.$$

Therefore, we have for $0 < t \leqslant \varphi_1(c/2, b, n, \varphi)$ that

$$c - \|A_t x\| \leqslant |Ax| - \|A_t x\| \leqslant c/2$$

so that $c/2 \leqslant \|A_t x\|$ and for $s \in (0, t)$, we get that

$$\eta(\alpha_{s,t})c \leqslant 2\eta(\alpha_{s,t})\|A_t x\| \leqslant 2\eta(\alpha_{s,t})\|A_s x\| \leqslant \|A_s x\| - \|A_t x\| \leqslant |Ax| - \|A_t x\|.$$

By Lemma 7.2.1, we have for any ε that

$$\forall t \in (0, \min\{\varphi_1(\varepsilon, b, n, \varphi), \varphi_1(c/2, b, n, \varphi)\}]\,\forall s \in (0, t)\,(\eta(\alpha_{s,t})c \leqslant \varepsilon)$$

which, in particular, implies

$$\forall t \in (0, \min\{\varphi_1(\eta(\min\{\varepsilon/2b, 2\})c/2, b, n, \varphi), \varphi_1(c/2, b, n, \varphi)\}]\,\forall s \in (0, t)\,(\eta(\alpha_{s,t}) \leqslant \eta(\min\{\varepsilon/2b, 2\})/2)$$

and using that η is nondecreasing, we get

$$\forall t \in (0, \min\{\varphi_1(\eta(\min\{\varepsilon/2b, 2\})c/2, b, n, \varphi), \varphi_1(c/2, b, n, \varphi), \lambda_0\}] \, \forall s \in (0, t) \, (\alpha_{s,t} \leqslant \varepsilon/2b).$$

Using Lemma 7.2.5 with $a = \|x - J_t x\|/t$ and $b = \|x - J_s x\|/s$ (noting that $\alpha_{s,t} = \alpha(a, b)$ for these a, b) together with $s < t$ as well as the triangle inequality, we now have

$$\begin{aligned}\left\| \frac{x - J_t x}{t} - \frac{x - J_s x}{s} \right\| &\leqslant \left| \frac{\|x - J_t x\|}{t} \alpha_{s,t} - \left\| \frac{x - J_t x}{t} - \frac{x - J_s x}{s} \right\| \right| + \frac{\|x - J_t x\|}{t} \alpha_{s,t} \\ &\leqslant \left| \left\| \frac{x - J_t x}{t} \right\| - \left\| \frac{x - J_s x}{s} \right\| \right| + \frac{\|x - J_t x\|}{t} \alpha_{s,t} \\ &\leqslant \left(|Ax| - \left\| \frac{x - J_t x}{t} \right\| \right) + b\alpha_{s,t}.\end{aligned}$$

Thus for $0 < t \leqslant \varphi_3'(\varepsilon, b, c, \eta, n, \varphi)$ and for $s \in (0, t)$, we have

$$\begin{aligned}\left\| \frac{x - J_t x}{t} - \frac{x - J_s x}{s} \right\| &\leqslant \varepsilon/2 + b\varepsilon/2b \\ &\leqslant \varepsilon\end{aligned}$$

by Lemma 7.2.1. □

Again, the case for $|Ax| = 0$ is trivial and yields the following quantitative version: if $|Ax| \leqslant \varepsilon/2$, then in particular

$$\left\| \frac{x - J_t^A x}{t} - \frac{x - J_s^A x}{s} \right\| \leqslant \|A_t x\| + \|A_s x\| \leqslant |Ax| + |Ax| \leqslant \varepsilon.$$

In that way, we get the following smoothening for both results combined.

Lemma 7.2.7. *Let X be a uniformly convex Banach space with a nondecreasing modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$. Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Let further $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Then*

$$\forall \varepsilon > 0 \forall t \in (0, \varphi_3(\varepsilon, b, \eta, n, \varphi)] \, \forall s \in (0, t) \left(\left\| \frac{x - J_t^A x}{t} - \frac{x - J_s^A x}{s} \right\| \leqslant \varepsilon \right)$$

where

$$\begin{aligned}\varphi_3(\varepsilon, b, \eta, n, \varphi) := \min\{&\varphi_1(\varepsilon/2, b, n, \varphi), \varphi_1(\eta(\min\{\varepsilon/2b, 2\})\varepsilon/4, b, n, \varphi), \\ &\varphi_1(\varepsilon/4, b, n, \varphi), \lambda_0/2\}\end{aligned}$$

with $\varphi_1(\varepsilon, b, n)$ as in Lemma 7.2.1.

Proof. Let ε be given. Then either $|Ax| \leqslant \varepsilon/2$ whereas

$$\left\| \frac{x - J_t^A x}{t} - \frac{x - J_s^A x}{s} \right\| \leqslant \varepsilon$$

for any $t, s < \lambda_0$ as discussed above. Otherwise we have $|Ax| \geqslant \varepsilon/2$ and thus by Lemma 7.2.6 with $c = \varepsilon/2$, it follows that

$$\left\| \frac{x - J_t^A x}{t} - \frac{x - J_s^A x}{s} \right\| \leqslant \varepsilon$$

for $s \in (0, t)$ and

$$t \leqslant \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \varphi_1(\eta(\min\{\varepsilon/2b, 2\})\varepsilon/4, b, n, \varphi), \varphi_1(\varepsilon/4, b, n, \varphi), \lambda_0/2\}.$$

□

Lemma 7.2.8 (Plant [170], Eq. (2.10))**.** *Let $x \in \mathrm{dom}A$ and $t, \lambda > 0$. Then*

$$\|J_\lambda x - S(t)x\| \leqslant \left(1 - \frac{t}{\lambda}\right) \|x - J_\lambda x\| + \frac{2}{\lambda} \int_0^t \|x - S(s)x\| \, \mathrm{ds}.$$

Lemma 7.2.9. *Let X be a uniformly convex Banach space with a nondecreasing modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$ and let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y\rangle_s \leqslant \langle z, x\rangle_s + \varepsilon\right).$$

Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Let further $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Suppose that $|Ax| \geqslant c$ for $c \in \mathbb{R}_{>0}$. Then

$$\forall \varepsilon > 0 \forall t, \frac{s}{t} \in (0, \varphi_4'(\varepsilon, b, c, \eta, n, \omega, \varphi)] \left(\left\| \frac{x - J_t^A x}{t} - \frac{x - S(s)x}{s} \right\| \leqslant \varepsilon \right)$$

where

$$\begin{aligned}\varphi_4'(\varepsilon, b, c, \eta, n, \omega, \varphi) := \min\{&\varphi_1(\varepsilon/3, b, n, \varphi), \varphi_2(\varepsilon/3, b, n, \omega, \varphi),\\ &\varphi_1(\eta(\min\{\varepsilon, 2\})c/4, b, n, \varphi), \sqrt{\varphi_2(c/2, b, n, \omega, \varphi)},\\ &\eta(\min\{\varepsilon, 2\})c/8b, 1, \lambda_0/2\}\end{aligned}$$

with φ_1, φ_2 as in Lemmas 7.2.1, 7.2.4, respectively.

Proof. As before, $x \neq S(s)x$ and $x \neq J_t x$ as $|Ax| \geqslant c > 0$. We write $\alpha'_{s,t} = \alpha(x - S(s)x, x - J_t x)$ for $t, s < \lambda_0$. Using Lemma 7.2.5, we again obtain

$$\|x - J_t x\| \leqslant \left(1 - 2\eta(\alpha'_{s,t})\right) \|x - S(s)x\| + \|J_t x - S(s)x\|.$$

Using Lemma 7.2.8, we get for $t, s \leqslant \min\{\lambda_0/2, 1\}$:

$$\begin{aligned}\|x - J_t x\| &\leqslant \left(1 - 2\eta(\alpha'_{s,t})\right)\|x - S(s)x\| + \left(1 - \frac{s}{t}\right)\|x - J_t x\| \\ &\quad + \frac{2}{t}\int_0^s \|x - S(\tau)x\|\,\mathrm{d}\tau \\ &\leqslant \left(1 - 2\eta(\alpha'_{s,t})\right)\|x - S(s)x\| + \left(1 - \frac{s}{t}\right)\|x - J_t x\| \\ &\quad + \frac{2}{t}\int_0^s s\frac{\|x - S(\tau)x\|}{\tau}\mathrm{d}\tau \\ &\leqslant \left(1 - 2\eta(\alpha'_{s,t})\right)\|x - S(s)x\| + \left(1 - \frac{s}{t}\right)\|x - J_t x\| + \frac{s^2}{t}2b\end{aligned}$$

which implies that

$$2\eta(\alpha'_{s,t})\frac{\|x - S(s)x\|}{s} \leqslant |Ax| - \frac{\|x - J_t x\|}{t} + \frac{s}{t}2b.$$

Now for

$$t \leqslant \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \sqrt{\varphi_2(c/2, b, n, \omega, \varphi)}\}$$

and

$$\frac{s}{t} \leqslant \min\{\varepsilon/4b, \sqrt{\varphi_2(c/2, b, n, \omega, \varphi)}\}$$

we obtain that

$$s \leqslant t\frac{s}{t} \leqslant \varphi_2(c/2, b, n, \omega, \varphi)$$

and thus (using Lemma 7.2.4), we obtain

$$\begin{aligned}\eta(\alpha'_{s,t})c &\leqslant 2\eta(\alpha'_{s,t})\left(|Ax| - c/2\right) \\ &\leqslant 2\eta(\alpha'_{s,t})\frac{\|x - S(s)x\|}{s} \\ &\leqslant |Ax| - \frac{\|x - J_t x\|}{t} + \frac{s}{t}2b \\ &\leqslant \varepsilon/2 + \frac{s}{t}2b \\ &\leqslant \varepsilon/2 + 2b\varepsilon/4b \\ &\leqslant \varepsilon.\end{aligned}$$

Dividing by c, we get $\eta(\alpha'_{s,t}) \leqslant \frac{\varepsilon}{c}$ for all such t, s. Thus, using that η is nondecreasing, we have $\alpha'_{s,t} \leqslant \varepsilon$ for

$$t \leqslant \min\{\varphi_1(\eta(\min\{\varepsilon, 2\})c/4, b, n, \varphi), \sqrt{\varphi_2(c/2, b, n, \omega, \varphi)}\}$$

and

$$\frac{s}{t} \leqslant \min\{\eta(\min\{\varepsilon, 2\})c/8b, \sqrt{\varphi_2(c/2, b, n, \omega, \varphi)}\}.$$

Using Lemma 7.2.5 and triangle inequality again, we now have similarly to before

$$\begin{aligned}
&\left\| \frac{x - J_t x}{t} - \frac{x - S(s)x}{s} \right\| \\
&\qquad \leqslant \left| \frac{\|x - J_t x\|}{t}\alpha'_{s,t} - \left\| \frac{x - J_t x}{t} - \frac{x - S(s)x}{s} \right\| \right| + \frac{\|x - J_t x\|}{t}\alpha'_{s,t} \\
&\qquad \leqslant \left| \left\| \frac{x - J_t x}{t} \right\| - \left\| \frac{x - S(s)x}{s} \right\| \right| + \frac{\|x - J_t x\|}{t}\alpha'_{s,t} \\
&\qquad \leqslant \left(|Ax| - \left\| \frac{x - J_t x}{t} \right\| \right) + \left(|Ax| - \left\| \frac{x - S(s)x}{s} \right\| \right) + b\alpha'_{s,t}.
\end{aligned}$$

Thus for $0 < t, \frac{s}{t} \leqslant \varphi'_4(\varepsilon, b, c, \eta, n, \omega, \varphi)$, we have

$$\begin{aligned}
\left\| \frac{x - J_t x}{t} - \frac{x - S(s)x}{s} \right\| &\leqslant \varepsilon/3 + \varepsilon/3 + b\varepsilon/3b \\
&\leqslant \varepsilon
\end{aligned}$$

by Lemma 7.2.1 and Lemma 7.2.4. □

As before, smoothening this result can be achieved by extracting from the proof for the case of $|Ax| = 0$ the following quantitative version: if $|Ax| \leqslant \varepsilon/2$, then

$$\left\| \frac{x - J_t^A x}{t} - \frac{x - S(s)x}{s} \right\| \leqslant \frac{\|x - J_t^A x\|}{t} + \frac{\|x - S(s)x\|}{s} \leqslant |Ax| + |Ax| \leqslant \varepsilon.$$

Therefore, we obtain the following result:

Lemma 7.2.10. *Let X be a uniformly convex Banach space with a nondecreasing modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$ and let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y \rangle_s \leqslant \langle z, x \rangle_s + \varepsilon \right).$$

Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Let further $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Then

$$\forall \varepsilon > 0 \forall t, \frac{s}{t} \in (0, \varphi_4(\varepsilon, b, \eta, n, \omega, \varphi)] \left(\left\| \frac{x - J_t^A x}{t} - \frac{x - S(s)x}{s} \right\| \leqslant \varepsilon \right)$$

where

$$\begin{aligned}
\varphi_4(\varepsilon, b, \eta, n, \omega, \varphi) := \min\{&\varphi_1(\varepsilon/3, b, n, \varphi), \varphi_2(\varepsilon/3, b, n, \omega, \varphi), \\
&\varphi_1(\eta(\min\{\varepsilon, 2\})\varepsilon/8, b, n, \varphi), \sqrt{\varphi_2(\varepsilon/4, b, n, \omega, \varphi)}, \\
&\eta(\min\{\varepsilon, 2\})\varepsilon/16b, 1, \lambda_0/2\}
\end{aligned}$$

with φ_1, φ_2 as in Lemmas 7.2.1, 7.2.4, respectively.

Proof. Let ε be given. Then either $|Ax| \leqslant \varepsilon/2$ which implies

$$\left\| \frac{x - J_t^A x}{t} - \frac{x - S(s)x}{s} \right\| \leqslant \varepsilon$$

as above for any such t and s or $|Ax| \geqslant \varepsilon/2$ where now the result is implies for any

$$t, \frac{s}{t} \in (0, \varphi_4'(\varepsilon, b, \varepsilon/2, \eta, n, \omega, \varphi)]$$

by Lemma 7.2.9 with $c = \varepsilon/2$. □

Finally, a combination of these two quantitative results yields a quantitative version of the theorem of Plant.

Theorem 7.2.11. *Let X be a uniformly convex Banach space with a nondecreasing modulus of uniform convexity $\eta : (0,2] \to (0,1]$ and let ω be such that*

$$\forall x, y, z \in X, b \in \mathbb{N}, \varepsilon > 0 \left(\|x\|, \|z\| \leqslant b \wedge \|x - y\| \leqslant \omega(b, \varepsilon) \to \langle z, y \rangle_s \leqslant \langle z, x \rangle_s + \varepsilon \right).$$

Let further φ be a modulus of uniform continuity for $|A \cdot|$ and let n be as in Lemma 7.2.1. Let further $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Then

$$\forall \varepsilon > 0 \forall t \in (0, \Phi(\varepsilon, b, \eta, \omega, \varphi, n)] \left(\frac{\|J_t^A x - S(t)x\|}{t} \leqslant \varepsilon \right)$$

where

$$\Phi(\varepsilon, b, \eta, \omega, \varphi, n) := (\min\{\varphi_3(\varepsilon/2, b, \eta, n, \varphi), \varphi_4(\varepsilon/2, b, \eta, n, \omega, \varphi)\})^2$$

with φ_1 - φ_4 as well as ψ defined by

$$\begin{aligned}
\varphi_1(\varepsilon, b, n, \varphi) &:= \min\{\varphi(\varepsilon, b + 2n + 3n^2)/b, \lambda_0/2\}, \\
\psi(\varepsilon, b, \omega) &:= \frac{\omega(2b, \varepsilon)}{2b}, \\
\varphi_2(\varepsilon, b, n, \omega, \varphi) &:= \psi(\varepsilon^2 \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \lambda_0/2\}/4, b + 2n + 3n^2, \omega), \\
\varphi_3(\varepsilon, b, \eta, n, \varphi) &:= \min\{\varphi_1(\varepsilon/2, b, n, \varphi), \varphi_1(\eta(\min\{\varepsilon/2b, 2\})\varepsilon/4, b, n, \varphi), \\
&\qquad \varphi_1(\varepsilon/4, b, n, \varphi), \lambda_0/2\}, \\
\varphi_4(\varepsilon, b, \eta, n, \omega, \varphi) &:= \min\{\varphi_1(\varepsilon/3, b, n, \varphi), \varphi_2(\varepsilon/3, b, n, \omega, \varphi), \\
&\qquad \varphi_1(\eta(\min\{\varepsilon, 2\})\varepsilon/8, b, n, \varphi), \sqrt{\varphi_2(\varepsilon/4, b, n, \omega, \varphi)}, \\
&\qquad \eta(\min\{\varepsilon, 2\})\varepsilon/16b, 1, \lambda_0/2\}.
\end{aligned}$$

Proof. Using the triangle inequality, we have

$$\frac{\|J_t x - S(t)x\|}{t} \leqslant \left\| \frac{x - J_t x}{t} - \frac{x - J_{\sqrt{t}}x}{\sqrt{t}} \right\| + \left\| \frac{x - J_{\sqrt{t}}x}{\sqrt{x}} - \frac{x - S(t)x}{t} \right\|$$

Then, for $t \leqslant 1$, we have $t \leqslant \sqrt{t}$ and $t/\sqrt{t} = \sqrt{t}$ so that for $t \leqslant \Phi(\varepsilon, b, \eta, \omega, \varphi, n)$, we obtain

$$\frac{\|J_t x - S(t)x\|}{t} \leqslant \varepsilon$$

using Lemmas 7.2.7 and 7.2.10. □

Remark 7.2.12. While the above result uses the construction of φ_1 from φ exhibited in Lemma 7.2.1, it is clear that if φ_1 is any other rate of convergence for $\|A_t x\|$ to $|Ax|$ as $t \to 0$, the above result nevertheless remains valid.

7.3 An analysis of Reich's result

Similar as in the context of Plant's result, in this section we fix a Banach space X and an accretive operator A that now satisfies the range condition (RC). As before, let S be the semigroup on $\overline{\mathrm{dom}A}$ generated by A using the Crandall-Liggett formula. The proof for Reich's result now proceeds by establishing

$$\lim_{t\to\infty} \frac{\|J_t x\|}{t} = d(0, \mathrm{ran}A)$$

and concluding from this that $J_t x/t$ is Cauchy for $t \to \infty$. This result is then in turn used to conclude the claim. While Reich actually establishes his result even for $x \in \overline{\mathrm{dom}A}$, we here focus for simplicity on the case where $x \in \mathrm{dom}A$. As mentioned in the brief outline at the beginning of this chapter, all the following results in the context of Reich's theorem take place for an operator A which satisfies the full range condition (RC) for the closure of the domain (recall Chapter 4).

The main object used in these proofs is the concrete value

$$d := d(0, \mathrm{ran}A) = \inf\{\|y\| \mid y \in \mathrm{ran}A\}$$

and for the quantitative results, the logical methodology implies (see the later logical remarks for a discussion of this) a dependence on a function f witnessing the above infimum quantitatively in the sense that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies

$$\forall \varepsilon > 0 \exists (y, z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon \right).$$

The proof of Reich's result then proceeds by a case distinction on whether $d > 0$ or $d = 0$ but, as before with the quantitative analysis of Plant's result, this case distinction can be smoothed as will be exhibited later. We at first begin with the following result which provides a rate of convergence for the limit $\|J_t x\|/t \to d$ for $t \to \infty$ (which can be obtained as the sequence is monotone).

Lemma 7.3.1. *Let $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ where $b \geqslant \|x\|, \|v\|$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y, z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon\right).$$

Then we have

$$\forall \varepsilon > 0 \forall t \geqslant \varphi(\varepsilon, b, f) \left(\left| \frac{\|J_t x\|}{t} - d \right| \leqslant \varepsilon \right)$$

where

$$\varphi(\varepsilon, b, f) := \frac{8(b + f(\varepsilon/2))}{\varepsilon}.$$

Proof. As $A_t x \in AJ_t x$ for any $t > 0$, we have $d \leqslant \|A_t x\|$. Let ε be given and let $z \in Ay$ such that $\|z\| - d \leqslant \varepsilon/2$ and $\|y\|, \|z\| \leqslant f(\varepsilon/2)$. Then using Proposition 3.3.3, (7), we have

$$\begin{aligned} \|A_t x\| &\leqslant \|A_t x - A_t y\| + \|A_t y\| \\ &\leqslant \frac{2}{t} \|x - y\| + \|z\| \\ &\leqslant \frac{2(b + f(\varepsilon/2))}{t} + d + \varepsilon/2. \end{aligned}$$

Thus, for $t \geqslant (\varepsilon/4(b + f(\varepsilon/2)))^{-1}$, we have

$$\|A_t x\| \leqslant \frac{2(b + f(\varepsilon/2))}{(\varepsilon/4(b + f(\varepsilon/2)))^{-1}} + d + \varepsilon/2 \leqslant d + \varepsilon$$

Now, for $t \geqslant (\varepsilon/8(b + f(\varepsilon/2)))^{-1}$, we obtain

$$\begin{aligned} \left| \frac{\|J_t x\|}{t} - d \right| &\leqslant \left| \frac{\|J_t x\|}{t} - \frac{\|x - J_t x\|}{t} \right| + \left| \|A_t x\| - d \right| \\ &\leqslant \frac{\|x\|}{t} + \left| \|A_t x\| - d \right| \\ &\leqslant \varepsilon \end{aligned}$$

as $t \geqslant (\varepsilon/8(b + f(\varepsilon/2)))^{-1}$ and thus $\|A_t x\| - d \leqslant \varepsilon/2$ as well as $t \geqslant (\varepsilon/2b)^{-1}$ and thus $\|x\|/t \leqslant \varepsilon/2$. □

The following result is a quantitative version of the well-known result due to Reich [173] that $d > 0$ implies that $\|J_t x\| \to \infty$ for $t \to \infty$ and $x \in \mathrm{dom}A$.

Lemma 7.3.2. *Assume that $d \geqslant D$ for $D \in \mathbb{R}_{>0}$. Let $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ where $b \geqslant \|x\|, \|v\|$. Then we have*

$$\forall K > 0 \forall t \geqslant \psi(K, b, D) \left(\|J_t x\| \geqslant K\right)$$

where

$$\psi(K, b, D) := \frac{b + K}{D}.$$

Proof. Suppose the claim is false, i.e. there is a K and a $t \geqslant \psi(K, b, D)$ such that $\|J_t x\| < K$. Then, we have

$$\begin{aligned} \|J_t x - J_1 J_t x\| &\leqslant |AJ_t x| \\ &\leqslant \|x - J_t x\| / t \\ &< (b + K)/D^{-1}(b + K) \\ &\leqslant D. \end{aligned}$$

Thus $\|A_1 J_t x\| < D \leqslant d$ which is a contradiction as $A_1 J_t x \in \mathrm{ran}A$. □

Lemma 7.3.3 (essentially Reich [174])**.** *Let X be uniformly convex with a modulus of uniform convexity η. Then, for $\varepsilon \in (0, 2]$, we have $2\eta(\varepsilon) \leqslant 1 - \langle y, j \rangle$ for all $j \in Jx$ with $\|x\| = \|y\| = 1$ and $\|x - y\| \geqslant \varepsilon$.*

Proof. Let x, y and $j \in Jx$ be given with $\|x\| = \|y\| = 1$ and $\|x - y\| \geqslant \varepsilon$. Then

$$\frac{\|x + y\|}{2} \leqslant 1 - \eta(\varepsilon)$$

by definition of η. Thus as $\langle x, j \rangle = \|x\|^2 = 1$ and $\|j\| = \|x\| = 1$, we have

$$1/2 + 1/2\langle y, j \rangle = \langle (x + y)/2, j \rangle \leqslant 1 - \eta(\varepsilon)$$

which yields the claim. □

Lemma 7.3.4. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies*

$$\forall \varepsilon > 0 \exists (y, z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon\right).$$

Let $\varepsilon > 0$ be given, assume that $D \in \mathbb{R}_{>0}$ with $d \geqslant D$ and let $z \in Ay$ be such that

$$\|z\| \leqslant d + 2d\eta(\min\{\varepsilon/2, 2\})$$

as well as $c \geqslant \|y\|, \|z\|$. *Let* $x \in \mathrm{dom}A$ *with* $v \in Ax$ *and* $b \in \mathbb{N}^*$ *with* $b \geqslant \|x\|, \|v\|$. *Then, for any* $t \geqslant \varphi_1(\varepsilon, b, D, c, \eta, f)$*:*

$$\left\| \frac{z}{\|z\|} + \frac{J_t x}{\|J_t x\|} \right\| \leqslant \varepsilon$$

where

$$\begin{aligned}\varphi_1(\varepsilon, b, D, c, \eta, f) := \max \Bigg\{ &\psi(c+1, b, D), \psi\left(\left(\frac{4}{\varepsilon}+1\right)c, b, D\right), \\ &\left(\frac{D\left(2\eta(\min\{\varepsilon/2, 2\})\right)^2}{18}\right)^{-1}(c+b), \\ &\varphi\left(\frac{D\left(2\eta(\min\{\varepsilon/2, 2\})\right)^2}{18}, b, f\right)\Bigg\}\end{aligned}$$

with φ *defined as in Lemma 7.3.1 and* ψ *as in Lemma 7.3.2*

Proof. As $A_t x \in AJ_t x$ and as A is accretive, there is a $j_t \in J(y - J_t x)$ such that $\langle z - A_t x, j_t \rangle \geqslant 0$. Therefore we have

$$\left\langle \frac{z}{\|z\|}, \frac{j_t}{\|y - J_t x\|} \right\rangle \geqslant \left\langle \frac{A_t x}{\|z\|}, \frac{j_t}{\|y - J_t x\|} \right\rangle$$

for any t such that $t \geqslant \psi(c+1, b, D)$ as then $\|J_t x\| > c$ which implies $y \neq J_t x$ and $z \neq 0$ follows by $\|z\| \geqslant d \geqslant D > 0$. Then further

$$\left\langle J_t x - x, \frac{j_t}{\|y - J_t x\|} \right\rangle \leqslant \|y - x\| - \|y - J_t x\|,$$

by (an argument similar to the proof of) Proposition 4.3.3 and we thus obtain

$$\begin{aligned}\left\langle A_t x, \frac{j_t}{\|y - J_t x\|} \right\rangle &\geqslant \left\| \frac{y}{t} - \frac{J_t x}{t} \right\| - \frac{\|y - x\|}{t} \\ &\geqslant \frac{\|J_t x\|}{t} - \frac{\|y\|}{t} - \frac{\|y - x\|}{t}.\end{aligned}$$

Thus for any $\delta \in \mathbb{R}_{>0}$ and any

$$t \geqslant \max\left\{(\delta/3)^{-1}(c+b), \varphi(\delta/3, b, f)\right\},$$

we obtain from Lemma 7.3.1 that

$$\begin{aligned}\left\langle \frac{z}{\|z\|}, \frac{j_t}{\|y - J_t x\|} \right\rangle &\geqslant \frac{d}{\|z\|} - \frac{\delta}{\|z\|} \\ &\geqslant \frac{1}{1 + 2\eta(\min\{\varepsilon/2, 2\})} - \frac{\delta}{\|z\|}.\end{aligned}$$

Now we get $1 - (2\eta(\min\{\varepsilon/2, 2\}))^2 + (2\eta(\min\{\varepsilon/2, 2\}))^2 = 1$ and therefore $1 = (1 + 2\eta(\min\{\varepsilon/2, 2\}))(1 - 2\eta(\min\{\varepsilon/2, 2\})) + (2\eta(\min\{\varepsilon/2, 2\}))^2$ which yields

$$\begin{aligned}\frac{1}{1+2\eta(\min\{\varepsilon/2,2\})} &= 1 - 2\eta(\min\{\varepsilon/2,2\}) + \frac{(2\eta(\min\{\varepsilon/2,2\}))^2}{1+2\eta(\min\{\varepsilon/2,2\})}\\ &\geqslant 1 - 2\eta(\min\{\varepsilon/2,2\}) + \frac{(2\eta(\min\{\varepsilon/2,2\}))^2}{3}.\end{aligned}$$

Thus for

$$t \geqslant \max\left\{\left(\frac{D\,(2\eta(\min\{\varepsilon/2,2\}))^2}{18}\right)^{-1}(c+b), \varphi\left(\frac{D\,(2\eta(\min\{\varepsilon/2,2\}))^2}{18}, b, f\right)\right\},$$

we obtain (using $z \in \mathrm{ran}A$) that

$$\begin{aligned}\left\langle \frac{z}{\|z\|}, \frac{j_t}{\|y-J_tx\|}\right\rangle &\geqslant 1 - 2\eta(\min\{\varepsilon/2,2\}) + \frac{(2\eta(\min\{\varepsilon/2,2\}))^2}{3}\\ &\qquad - \frac{D\,(2\eta(\min\{\varepsilon/2,2\}))^2}{6\,\|z\|}\\ &\geqslant 1 - 2\eta(\min\{\varepsilon/2,2\}) + \frac{(2\eta(\min\{\varepsilon/2,2\}))^2}{6}\\ &> 1 - 2\eta(\min\{\varepsilon/2,2\}).\end{aligned}$$

Then in particular

$$\left\|\frac{z}{\|z\|} - \frac{y-J_tx}{\|y-J_tx\|}\right\| \leqslant \frac{\varepsilon}{2}$$

by Lemma 7.3.3 for all such t.

Now, secondly:

$$\left\|\frac{y-J_tx}{\|y-J_tx\|} + \frac{J_tx}{\|J_tx\|}\right\| \leqslant \frac{\|y\|}{|\|y\|-\|J_tx\||} + \left|1 - \frac{\|J_tx\|}{\|y-J_tx\|}\right|$$

For $\delta > 0$ and $t \geqslant \psi((\delta^{-1}+1)c, b, D)$, we immediately have

$$\frac{\|y\|}{|\|y\|-\|J_tx\||} \leqslant \frac{c}{(\delta^{-1}+1)c - \|y\|} \leqslant \delta$$

by Lemma 7.3.2. Similarly, we get for $t \geqslant \psi((\delta^{-1}+1)c, b, D)$, as $(\delta^{-1}+1)c \geqslant \delta^{-1}c$, that

$$\frac{\|y\|}{\|y\|+\|J_tx\|} \leqslant \frac{c}{\|y\|+\delta^{-1}c} \leqslant \frac{c}{\delta^{-1}c} = \delta.$$

Further, we have

$$1 - \frac{\|y\|}{\|y\|+\|J_tx\|} \leqslant \frac{\|J_tx\|}{\|y-J_tx\|} \leqslant 1 + \frac{\|y\|}{|\,\|y\|-\|J_tx\|\,|}$$

and thus for $t \geqslant \psi((\delta^{-1}+1)c, b, D)$, we get

$$\left|1 - \frac{\|J_t x\|}{\|y - J_t x\|}\right| \leqslant \delta.$$

Combining the above, we have that for any $t \geqslant \psi(((\varepsilon/4)^{-1}+1)c, b, D)$:

$$\left\|\frac{y - J_t x}{\|y - J_t x\|} + \frac{J_t x}{\|J_t x\|}\right\| \leqslant \frac{\varepsilon}{2}.$$

Thus, finally for $t \geqslant \varphi_1(\varepsilon, b, D, c, \eta, f)$ we obtain the desired result by triangle inequality. □

Lemma 7.3.5. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0,2] \to (0,1]$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y,z) \in A\left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon\right).$$

Let $\varepsilon > 0$ be given, assume that $E \in \mathbb{N}^*$, $D \in \mathbb{R}_{>0}$ where $E \geqslant d \geqslant D$. Let $x \in \mathrm{dom}A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ where $b \geqslant \|x\|, \|v\|$. Then, for any $t, s \geqslant \varphi_2'(\varepsilon, b, D, \eta, E, f)$:

$$\left\|\frac{J_s x}{s} - \frac{J_t x}{t}\right\| \leqslant \varepsilon$$

where

$$\varphi_2'(\varepsilon, b, D, \eta, E, f) := \max\{\varphi(\varepsilon/3, b, f), \varphi_1(\varepsilon/6E, b, D, f(2D\eta(\min\{\varepsilon/12E, 2\})), \eta, f)\}$$

with φ as in Lemma 7.3.1 and φ_1 as in Lemma 7.3.4.

Proof. We have that there exits $z \in Ay$ such that $\|z\| \leqslant d + 2d\eta(\min\{\varepsilon/4, 2\})$ with $\|y\|, \|z\| \leqslant f(2D\eta(\min\{\varepsilon/4, 2\}))$. Thus, using Lemma 7.3.4, we have that for $t, s \geqslant \varphi_1(\varepsilon/2, b, D, f(2D\eta(\min\{\varepsilon/4, 2\})), \eta, f)$:

$$\left\|\frac{J_s x}{\|J_s x\|} - \frac{J_t x}{\|J_t x\|}\right\| \leqslant \left\|\frac{z}{\|z\|} + \frac{J_s x}{\|J_s x\|}\right\| + \left\|-\frac{z}{\|z\|} - \frac{J_t x}{\|J_t x\|}\right\| \leqslant \varepsilon.$$

Therefore, we in particular have that

$$\begin{aligned}
\left\|\frac{J_s x}{s} - \frac{J_t x}{t}\right\| &= \left\|\frac{J_s x}{\|J_s x\|}\frac{\|J_s x\|}{s} - \frac{J_t x}{\|J_t x\|}\frac{\|J_t x\|}{t}\right\| \\
&\leqslant \left\|\frac{J_s x}{\|J_s x\|}\frac{\|J_s x\|}{s} - \frac{J_s x}{\|J_s x\|}d\right\| + \left\|\frac{J_s x}{\|J_s x\|}d - \frac{J_t x}{\|J_t x\|}d\right\| \\
&\qquad + \left\|\frac{J_t x}{\|J_t x\|}d - \frac{J_t x}{\|J_t x\|}\frac{\|J_t x\|}{t}\right\| \\
&\leqslant \left|\frac{\|J_s x\|}{s} - d\right| + d\left\|\frac{J_s x}{\|J_s x\|} - \frac{J_t x}{\|J_t x\|}\right\| + \left|d - \frac{\|J_t x\|}{t}\right|.
\end{aligned}$$

Thus, for $t, s \geqslant \varphi_2'(\varepsilon, b, D, \eta, E, f)$, we get the claim by Lemma 7.3.1 together with the above. □

This result, which presents the quantitative version of the Cauchyness of $J_t x/t$ in the case that $d > 0$, can now be smoothed to omit this assumption. For this, note that through the trivial proof of the case of $d = 0$, one obtains the following quantitative version of the full result:

Lemma 7.3.6. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0,2] \to (0,1]$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y,z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon \right).$$

Let $\varepsilon > 0$ be given, assume that $E \in \mathbb{N}^*$ where $E \geqslant d$. Let $x \in \mathrm{dom} A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ where $b \geqslant \|x\|, \|v\|$. Then, for any $t, s \geqslant \varphi_2(\varepsilon, b, \eta, E, f)$:

$$\left\| \frac{J_s x}{s} - \frac{J_t x}{t} \right\| \leqslant \varepsilon$$

where

$$\varphi_2(\varepsilon, b, \eta, E, f) := \max\{\varphi(\varepsilon/4, b, f), \varphi(\varepsilon/3, b, f), \\ \varphi_1(\varepsilon/6E, b, \varepsilon/4, f(\varepsilon\eta(\min\{\varepsilon/12E, 2\})/2), \eta, f)\}$$

with φ as in Lemma 7.3.1 and φ_1 as in Lemma 7.3.4.

Proof. Suppose that $d \leqslant \varepsilon/4$. By Lemma 7.3.1, we have that

$$\left| \frac{\|J_t x\|}{t} - d \right| \leqslant \varepsilon/4$$

for any $t \geqslant \varphi(\varepsilon/4, b, f)$. Thus in particular we have that $\|J_t x\|/t \leqslant \varepsilon/2$ for all such t and thus

$$\left\| \frac{J_s x}{s} - \frac{J_t x}{t} \right\| \leqslant \left\| \frac{J_s x}{s} \right\| + \left\| \frac{J_t x}{t} \right\| \leqslant \varepsilon$$

for all $t, s \geqslant \varphi(\varepsilon/4, b, f)$ in that case. Otherwise $d \geqslant \varepsilon/4$ and thus the above result holds for $t, s \geqslant \varphi_2'(\varepsilon, b, \varepsilon/4, \eta, E, f)$ by Lemma 7.3.5 with $D = \varepsilon/4$. □

The rest of the proof given in [174] now relies on the use of the limit $-v_x$ of $J_t x/t$ for $t \to \infty$. By the above Lemma, this limit exists as X is complete. While we emphasized that this limit a priori depends on the starting point x, the following lemma (which provides a concrete quantitative version of Lemma 3.2 given in [174]) shows that this limit is actually unique, i.e all the v_x coincide.

Lemma 7.3.7. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0,2] \to (0,1]$ with (w.l.o.g.) $\eta(\varepsilon) \leqslant \varepsilon$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y,z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon \right).$$

Let $\varepsilon > 0$ be given, assume that $E \in \mathbb{N}^*$ and $D \in \mathbb{R}_{>0}$ where $E \geqslant d \geqslant D$ and let $z \in Ay$ be such that

$$\|z\| \leqslant d + 2d\eta\left(\min\left\{\frac{\varepsilon}{16(E+1)}, 2\right\}\right).$$

If $x \in \mathrm{dom}A$, then $\|z - v_x\| \leqslant \varepsilon$.

Proof. We write $\delta_\varepsilon = 2d\eta(\min\{\varepsilon/16(E+1), 2\})$. Then, for $\|z\| \leqslant d + \delta_\varepsilon$, we have

$$\begin{aligned}
\left\| z + \frac{J_t x}{t} \right\| &= \left\| z + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\| \\
&\leqslant \left\| z - \frac{d}{\|z\|} z \right\| + \left\| \frac{d}{\|z\|} z + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\| \\
&\leqslant \|z\| - d + \left\| \frac{d}{\|z\|} z + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\| \\
&\leqslant \delta_\varepsilon + \left\| \frac{d}{\|z\|} z + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\|.
\end{aligned}$$

Similar to before, we have

$$\begin{aligned}
\left\| \frac{d}{\|z\|} z + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\| &\leqslant \left\| \frac{d}{\|z\|} z - \frac{z}{\|z\|} \frac{\|J_t x\|}{t} \right\| + \left\| \frac{z}{\|z\|} \frac{\|J_t x\|}{t} + \frac{J_t x}{\|J_t x\|} \frac{\|J_t x\|}{t} \right\| \\
&= \left| d - \frac{\|J_t x\|}{t} \right| + \frac{\|J_t x\|}{t} \left\| \frac{z}{\|z\|} + \frac{J_t x}{\|J_t x\|} \right\|.
\end{aligned}$$

From this we obtain that

$$\left\| z + \frac{J_t x}{t} \right\| \leqslant \delta_\varepsilon + \frac{\varepsilon}{4}$$

for all

$$t \geqslant \max\{\varphi(\min\{\varepsilon/8, 1\}, b, f), \varphi_1(\varepsilon/8(E+1), b, D, c, \eta, f)\}$$

where $c, b \in \mathbb{N}^*$ are such that $c \geqslant \|y\|, \|z\|$ and $b \geqslant \|x\|, \|v\|$ for $v \in Ax$ as, for one, $t \geqslant \varphi(\min\{\varepsilon/8, 1\}, b, f)$ and thus

$$\left| d - \frac{\|J_t x\|}{t} \right| \leqslant \min\{\varepsilon/8, 1\}$$

by Lemma 7.3.1 as well as $\frac{\|J_t x\|}{t} \leqslant d + 1 \leqslant E + 1$ and, for another, $t \geqslant \varphi_1(\varepsilon/8(E+1), b, D, c, \eta, f)$ and thus

$$\frac{\|J_t x\|}{t} \left\| \frac{z}{\|z\|} + \frac{J_t x}{\|J_t x\|} \right\| \leqslant \varepsilon/8.$$

by Lemma 7.3.4. Then the properties of η imply that $\delta_\varepsilon \leqslant \varepsilon/4$ and thus

$$\left\| z + \frac{J_t x}{t} \right\| \leqslant \varepsilon/2$$

for all such t. Then

$$\|z - v_x\| \leqslant \left\| z + \frac{J_t x}{t} \right\| + \left\| v_x + \frac{J_t x}{t} \right\|$$

for all t and thus choosing

$$t = \max\{\varphi(\min\{\varepsilon/8, 1\}, b, f), \varphi_1(\varepsilon/8(E+1), b, D, c, \eta, f), \varphi_2(\varepsilon/2, b, \eta, E, f)\}$$

implies $\|z - v_x\| \leqslant \varepsilon$ by definition of v_x (which yields that φ_2 is a rate of convergence for $J_t x/t$ towards $-v_x$) and Lemma 7.3.6. □

Lemma 7.3.8. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0, 2] \to (0, 1]$ with (w.l.o.g.) $\eta(\varepsilon) \leqslant \varepsilon$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y, z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon \right).$$

Let $\varepsilon > 0$ be given, assume that $E \in \mathbb{N}^*$ where $E \geqslant d$ and let $z \in Ay$ be such that

$$\|z\| \leqslant d + \min\left\{ \varepsilon\eta\left(\min\left\{ \frac{\varepsilon}{16(E+1)}, 2 \right\} \right)/4, \varepsilon/8 \right\}.$$

If $x \in \mathrm{dom}A$, then $\|z - v_x\| \leqslant \varepsilon$.

Proof. Let ε be given. Then either $d \leqslant \varepsilon/8$ which, since $\|z\| \leqslant d + \varepsilon/8$, implies $\|z\| \leqslant \varepsilon/4$. For

$$t \geqslant \max\{\varphi(\varepsilon/4, b, f), \varphi_2(\varepsilon/4, b, \eta, E, f)\},$$

we then get

$$\begin{aligned} \|z - v_x\| &\leqslant \|z\| + \|v_x\| \\ &\leqslant \|z\| + d + \left| d - \frac{\|J_t x\|}{t} \right| + \left\| v_x + \frac{J_t x}{t} \right\| \\ &\leqslant \varepsilon. \end{aligned}$$

Otherwise we have $d \geqslant \varepsilon/8$ and thus, we get the same result for

$$\|z\| \leqslant d + \varepsilon\eta\left(\min\left\{ \frac{\varepsilon}{16(E+1)}, 2 \right\} \right)/4 \leqslant d + 2d\eta\left(\min\left\{ \frac{\varepsilon}{16(E+1)}, 2 \right\} \right)$$

by Lemma 7.3.7 with $D = \varepsilon/8$. □

Theorem 7.3.9. *Let X be a uniformly convex Banach space with a modulus of uniform convexity $\eta : (0,2] \to (0,1]$. Suppose that $f : \mathbb{R}_{>0} \to \mathbb{N}$ satisfies $f(\varepsilon) \geqslant f(\delta)$ for $\varepsilon \leqslant \delta$ and*

$$\forall \varepsilon > 0 \exists (y,z) \in A \left(\|y\|, \|z\| \leqslant f(\varepsilon) \wedge \|z\| - d \leqslant \varepsilon \right).$$

Assume that $E \in \mathbb{N}^*$ where $E \geqslant d$ and further that $x \in \mathrm{dom} A$ with $v \in Ax$ and $b \in \mathbb{N}^*$ with $b \geqslant \|x\|, \|v\|$. Then

$$\forall \varepsilon > 0 \forall t \geqslant \Phi(\varepsilon, b, \eta, E, f) \left(\frac{\left\| J_t^A x - S(t)x \right\|}{t} \leqslant \varepsilon \right)$$

where

$$\begin{aligned}\Phi(\varepsilon, b, \eta, E, f) := \max \Bigg\{ & \frac{4}{\varepsilon} \left(b + f \left(\min \left\{ \varepsilon \eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\} \right) \right), \\ & \frac{8}{\varepsilon} f \left(\min \left\{ \varepsilon \eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\} \right), \\ & \varphi_2(\varepsilon/2, b, \eta, E, f) \Bigg\}\end{aligned}$$

with

$$\begin{aligned}\varphi(\varepsilon, b, f) &:= \frac{8(b + f(\varepsilon/2))}{\varepsilon}, \\ \psi(K, b, D) &:= \frac{b+K}{D}, \\ \varphi_1(\varepsilon, b, D, c, \eta, f) &:= \max \Bigg\{ \psi(c+1, b, D), \psi \left(\left(\frac{4}{\varepsilon} + 1 \right) c, b, D \right), \\ &\qquad \left(\frac{D \left(2\eta(\min\{\varepsilon/2, 2\}) \right)^2}{18} \right)^{-1} (c+b), \\ &\qquad \varphi \left(\frac{D \left(2\eta(\min\{\varepsilon/2, 2\}) \right)^2}{18}, b, f \right) \Bigg\}, \\ \varphi_2(\varepsilon, b, \eta, E, f) &:= \max\{ \varphi(\varepsilon/4, b, f), \varphi(\varepsilon/3, b, f), \\ &\qquad \varphi_1(\varepsilon/6E, b, \varepsilon/4, f(\varepsilon\eta(\min\{\varepsilon/12E, 2\})/2), \eta, f) \}.\end{aligned}$$

Proof. Given ε, there are $z \in Ay$ such that

$$\|z\| \leqslant d + \min \left\{ \varepsilon \eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\}$$

and such that $\|y\|, \|z\| \leqslant f \left(\min \left\{ \varepsilon \eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\} \right)$. Now, we in particular have $\|A_r J_r a\| \leqslant |AJ_r a| \leqslant \|A_r a\|$ for all $a \in \mathrm{dom} A$ and thus $\left\| A_{t/n} J_{t/n}^i y \right\| \leqslant$

$\left\|A_{t/n}J_{t/n}^{i-1}y\right\|$. Iterating this gives

$$\left\|A_{t/n}J_{t/n}^{i}y\right\| \leqslant \left\|A_{t/n}y\right\| \leqslant \|z\| \leqslant d + \min\left\{\varepsilon\eta\left(\min\left\{\frac{\varepsilon/8}{16(E+1)}, 2\right\}\right)/32, \varepsilon/64\right\}$$

for all $i \in [0; n-1]$. Now we get

$$\left\|A_{t/n}J_{t/n}^{i}y - v_x\right\| \leqslant \varepsilon/8$$

for any $i \in [0; n-1]$ by Lemma 7.3.8 which implies $\left\|(y - J_{t/n}^{n}y)/t - v_x\right\| \leqslant \varepsilon/8$ for all t and all n as

$$\begin{aligned}
\left\|\frac{y - J_{t/n}^{n}y}{t} - v_x\right\| &= \left\|\frac{\sum_{i=0^{n-1}} J_{t/n}^{i}y - J_{t/n}^{i+1}y}{nt/n} - \frac{\sum_{i=0}^{n-1} v_x}{n}\right\| \\
&= \left\|\frac{\sum_{i=0}^{n-1}\left(\frac{J_{t/n}^{i}y - J_{t/n}^{i+1}y}{t/n} - v_x\right)}{n}\right\| \\
&\leqslant \frac{\sum_{i=0}^{n-1}\left\|A_{t/n}J_{t/n}^{i}y - v_x\right\|}{n}.
\end{aligned}$$

Thus

$$\left\|\frac{y - S(t)y}{t} - v_x\right\| \leqslant \varepsilon/8$$

for all t.

Then in particular

$$\left\|\frac{S(t)y}{t} + v_x\right\| \leqslant \frac{\|y\|}{t} + \left\|\frac{y - S(t)y}{t} - v_x\right\| \leqslant \frac{\|y\|}{t} + \varepsilon/8$$

for all t. In particular, for

$$t \geqslant (\varepsilon/8)^{-1} f\left(\min\left\{\varepsilon\eta\left(\min\left\{\frac{\varepsilon/8}{16(E+1)}, 2\right\}\right)/32, \varepsilon/64\right\}\right),$$

we have

$$\left\|\frac{S(t)y}{t} + v_x\right\| \leqslant \varepsilon/4.$$

Continuing, we obtain

$$\left\|\frac{S(t)x}{t} + v_x\right\| \leqslant \left\|\frac{S(t)y}{t} + v_x\right\| + \frac{\|x - y\|}{t}$$

which implies

$$\left\|\frac{S(t)x}{t} + v_x\right\| \leqslant \varepsilon/2$$

for all

$$t \geqslant \max \left\{ (\varepsilon/8)^{-1} f \left(\min \left\{ \varepsilon\eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\} \right), \right.$$
$$\left. (\varepsilon/4)^{-1} \left(b + f \left(\min \left\{ \varepsilon\eta \left(\min \left\{ \frac{\varepsilon/8}{16(E+1)}, 2 \right\} \right) /32, \varepsilon/64 \right\} \right) \right) \right\}.$$

Finally, we get

$$\left\| \frac{S(t)x - J_t x}{t} \right\| \leqslant \left\| \frac{S(t)x}{t} + v_x \right\| + \left\| v_x + \frac{J_t x}{t} \right\|$$

and thus

$$\left\| \frac{S(t)x - J_t x}{t} \right\| \leqslant \varepsilon$$

for all $t \geqslant \Phi(\varepsilon, b, \eta, E, f)$ by Lemma 7.3.6 and the definition of v_x (which yields that φ_2 is a rate of convergence as before). □

7.4 Logical remarks on the above results

Lastly, we want to outline the additional modifications to H_p^ω necessary for formalizing the proofs of the theorems of Plant and Reich. These modifications in that way give rise to the systems and bound extraction results underlying the extractions outlined in this chapter. In that context, we here in particular move away from the use of arbitrary real errors ε and again consider representations of errors via natural numbers through 2^{-k}.

At first, both results are formulated for points $x \in \mathrm{dom}A$ and by the logical methodology, this stands for the existential assumption $\exists y(y \in Ax)$ which yields that at least a priori the extracted rates will in particular depend on an upper bound on the norm of this witness which is also the case for the above rates.

The second prominent assumption in both results is that of uniform convexity which was quantitatively treated above via the modulus of uniform convexity η. Formally, this can be achieved by adding an additional constant η of type 1 together with a corresponding axiom stating that it represents a modulus of uniform convexity for X (see [96] for more details):

$$\forall x^X, y^X, k^0 \left(\|x\|_X, \|y\|_X <_{\mathbb{R}} 1 \wedge \left\| \frac{x +_X y}{2} \right\|_X >_{\mathbb{R}} 1 - 2^{-\eta(k)} \to \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-k} \right).$$

To really formally encapsulate the previous proof where η was applied to various reals, one would first have to extend η to $\mathbb{Q} \cap (0, 2]$ via

$$\eta(\varepsilon) := 2^{-\eta(\min k[2^{-k} \leqslant \varepsilon])}$$

for $\varepsilon \in \mathbb{Q} \cap (0, 2]$ and then move to rational approximations of the reals in question. We avoid spelling this out any further.

We now first focus on the theorem of Plant. The main object featuring in Plant's proof (and consequently in the above results as well) is the limit functional $|Ax|$. The use of this functional can now be emulated in the context of H_p^ω by extending the underlying language with a further constant of type $1(X)$ which we denote by $|A \cdot|$ (where we correspondingly denote $|A \cdot|x$ by $|Ax|$ for simplicity). One first natural axiom for this constant is induced by the natural bound on $|Ax|$ by $\|v\|$ for $v \in Ax$ witnessing $x \in \mathrm{dom}A$:

$$\forall x^X, v^X, \lambda^1 \left(v \in Ax \wedge 0 <_{\mathbb{R}} \lambda <_{\mathbb{R}} \lambda_0 \to \|A_\lambda x\|_X \leqslant_{\mathbb{R}} |Ax| \leqslant_{\mathbb{R}} \|v\|_X\right). \qquad (L1)$$

As shortly mentioned in the above quantitative results, the convergence of $\|A_\lambda x\|$ to $|Ax|$ for $x \in \mathrm{dom}A$ as $\lambda \to 0$ is "equivalent" to the lower semi-continuity of $|Ax|$ on $\mathrm{dom}A$. This vague "equivalence" can now be made precise through the system $H_p^\omega + (L1)$ in the following sense:

Proposition 7.4.1. *Over $H_p^\omega + (L1)$, the following are equivalent:*

1. *$\|A_t x\| \to |Ax|$ while $t \to 0^+$ for all $x \in \mathrm{dom}A$, i.e.*

$$\forall x^X, k^0 \exists n^0 \left(x \in \mathrm{dom}A \to |Ax| - \|A_{2^{-n}} x\|_X \leqslant_{\mathbb{R}} 2^{-k}\right);$$

2. *lower semi-continuity for $|Ax|$ for all $x \in \mathrm{dom}A$, i.e.*

$$\forall k^0, x^X \exists m^0 \forall y^X \Big(x \in \mathrm{dom}A \wedge y \in \mathrm{dom}A \\ \wedge \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-m} \to |Ax| - |Ay| \leqslant_{\mathbb{R}} 2^{-k}\Big).$$

Proof. From (1) to (2), let $x \in \mathrm{dom}A$ and k be given. For $y \in \mathrm{dom}A$, we have

$$\begin{aligned} |Ax| - |Ay| &\leqslant |Ax| - \|A_\lambda y\| \\ &\leqslant |Ax| - \|A_\lambda x\| + |\, \|A_\lambda x\| - \|A_\lambda y\| \,| \\ &\leqslant |Ax| - \|A_\lambda x\| + 2/\lambda \, \|x - y\| \end{aligned}$$

for any $\lambda \in (0, \lambda_0)$. Now, using (1) we pick n such that $|Ax| - \|A_{2^{-n}}x\| \leqslant 2^{-(k+1)}$ and then pick $m = n+k+2$ such that $2^{n+1}\|x-y\| \leqslant 2^{-(k+1)}$ which yields $|Ax| - |Ay| \leqslant 2^{-k}$.

From (2) to (1), let $x \in \mathrm{dom}A$ and k be given. Using (2), we pick an m such that $|Ax| - |Ay| \leqslant 2^{-k}$ for all $y \in \mathrm{dom}A$ such that $\|x-y\| \leqslant 2^{-m}$. Now, for $n \geqslant b+m$ for $b \geqslant \|v\|$ for some $v \in Ax$, we then get

$$\|x - J_{2^{-n}}x\| \leqslant 2^{-n}\|v\| \leqslant 2^{-m}$$

which in particular implies

$$|Ax| - \|A_{2^{-n}}x\| \leqslant |Ax| - |AJ_{2^{-n}}x| \leqslant 2^{-k}$$

using $A_{2^{-n}}x \in AJ_{2^{-n}}x$. □

In that way, the convergence of $\|A_\lambda x\|$ to $|Ax|$ on $\mathrm{dom}A$ relates to an extensionality principle of $|A\cdot|$. Now, in the context of set-valued operators, these continuity and extensionality principles can be logically complicated and their study actually gives rise to a hierarchy of fragments of extensionality with a corresponding hierarchy of continuity principles with various intricacies (which will be discussed in Chapter 11). In any way, as in the case of the functional $\langle\cdot,\cdot\rangle_s$, the logical methodology based on the monotone Dialectica interpretation now implies the following quantitative version of the statement of item (2): under this interpretation, the statement (2) is upgraded to the existence of a "modulus of uniform lower-semicontinuity" which, as with $\langle\cdot,\cdot\rangle_s$, by the uniformity on x induced by majorization, is essentially a modulus of uniform continuity. Concretely, this uniformized version of item (2) can be formally hardwired into the system by extending it with an additional constant φ of type $0(0)(0)$ together with the axiom

$$\forall b^{\mathbb{N}}, k^{\mathbb{N}}, x^X, y^X, u^X, v^X \Big(\Big(u \in Ax \wedge v \in Ay \wedge \|x\|_X, \|y\|_X, \|u\|_X, \|v\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\varphi(k,b)}\Big) \to |Ax| - |Ay| \leqslant_{\mathbb{R}} 2^{-k}\Big). \tag{L2}$$

Under this extension, Lemma 7.2.1 is then the natural extraction of a corresponding rate of convergence from the above equivalence proof, under this (therefore) necessary assumption of a modulus of uniform continuity for $|A\cdot|$, following the previous metatheorems. Note however that these metatheorems in general, through this treatment of $|A\cdot|$, imply a dependence of the extracted bounds on a majorant for the constant $|A\cdot|$, i.e. on a function $f : \mathbb{N} \to \mathbb{N}$ such that

$$\|x\| \leqslant b \to |Ax| \leqslant f(b) \text{ for all } x \in \mathrm{dom}A.$$

Only under this additional dependence on a majorant for $|A\cdot|$ do the previous metatheorems contained in Theorems 4.4.6 and 4.4.7 extend to $H^\omega_p + (L1) + (L2)$.

Here, we shortly want to make a note on the strength of the existence of such a majorant. For this, recall the notion of majorizability for set-valued operators from Chapter 3: an operator $A : X \to 2^X$ is called majorizable if there exists a function $f : \mathbb{N} \to \mathbb{N}$ such that

$$\forall x \in \mathrm{dom}A, b \in \mathbb{N} \left(\|x\| \leqslant b \to \exists y \in Ax \left(\|y\| \leqslant f(b) \right) \right).$$

As discussed in [165], there are non-majorizable operators and so the assumption that A is majorizable is a proper restriction. In particular, note now that if A is such that the minimal selection $A^\circ x = \mathrm{argmin}\{\|y\| \mid y \in Ax\}$ exists, then $|Ax| = \|A^\circ x\|$ and thus majorizability of A is equivalent to majorizability of $|A\cdot|$. Thus, while in general potentially a bit weaker, the assumption of majorizability of $|A\cdot|$ in particular also seems to carry additional strength similar to that of majorizability of A in most cases.

However, as apparent from the result in Lemma 7.2.1, such a majorant however does not feature in the extracted bounds and we actually find that such a majorant also does not feature in any of the other quantitative results in the context of Plant's theorem. While this seems to be a particular coincidence in the context of Lemma 7.2.1, there is actually a logical reason which guarantees this "non-dependence" a priori for all the other results. Concretely, the reason is that all the other proofs analyzed have the two crucial properties that, for one, they can be formalized under the assumption of a rate of convergence for $\|A_\lambda x\|$ toward $|Ax|$ which can similarly be added to the system and that, for another, they are "pointwise" results in x in the sense that they do not require knowledge of $|A\cdot|$ for any point other than x. In that way, instead of following the above route of formalizing the whole functional $|A\cdot|$, one can add two constant representing this "particular" x and a witness $v \in Ax$ as well as a constant $|Ax|$ of type 1 for this single value of $|A\cdot|$ at the constant x and a constant φ representing a rate of convergence for $\|A_\lambda x\|$ to $|Ax|$ for this single constant x. Then, the other proofs still formalize and in particular depend only on majorants for φ, x, v and $|Ax|$ and the one for the latter three can be assumed to coincide and to be represented in the above results by b. In particular, the strong assumption of majorizability of $|A\cdot|$ can be avoided a priori in that way. That the extracted rates are true for all x then is drawn as a conclusion on the metalevel as the additional constants were generic. In this way, this also provides a logical insight on why all the other results in the context of Plant's theorem remain true if φ_1 represents any other rate of convergence besides the one constructed from the modulus of uniform continuity for $|A\cdot|$ as commented on before.

As a last comment on the logical particularities of the proofs towards Plant's theorem, we want to note in the context of Miyadera's lemma from [148] that the only properties of $\langle\cdot,\cdot\rangle_s$ required in the proof given in [148] are the properties discussed in Section 4.3.3. Further, by the fact that the proof given by Crandall in [48] of his respective result actually only invokes Miyadera's lemma for $x \in \mathrm{dom}A$ and for *some* $\zeta^* \in J(x - x_0)$, this ζ^* can thus for simplicity be assumed to coincide with j_{v,y_0} for $v \in Ax$ witnessing $x \in \mathrm{dom}A$ and $y_0 \in Ax_0$ as in Miyadera's lemma. So, combined we have that this use of Miyadera's lemma in the context of the proof of Plant's result immediately formalizes in the system $H^\omega_p + (+)$.

We now consider the theorem of Reich (which features less logical subtleties). The main object featuring in Reich's proof is the value d, the infimum over norms of all elements in the range of the operator. Internally in H^ω_p, this value can be represented by adding a further constant of type 1 which we, for simplicity, also denote by d together with a further constant f of type 1 representing the witness for the monotone Dialectica interpretation of the property

$$\forall k \exists y, z \left(z \in Ay \wedge \|z\| - d \leqslant 2^{-k}\right)$$

expressing that d indeed is the said infimum. So, we can concretely facilitate the use of d by adding the following two axioms for d:

$$\begin{cases} \forall y^X, z^X \left(z \in Ay \to d \leqslant_{\mathbb{R}} \|z\|_X\right), \\ \forall k^0 \exists y, z \leqslant_X f(k)1_X \left(z \in Ay \wedge \|z\|_X - d \leqslant_{\mathbb{R}} 2^{-k}\right). \end{cases} \tag{d}$$

The additional constants are immediately majorizable: f is majorized by f^M as it is of type 1 and d is just majorized by $(n)_\circ$ for $n \geqslant \|d_X\|$. Therefore the bound extraction theorems extend to this augmentation of H^ω_p in an immediate way where, in particular, the extracted bounds will in general depend on an upper bound on d as can be seen from some of the bounds extracted in the context of Reich's theorem.

The second particularity of the formalization of the proof of Reich's result is that one actually needs to work with the limit of $J_t x/t$, called $-v_x$ in the above, as a concrete object. In the context of the limit operator C however, we can formally deal with this object in the context of the formal systems underlying this extraction rather immediately by utilizing the previously extracted rate $\varphi_2(k)$ (where we for simplicity omit the other parameters for now and switched to a representation of errors via 2^{-k}) to then address the limit v_x in the system by considering

$$v_x = -C\left(\left(\frac{J_{\varphi_2(k)}x}{\varphi_2(k)}\right)_k\right).$$

In particular, with this definition of v_x, the other proofs in the context of Reich's theorem immediately formalize.

As a last logical comment, we shortly want to discuss on the particular use of the law of excluded middle (and thus of classical logic) in the proofs for the results of Reich and Plant and how this features in the extractions, considering the fact that rates of convergence were nevertheless extracted in the absence of monotonicity. This in fact relates to the circumstances of the (previously called "smoothable') case distinctions. Namely, as can be observed by closer inspection of the corresponding proofs, the only part where classical logic actually features in the proofs of Reich and Plant is through the use of multiple case distinctions which, in the case of Reich's result, takes the form on dividing the proof between whether

$$d = 0 \text{ or } d > 0$$

and, in the case of Plant's result, takes the form of dividing the proof between whether

$$|Ax| = 0 \text{ or } |Ax| > 0.$$

The deductions of the main results from both parts of this case distinction are essentially constructive (where the $=0$-case is almost trivial in both cases) and in that way, the constructive metatheorems actually allow for the extraction of a rate of convergence from the >0-cases as the corresponding results are of the form

$$d >_{\mathbb{R}} 0 \to C \equiv \forall c^0 \left(d \geqslant_{\mathbb{R}} 2^{-c} \to C\right) \text{ and } |Ax| >_{\mathbb{R}} 0 \to C \equiv \forall c^0 \left(|Ax| \geqslant_{\mathbb{R}} 2^{-c} \to C\right)$$

where C is any of the respective convergence statements. These rates will moreover depend on the parameter c. For the $=0$-cases, being of the form

$$d =_{\mathbb{R}} 0 \to C \text{ and } |Ax| =_{\mathbb{R}} 0 \to C$$

where C is any of the respective convergence statements, we find that in these cases one can actually find different constructive proofs (compared to the ones given in the literature) of the classically equivalent but constructively stronger statements

$$\exists c'^0 \left(d \leqslant_{\mathbb{R}} 2^{-c'} \to C\right) \text{ and } \exists c'^0 \left(|Ax| \leqslant_{\mathbb{R}} 2^{-c'} \to C\right).$$

These new proofs of said statements (which were presented and analyzed in the previous section) are again essentially constructive so that the constructive metatheorems guarantee the extraction of a rate again, now together with the extraction of an upper bound on (and thus a realizer of) the value c'. The previously mentioned "smoothening" is now just a combination of these two cases by instantiating the former rate with $c = c'$ and combining the two resulting rates.

8 Proof mining for the dual of a Banach space with extensions for uniformly Fréchet differentiable functions

8.1 Introduction

In this chapter, we move away from the theory of nonlinear semigroups and strive to extend the current logical methods used in proof mining so that they become applicable to proofs which involve some of the most fundamental notions from convex and nonlinear functional analysis, including the dual space of a Banach space and its norm as well as uniformly Fréchet differentiable functions and their gradients and Fenchel conjugates.

In more detail, since the first modern metatheorems of proof mining were developed in [71, 95], a focus for applications of proof mining has been placed on the areas of convex and functional analysis. Interestingly, one of the most fundamental objects in the context of the latter, the continuous dual of a Banach space, has not yet received a proper treatment (due to various difficulties arising in that context which will be discussed further below). Similarly, many if not most applications to convex analysis have been concerned with fixed point iterations for nonexpansive maps and their cousins as well as abstract monotone and accretive operator theory and so, also here, some of the main objects in convex analysis have not been treated so far, in particular including the gradients of differentiable convex functions as well as their Fenchel conjugates. In that way, proof mining has so far missed out on some of the most promising areas of applications which rely on these objects. For two prominent examples, we want to mention the theory and applications of the prominent Bregman distances (going back to the seminal work [22]) as well as the theory of von Neumann algebras.[1]

[1]For the latter, an approach for extending proof mining methods to the context of tracial von

Now, the main results of this chapter are logical metatheorems that quantify and allow for the extraction of the computational content of theorems pertaining to the use of the continuous dual of an abstract normed space together with the associated dual norm. This is achieved by extending the systems currently in use for proof mining in the context of normed linear spaces by carefully selected constants and corresponding axioms that govern the use of the involved objects. In particular, a novel approach is used in this context to circumvent some of the difficulties which are a priori present when treating the dual space: The dual space is a concretely defined object relative to the underlying normed space represented by, say, an abstract type X. Naively, elements of the continuous dual therefore live in the type $1(X)$ and, in that way, singling out the continuous linear maps from all functionals of that type requires the use of a predicate which is of high quantifier complexity and which thus makes essentially all attempts at a direct specification futile if one wants to retain meaningful bound extraction results as the high computational strength of the comprehension needed to deal with the predicate would distort the complexity of bounds extracted from proofs which discuss these objects only in an abstract way while not carrying any apparent computational strength in the principles used in the proof. A second issue is that the norm of the continuous dual is also a concrete object that derives from the norm of the underlying space X via the use of a supremum over elements from this abstract space and such suprema cannot be represented in the pure underlying language of the systems commonly used in proof mining. We avoid these problems in the following ways: Instead of specifying the continuous dual as the subspace of *all* continuous and linear functionals of type $1(X)$, we present an abstract approach using an additional abstract base type X^* and then axiomatically specify that all elements of this abstract space represented by X^* behave like continuous linear functionals. However, there are no axioms specifying that this abstract space really contains representations for *all* elements from the continuous dual associated with X as represented by a set of functionals of type $1(X)$. Instead, we only include a corresponding rule that facilitates the closure of the space as represented by the new abstract type X^* under functionals which are provably linear and continuous. In this way, our approach is intensional (and in some way similar to the treatment of set-valued operators in the context of proof mining developed in [165] as will be discussed later). This intensional treatment of the dual then allows us to utilize a proof-theoretically tame approach for treating

Neumann algebras has recently been given in [157].

suprema over (certain) bounded sets in abstract spaces, developed in the first part of this chapter, to provide defining axioms for the norm of the dual.

As discussed before already, the success of applications of proof mining to concrete mathematical proofs in many ways relies on a modularity of this logical approach in the sense that the main logical systems can be extended and adapted with specific mathematical objects or notions and associated axioms to fit specific problems, all the while guaranteeing that our metatheorems still hold. As examples of such extensions, we shall discuss how one can utilize the new system for the dual of a normed space to provide a novel treatment of the reflexivity property of a Banach space (in certain circumstances) and with that the second dual. Further, we extend these systems to deal with various notions from convex analysis that utilize the dual of a normed space, including uniformly Fréchet differentiable functions and their gradients as well as corresponding Fenchel conjugates, where in particular the treatment of the latter is made possible by again utilizing the intensional approach to the dual which allows for a treatment of the supremum defining the Fenchel conjugate via the proof-theoretically tame approach to suprema over bounded sets mentioned before. So also in those cases, we find that the intensional approach provides mathematically strong systems for treating very concrete objects in the context of systems that allow for bound extraction metatheorems which accurately reflect the complexity of the principles used in proofs by the complexity of the extracted bounds.

The applicability of the metatheorems for these systems as established in this chapter will then in particular be justified by the fact that they allow for many new case studies to be carried out in the areas discussed above and examples for such applications will be given in the next Chapter 9. However, we want to also mention the works [1, 7, 12, 22, 38, 40, 83, 121, 191, 213] as promising future applications as, by inspection of the proofs, they seem to be formalizable in (suitable extensions of) the systems introduced here. Lastly, we also strongly believe that the general approach to suprema over bounded sets introduced here as well as the tame intensional approach to the dual space and to convex functions and their gradients and conjugates will be useful in inspiring further developments in the realm of the logical metatheorems of proof mining. As an initial indication of this, in Chapter 10, we will extend the systems introduced in this chapter to provide a suitable base for a treatment of monotone operators on Banach spaces as introduced by Browder [28, 30]. Further, in Chapter 11, we treat the Hausdorff-metric in systems which are amenable to proof mining metatheorems by using an intensional approach to the sets measured by the Hausdorff-metric together with the tame approach to suprema over bounded sets presented here and we then illustrate the applicability of this treatment in the last Chapter 12 where we provide quantitative results on Mann-type iterations of set-valued nonexpansive mappings.[2]

8.2 Proof-theoretically tame suprema over bounded sets

In this section, we now want to present a way that suprema over (certain) bounded sets in abstract spaces can be treated in the context of finite type arithmetic such that one retains meaningful bound extraction theorems in the sense that the treatment of the supremum in question does not result in any change in the computational strength of extracted bounds (besides of that caused by the other principles used in the proof). The presentation is conceptual and in that way to some degree informal. We will later discuss concrete instantiating examples for suprema where such a treatment can be utilized. In the following, we focus on the case of normed spaces and consequently work over (possibly extensions of) the language of $\mathcal{A}^\omega[X,\|\cdot\|]$. The same considerations can however be immediately applied in the context of metric spaces too.

Assume for this that we have a predicate $C(x,\underline{p})$ specifying a subset of X (possibly in an extension of the underlying language) in terms of external parameters $\underline{p}$ with types $\underline{\sigma}=\sigma_1,\dots,\sigma_k$ from a second set specified by a predicate $D(\underline{p})$. Write $\underline{\sigma}^t=\sigma_k,\dots,\sigma_1$ as in Chapter 2. Then, stating for an additional term s of type $1(\underline{\sigma}^t)$ that it represents the supremum of a function f of type $1(\underline{\sigma}^t)(X)$ over the set specified by C (if existent), i.e.

$$\sup_{x^X:C(x,\underline{p})} f(x,\underline{p}) =_{\mathbb{R}} s(\underline{p}) \text{ for all } \underline{p}^{\underline{\sigma}} \text{ with } D(\underline{p}),$$

can be facilitated by two axioms: one stating that $s(\underline{p})$ is an upper bound, i.e.

$$\forall \underline{p}^{\underline{\sigma}}, x^X \left(D(\underline{p}) \wedge C(x,\underline{p}) \to f(x,\underline{p}) \leqslant_{\mathbb{R}} s(\underline{p})\right), \tag{$(S)_1$}$$

[2]Besides these examples listed here, we also want to mention that intensional methods together with the tame approach to suprema over bounded sets may in particular be useful to treat the so-called generalized Bregman distances recently introduced by Burachik, Dao and Lindstrom [33] and that the approach to the dual space may be adapted to treat function spaces between general vector spaces in order to treat associated operator algebras.

as well as an axiom stating that the values of $f(x,\underline{p})$ get arbitrarily close to $s(\underline{p})$ over the specified set, i.e.

$$\forall \underline{p}^{\underline{\sigma}} \left(D(\underline{p}) \to \forall k^0 \exists x^X \left(C(x,\underline{p}) \wedge s(\underline{p}) - 2^{-k} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right). \qquad (S)'_2$$

Remark 8.2.1. Note that it is a rather immediate consequence of $(S)'_2$ that $s(\underline{p})$ also satisfies the usual defining property of being a supremum in the sense that $s(\underline{p})$ is the least upper bound of all $f(x,\underline{p})$ over the specified set, i.e.

$$\forall \underline{p}^{\underline{\sigma}}, M^1 \left(D(\underline{p}) \wedge M <_{\mathbb{R}} s(\underline{p}) \to \exists x^X \left(C(x,\underline{p}) \wedge M <_{\mathbb{R}} f(x,\underline{p})\right)\right), \qquad (+)$$

as by unraveling the quantifiers hidden in the real inequalities in the above statement and prenexing accordingly, we get that (+) is in fact equivalent to

$$\forall \underline{p}^{\underline{\sigma}}, M^1, k^0 \exists x^X, j^0 \left(D(\underline{p}) \wedge M + 2^{-k} <_{\mathbb{R}} s(\underline{p}) \to \left(C(x,\underline{p}) \wedge M + 2^{-j} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right), \qquad (++)$$

and so, assuming $M + 2^{-k} < s(\underline{p})$, we pick an x using $(S)'_2$ that satisfies $s(\underline{p}) - 2^{-(k+1)} \leqslant f(x,\underline{p})$. This x therefore also satisfies $M + 2^{-(k+1)} \leqslant f(x,\underline{p})$. So (++) holds true with this x and $j = k+1$.

In and of themselves, these schemes are not amenable to proof mining methods without resulting in additional computational strength. We now want to discuss situations in which the above two axioms do become admissible a priori in the context of bound extraction theorems (in the sense that they do not result in additional computational strength). In particular, we want to consider what happens if the set specified by $C(x,\underline{p})$ is such that every element x satisfies (not necessarily provably) that $\|x\| \leqslant b(\underline{p})$ for some additional term b of type $1(\underline{\sigma}^t)$, i.e. the elements x such that $C(x,\underline{p})$ holds true are bounded in terms of the parameters $\underline{p}$. In that case, the existential quantifier in $(S)'_2$ becomes bounded and, after prenexing the inner quantifiers accordingly, the statement can therefore be equivalently written as

$$\forall \underline{p}^{\underline{\sigma}}, k^0 \exists x^X \leqslant_X b(\underline{p})1_X \left(D(\underline{p}) \to \left(C(x,\underline{p}) \wedge s(\underline{p}) - 2^{-k} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right). \qquad (S)_2$$

Now, in the case of a quantifier-free C and an existential D, the above statement is of the form Δ exhibited in [76, 96] (and discussed before at various places, see e.g. Chapter 3) which is a priori permissible in the bound extraction theorems based on the monotone functional interpretation. Even further, the statement is still of the form Δ if C is purely universal. In that case however, the boundedness statement $(S)_1$ can only be rephrased in an admissible way if C can be equivalently written as an existential statement or if the universal quantifiers can themselves be bounded.

By generalizing this pattern of the duality of the requirements on C induced by $(S)_1$ and $(S)_2$, we can immediately exhibit a much larger class of statements which are a priori permissible for C: the above approach indeed yields admissible ways of phrasing suprema if C can be simultaneously written as a formula of the form

$$\forall \underline{a}_1^{\underline{\delta}_1} \exists \underline{b}_1 \leqslant_{\underline{\sigma}_1} \underline{r}_1 \underline{a}_1 \ldots \forall \underline{a}_n^{\underline{\delta}_n} \exists \underline{b}_n \leqslant_{\underline{\sigma}_n} \underline{r}_n \underline{a}_1 \ldots \underline{a}_n \forall \underline{c}^{\underline{\gamma}} D_{qf}(x, \underline{p}, \underline{a}_1, \ldots, \underline{a}_n, \underline{b}_1, \ldots, \underline{b}_n, \underline{c})$$

which is a kind of generalized form Δ which we, following Remark 10.24 in [96], denote by Δ^* as well as equivalently as a formula of the form

$$\exists \underline{\widetilde{a}}_1^{\underline{\widetilde{\delta}}_1} \forall \underline{\widetilde{b}}_1 \leqslant_{\underline{\widetilde{\sigma}}_1} \underline{\widetilde{r}}_1 \underline{\widetilde{a}}_1 \ldots \exists \underline{\widetilde{a}}_m^{\underline{\widetilde{\delta}}_m} \forall \underline{\widetilde{b}}_m \leqslant_{\underline{\widetilde{\sigma}}_m} \underline{\widetilde{r}}_m \underline{\widetilde{a}}_1 \ldots \underline{\widetilde{a}}_m \exists \underline{\widetilde{c}}^{\underline{\widetilde{\gamma}}} \widetilde{D}_{qf}(x, \underline{p}, \underline{\widetilde{a}}_1, \ldots, \underline{\widetilde{a}}_m, \underline{\widetilde{b}}_1, \ldots, \underline{\widetilde{b}}_m, \underline{\widetilde{c}})$$

which we want to denote by $\overline{\Delta^*}$. In more suggestive words, the statements $(S)_1$ and $(S)_2$ are a priori admissible in particular if C is a '$\Delta_1(\Delta^*)$' formula. Further, it is clear that D can also be of the form $\overline{\Delta^*}$ as it is immediate to see that also in that case, both statements $(S)_1$ and $(S)_2$ have a monotone functional interpretation.

However, in many cases the mathematical particularities of a situation at hand actually yield that such a representation of C is not even necessary for specifying a concrete supremum in an admissible way since other facts about it sometimes allow one to equivalently express that $s(\underline{p})$ is an upper bound for the given function over the given set in a way that does not require the above format of $(S)_1$. An immediate example where the above formulation of $(S)_1$ can be avoided is when the bounded subset specified by C is just $\overline{B}_r(0)$ in, say, a given normed space $(X, \|\cdot\|)$ and $D(\underline{p}, r)$ specifies a set of parameters $\underline{p}, r$ as before (now with types $\underline{\sigma}, 1$). If f is additionally extensional in that case, then the statement $(S)_1$ can be replaced by

$$\forall r^1, \underline{p}^{\underline{\sigma}} \forall x^X \left(D(\underline{p}, r) \to f(\widetilde{x}^r, \underline{p}, r) \leqslant_{\mathbb{R}} s(\underline{p}, r) \right)$$

where we make use of the functional[3]

$$\widetilde{x}^r = \frac{rx}{\max_{\mathbb{R}}\{\|x\|_X, r\}}$$

which allows for implicit quantification over elements from $\overline{B}_r(0)$.

[3]This functional seems to have first been used for $r = 1$ in [111].

In that way, it is in many cases in particular the complexity of $D(\underline{p})$, specifying the set of parameters, that is crucial for the admissibility of the above axioms. However, even in situations where a natural $D(\underline{p})$ is not of the right complexity, one can sometimes mitigate the resulting issues by providing a suitable quantifier-free *intensional description* of the set specified by $D(\underline{p})$ (potentially over an extended language). The case that we want to make in this chapter (as well as in this thesis for that matter) is that such situations, where the circumstances allow for an intensional treatment of the set specified by $D(\underline{p})$ such that the above treatment is applicable so that one can deal with certain suprema in that context but one nevertheless retains meaningful and mathematically strong systems that allow for the formalization of theorems and proofs from the respective areas that one wants to treat, occur rather frequently in the mainstream mathematical literature. We therefore want to make the case that this perspective thus provides a suitable way of approaching many previously untreated objects from (nonlinear) analysis. Concretely, the following sections will present some prime examples for such situations where we will in particular see that, in the context of an intensional formulation of the dual space of a Banach space, both the norm of that dual as well as the conjugate of a convex function can be treated in such a manner which results in proof-theoretically tame but mathematically strong systems for these areas, unlocking these branches for methods from proof mining for the first time.

8.3 A formal system for a normed space and its dual

In this section, we will now define the respective extensions of $\mathcal{A}^\omega[X, \|\cdot\|]$ that allow us to deal with notions in the context of the dual space of the normed space represented by X. For this, given a real normed space $(X, \|\cdot\|)$, we write X^* for the continuous dual of X and we write $\langle x, x^* \rangle$ for application of an $x^* \in X^*$ to an $x \in X$.

The main object associated with X^* is of course the norm $\|\cdot\|$ that turns X^* into a normed space which in particular will be a Banach space. The norm on X^* is concretely defined as

$$\|x^*\| = \sup\{|\langle x, x^* \rangle| \mid x \in X, \|x\| \leqslant 1\}$$

for $x^* \in X^*$. Any other basic notions from functional analysis will be introduced as needed throughout the chapter but we in general refer to [185, 202] for standard references on the subject.

The formal approach we choose towards the dual space is now as discussed in the introduction: We treat the dual space as an intensional object and so, instead of specifying the dual space as those objects with type $1(X)$ which indeed represent continuous linear functionals $X \to \mathbb{R}$, we introduce a new abstract type X^* into the language and correspondingly consider the extended set of types T^{X,X^*} defined as

$$0, X, X^* \in T^{X,X^*}, \quad \rho, \tau \in T^{X,X^*} \Rightarrow \tau(\rho) \in T^{X,X^*}.$$

This new type X^* is used to abstractly signify a space which we consider to be the dual space of X.

In and of itself, the immediate issue with this is that elements of type X^* have no relationship with elements of type X. To restore the application character of elements of type X^*, i.e. that they shall represent functionals that can be applied to elements of type X, we then need to further introduce a functional $\langle \cdot, \cdot \rangle_{X^*}$ of type $1(X)(X^*)$ by means of a new constant with suitable axioms that facilitates an abstract account of this application in the sense that $\langle x, x^* \rangle_{X^*}$ is a formal representation of the resulting real value. Also, we need constants to restore the linear structure on X^*.

Once these extensions are in place, we will be able to introduce the norm into the system by another additional constant which is specified to be the true dual norm on X^* induced by the norm on X by using the tame approach to suprema over bounded sets in abstract spaces outlined before.

Concretely, we thus add the following constants to the underlying language of the system $\mathcal{A}^\omega[X, \|\cdot\|]$ extended with the new base type X^*:

1. $+_{X^*}$ of type $X^*(X^*)(X^*)$,

2. $-_{X^*}$ of type $X^*(X^*)$,

3. $\cdot_{X^*}$ of type $X^*(X^*)(1)$,

4. 0_{X^*} of type X^*,

5. 1_{X^*} of type X^*,

6. $\langle \cdot, \cdot \rangle_{X^*}$ of type $1(X)(X^*)$.

For treating X^* as a normed vector space, we add another constant $\|\cdot\|_{X*}$ of type $1(X^*)$ for dealing with the dual norm. Indeed, the defining property of that norm being a certain supremum now has to be appropriately stated by suitable axioms which we obtain by instantiating the previous schemes $(S)_1$ and $(S)_2$. The first part of the supremum, i.e. that $\|x^*\|_{X*}$ is an upper bound on the function values of x^*, can be equivalently stated by the axiom

$$\forall x^{*X^*}, x^X \left(|\langle x, x^*\rangle_{X*}| \leqslant_{\mathbb{R}} \|x^*\|_{X*} \|x\|_X\right), \tag{$*$)$_1$}$$

essentially stating that a Cauchy-Schwarz type inequality holds. In that way, we avoid the otherwise necessary task of removing the premise $\|x\|_X \leqslant_{\mathbb{R}} 1$ suggested by the general scheme $(S)_1$ as mentioned before (e.g. via implicitly quantifying over $\overline{B}_1(0)$ through the use of $\widetilde{x}^1$). For the other part of the supremum, i.e. the statement that $\|x^*\|_{X*}$ is indeed the least such upper bound, we follow the general approach outlined in the previous section by instantiating $(S)_2$ and we thus opt for the axiom

$$\forall x^{*X^*}, k^0 \exists x \leqslant_X 1_X \left(\|x^*\|_{X*} - 2^{-k} \leqslant_{\mathbb{R}} |\langle x, x^*\rangle_{X*}|\right), \tag{$*$)$_2$}$$

expressing that $\langle x, x^*\rangle$ gets arbitrarily close to $\|x^*\|$ on the unit ball. This axiom $(*)_2$ is of the form Δ and thus a priori permissible when aiming for bound extraction theorems. We will later see that the usual norm axioms can be immediately derived from these two axioms. For now, just note that the intensional approach to X^* via an abstract type was crucially used here to provide quantification over elements from the dual in a quantifier-free way and thus to guarantee that the previous predicate D can be avoided so that the axioms resulting from instantiating the schemes $(S)_1, (S)_2$ have a monotone functional interpretation.

Remark 8.3.1. Similar to Remark 8.2.1, in the context $(*)_2$, it can be easily seen that $\|x^*\|_{X*}$ also (provably) satisfies the usual definition of being a supremum in the sense that it is the least upper bound of all values $|\langle x, x^*\rangle_{X*}|$, i.e.

$$\forall x^{*X^*}, M^1 \left(M <_{\mathbb{R}} \|x^*\|_{X*} \to \exists x \leqslant_X 1_X \left(M <_{\mathbb{R}} |\langle x, x^*\rangle_{X*}|\right)\right),$$

and, as also similar to the discussion in Remark 8.2.1, that $(*)_2$ actually even (provably) implies the following 'instantiated' version of that statement:

$$\forall x^{*X^*}, M^1, k^0 \exists x \leqslant_X 1_X \left(M + 2^{-k} <_{\mathbb{R}} \|x^*\|_{X*} \to \left(M + 2^{-(k+1)} \leqslant_{\mathbb{R}} |\langle x, x^*\rangle_{X*}|\right)\right).$$

It should be noted that this consequence of $(*)_2$ formalizes the defining property of $\|x^*\|_{X*}$ being a supremum in a way as it is often used in proofs from the literature (which we will see in the various formal proofs given later).

Using the norm, we can now provide an internal definition of equality on X^* via the abbreviation[4]

$$x^* =_{X^*} y^* := \|x^* -_{X^*} y^*\|_{X^*} =_{\mathbb{R}} 0$$

for x^{*X^*}, y^{*X^*}.

We now turn to the axioms for the application constant $\langle \cdot, \cdot \rangle_{X^*}$ which essentially just state that the map is bilinear:[5]

$$\begin{cases} \forall x^X, x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1 \left(\langle x, \alpha x^* +_{X^*} \beta y^* \rangle_{X^*} =_{\mathbb{R}} \alpha \langle x, x^* \rangle_{X^*} + \beta \langle x, y^* \rangle_{X^*} \right), \\ \forall x^X, x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1 \left(\langle x, \alpha x^* -_{X^*} \beta y^* \rangle_{X^*} =_{\mathbb{R}} \alpha \langle x, x^* \rangle_{X^*} - \beta \langle x, y^* \rangle_{X^*} \right), \end{cases} \qquad (*)_3$$

$$\begin{cases} \forall x^X, y^X, x^{*X^*}, \alpha^1, \beta^1 \left(\langle \alpha x +_X \beta y, x^* \rangle_{X^*} =_{\mathbb{R}} \alpha \langle x, x^* \rangle_{X^*} + \beta \langle y, x^* \rangle_{X^*} \right), \\ \forall x^X, y^X, x^{*X^*}, \alpha^1, \beta^1 \left(\langle \alpha x -_X \beta y, x^* \rangle_{X^*} =_{\mathbb{R}} \alpha \langle x, x^* \rangle_{X^*} - \beta \langle y, x^* \rangle_{X^*} \right), \end{cases} \qquad (*)_4$$

Lastly, we specify the vector space structure of X^* further, akin to [95]:[6]

$$\text{The vector space axioms for } +_{X^*},\ -_{X^*},\ \cdot_{X^*},\ 0_{X^*},\ 1_{X^*} \text{ w.r.t. } =_{X^*}. \qquad (*)_5$$

With this abstract approach, an issue of course arises regarding the connection between the bounded linear functionals represented in $1(X)$ and the elements of X^*. Concretely, it is clear just by examination of the quantifier complexity that an axiom stating that every element of $1(X)$ which is a continuous linear functional is indeed represented by some corresponding element of X^* will not be permissible meanwhile aiming for bound extraction theorems due to the complex premise of linearity and continuity (which is why we opted for an intensional treatment in the first place). In that way, we resort to the next best thing available in this situation: we include a rule guaranteeing that at least all terms of type $1(X)$ which provably belong to the dual of X are represented by an element of X^*. Concretely, we consider the following quantifier-free linearity rule[7]

$$\frac{F_0 \to \left(\forall x^X, y^X, \alpha^1, \beta^1 \left(t(\alpha x +_X \beta y) =_{\mathbb{R}} \alpha t x + \beta t y \right) \land \forall x^X \left(|tx| \leqslant_{\mathbb{R}} M \|x\|_X \right) \right)}{F_0 \to \exists x^* \leqslant_{X^*} M 1_{X^*} \forall x^X \left(tx =_{\mathbb{R}} \langle x, x^* \rangle_{X^*} \right)} \qquad \text{(QF-LR)}$$

[4] Similar as in the context of $\mathcal{A}^\omega[X, \|\cdot\|]$ with $-_X$, we write $x^* -_{X^*} y^*$ for $x^* +_{X^*} (-_{X^*} y^*)$.

[5] In the following, we omit the types from $\cdot_{X^*}$ or $\cdot_{X^*}$ altogether, similar as with $\cdot_X$.

[6] In particular, by including 1_{X^*} in the list of constants in the description of this collection of axioms, we want to indicate that these axioms include $\|1_{X^*}\|_{X^*} =_{\mathbb{R}} 1$.

[7] Similar to before, given objects x^*, y^* of type X^*, we here write $x^* \leqslant_{X^*} y^*$ for $\|x^*\|_{X^*} \leqslant_{\mathbb{R}} \|y^*\|_{X^*}$.

where F_0 is a quantifier-free formula and where t and M are terms of type $1(X)$ and 1, respectively.

But of course even in the context of this rule, the treatment of X^* can be regarded as an intensional one and the type X^* will also be interpretable by a suitable subspace of X^* (see also Remark 8.3.5 later on). What we want to argue with this approach outlined here is that full knowledge of X^* from the perspective of X seems seldom necessary for many applications and it often suffices if the subset specified by X^* is populated "enough" (with "enough" being relative to a certain application). For this, the above rule provides a minimal population of X^* which we now further extend by the following axiom which guarantees the existence of certain elements in X^* that will later be convenient to have so that we can develop the main aspects of the basic theory of X^* formally with ease. Concretely, this axiom codes a central consequence of the Hahn-Banach theorem for X^* by which it follows that $J(x) \neq \emptyset$ for any $x \in X$ where J is the normalized duality map of X, i.e. that for any $x \in X$:

$$\exists x^* \in X^* \left(\langle x, x^* \rangle_{X^*} = \|x\|^2 = \|x^*\|^2\right).$$

Instead of arguing that this statement is provable on the level of X using types $1(X)$ and then using the above rule (QF-LR) to transfer the existence of such functionals to the type X^*, we can just state this inclusion via an axiom of type Δ:

$$\forall x^X \exists x^* \leqslant_{X^*} \|x\|_X 1_{X^*} \left(\langle x, x^* \rangle_{X^*} =_{\mathbb{R}} \|x\|_X^2 =_{\mathbb{R}} \|x^*\|_{X^*}^2\right). \qquad (*)_6$$

Definition 8.3.2. We define the system $\mathcal{A}^\omega[X, \|\cdot\|_X, X^*, \|\cdot\|_{X^*}]$ for the abstract dual space of an abstract normed space as the extension of $\mathcal{A}^\omega[X, \|\cdot\|]$, formulated over the extended language using the types T^{X,X^*}, by the constants $+_{X^*}$, $-_{X^*}$, $\cdot_{X^*}$, 0_{X^*}, 1_{X^*}, $\langle \cdot, \cdot \rangle_{X^*}$, $\|\cdot\|_{X^*}$, the axioms $(*)_1$ - $(*)_6$ and the rule (QF-LR).

Remark 8.3.3. In the spirit of the above discussion preceding the rule (QF-LR), we want to mention that the use of a new abstract type for treating X^* intensionally can be avoided while achieving a system of similar strength. Concretely, we could alternatively have introduced a characteristic function χ_{X^*} of type $0(1(X))$ into the language of $\mathcal{A}^\omega[X, \|\cdot\|]$ together with a constant for the norm on X^*, now formulated using the type $1(X)$ instead of X^*. The respective axioms for the norm then could have been formulated with a quantification over X^* facilitated by the additional premise $\chi_{X^*}(x^*) =_0 0$ for elements x^* of type $1(X)$ (i.e. by similarly instantiating the schemes $(S)_1, (S)_2$ but where one now uses χ_{X^*} to instantiate D). In particular, in this context, the arithmetical operations on X^* would be definable by λ-abstraction together with the arithmetical operations on X and $\mathbb{R}$ and application of elements from X^* to elements from X would not require a new functional but would just be represented by a proper application of terms. This would be a kind of intensional treatment in the spirit of the previous approaches to set-valued operators from Chapter 3. However, the above approach via a new abstract type together with an application functional seemed to us more adherent to the abstract character that the dual space seems to have in many application scenarios (which is in particular further substantiated through the perspective of the notion of dual systems from the theory of topological vector spaces as will be discussed later in Remark 8.3.5) and also seemed to confine a bit better to the general abstract nature of the whole approach to normed spaces using abstract types in proof mining.

In the following, for simplicity, we abbreviate $\mathcal{A}^\omega[X, \|\cdot\|_X, X^*, \|\cdot\|_{X^*}]$ by $\mathcal{D}^\omega$. It can be immediately shown that, in this system, the bilinear application form $\langle\cdot,\cdot\rangle_{X^*}$ is non-degenerate (in the sense of dual systems, see the later Remark 8.3.5) and extensional:

Lemma 8.3.4. *The system $\mathcal{D}^\omega$ proves:*

1. *The bilinear form $\langle\cdot,\cdot\rangle_{X^*}$ is extensional, i.e.*

$$\forall x^X, y^X, x^{*X^*}, y^{*X^*} \left(x =_X y \wedge x^* =_{X^*} y^* \to \langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \langle y, y^*\rangle_{X^*}\right).$$

2. *The bilinear form $\langle\cdot,\cdot\rangle_{X^*}$ is non-degenerate, i.e.*
   (a) $\forall x^X \left(\forall x^{*X^*} \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} 0\right) \to x =_X 0_X\right)$,
   (b) $\forall x^{*X^*} \left(\forall x^X \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} 0\right) \to x^* =_{X^*} 0_{X^*}\right)$.

Proof. We begin with item (1): Let x, y and x^*, y^* be given and suppose that $x = y$ as well as $x^* = y^*$. Then note that $1v = v$ is a vector space axiom (and corresponding instantiations for x, y, x^*, y^* thus follow from the axioms of $\mathcal{A}^\omega[X, \|\cdot\|]$ and axiom $(*)_5$) and thus we have

$$\begin{aligned}
|\langle x, x^*\rangle - \langle y, y^*\rangle| &\leqslant |\langle x, x^*\rangle - \langle y, x^*\rangle| + |\langle y, x^*\rangle - \langle y, y^*\rangle| \\
&= |1\langle x, x^*\rangle - 1\langle y, x^*\rangle| + |1\langle y, x^*\rangle - 1\langle y, y^*\rangle| \\
&= |\langle 1x - 1y, x^*\rangle| + |\langle y, 1x^* - 1y^*\rangle| \\
&= |\langle x - y, x^*\rangle| + |\langle y, x^* - y^*\rangle| \\
&\leqslant \|x - y\| \|x^*\| + \|y\| \|x^* - y^*\| = 0
\end{aligned}$$

where the the third line follows from axioms $(*)_{3,4}$, the fourth line follows from multiple applications of the quantifier-free extensionality rule together with the previously mentioned vector space axiom and the last line follows from axiom $(*)_1$ and the assumptions that $x = y$ and $x^* = y^*$.

For item (2), we begin with (a). For this, we actually show

$$\forall x^X, k^0 \exists x^* \leqslant_{X^*} \|x\|_X 1_{X^*} \left(|\langle x, x^*\rangle_{X^*}| \leqslant_{\mathbb{R}} (2^{-k})^2 \to \|x\|_X \leqslant_{\mathbb{R}} 2^{-k}\right).$$

Let x be given and pick x^* via axiom $(*)_6$ such that $\|x^*\| = \|x\|$ as well as $\langle x, x^*\rangle = \|x\|^2$. Thus in particular if $|\langle x, x^*\rangle| \leqslant (2^{-k})^2$, then $\|x\| \leqslant 2^{-k}$.

For (b), we actually show

$$\forall x^{*X^*}, k^0 \exists x \leqslant_X 1_X \left(|\langle x, x^*\rangle_{X^*}| \leqslant_{\mathbb{R}} 2 \cdot 2^{-(k+2)} \to \|x^*\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k}\right).$$

Thus, let x^* be given and suppose $\|x^*\| > 2^{-k} = 2^{-(k+1)} + 2^{-(k+1)}$. By axiom $(*)_2$ (recall Remark 8.3.1), we get that there exists an x with $\|x\| \leqslant 1$ and such that $|\langle x, x^*\rangle| \geqslant 2^{-(k+1)} + 2^{-(k+2)}$, i.e. $|\langle x, x^*\rangle| > 2 \cdot 2^{-(k+2)}$. □

Remark 8.3.5. The above treatment of X^* ties to the notion of dual systems from the context of topological vector spaces (see e.g. [187]). Concretely, a dual system is a triple (X, Y, f) consisting of real vector spaces X, Y together with a bilinear form $f : X \times Y \to \mathbb{R}$. The dual system is called non-degenerate if

1. $f(x, y) = 0$ for all $y \in Y$ implies $x = 0$,
2. $f(x, y) = 0$ for all $x \in X$ implies $y = 0$.

In that way, the idea of the above approach using axioms $(*)_1$ - $(*)_5$ is to essentially axiomatize that X and X^* with $\langle \cdot, \cdot\rangle_{X^*}$ form a dual system. In particular, also the idea of an additional application functional is influenced by that perspective.

The linearity rule (QF-LR) and the axiom $(*)_6$ then guarantee that this subspace of the dual coded by X^* is at least in a certain way "close enough" to the full dual space and together with potential additional axioms they can serve to make sure that the subspace is rich enough for the application at hand. In particular, $(*)_6$ yields that the dual system thus axiomatized is non-degenerate which is exactly what was shown in the above lemma.

It still remains to be seen that the function specified by $\|\cdot\|_{X^*}$ is indeed a norm on X^*. For that, we show in the following lemma that the axioms for norms commonly in place for systems used in proof mining (as e.g. in the case of $\mathcal{A}^\omega[X, \|\cdot\|]$) are provable for the constant $\|\cdot\|_{X^*}$ in $\mathcal{D}^\omega$. In contrast to the usual norm axioms, these norm axioms are chosen such that it immediately follows that the arithmetical operations and the norm are extensional. In that way, we also find here that all the new constants that we added for the dual space are provably extensional in our system and that the system proves the same facts about the normed linear structure of X^* that it also proves of X.

Lemma 8.3.6. *The system $\mathcal{D}^\omega$ proves the norm axioms exhibited in [95], now formulated for $\|\cdot\|_{X^*}$:*

1. $\forall x^{*X^*} (\|x^* -_{X^*} x^*\|_{X^*} =_{\mathbb{R}} 0)$,
2. $\forall x^{*X^*}, y^{*X^*} (\|x^* -_{X^*} y^*\|_{X^*} =_{\mathbb{R}} \|y^* -_{X^*} x^*\|_{X^*})$,
3. $\forall x^{*X^*}, y^{*X^*}, z^{*X^*} (\|x^* -_{X^*} y^*\|_{X^*} \leqslant_{\mathbb{R}} \|x^* -_{X^*} z^*\|_{X^*} +_{\mathbb{R}} \|z^* -_{X^*} y^*\|)$,
4. $\forall x^{*X^*}, y^{*X^*}, \alpha^1 (\|\alpha x^* -_{X^*} \alpha y^*\|_{X^*} =_{\mathbb{R}} |\alpha| \, \|x^* -_{X^*} y^*\|_{X^*})$,
5. $\forall x^{*X^*}, \alpha^1, \beta^1 (\|\alpha x^* -_{X^*} \beta x^*\|_{X^*} = |\alpha - \beta| \, \|x^*\|_{X^*})$,
6. $\begin{cases} \forall x^{*X^*}, y^{*X^*}, u^{*X^*}, u^{*X^*} (\|(x^* +_{X^*} y^*) -_{X^*} (u^* +_{X^*} v^*)\|_{X^*} \\ \qquad \leqslant_{\mathbb{R}} \|x^* -_{X^*} u^*\|_{X^*} +_{\mathbb{R}} \|y^* -_{X^*} v^*\|_{X^*}), \end{cases}$
7. $\forall x^{*X^*}, y^{*X^*} (\|(-_{X^*} x^*) -_{X^*} (-_{X^*} y^*)\|_{X^*} =_{\mathbb{R}} \|x^* -_{X^*} y^*\|_{X^*})$,
8. $\forall x^{*X^*}, y^{*X^*} (| \, \|x^*\|_{X^*} - \|y^*\|_{X^*} \, | \leqslant_{\mathbb{R}} \|x^* -_{X^*} y^*\|_{X^*})$.

Proof. We only show items (1), (3), (4), (6) as well as (8) to exhibit the general pattern of proof used here. The other items can be done similarly. For items (4), (6) and (8), we will omit mentioning the use of axiom $(*)_5$ and freely manipulate algebraic expressions in X^*.[8] Also, in the context of the use of axiom $(*)_2$, recall Remark 8.3.1 for the particular consequence of $(*)_2$ that formalizes the usual least upper bound property of the supremum for $\|\cdot\|_{X^*}$.

[8]For this, some care of course needs to be exerted in order to guarantee that we do not require extensionality of these operations in the first place. By making the following arguments more precise, this can actually be verified for the given proofs (using e.g. Lemma 8.3.4) but we are here content with just sketching the arguments without this care. If one does not want to deal with this careful exercise, one could also just add the above statements about the norm as additional universal axioms.

(1) Since $|\langle x, x^*\rangle| \leqslant \|x^*\| \, \|x\|$, we have $\|x^*\| \geqslant 0$ for any x^* (by instantiating x with 1_X). Suppose now that $\|x^* - x^*\| > 0$. By the axiom $(*)_2$, we get an x such that $0 < |\langle x, x^* - x^*\rangle|$. Now, using $(*)_5$, we get $1x^* = x^*$ and so the quantifier-free extensionality rule yields $0 < |\langle x, 1x^* - 1x^*\rangle|$. By axiom $(*)_3$, we have $0 < |1\langle x, x^*\rangle - 1\langle x, x^*\rangle| = 0$ which is a contradiction. This gives $\|x^* - x^*\| = 0$.

(3) Suppose that $\|x^* - y^*\| > \|x^* - z^*\| + \|z^* - y^*\|$. Then by axiom $(*)_2$, we get an x with $\|x\| \leqslant 1$ and

$$|\langle x, x^* - y^*\rangle| > \|x^* - z^*\| + \|z^* - y^*\| \, .$$

Now, instantiating the vector space axioms $(*)_5$, we get $z^* + (-z^*) = 0$ and $x^* + 0 = x^*$ so that by two applications of the quantifier-free rule of extensionality, we have

$$|\langle x, x^* - y^*\rangle| = |\langle x, (x^* + (z^* + (-z^*))) + (-y^*)\rangle|.$$

By instantiating the associativity and commutativity axioms for $+$ from $(*)_5$, we get through multiple applications of the quantifier-free extensionality rule that

$$|\langle x, x^* - y^*\rangle| = |\langle x, (x^* - z^*) + (z^* - y^*)\rangle|.$$

At last, we get

$$\begin{aligned} \|x^* - z^*\| + \|z^* - y^*\| &< |\langle x, x^* - y^*\rangle| \\ &= |\langle x, 1(x^* - z^*) + 1(z^* - y^*)\rangle| \\ &= |\langle x, x^* - z^*\rangle + \langle x, z^* - y^*\rangle| \\ &\leqslant \|x\| \, \|x^* - z^*\| + \|x\| \, \|z^* - y^*\| \\ &\leqslant \|x^* - z^*\| + \|z^* - y^*\| \end{aligned}$$

where the second line follows from the previous by further instantiating the vector space axiom $1v = v$ from $(*)_5$ and using the quantifier-free extensionality rule, the third line follows from axiom $(*)_3$ and real arithmetic, the fourth line follows from real arithmetic and axiom $(*)_1$ and the last line follows as $\|x\| \leqslant 1$. Clearly, the above is a contradiction and so

$$\|x^* - y^*\| \leqslant \|x^* - z^*\| + \|z^* - y^*\|$$

holds after all.

(4) Suppose first that $\|\alpha x^* - \alpha y^*\| > |\alpha| \, \|x^* - y^*\|$. Then by axiom $(*)_1$, $(*)_2$ and $(*)_3$, we get an x with $\|x\| \leqslant 1$ such that

$$\begin{aligned} |\alpha| \, \|x^* - y^*\| &< |\langle x, \alpha x^* - \alpha y^* \rangle| \\ &= |\alpha| \cdot |\langle x, x^* - y^* \rangle| \\ &\leqslant |\alpha| \, \|x\| \, \|x^* - y^*\| \\ &\leqslant |\alpha| \, \|x^* - y^*\| \end{aligned}$$

which is a contradiction. On the other hand, if $\|\alpha x^* - \alpha y^*\| < |\alpha| \, \|x^* - y^*\|$, then $|\alpha| > 0$ since otherwise $0 \leqslant \|\alpha x^* - \alpha y^*\| < 0$. Thus in particular we have

$$\frac{\|\alpha x^* - \alpha y^*\|}{|\alpha|} < \|x^* - y^*\| \, .$$

Again, by axioms $(*)_1$, $(*)_2$ and $(*)_3$, we get an x with $\|x\| \leqslant 1$ such that

$$\begin{aligned} \frac{\|\alpha x^* - \alpha y^*\|}{|\alpha|} &< |\langle x, x^* - y^* \rangle| \\ &= \frac{1}{|\alpha|} |\langle x, \alpha x^* - \alpha y^* \rangle| \\ &\leqslant \frac{1}{|\alpha|} \, \|\alpha x^* - \alpha y^*\| \end{aligned}$$

which is a contradiction.

(6) Assume $\|(x^* + y^*) - (u^* + v^*)\| > \|x^* - u^*\| + \|y^* - v^*\|$. Then by axioms $(*)_1$, $(*)_2$ and $(*)_3$ there exists an x with $\|x\| \leqslant 1$ such that

$$\begin{aligned} \|x^* - u^*\| + \|y^* - v^*\| &< |\langle x, (x^* + y^*) - (u^* + v^*) \rangle| \\ &\leqslant |\langle x, x^* - u^* \rangle| + |\langle x, y^* - v^* \rangle| \\ &\leqslant \|x\| \, \|x^* - u^*\| + \|x\| \, \|y^* - v^*\| \\ &\leqslant \|x^* - u^*\| + \|y^* - v^*\| \end{aligned}$$

which is a contradiction.

(8) We show

$$\|x^*\| \leqslant \|x^* - y^*\| + \|y^*\| \text{ and } \|y^*\| \leqslant \|x^* - y^*\| + \|x^*\| \, .$$

For the former, suppose $\|x^*\| > \|x^* - y^*\| + \|y^*\|$. By axiom $(*)_1$, $(*)_2$ and $(*)_3$, we get that there exists an x with $\|x\| \leqslant 1$ and

$$\begin{aligned} \|x^* - y^*\| + \|y^*\| &< \langle x, x^* \rangle \\ &= \langle x, x^* - y^* \rangle + \langle x, y^* \rangle \\ &\leqslant \|x^* - y^*\| + \|y^*\| \end{aligned}$$

which is a contradiction.

For the latter, similarly suppose $\|y^*\| > \|x^* - y^*\| + \|x^*\|$ where we again get an x with $\|x\| \leqslant 1$ such that

$$\begin{aligned}\|x^* - y^*\| + \|x^*\| &< \langle x, y^* \rangle \\ &= -\langle x, -y^* \rangle \\ &= -\langle x, x^* - y^* \rangle + \langle x, x^* \rangle \\ &\leqslant \|x^* - y^*\| + \|x^*\|\end{aligned}$$

which is again a contradiction.

□

Remark 8.3.7. A simple property of Banach spaces (see e.g. [185]) is that being a Banach space is inherited from a space Y to all spaces $B(X, Y)$ of continuous linear functionals mapping into Y from a normed space X. In that way, the dual $X^* = B(X, \mathbb{R})$ of a normed space X is always a Banach space as $\mathbb{R}$ is itself complete. The latter property of completeness of $\mathbb{R}$ is formally represented in WE-PA$^\omega$ in the following way (where we follow the discussion given in [96]): provably in WE-PA$^\omega$ (and already in weak fragments thereof), we have

$$\forall \Phi^{1(0)} \left(\forall n^0 \forall m, k \geqslant_0 n \left(|\Phi k - \Phi m| \leqslant_{\mathbb{R}} 2^{-n} \right) \to \exists f^1 \forall n^0 \left(|\Phi n - f| \leqslant_{\mathbb{R}} 2^{-n} \right) \right)$$

where, in fact, f can be given by $fk := \widehat{\Phi(k+3)}(k+3)$. In that way, also the Cauchy completeness of X^* can be represented: provably in $\mathcal{D}^\omega$, given a sequence $x^{*X^*(0)}$ with

$$\forall n^0 \forall m, k \geqslant_0 n \left(\|x^* k -_{X^*} x^* m\|_{X^*} \leqslant_{\mathbb{R}} 2^{-n} \right),$$

we have for any x^X that

$$|\langle x, x^* k \rangle_{X^*} - \langle x, x^* m \rangle_{X^*}| \leqslant_{\mathbb{R}} \|x^* k -_{X^*} x^* m\|_{X^*} \|x\|_X$$

and thus we immediately get[9]

$$\forall x^X \forall n^0 \forall m, k \geqslant_0 (n + [\|x\|_X](0) + 1) \left(|\langle x, x^* k \rangle_{X^*} - \langle x, x^* m \rangle_{X^*}| \leqslant_{\mathbb{R}} 2^{-n} \right).$$

By the above completeness of $\mathbb{R}$ we can define its limit by a term in x in the sense that provably

$$\forall x^X \forall n \geqslant_0 ([\|x\|_X](0) + 1) \left(|\langle x, x^* n \rangle_{X^*} - fx| \leqslant_{\mathbb{R}} 2^{-n} \right)$$

[9]Here, we write $[a](n)$ for the n-th number in the type 1 representation of the real number a as before.

for fx of type 1 defined by

$$fxk := (\langle x, x^*(k + 3 + [\|x\|_X](0) + 1)\rangle_{X^*})\,\hat{}\,(k + 3)$$

where we wrote $(\cdot)\hat{}$ for the $\widehat{\cdot}$-operation. So f is a functional of type $1(X)$ and by formalizing a standard textbook proof it is now provable that this functional is linear and that it indeed has a bounded norm (in the sense that there is a K with $|fx| \leqslant K\|x\|$). The fact that this is indeed the limit of the sequence (x^*_n) w.r.t. the norm of X^* also has a trivial proof but this proof cannot be formalized in the underlying system and the reason for this is the basic issue with this whole approach: while the limit of the sequence can be pinpointed by a closed term, this term is of type $1(X)$. We however have no immediate way of inferring that this limit is indeed represented in X^* in general. Only if (x^*_n) is provably Cauchy in the above sense (i.e. with the given rate), then f is provably and without any assumptions linear and bounded. Then the quantifier-free linearity rule (QF-LR) can be used to conclude the existence of an x^*_f of type X^* such that provably

$$\forall x^X \left(fx =_{\mathbb{R}} \langle x, x^*_f\rangle_{X^*}\right).$$

This x^*_f can then be shown to be the limit. But if the sequence is not provably Cauchy in the above sense, the use of this rule is not permitted. Note that this issue is also not avoided by using a characteristic function χ_{X^*} to single out X^* from all functionals of type $1(X)$ as discussed in Remark 8.3.3 since also here, only a corresponding rule could be formulated which states the closure of χ_{X^*} under functionals which are provably linear and bounded. However, if we would be working with χ_{X^*}, we could add an axiom stating that the above term is included for any such sequence x^* which would require implicit quantification over Cauchy sequences in X^* akin to the methods employed in the context of the limit functional C of Kohlenbach (see [96] and Chapter 4). But in that case, we can also achieve the same result in the context of the abstract type X^* by formulating C and its axiom over this language. We do not explore this here any further.

Remark 8.3.8. By formalizing a standard argument (see e.g. Chapter 2, §4, Theorem 1 in [54]), one can also show in $\mathcal{D}^\omega$ that the uniform smoothness of X, formulated using a so-called modulus of uniform smoothness τ of type 1 (see [111]), i.e.[10]

$$\forall x^X, y^X, k^0 \big(\|x\|_X >_{\mathbb{R}} 1 \wedge \|y\|_X <_{\mathbb{R}} 2^{-\tau(k)} \\ \rightarrow \left\|\widetilde{x}^1 +_X y\right\|_X + \left\|\widetilde{x}^1 -_X y\right\|_X \leqslant_{\mathbb{R}} 2 + 2^{-k}\|y\|_X \big),$$

[10] Here, $\widetilde{x}^1$ is defined as in Section 8.2.

is equivalent to the uniform convexity of X^*, formulated using a modulus of uniform convexity η of type 1 (recall Chapter 7), i.e.

$$\forall x^{*X^*}, y^{*X^*}, k^0 \Bigg(\|x^*\|_{X^*}, \|y^*\|_{X^*} <_{\mathbb{R}} 1 \wedge \left\| \frac{x^* +_{X^*} y^*}{2} \right\|_{X^*} >_{\mathbb{R}} 1 - 2^{-\eta(k)} \\ \to \|x^* -_{X^*} y^*\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k} \Bigg).$$

We do not spell this out here any further.

8.4 Reflexivity of Banach spaces

8.4.1 The evaluation map and reflexivity

In the following, we write X^{**} for the bidual of X. We begin with the central notion of reflexivity.

Definition 8.4.1. Define the evaluation map $\phi : X \to X^{**}$ by

$$\phi(x)(x^*) = \langle x, x^* \rangle$$

for $x^* \in X^*$ and $x \in X$. The space X is called reflexive if ϕ is surjective.

Basic properties of the evaluation map needed in formal discussions later are the following: At first, using the Hahn-Banach theorem, it is immediate that the mapping ϕ is injective and preserves norms, i.e.

$$\|\phi(x)\| = \|x\| \text{ for all } x \in X.$$

In that way, ϕ maps X isometrically into X^{**} and X is reflexive if, equivalently, ϕ is an isometric isomorphism between X and X^{**}. Further, the following result is central for reflexive spaces:

Proposition 8.4.2 (James' theorem [81])**.** *A Banach space X is reflexive if, and only if, for any $x^* \in X^*$ with $\|x^*\| = 1$, there is an $x \in X$ with $\|x\| = 1$ and $\langle x, x^* \rangle = 1$.*

8.4.2 Treating reflexivity

To treat reflexivity in its version given by Definition 8.4.1, we will need access to the bidual X^{**}. Similarly to our abstract approach to X^*, we do not define this space from the objects from X^* but treat it in an abstract way as we did with X^*. Concretely, we first extend the underlying language by a third abstract type X^{**}, moving to a further extended set of types $T^{X,X^*,X^{**}}$ and to the resulting extended language similar to before. We then utilize this type to further introduce, as before, constants for the linear and normed structure on X^{**} as well as for the application of elements from X^{**} to elements from X^*, i.e.[11]

1. $+_{X^{**}}$ of type $X^{**}(X^{**})(X^{**})$,
2. $-_{X^{**}}$ of type $X^{**}(X^{**})$,
3. $\cdot_{X^{**}}$ of type $X^{**}(X^{**})(1)$,
4. $0_{X^{**}}$ of type X^{**},
5. $1_{X^{**}}$ of type X^{**},
6. $\langle\cdot,\cdot\rangle_{X^{**}}$ of type $1(X^*)(X^{**})$,
7. $\|\cdot\|_{X^{**}}$ of type $1(X^{**})$.

These constants are then used to formulate the previous axioms $(*)_1$ - $(*)_6$ and the rule (QF-LR) for the bidual:[12]

$$\forall x^{**X^{**}}, x^{*X^*} \left(|\langle x^*, x^{**}\rangle_{X^{**}}| \leqslant_{\mathbb{R}} \|x^{**}\|_{X^{**}} \|x^*\|_{X^*}\right), \tag{$(**)_1$}$$

$$\forall x^{**X^{**}}, k^0 \exists x^* \leqslant_{X^*} 1_{X^*} \left(\|x^{**}\|_{X^{**}} - 2^{-k} \leqslant_{\mathbb{R}} |\langle x^*, x^{**}\rangle_{X^{**}}|\right), \tag{$(**)_2$}$$

$$\begin{cases} \forall x^{*X^*}, x^{**X^{**}}, y^{**X^{**}}, \alpha^1, \beta^1 \\ \qquad \left(\langle x^*, \alpha x^{**} +_{X^{**}} \beta y^{**}\rangle_{X^{**}} =_{\mathbb{R}} \alpha\langle x^*, x^{**}\rangle_{X^{**}} + \beta\langle x^*, y^{**}\rangle_{X^{**}}\right), \\ \forall x^{*X^*}, x^{**X^{**}}, y^{**X^{**}}, \alpha^1, \beta^1 \\ \qquad \left(\langle x^*, \alpha x^{**} -_{X^{**}} \beta y^{**}\rangle_{X^{**}} =_{\mathbb{R}} \alpha\langle x^*, x^{**}\rangle_{X^{**}} - \beta\langle x^*, y^{**}\rangle_{X^{**}}\right), \end{cases} \tag{$(**)_3$}$$

$$\begin{cases} \forall x^{*X^*}, y^{*X^*}, x^{**X^{**}}, \alpha^1, \beta^1 \\ \qquad \left(\langle \alpha x^* +_{X^*} \beta y^*, x^{**}\rangle_{X^{**}} =_{\mathbb{R}} \alpha\langle x^*, x^{**}\rangle_{X^{**}} + \beta\langle y^*, x^{**}\rangle_{X^{**}}\right), \\ \forall x^{*X^*}, y^{*X^*}, x^{**X^{**}}, \alpha^1, \beta^1 \\ \qquad \left(\langle \alpha x^* -_{X^*} \beta y^*, x^{**}\rangle_{X^{**}} =_{\mathbb{R}} \alpha\langle x^*, x^{**}\rangle_{X^{**}} - \beta\langle y^*, x^{**}\rangle_{X^{**}}\right), \end{cases} \tag{$(**)_4$}$$

$$\text{The vector space axioms for } +_{X^{**}}, -_{X^{**}}, \cdot_{X^{**}}, 0_{X^{**}}, 1_{X^{**}} \text{ w.r.t. } =_{X^{**}}. \tag{$(**)_5$}$$

$$\forall x^{*X^*} \exists x^{**} \leqslant_{X^{**}} \|x^*\|_{X^*} 1_{X^{**}} \left(\langle x^*, x^{**}\rangle_{X^{**}} =_{\mathbb{R}} \|x^*\|^2_{X^*} =_{\mathbb{R}} \|x^{**}\|^2_{X^{**}}\right). \tag{$(**)_6$}$$

[11] As before, in formulas, we often omit the types around the $\cdot_{X^{**}}$-operation or we omit the operation entirely.

[12] Similar to before, by including $1_{X^{**}}$ in the list of constants in the description of axiom $(**)_5$, we want to indicate that these axioms include $\|1_{X^{**}}\|_{X^{**}} =_{\mathbb{R}} 1$.

For the rule, we opt for the formulation[13]

$$\frac{\begin{cases} F_0 \to (\forall x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1\, (t(\alpha x^* +_{X^*} \beta y^*) =_{\mathbb{R}} \alpha t x^* + \beta t y^*) \\ \qquad \wedge \forall x^{*X^*}\, (|tx^*| \leqslant_{\mathbb{R}} M \, \|x^*\|_{X^*})) \end{cases}}{F_0 \to \exists x^{**} \leqslant_{X^{**}} M 1_{X^{**}} \forall x^{*X^*}\, (tx^* =_{\mathbb{R}} \langle x^*, x^{**}\rangle_{X^{**}})} \qquad (\text{QF-LR}^{**})$$

where F_0 is a quantifier-free formula as before and t is a term of type $1(X^*)$ and M a term of type 1. We write $\mathcal{D}^\omega[X^{**}, \|\cdot\|_{X^{**}}]$ for the system $\mathcal{D}^\omega$ extended by the above constants, axioms and the rule.

In that formalism, reflexivity of the space – defined by means of the surjectivity of the evaluation map – can be easily expressed:

$$\forall x^{**X^{**}} \exists x^X \forall x^{*X^*} \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \langle x^*, x^{**}\rangle_{X^{**}}\right).$$

As discussed above, the map ϕ is an isometry and thus any such x naturally satisfies $\|x\| = \|x^{**}\|$. Therefore, the above statement is naturally equivalent to one of the form Δ which we henceforth adopt as our axiom for reflexivity:

$$\begin{aligned} &\forall x^{**X^{**}} \exists x \leqslant_X \|x^{**}\|_{X^{**}} 1_X \forall x^{*X^*} \\ &\qquad \left(\|x\|_X =_{\mathbb{R}} \|x^{**}\|_{X^{**}} \wedge \langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \langle x^*, x^{**}\rangle_{X^{**}}\right). \end{aligned} \qquad (\text{R})$$

As a simple example for the use of the axiom (R), we now consider the formal provability of one direction of James' theorem.

Lemma 8.4.3. *The system* $\mathcal{D}^\omega[X^{**}, \|\cdot\|_{X^{**}}] + (\text{R})$ *proves:*

$$\forall x^{*X^*} \exists x^X \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \|x\|_X^2 =_{\mathbb{R}} \|x^*\|_{X^*}^2\right).$$

In particular, $\mathcal{D}^\omega[X^{**}, \|\cdot\|_{X^{**}}] + (\text{R})$ *proves*

$$\forall x^{*X^*} \exists x^X \left(\|x^*\|_{X^*} =_{\mathbb{R}} 1 \to \|x\|_X =_{\mathbb{R}} 1 \wedge \langle x, x^*\rangle_{X^*} =_{\mathbb{R}} 1\right)$$

as in James' theorem.

Proof. Let x^* be given. By axiom $(**)_6$, we have that there exists an x^{**} with

$$\langle x^*, x^{**}\rangle = \|x^{**}\|^2 = \|x^*\|^2.$$

By axiom (R), we obtain that there exists an x with $\|x\| = \|x^{**}\| = \|x^*\|$ and $\langle x, x^*\rangle = \langle x^*, x^{**}\rangle = \|x^*\|^2$. □

[13] Also here, given objects x^{**}, y^{**} of type X^{**}, we write $x^{**} \leqslant_{X^{**}} y^{**}$ for $\|x^{**}\|_{X^{**}} \leqslant_{\mathbb{R}} \|y^{**}\|_{X^{**}}$ similarly to before.

Note that the above version of the characterization of reflexive spaces as in James' Theorem is easily formulated as an axiom of type Δ via

$$\forall x^{*X^*} \exists x^X \leqslant_X \|x^*\|_{X^*} 1_X \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \|x\|_X^2 =_{\mathbb{R}} \|x^*\|_{X^*}^2\right). \tag{JT}$$

If the bidual is not used in the context of reflexivity but one only needs to rely on the dual and the characterization via James' theorem, then the system $\mathcal{D}^\omega$ + (JT) can be used instead.[14]

However, there is a central issue surrounding this treatment of reflexivity. Namely, the axioms inherit a potential weakness through the intensionality used in the approach: the strength of the axioms (R) and (JT) is determined by the degree of how populated X^{**} and X^* are, respectively, i.e. how concretely they are specified. The more functionals the systems can provably determine to belong to these spaces, the stronger the axioms get. In that way, if a proof relies on the use of reflexivity on a specific complicated object x^{**} from X^{**}, then this complexity will be reflected by a potential analysis as, to formalize this use, one first has to provide formal means to hardwire this object into X^{**} via corresponding axioms which have a monotone functional interpretation.

Remark 8.4.4. By formalizing a standard argument (see e.g. Chapter 2, §4, Theorem 2 in [54]), one can show that $\mathcal{D}^\omega$ together with an axiom specifying that X is uniformly convex (using a corresponding modulus η) proves the above axiom (JT).

8.5 Extensions for uniformly Fréchet differentiable functions, their gradients and conjugates

We will now discuss the main extension of the above system for the dual of a normed space which provides a firm basis for the treatment of uniformly Fréchet differentiable convex functions, their gradients and in particular their Fenchel conjugates in Banach spaces. In that way, as we will further discuss later on, these extensions then allow for a formal treatment of Bregman distances associated with the respective convex function. This provides the first proper foray of proof mining into this part of convex analysis and also provides a first approach to deal with these rather concrete and complex objects. The bound extraction results established later for these extensions then also

[14]Note that this system is conservative over the base system by relativizing the quantifiers over elements of X^* accordingly.

form the basis for the extraction of quantitative results on the asymptotic regularity and convergence of iterations involving Bregman strongly nonexpansive operators given in Chapter 9. We refer to the references given in the introduction for further examples from the vast array of potential future applications of these systems.

8.5.1 Basic properties of Fréchet differentiable functions

We here shortly survey the (very minimal) essential definitions from the realm of convex analysis. Further definitions are given throughout the sections as needed. For any other details, we refer to the standard works [11, 182, 184, 212].

Let $f : X \to (-\infty, +\infty]$ be a given function with extended real values. In the following analytical section, we will assume that

1. f is proper, i.e.
$$\mathrm{dom}f := \{x \in X \mid f(x) < +\infty\} \neq \emptyset,$$

2. f is lower-semicontinuous, i.e.
$$\forall x \in \mathrm{dom}f \forall y < f(x) \exists \delta > 0 \forall z \in B_\delta(x) \left(f(z) > y\right),$$

3. f is convex, i.e.
$$\forall x, y \in \mathrm{dom}f \forall \lambda \in [0,1] \left(f\left(\lambda x + (1-\lambda)\,y\right) \leqslant \lambda f(x) + (1-\lambda)\,f(y)\right).$$

One of the central tools to study convex functions analytically are so-called generalized gradients. The central kind of these generalized gradients are the so-called subgradients as prominently already used in earliest works on modern convex analysis by Brøndsted and Rockafellar (see e.g. [26, 180]). For this, we write $\mathrm{intdom}f$ for the interior of $\mathrm{dom}f$.

Definition 8.5.1 (Subdifferential)**.** Let $x \in \mathrm{intdom}f$. We define

$$\partial f(x) := \{x^* \in X^* \mid f(x) + \langle y - x, x^* \rangle \leqslant f(y) \text{ for all } y \in X\}.$$

In this work, the focus will be on convex functions which are also Fréchet differentiable.

Definition 8.5.2 (Gâteaux and Fréchet differentiability)**.** A function f is called Gâteaux differentiable at x if there exists an element $\nabla f(x) \in X^*$ such that

$$\lim_{t \to 0} \frac{f(x + ty) - f(x)}{t} = \langle y, \nabla f(x) \rangle$$

for any y. It is called Gâteaux differentiable if it is Gâteaux differentiable at every $x \in \mathrm{intdom} f$. Further, f is called Fréchet differentiable if this limit is uniform in $\|y\| = 1$ and uniformly Fréchet differentiable if the limit is also uniform in x. We call ∇f the Gâteaux or Fréchet derivative, respectively.

The simplest example of a Fréchet derivative is obtained in uniformly smooth Banach spaces where for $f = \|\cdot\|^2/2$, we obtain $\nabla f = J$ for the normalized duality map J (see e.g. [212]). In particular, in Hilbert spaces, this reduces to the identity after identifying X^* with X.

The following properties connect the Fréchet derivative with the subgradients discussed before and will be essential for our treatment of the gradient for uniformly Fréchet differentiable functions. Their proofs can be found e.g. in [212] (or in [11] for the case of Hilbert spaces where the proofs are rather similar).

Proposition 8.5.3. *Let $x \in \mathrm{intdom} f$. Then, the following are equivalent:*

1. *f is Fréchet differentiable at x.*
2. *Every selection of ∂f is norm-to-norm continuous at x.*
3. *There exists a selection of ∂f that is norm-to-norm continuous at x.*

Further it holds that:

1. *If f is Gâteaux differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.*
2. *If f is continuous at x and $\partial f(x) = \{u\}$, then f is Gâteaux differentiable at x and $u = \nabla f(x)$.*

By the following result due to Reich and Sabach [176], being uniformly Fréchet differentiable (essentially) implies being Fréchet differentiable with a gradient that is uniformly norm-to-norm continuous on bounded sets.

Proposition 8.5.4 ([176])**.** *If f is uniformly Fréchet differentiable and ∇f is bounded on bounded sets (which in particular holds if X is reflexive and f is bounded on bounded sets), then ∇f is uniformly norm-to-norm continuous on bounded sets.*

The focus of the following sections will now be on providing logical systems for the treatment of convex functions f with uniformly continuous gradients as well as their conjugate functions and their corresponding gradients. By the above proposition, this therefore in particular treats uniformly Fréchet differentiable functions where ∇f is bounded on bounded sets.

8.5.2 A first formal treatment of gradients for uniformly Fréchet differentiable functions

To treat a convex function, we add a constant f of type $1(X)$ to the language. In the following discussions, we will for simplicity disregard the potential "partialness" of the function (induced by it taking values in the extended real line) and only treat total functions $f : X \to \mathbb{R}$ and their properties. Note the longer Remark 8.5.6 for a discussion on how the treatment presented below can be adapted to also handle the general setting.

The first immediate axiom for f is the following:

$(f)_1$ That f is convex, i.e.

$$\forall x^X, y^X, \lambda^1 \left(f\left(\widetilde{\lambda}x +_X \left(1 - \widetilde{\lambda}\right) y\right) \leqslant_{\mathbb{R}} \widetilde{\lambda} f(x) + \left(1 - \widetilde{\lambda}\right) f(y)\right).$$

Here, we have used the operation $\widetilde{\cdot}$ as e.g. defined in [96] for implicit quantification over $[0,1]$.

The lower-semicontinuity will not be added formally to the system as it will be derivable (in the form of uniform continuity on bounded subsets) from the axioms on the gradient.

Note that therefore, some caution is warranted for the use of the axiom $(f)_1$ as the use of $\widetilde{\lambda}$ for formulating convexity requires the extensionality of f to work properly. However, Lemma 8.5.5 establishes the uniform continuity of f as mentioned above and thus the extensionality of f and this lemma does not rely on $(f)_1$ so that no issues arise here.

Regarding the gradient, we add another constant ∇f of type $X^*(X)$ to the system. The relevant axioms for this constant will now stipulate that ∇f is a selection function for ∂f together with the fact that ∇f is uniformly continuous on bounded subsets.

Since the main emphasis will later be on systems which treat Legendre functions and since these functions naturally satisfy $\mathrm{dom}\nabla f = \mathrm{intdom}\, f$ and since we have before assumed that $\mathrm{dom}\, f = X$, we also consider ∇f to be totally defined.

We thus arrive at the following axioms:

$(\nabla f)_1$ That ∇f is a selection of ∂f, i.e.

$$\forall x^X, y^X \left(f(x) + \langle y -_X x, \nabla f(x)\rangle_{X^*} \leqslant_{\mathbb{R}} f(y)\right).$$

$(\nabla f)_2$ That ∇f is uniformly continuous on bounded subsets, i.e.

$$\forall x^X, y^X, b^0, k^0 \Big(\|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\omega^{\nabla f}(k,b)} \to \|\nabla f(x) -_{X^*} \nabla f(y)\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k}\Big).$$

Here, $\omega^{\nabla f}$ is another additional constant of type $0(0)(0)$.

We write $\mathcal{D}^\omega[f, \nabla f]$ for the theory resulting from $\mathcal{D}^\omega$ by extending it with the previous constants as well as the axioms $(f)_1$, $(\nabla f)_1$ and $(\nabla f)_2$. By the results contained in Proposition 8.5.3, any model of this system has to interpret the constant ∇f via the true gradient and what we want to argue is that this system is indeed sufficient to develop a large part of the theory of these gradients. As an initial litmus test, we in the following consider formalizations of various basic but central results on the function f and its gradient if the latter is uniformly continuous.

Lemma 8.5.5. *The theory $\mathcal{D}^\omega[f, \nabla f]$ proves:*

1. *f is uniformly Fréchet differentiable on bounded subsets, i.e.*

$$\forall b^0, k^0 \exists j^0 \forall x^X, y^X \left(\|x\|_X <_{\mathbb{R}} b \wedge 0 <_{\mathbb{R}} \|y\|_X <_{\mathbb{R}} 2^{-j} \to \frac{|f(x+y) - f(x) - \langle y, \nabla f(x)\rangle_{X^*}|}{\|y\|_X} \leqslant_{\mathbb{R}} 2^{-k}\right),$$

where in fact one can choose

$$j = \omega^{\nabla f}(k, b+1).$$

2. *∇f is bounded on bounded subsets, i.e.*

$$\forall b^0 \exists c^0 \forall x^X \left(\|x\|_X <_{\mathbb{R}} b \to \|\nabla f(x)\|_{X^*} \leqslant_{\mathbb{R}} c\right),$$

where in fact one can choose

$$c = C(b) = b2^{\omega^{\nabla f}(0,b)} + [\|\nabla f(0)\|_{X^*}](0) + 2.$$

3. f is uniformly continuous on bounded subsets, i.e.

$$\forall k^0, b^0 \exists j^0 \forall x^X, y^X (\|x\|_X, \|y\|_X <_{\mathbb{R}} b \\ \wedge \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-j} \to |f(x) - f(y)| \leqslant_{\mathbb{R}} 2^{-k}),$$

where in fact one can choose

$$j = \omega^f(k, b) = k + C(b).$$

4. f is bounded on bounded sets, i.e.

$$\forall b^0 \exists d^0 \forall x^X (\|x\|_X <_{\mathbb{R}} b \to |f(x)| \leqslant_{\mathbb{R}} d),$$

where in fact one can choose

$$d = D(b) = b2^{\omega^f(0,b)} + [|f(0)|](0) + 2.$$

Proof. 1. Using $(\nabla f)_1$ and extensionality of $\langle \cdot, \cdot \rangle$, we get

$$f(x+y) - f(x) \geqslant \langle x + y - x, \nabla f(x) \rangle \\ = \langle y, \nabla f(x) \rangle.$$

Similarly we derive

$$f(x) - f(x+y) \geqslant \langle -y, \nabla f(x+y) \rangle.$$

Together, we get

$$0 \leqslant f(x+y) - f(x) - \langle y, \nabla f(x) \rangle \\ \leqslant \langle y, \nabla f(x+y) \rangle - \langle y, \nabla f(x) \rangle \\ \leqslant \|y\| \, \|\nabla f(x+y) - \nabla f(x)\|.$$

Therefore we get

$$\frac{|f(x+y) - f(x) - \langle y, \nabla f(x) \rangle|}{\|y\|} \leqslant \|\nabla f(x+y) - \nabla f(x)\|.$$

So, for $\|x\| < b$ and y with $\|y\| < 2^{-\omega^{\nabla f}(k,b+1)}$, we get $\|x + y\| < b + 1$ and as $\|x + y - x\| = \|y\| < 2^{-\omega^{\nabla f}(k,b+1)}$, this yields

$$\frac{|f(x+y) - f(x) - \langle y, \nabla f(x) \rangle|}{\|y\|} \leqslant 2^{-k}$$

by $(\nabla f)_2$.

2. We have

$$\forall x, y \left(\|x\|, \|y\| \leqslant b \wedge \|x - y\| \leqslant 2^{-\omega^{\nabla f}(0,b)} \to \|\nabla f(x) - \nabla f(y)\| \leqslant 1 \right).$$

One can then inductively construct $b2^{\omega^{\nabla f}(0,b)}$-many points $x_1, \ldots, x_{k-1}$ with $\|x_i\| < b$ and

$$\|x_1\|, \|x_1 - x_2\|, \ldots, \|x_{k-1} - x\| < 2^{-\omega^{\nabla f}(0,b)}.$$

This yields

$$\|\nabla f(0) - \nabla f(x_1)\|, \|\nabla f(x_1) - \nabla f(x_2)\|, \ldots, \|\nabla f(x_{k-1}) - \nabla f(x)\| \leqslant 2^{-0} = 1$$

so that, using the triangle inequality, we derive

$$\|\nabla f(x)\| \leqslant b2^{\omega^{\nabla f}(0,b)} + 1 + \|\nabla f(0)\|.$$

The claim now follows from the fact that $[\|\nabla f(0)\|_{X^*}](0) + 1 \geqslant \|\nabla f(0)\|$.

3. We have

$$\begin{aligned} f(x) - f(y) &\leqslant \langle x - y, \nabla f(x) \rangle \\ &\leqslant \|x - y\| \|\nabla f(x)\| \end{aligned}$$

and similarly, we get

$$f(y) - f(x) \leqslant \|x - y\| \|\nabla f(y)\|.$$

Using the fact that ∇f is bounded on bounded sets with $\|\nabla f(x)\| \leqslant C(b)$ for $\|x\| < b$, we then get that

$$|f(x) - f(y)| \leqslant 2^{-k}$$

for $\|x\|, \|y\| < b$ with

$$\|x - y\| \leqslant 2^{-(k + C(b))}.$$

4. Similar to item (2).

□

Remark 8.5.6. We can incorporate functions $f : X \to (-\infty, +\infty]$ into the above framework by using an intensional account of f's domain. Concretely, to deal with such an f, we may introduce a new constant χ_f of type $0(X)$ into the language and then formulate all statements regarding $f(x)$ by relativizing x to

$$x \in \mathrm{dom} f := \chi_f x =_0 0.$$

The problem with this approach is now that the gradient ∇f also requires a treatment for its domain $\mathrm{dom}\nabla f \subseteq \mathrm{intdom} f$ and it is further crucial that this inclusion can be recognized by the system. For this, we can further modify the above intensional approach to domains of partial functions on X by incorporating the information required by the "openness" of the domain into the characteristic function. Concretely, the domain of ∇f can be treated by considering a slightly augmented characteristic function represented by a constant $\chi_{\nabla f}$ of type $0(0)(X)$ together with the defining universal axiom[15]

$$\forall x^X, k^0 \left(\chi_{\nabla f} xk =_0 0 \to \forall y^k \left(\left(x -_X \widetilde{y}^{(2^{-k})}\right) \in \mathrm{dom} f\right)\right)$$

expressing that $\mathrm{dom}\nabla f \subseteq \mathrm{intdom} f$ indeed holds by encoding the radius witnessing that $x \in \mathrm{intdom} f$ with x in $\chi_{\nabla f}$. It is now an easy exercise to generalize the above formal forays into the theory of f and its gradient ∇f to this modification by also relativizing statements regarding $\nabla f(x)$ using

$$(x, k) \in \mathrm{dom}\nabla f := \chi_{\nabla f} xk =_0 0$$

and

$$x \in \mathrm{dom}\nabla f := \exists k^0((x, k) \in \mathrm{dom}\nabla f).$$

Note further that this approach is very flexible not only regarding applications but also regarding formalizations of further properties of these domains which may be required in certain contexts. For example, as mentioned before, in the context of Legendre functions, a characterizing condition for these domains is in fact that the full equality $\mathrm{dom}\nabla f = \mathrm{intdom} f$ holds. This property can be further expressed by an axiom of type Δ. For this, note that the naive formulation of the reverse inclusion $\mathrm{intdom} f \subseteq \mathrm{dom}\nabla f$ can be formally expressed as

$$\forall x^X \left(\exists k^0 \forall y^X \left(\left(x -_X \widetilde{y}^{(2^{-k})}\right) \in \mathrm{dom} f\right) \to \exists j^0 \left((x, j) \in \mathrm{dom}\nabla f\right)\right)$$

But now, if $x \in \mathrm{intdom} f$ with a radius 2^{-k} is already supposed to hold, we can just simplify the above expression by instantiating it with $j = k$ which, after prenexing accordingly, brings us to the following axiom

$$\forall x^X, k^0 \exists y^X \leqslant_X 2^{-k} 1_X \left(\left(x -_X \widetilde{y}^{(2^{-k})}\right) \in \mathrm{dom} f \to (x, k) \in \mathrm{dom}\nabla f\right)$$

which is of type Δ by the restriction $\|y\| \leqslant 2^{-k}$ which does not restrict the meaning of the original statement as we anyhow move to $\widetilde{y}^{(2^{-k})}$.

[15]Here, we use the $\widetilde{\cdot}^{r}$ operation on elements of type X as defined in Section 8.2.

8.5.3 The Fenchel conjugate and its formal treatment

In the following, we will work over a reflexive space X. A main object in nonlinear analysis, in particular lying at the heart of the main approach to duality theory in Banach spaces, is the Fenchel conjugate f^* of a convex function f (as introduced in [61], see also [27, 181]): concretely, $f^* : X^* \to (-\infty, +\infty]$ is defined by

$$f^*(x^*) = \sup_{x \in X} \left(\langle x, x^* \rangle - f(x) \right).$$

The first immediate result from the definition is the following Young-Fenchel inequality: for any $x \in X$ and any $x^* \in X^*$, it holds that

$$f(x) + f^*(x^*) \geqslant \langle x, x^* \rangle.$$

If f^* is to be treated in any formal way in the underlying systems, we will have to require that f^* is majorizable which amounts to it being bounded on bounded sets. This requirement is linked with coercivity conditions on f by the following result:

Proposition 8.5.7 ([8]). *Call f supercoercive (or strongly coercive) if*

$$\lim_{\|x\| \to +\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Then, the following are equivalent:

1. *f is supercoercive.*

2. *f^* is bounded on bounded subsets.*

In particular, both imply that $\mathrm{dom} f^* = X^*$.

In that way, any metatheorem treating f^* via a constant (say of type $1(X^*)$) is in essence restricted to requiring that f is supercoercive. In that situation, however, the treatment of the supremum defining f^* is possible, following the tame approach to suprema outlined in the preceding sections. This in particular follows from the fact that if f is supercoercive, then the set on which the supremum is approached is bounded without loss of generality. This is formalized in the following lemma.

Lemma 8.5.8. *Let $\alpha : \mathbb{N} \to \mathbb{N}$ be a modulus of supercoercivity, i.e.*

$$\forall K \in \mathbb{N}, x \in X \left(\|x\| > \alpha(K) \to f(x)/\|x\| \geqslant K \right)$$

and let $F^*: \mathbb{N} \to \mathbb{N}$ be a function witnessing that f^* is bounded below on bounded sets, i.e.

$$\forall b \in \mathbb{N}, x^* \in X^* \left(\|x^*\| \leqslant b \to f^*(x^*) \geqslant -F^*(b)\right).$$

Then for $x^* \in X^*$ with $\|x^*\| \leqslant b$, we have

$$f^*(x^*) = \sup_{x \in \overline{B}_{r(\alpha,F^*,b)}(0)} \left(\langle x, x^*\rangle - f(x)\right)$$

where

$$r(\alpha, F^*, b) = \max\{\alpha(b+1)+1, F^*(b)+1\}.$$

Proof. Let $x \in X$ be given such that $\|x\| \geqslant \alpha(b+1)+1$. Then $f(x) \geqslant (b+1)\|x\|$. Naturally, we then have

$$\begin{aligned}\langle x, x^*\rangle - f(x) &\leqslant \|x\|\,\|x^*\| - (b+1)\|x\| \\ &= (\|x^*\| - (b+1))\|x\| \\ &\leqslant -\|x\|.\end{aligned}$$

Thus, if $\|x\| \geqslant F^*(b)+1$ also holds, then we have

$$\langle x, x^*\rangle - f(x) \leqslant -F^*(b) - 1 \leqslant f^*(x^*) - 1$$

and therefore, we get the claim. □

The lower bound F^* featured in the above result is naturally computed from f. Concretely, using the totality of f, we get

$$f^*(x^*) \geqslant \langle 0, x^*\rangle - f(0) \geqslant -|f(0)| \geqslant -([|f(0)|](0)+1).$$

So, in our concrete situation for a total f, we even have that

$$r(\alpha, b) = \max\{\alpha(b+1)+1, [|f(0)|](0)+2\}$$

suffices. Majorizing f^* can now also be trivially achieved by just noting that

$$|\langle x, x^*\rangle - f(x)| \leqslant \|x\|\,\|x^*\| + |f(x)|$$

and thus, knowing that there is an x with $\|x\| < r(\alpha, b)$ and such that $\langle x, x^*\rangle - f(x)$ approximates the supremum $f^*(x^*)$ with error 1, we get

$$f^*(x^*) \leqslant r(\alpha, b)\,\|x^*\| + r(\alpha, b)2^{\omega^f(0, r(\alpha,b))} + [|f(0)|](0) + 3$$

using Lemma 8.5.5 which immediately allows us to compute a majorant for f^*.

The axioms for f^* are now readily presented:

$(f)_2$ That f supercoercive with modulus α^f, i.e.

$$\forall K^0, x^X \left(\|x\|_X >_{\mathbb{R}} \alpha^f(K) \to f(x)/\|x\|_X \geqslant_{\mathbb{R}} K\right).$$

Here, α^f is an additional constant of type 1.

$(f^*)_1$ That f^* is a pointwise upper bound for all affine functionals $g_x(x^*) = \langle x, x^*\rangle - f(x)$, i.e.

$$\forall x^{*X^*}, x^X \left(\langle x, x^*\rangle_{X^*} - f(x) \leqslant_{\mathbb{R}} f^*(x^*)\right).$$

$(f^*)_2$ That f^* is indeed the pointwise supremum of all these affine functionals, i.e.

$$\begin{aligned}\forall x^{*X^*}, b^0, k^0 \exists x^X \leqslant_X \max\{\alpha^f(b+1)+1, [|f(0)|](0)+2\}1_X \\ \left(\|x^*\|_{X^*} <_{\mathbb{R}} b \to \left(f^*(x^*) - 2^{-k} \leqslant_{\mathbb{R}} \langle x, x^*\rangle_{X^*} - f(x)\right)\right).\end{aligned}$$

Note that also here, we have a natural benefit in approaching this supremum as we can avoid instantiating C in the schema $(S)_1$ since the corresponding claim that f^* is an upper bound actually holds in an unrestricted form.

Remark 8.5.9. Similar to Remark 8.2.1 (recall also Remark 8.3.1), in the context $(f^*)_2$, also $f^*(x^*)$ satisfies the usual definition of being a supremum in the sense that it is the least upper bound of all values $\langle x, x^*\rangle - f(x)$ and, also similar to before, $(f^*)_2$ even implies the following statement:

$$\begin{aligned}\forall x^{*X^*}, b^0, M^1, k^0 \exists x^X \leqslant_X \max\{\alpha^f(b+1)+1, [|f(0)|](0)+2\}1_X \\ \left(\|x^*\|_{X^*} <_{\mathbb{R}} b \wedge M + 2^{-k} <_{\mathbb{R}} f^*(x^*) \to \left(M + 2^{-(k+1)} \leqslant_{\mathbb{R}} \langle x, x^*\rangle_{X^*} - f(x)\right)\right).\end{aligned}$$

A first immediate property that can be derived for f^* is its convexity:

Lemma 8.5.10. *The system $\mathcal{D}^\omega$ extended with constants for f, α^f and f^* together with the axioms $(f^*)_1$ and $(f^*)_2$ proves that f^* is convex.*

Proof. Suppose that f^* is not convex, i.e. that there are x^*, y^* and $\alpha \in [0,1]$ such that

$$\alpha f^*(x^*) + (1-\alpha)f^*(y^*) < f^*(\alpha x^* + (1-\alpha)y^*)$$

Then by $(f^*)_2$ (recall Remark 8.5.9), we get a z such that

$$\begin{aligned}\alpha f^*(x^*) + (1-\alpha)f^*(y^*) &< \langle z, \alpha x^* + (1-\alpha)y^*\rangle - f(z) \\ &= \alpha(\langle z, x^*\rangle - f(z)) + (1-\alpha)(\langle z, y^*\rangle - f(z)) \\ &\leqslant \alpha f^*(x^*) + (1-\alpha)f^*(y^*).\end{aligned}$$

where the last line follows from $(f^*)_1$. This is a contradiction. ☐

Note that not even the convexity of f is necessary for this.

If f^* is uniformly Fréchet differentiable as well, its gradient can now be introduced as before: we add a constant ∇f^* of type $X(X^*)$ and consider the following axioms.

$(\nabla f^*)_1$ That ∇f^* is a selection of ∂f^*, i.e.

$$\forall x^{*X^*}, y^{*X^*} \left(f^*(x^*) + \langle \nabla f^*(x^*), y^* -_{X^*} x^* \rangle_{X^*} \leqslant_{\mathbb{R}} f^*(y^*) \right).$$

$(\nabla f^*)_2$ That ∇f^* is uniformly continuous on bounded subsets, i.e.

$$\forall x^{*X^*}, y^{*X^*}, b^0, k^0 \Bigg(\bigg(\|x^*\|_{X^*}, \|y^*\|_{X^*} <_{\mathbb{R}} b \\ \wedge \|x^* -_{X^*} y^*\|_{X^*} <_{\mathbb{R}} 2^{-\omega^{\nabla f^*}(k,b)} \bigg) \to \|\nabla f^*(x^*) -_X \nabla f^*(y^*)\|_X \leqslant_{\mathbb{R}} 2^{-k} \Bigg).$$

Here, $\omega^{\nabla f^*}$ is another additional constant of type $0(0)(0)$.

We want to note that the gradients of f and f^* are simultaneously well-defined only if f is Legendre in the sense of the following influential definition of Bauschke, Borwein and Combettes.

Definition 8.5.11 ([8])**.** A function f is called:

1. essentially smooth if ∂f is locally bounded and single-valued on its domain,

2. essentially strictly convex if $(\partial f)^{-1}$ is locally bounded and f is strictly convex on every convex subset of $\mathrm{dom}\partial f$,

3. Legendre if it is both essentially smooth and essentially strictly convex.

Over reflexive spaces, these properties can be recognized as equivalently stating a particularly nice differentiability property for both f and its conjugate f^*.

Proposition 8.5.12 ([8])**.** *If X is reflexive, then f is Legendre if, and only if*

1. *It holds that* $\mathrm{intdom} f \neq \emptyset$*, that f is Gâteaux differentiable on* $\mathrm{intdom} f$*, and* $\mathrm{dom}\nabla f = \mathrm{intdom} f$*.*

2. *It holds that* $\mathrm{intdom} f^* \neq \emptyset$*, that f^* is Gâteaux differentiable on* $\mathrm{intdom} f^*$*, and* $\mathrm{dom}\nabla f^* = \mathrm{intdom} f^*$*.*

Therefore, the above axioms can only be satisfied if f is already Legendre since any f and f^* satisfying them are even uniformly Fréchet differentiable.

Remark 8.5.13. While reflexivity features as a key assumption in the above proposition, if further differentiability assumptions are made regarding f and f^*, then reflexivity is an inherent property in that context. Concretely, by a result of Borwein and Vanderwerff [19], any space where f and f^* are Fréchet differentiable, f is continuous and $\mathrm{dom} f^* = X^*$ is already reflexive and it follows from results by Borwein, Guirao, Hájek and Vanderwerff [18] that if f and f^* are uniformly Fréchet differentiable and $\mathrm{dom} f^* = X^*$, then X is even superreflexive. In that way, in the context of the continuity assumptions formalized by the above axioms, we are always conceptually working over (super-)reflexive spaces and we used this reflexivity here already to formalize ∇f^* via an object of type $X(X^*)$, using X as the type for the images in order to formally avoid X^{**}.

Further, the following relation between the gradient of a function and of its conjugate holds for Legendre functions:

Proposition 8.5.14 ([8]). *If X is reflexive and f is Legendre, then ∇f is a bijection with* $\mathrm{ran}\nabla f = \mathrm{dom}\nabla f^*$, $\mathrm{ran}\nabla f^* = \mathrm{dom}\nabla f = \mathrm{intdom} f$ *and*

$$\nabla f = (\nabla f^*)^{-1}.$$

Instead of formalizing the corresponding proof to verify whether the previous axioms already suffice for proving this relation, we can just hardwire this property into the system by adding the following corresponding axiom:

(L) $\forall x^X, x^{*X^*} \left(\nabla f \nabla f^*(x^*) =_{X^*} x^* \wedge \nabla f^* \nabla f(x) =_X x\right)$.

We write $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ for the system $\mathcal{D}^\omega[f, \nabla f]$ extended with the above constants and axioms $(f)_2$, $(f^*)_1$, $(f^*)_2$ as well as $(\nabla f^*)_1$, $(\nabla f^*)_2$, (L).

Remark 8.5.15. Note that the previous Lemma 8.5.5, if suitably adapted, also holds for f^* and ∇f^* in this new theory $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$. We therefore do not replicate this here.

Remark 8.5.16. It is well-known in the literature on convex analysis that differentiability properties of the conjugate f^* are related to convexity properties of the original function f (see e.g. [34, 35, 36] among many others). In that way, any function f that induces a model of the theory $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ actually is even totally convex on bounded subsets as well as uniformly strictly convex. We refer to Chapters 9 and 10 for further (formal) investigations into the interrelations of these properties and their quantitative analogues as guided by the logical methodology introduced in this chapter.

8.5.4 Bregman distances and their formal treatment

As a small indication for the applicability of the above formal systems, we just want to note that the language is already expressive enough to deal with some of the central objects in the modern realm of convex analysis. The object that we want to focus on here is the central Bregman distance introduced in [22] which features in many algorithmic approaches in that field (see in particular again the references in the introduction as well as the references in [9]).

These Bregman distances are defined relative to a convex function in terms of its gradient:

Definition 8.5.17 ([22])**.** Let f be Gâteaux differentiable. The function $D_f : \mathrm{dom} f \times \mathrm{intdom} f \to [0, +\infty)$ is defined as follows:

$$D_f(x, y) := f(x) - f(y) - \langle x - y, \nabla f(y)\rangle.$$

As such, a benefit of the above treatment of f and ∇f is that in the context of the system $\mathcal{D}^\omega[f, \nabla f]$, this function can just be given by a closed term.

The same is true for the function $W_f : \mathrm{dom}\ f \times \mathrm{dom}\ f^* \to [0, +\infty)$ defined by

$$W_f(x, x^*) = f(x) - \langle x, x^*\rangle + f^*(x^*).$$

which often provides a medium through which D_f is studied (see e.g. [143, 144]). Also this function can be represented by a closed term in the underlying system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ and the basic properties of both are immediately provable. We just mention two of these here:

Lemma 8.5.18. *The system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ proves the three and four point identities (see e.g. [9]):*

1. $$\begin{cases} \forall x^X, y^X, z^X \big(D_f(x, y) + D_f(y, z) - D_f(x, z) \\ \qquad =_{\mathbb{R}} \langle x -_X y, \nabla f(z) -_{X^*} \nabla f(y)\rangle_{X^*}\big). \end{cases}$$

2. $$\begin{cases} \forall x^X, y^X, z^Z, w^X \big(D_f(y, x) - D_f(y, z) - D_f(w, x) + D_f(w, z) \\ \qquad =_{\mathbb{R}} \langle y -_X w, \nabla f(z) -_{X^*} \nabla f(x)\rangle_{X^*}\big). \end{cases}$$

Not only does the system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ provide a framework for adequately expressing the central objects and theorems in the theory of these Bregman distances but, as common in proof mining, the metatheorems for this system established in the upcoming section can be used to provide a finitary quantitative account on some of the central assumptions used in the context of applications of these Bregman distances like that of consistency of the Bregman distance, i.e.

$$\forall x^X, y^X \left(x =_X y \leftrightarrow D_f(x,y) =_{\mathbb{R}} 0\right),$$

as well as total convexity and sequential consistency (see e.g. [36]), among many others, where the metatheorems suggest appropriate moduli that witness the quantitative content of these statements. These moduli are then crucially used in applications as will also be the case in the forthcoming work [164] as well as in Chapter 9.

8.6 A bound extraction theorem

We now establish the bound extraction theorems for the system $\mathcal{D}^\omega$ and the extensions discussed previously. Our proof follows the approach of [71, 95, 96] as presented in Chapter 3 and in that way is rather standard. Consequently, we will omit some proofs (only giving those details that concern new material) and sometimes be brief about the presentation, occasionally only sketching the general outline of the arguments. For the following, recall the definition of Gödel's functional interpretation and the negative translation from Chapter 3. These naturally extend to the new languages from this chapter.

Also recall Lemma 3.7.3, formulated for $\mathcal{A}^\omega[X, \|\cdot\|]$ and extensions of that theory by universal sentences, for the soundness result for the combination of both the Dialectica interpretation and the negative translation which forms the basis for the upcoming metatheorems. Similar to that context, we write $\mathcal{D}^{\omega-}$ for the respective system *without* the axiom schemes QF-AC and DC.

Besides Gödel's functional interpretation, the other central notion used in the bound extraction results is that of (strong) majorizability and the associated structure $\mathcal{M}^{\omega,X}$. In this chapter, based on the use of a second abstract type X^* (and potentially a third with X^{**}), we have to further extend these notions to this second (and third) type (similar to the discussion in [96], Section 17.6). We here only focus on the case of a single additional type X^* and do not explicitly discuss the extension with X^{**} which can be treated analogously. In our context, the majorants for objects of types from

T^{X,X^*} will still be objects with a type from T according to the following extended projection:

Definition 8.6.1 (essentially [71]). Define $\widehat{\tau} \in T$, given $\tau \in T^{X,X^*}$, by recursion on the structure via

$$\widehat{0} := 0,\ \widehat{X} := 0,\ \widehat{X^*} := 0,\ \widehat{\tau(\xi)} := \widehat{\tau}(\widehat{\xi}).$$

The majorizability relation for the types T^{X,X^*} is then defined recursively along with the structure $\mathcal{M}^{\omega,X,X^*}$ of all majorizable functionals over a given normed space X with dual X^*:

Definition 8.6.2 (essentially [71, 95]). Let $(X, \|\cdot\|)$ be a non-empty normed space with dual X^*. The structure $\mathcal{M}^{\omega,X,X^*}$ and the majorizability relation $\gtrsim_\rho$ are defined by

$$\begin{cases} M_0 := \mathbb{N}, n \gtrsim_0 m := n \geqslant m \wedge n, m \in \mathbb{N}, \\ M_X := X, n \gtrsim_X x := n \geqslant \|x\| \wedge n \in M_0, x \in M_X, \\ M_{X^*} := X^*, n \gtrsim_{X^*} x^* := n \geqslant \|x^*\| \wedge n \in M_0, x^* \in M_{X^*}, \\ f \gtrsim_{\tau(\xi)} x := f \in M_{\widehat{\tau}}^{M_{\widehat{\xi}}} \wedge x \in M_\tau^{M_\xi} \\ \qquad\qquad \wedge \forall g \in M_{\widehat{\xi}}, y \in M_\xi (g \gtrsim_\xi y \to fg \gtrsim_\tau xy) \\ \qquad\qquad \wedge \forall g, y \in M_{\widehat{\xi}} (g \gtrsim_{\widehat{\xi}} y \to fg \gtrsim_{\widehat{\tau}} fy), \\ M_{\tau(\xi)} := \left\{ x \in M_\tau^{M_\xi} \mid \exists f \in M_{\widehat{\tau}}^{M_{\widehat{\xi}}} : f \gtrsim_{\tau(\xi)} x \right\}. \end{cases}$$

Correspondingly, the full set-theoretic type structure $\mathcal{S}^{\omega,X,X^*}$ is defined via $S_0 := \mathbb{N}$, $S_X := X$, $S_{X^*} := X^*$ and

$$S_{\tau(\xi)} := S_\tau^{S_\xi}.$$

These structures later turn into models of our systems if equipped with corresponding interpretations for the additional constants.

The general high-level outline of the proof of the bound extraction theorem is now as before: we use functional interpretation and negative translation to extract realizers from (essentially) $\forall\exists$-theorems which have types that belong to T^{X,X^*}. Using majorizability, we then construct bounds for these realizers which are moreover valid in a model based on $\mathcal{M}^{\omega,X,X^*}$. If the types occurring in the axioms and the theorem are "low enough", we can then in a final step recover to the truth in the usual full set-theoretic structure $\mathcal{S}^{\omega,X,X^*}$.

For the concrete implementation of "low enough", we need to extend the previous definitions of small and admissible types to the new base type. This can be done in complete analogy to before (see Chapter 3): We call ξ *small* if it is of the form $\xi = \xi_0(0)\dots(0)$ (including $0, X, X^*$) for $\xi_0 \in \{0, X, X^*\}$ and call it *admissible* if it is of the form $\xi = \xi_0(\tau_k)\dots(\tau_1)$ (including $0, X, X^*$) where each τ_i is small and $\xi_0 \in \{0, X, X^*\}$ as before.

Similarly, take the notions of $\forall$-/$\exists$-formulas to be now defined by also considering the new abstract type and the same also holds for the class Δ where the type restrictions are now to be understood in this extended sense. For this, we in particular also rely on the following extension of the relation $\leqslant$ which is now defined by recursion on the type via

1. $x \leqslant_0 y := x \leqslant_0 y$,
2. $x \leqslant_X y := \|x\|_X \leqslant_{\mathbb{R}} \|y\|_X$,
3. $x^* \leqslant_{X^*} y^* := \|x^*\|_{X^*} \leqslant_{\mathbb{R}} \|y^*\|_{X^*}$,
4. $x \leqslant_{\tau(\xi)} y := \forall z^\xi (xz \leqslant_\tau yz)$.

Given a set Δ of such formulas, we write $\widetilde{\Delta}$ for the set of all Skolem normal forms as before.

In the bound extraction theorems, axioms of type Δ are also treated as before "in spirit" of the monotone functional interpretation. Here however, we want to exert a bit more care as sentences of type Δ already occur in the axioms of $\mathcal{D}^\omega$ (and its extensions). Further, the treatment of the rule (QF-LR) relies crucially on the treatment of sentences of type Δ as well. Write $\widehat{\mathcal{D}}^\omega$ for $\mathcal{D}^\omega$ without any of its axioms of type Δ and without the rule (QF-LR). Then, given a set Δ of additional axioms of type Δ, we treat all axioms of type Δ present in $\mathcal{D}^\omega + \Delta$ together with (QF-LR) by forming a new theory $\overline{\mathcal{D}}^\omega_\Delta$ which arises from $\widehat{\mathcal{D}}^\omega$ by adding the Skolem functionals $\underline{B}$ for any axiom of type Δ, say of the form

$$\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c}),$$

as new constants to the language and adding its "instantiated Skolem normal form", i.e. the sentence

$$\underline{B} \leqslant_{\underline{\sigma}(\underline{\delta})} \underline{r} \wedge \forall \underline{a}^{\underline{\delta}} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{B}\underline{a}, \underline{c}),$$

as a new axiom. Further, we do the same with all conclusions of the rule (QF-LR): for any provable premise

$$\mathcal{D}^\omega + \Delta \vdash F_0 \to \left(\forall x^X, y^X, \alpha^1, \beta^1 \left(t(\alpha x +_X \beta y) =_{\mathbb{R}} \alpha tx + \beta ty\right) \wedge \forall x^X \left(|tx| \leqslant_{\mathbb{R}} M \left\|x\right\|_X\right)\right)$$

with terms t and M, we add a new constant x^*_t of type X^* to the language of $\overline{\mathcal{D}}^\omega_\Delta$ together with the corresponding axiom

$$\left\|x^*_t\right\|_{X^*} \leqslant_{\mathbb{R}} M \wedge \left(F_0 \to \forall x^X \left(tx =_{\mathbb{R}} \langle x, x^*_t\rangle_{X^*}\right)\right).$$

This new theory $\overline{\mathcal{D}}^\omega_\Delta$ extends $\mathcal{A}^\omega[X, \|\cdot\|]$ only by new types, constants and universal axioms and, consequently, Lemma 3.7.3 also applies to this theory $\overline{\mathcal{D}}^\omega_\Delta$ where the conclusion is proved in $\overline{\mathcal{D}}^{\omega-}_\Delta + (\mathrm{BR})$ where $\overline{\mathcal{D}}^{\omega-}_\Delta$ arises from $\overline{\mathcal{D}}^\omega_\Delta$ by removing the principles QF-AC and DC.

Similar constructions can also be made for the respective extensions of $\mathcal{D}^\omega$.

The central majorizability result is now the following, guaranteeing the majorizability of all closed terms in $\overline{\mathcal{D}}^\omega_\Delta$ (and its extensions). In that way, the result extends the central Lemma 9.11 in [71] and is analogous in spirit to Lemma 3.7.7 from Chapter 3.

Lemma 8.6.3. *Let Δ be a set of additional axioms of type Δ. Let $(X, \|\cdot\|)$ be a (nontrivial) Banach space with its dual X^*. Then $\mathcal{M}^{\omega,X,X^*}$ is a model of $\overline{\mathcal{D}}^{\omega-}_\Delta + (\mathrm{BR})$, provided $\mathcal{S}^{\omega,X,X^*} \models \Delta$ (with $\mathcal{M}^{\omega,X,X^*}$ and $\mathcal{S}^{\omega,X,X^*}$ defined via suitable interpretations of the additional constants). Moreover, for any closed term t of $\overline{\mathcal{D}}^{\omega-}_\Delta + (\mathrm{BR})$, one can construct a closed term t^* of $\mathcal{A}^\omega + (\mathrm{BR})$ such that*

$$\mathcal{M}^{\omega,X,X^*} \models (t^* \gtrsim t).$$

Further, the same claim holds for the following extensions of $\mathcal{D}^\omega$:

1. *The theory $\mathcal{D}^\omega[X^{**}, \|\cdot\|_{X^{**}}]$ over the language with the additional abstract type X^{**} or its extension with the reflexivity axiom where the model and the majorizability relation have to be extended to also incorporate this type (and the space has to be reflexive in the latter case). In any case, one then has to employ a similar construction as with* (QF-LR) *to also eliminate the rule* (QF-LR**) *and any other potential axioms of type Δ for these new systems.*

2. *Assume a convex and Fréchet differentiable function $f : X \to \mathbb{R}$ where ∇f is uniformly continuous on bounded subsets with modulus $\omega^{\nabla f}$. Then the result*

holds for $\mathcal{D}^\omega[f, \nabla f]$ *where, in that case, we will have the modified conclusion that there exists a term* t^* *such that*

$$\mathcal{M}^{\omega,X,X^*} \models \forall \omega^{0(0)(0)}, n^0 \left(\omega \gtrsim \omega^{\nabla f} \wedge n \geqslant_{\mathbb{R}} |f(0)|, \|\nabla f(0)\|_{X^*} \to t^*(\omega, n) \gtrsim t\right)$$

holds. If f *is additionally supercoercive with a modulus* α^f *and* f^* *is Fréchet differentiable with a gradient* ∇f^* *that is uniformly continuous on bounded subsets with a modulus* $\omega^{\nabla f^*}$, *the same claim also holds for* $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ *if we further require that* $\omega \gtrsim \omega^{\nabla f^*}, \alpha^f$ *and* $n \geqslant_{\mathbb{R}} |f^*(0)|, \|\nabla f^*(0)\|_X$. *In any case, one then has to employ a similar construction as before to also eliminate the rule* (QF-LR) *and any other potential axioms of type* Δ *for these new systems.*

Proof. As with the proof of Lemma 3.7.7, the structure of the proof is very much standard and follows that of the proof of Lemma 17.85 in [96]. In particular, many parts of that proof carry over and we in that vein only discuss the interpretations and verify the majorizability of the new constants contained in $\mathcal{D}^\omega$ and its extensions together with their validity in the resulting models. In particular, we at first do not explicitly deal with the additional constants induced by the axioms of type Δ in $\mathcal{D}^\omega + \Delta$ (and its extensions) through forming the theory $\overline{\mathcal{D}}^\omega_\Delta$ and only discuss these at the end of the proof.

We now first focus on $\mathcal{D}^\omega$ and assume that there are no further axioms of type Δ beyond those in $\mathcal{D}^\omega$. For that, we initially provide the corresponding interpretations of the constants of $\mathcal{D}^\omega$. For the constants already contained in $\mathcal{A}^\omega[X, \|\cdot\|]$, we may choose suitable interpretations as in [96] (which are anyhow analogous to the interpretation for the constants related to X^* chosen below). For the new constants added to $\mathcal{A}^\omega[X, \|\cdot\|]$ to form $\mathcal{D}^\omega$, we consider the following interpretations (writing $\mathcal{M}$ for $\mathcal{M}^{\omega,X,X^*}$):

1. $[+_{X^*}]_\mathcal{M} :=$ addition in X^*,
2. $[-_{X^*}]_\mathcal{M} :=$ inverse of $+$ in X^*,
3. $[\cdot_{X^*}]_\mathcal{M} := \lambda \alpha \in \mathbb{N}^\mathbb{N}, x^* \in X^*.(r_\alpha \cdot x^*)$ where $\cdot$ is the scalar multiplication in X^*,
4. $[0_{X^*}]_\mathcal{M} :=$ the zero vector in X^*,
5. $[1_{X^*}]_\mathcal{M} :=$ some canonically chosen unit vector $a^* \in X^*$,
6. $[\langle \cdot, \cdot \rangle_{X^*}]_\mathcal{M} := \lambda x \in X, x^* \in X^*.(\langle x, x^* \rangle)_\circ$ where $\langle x, x^* \rangle$ is the value of x under x^*,
7. $[\|\cdot\|_{X^*}]_\mathcal{M} := \lambda x^* \in X^*.(\|x^*\|)_\circ$ where $\|x^*\|$ denotes the norm of x^* in X^*.

Note that the element a^* in item (5) exists since X and thus X^* is non-trivial.

This is only well-defined in $\mathcal{M}^{\omega,X,X^*}$ if we can construct majorants of these objects. This we can do as follows:

1. $\lambda x^0, y^0.(x + y) \gtrsim +_{X^*}$,
2. $\lambda x^0.x \gtrsim -_{X^*}$,
3. $\lambda \alpha^1, x^0.\left((\alpha(0) + 1)x\right) \gtrsim \cdot_{X^*}$,
4. $0^0 \gtrsim 0_{X^*}$,
5. $1^0 \gtrsim 1_{X^*}$,
6. $\lambda x^0, y^0, n^0.j((x \cdot y)2^{n+2}, 2^{n+1} - 1) \gtrsim \langle \cdot, \cdot \rangle_{X^*}$,
7. $\lambda x^0, n^0.j(x2^{n+2}, 2^{n+1} - 1) \gtrsim \|\cdot\|_{X^*}$.

The justifications that those terms listed in item (1) - (5) and (7) really are majorants are completely analogous to the usual normed case of X alone (see e.g. the proof of Lemma 17.85 in [96]) and we thus omit the details for them (note that item (7), similar to item (6) discussed below, relies on Lemma 2.1.2). We thus only discuss item (6) explicitly: to show that $\lambda x^0, y^0, n^0.j((x \cdot y)2^{n+2}, 2^{n+1} - 1) \gtrsim \langle \cdot, \cdot \rangle_{X^*}$, note first that

$$\lambda n^0.j((x \cdot y)2^{n+2}, 2^{n+1} - 1) = (x \cdot y)_\circ$$

for the natural numbers x, y. Now, we need to show that if $n \gtrsim x^*$ and $m \gtrsim x$ (i.e. $n \geqslant \|x^*\|$ and $m \geqslant \|x\|$), then $(n \cdot m)_\circ \gtrsim (\langle x, x^* \rangle)_\circ$ and if $n' \geqslant n$, $m' \geqslant m$, then $(n' \cdot m')_\circ \gtrsim (n \cdot m)_\circ$. For the former, note that by axiom $(*)_1$, we have $|\langle x, x^* \rangle| \leqslant \|x^*\| \, \|x\| \leqslant n \cdot m$ and thus Lemma 2.1.2 implies $(n \cdot m)_\circ \gtrsim (\langle x, x^* \rangle)_\circ$. The latter follows immediately from Lemma 2.1.2 as well.

The above arguments can be similarly used for treating X^{**} and we thus do not spell this out in any more detail here.

Lastly, we consider the extensions $\mathcal{D}^\omega[f, \nabla f]$ and $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ where we focus only on the latter. For this, we fix the interpretation of the constants ∇f and ∇f^* as well as α^f, $\omega^{\nabla f}$ and $\omega^{\nabla f^*}$ just by their respective counterparts fixed in the formulation of item (2). Further, we set

1. $[f]_{\mathcal{M}} := \lambda x \in X.(f(x))_{\circ}$,

2. $[f^*]_{\mathcal{M}} := \lambda x^* \in X^*.(f^*(x^*))_{\circ}$.

Given $\omega \in \mathbb{N}^{\mathbb{N}}$ and $n \in \mathbb{N}$ with $\omega \gtrsim \omega^{\nabla f}$, $\omega^{\nabla f^*}, \alpha^f$ as well as $n \geqslant |f(0)|$, $|f^*(0)|$, $\|\nabla f(0)\|$, $\|\nabla f^*(0)\|$, majorizability of the other constants follows rather immediately according to the following constructions:

1. $\lambda x^0, m^0.j((x2^{C(x)} + n + 1)2^{m+2}, 2^{m+1} - 1) \gtrsim f$,

2. $\lambda x^0, m^0.j((x2^{C(x)} + n + 1)2^{m+2}, 2^{m+1} - 1) \gtrsim f^*$,

3. $\lambda x^0.(C(x)) \gtrsim \nabla f$,

4. $\lambda x^0.(C(x)) \gtrsim \nabla f^*$,

where $C(x) = x2^{\omega(0,x)} + n + 1$. Justifications that those terms really are majorants can again be given in a completely analogous way as before (utilizing Lemma 2.1.2 as before but also Lemma 8.5.5 and its variant for f^* and ∇f^* as in Remark 8.5.15) and we thus omit the details.

That $\mathcal{M}^{\omega,X,X^*}$ with these chosen interpretations is a model of $\mathcal{D}^{\omega-}$ + (BR) (and its extensions) can be shown similarly as in analogous results (see e.g. [96]). The intended interpretations of the constants of $\mathcal{D}^{\omega}$ and its extensions in $\mathcal{S}^{\omega,X,X^*}$, turning $\mathcal{S}^{\omega,X,X^*}$ into a model of these systems, are defined in analogy to the corresponding model $\mathcal{M}^{\omega,X,X^*}$ defined above.

For treating the other additional axioms in $\mathcal{D}^{\omega} + \Delta$ (or its extensions) of type Δ beyond the axioms already contained in $\mathcal{D}^{\omega}$ (or its extensions), we rely on the following argument (akin to [76], Lemma 5.11) showing that $\mathcal{S}^{\omega,X,X^*} \models \Delta$ implies $\mathcal{M}^{\omega,X,X^*} \models \widetilde{\Delta}$. For this, the proof given in [76] for Lemma 5.11 carries over which we sketch here: While $\mathcal{M}^{\omega,X,X^*}$ in general is not a model of the axiom of choice [88], one can show (similar to [88]) that $\mathcal{M}^{\omega,X,X^*} \models$ b-AC$_{X,X^*}$ where

$$\text{b-AC}_{X,X^*} := \bigcup_{\delta,\rho \in T^{X,X^*}} \text{b-AC}^{\delta,\rho}$$

with

$$\text{b-AC}^{\delta,\rho} := \forall Z^{\rho(\delta)} \left(\forall x^{\delta} \exists y \leqslant_{\rho} ZxA(x,y,Z) \to \exists Y \leqslant_{\rho(\delta)} Z \forall x^{\delta} A(x,Yx,Z) \right).$$

Further, we now can see the significance of the notions of small and admissible types in axioms of type Δ: for small types ρ, we have $M_\rho = S_\rho$ while for admissible types ρ, we have $M_\rho \subseteq S_\rho$ (for which it is important that admissible types take arguments of small types). For this, the proof given in [71] carries over. Further, we need that it is provable in $\mathcal{D}^{\omega-}$ that

$$\forall x', x, y\, (x' \gtrsim_\rho x \wedge x \geqslant_\rho y \rightarrow x' \gtrsim_\rho y) \tag{+}$$

holds for all types ρ which can be shown similar as e.g. in [96].

Suppose now that

$$\mathcal{S}^{\omega,X,X^*} \models \forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{ra} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c}).$$

Then also $\mathcal{M}^{\omega,X,X^*}$ is a model of this sentence: First the types of the variables which are universally quantified are admissible, so over $\mathcal{M}^{\omega,X,X^*}$ the domain of the universal quantifiers is reduced. For the witnesses for $\underline{b}$, which exist in $\mathcal{S}^{\omega,X,X^*}$, note first that these could potentially live in $\mathcal{M}^{\omega,X,X^*}$ as the types of the variables in $\underline{b}$ are admissible, i.e. they take arguments of small types and map into small types. It thus only remains to be seen whether such a witness is majorizable for majorizable inputs $\underline{a}$. However, by the above argument, the terms in $\underline{r}$ are all majorizable and if $\underline{a}$ comes from $\mathcal{M}^{\omega,X,X^*}$, then $\underline{ra}$ is majorizable. That we have $\underline{b} \leqslant_{\underline{\sigma}} \underline{ra}$ now implies that $\underline{b}$ is majorizable by (+) (and consequently the corresponding interpretations exist in $\mathcal{M}^{\omega,X,X^*}$ too). Lastly, it is rather immediate to see that $\mathcal{M}^{\omega,X,X^*} \models \Delta$ implies $\mathcal{M}^{\omega,X,X^*} \models \widetilde{\Delta}$ using b-AC$_{X,X^*}$.

From $\mathcal{M}^{\omega,X,X^*} \models \widetilde{\Delta}$, we immediately get that the above majorizability result extends to those variants of the systems where the corresponding Skolem functionals of these axioms are added and where the axioms themselves are replaced by their instantiated Skolem normal forms (i.e. $\overline{\mathcal{D}}^{\omega-}_{\Delta}$ and its extensions) and we also immediately get that the corresponding structures defined by canonical interpretations of those additional constants are indeed models of the corresponding systems.

Note that, technically, these arguments were already needed in the above considerations to see that $\mathcal{M}^{\omega,X,X^*}$ really is a model of $\mathcal{D}^{\omega-}$ (and its extensions). However, we did not discuss this there explicitly as for those specific axioms of type Δ belonging to $\mathcal{D}^{\omega-}$ (and its extensions), the types of the variables occurring in them are not only small but actually all among $\{0, 1, X, X^*\}$ so that it was immediately clear that the models coincide at that level (essentially just by definition) and we thus omitted such a general discussion there. □

Combined with the Dialectica interpretation, the main result we then arrive at is the following bound extraction result for classical proofs:

Theorem 8.6.4. *Let τ be admissible, δ be of degree 1 and s be a closed term of $\mathcal{D}^\omega$ of type $\sigma(\delta)$ for admissible σ. Let Δ be a set of formulas of the form $\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c})$ where F_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and where $\underline{r}$ is a tuple of closed terms of appropriate type. Let $B_\forall(x, y, z, u)/C_\exists(x, y, z, v)$ be $\forall$-/$\exists$-formulas of $\mathcal{D}^\omega$ with only $x, y, z, u/x, y, z, v$ free. If*

$$\mathcal{D}^\omega + \Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x, y, z, u) \rightarrow \exists v^0 C_\exists(x, y, z, v) \right),$$

then one can extract a partial functional $\Phi : S_\delta \times S_{\widehat{\tau}} \rightharpoonup \mathbb{N}$ which is total and (bar-recursively) computable on $M_\delta \times M_{\widehat{\tau}}$ and such that for all $x \in S_\delta$, $z \in S_\tau$, $z^* \in S_{\widehat{\tau}}$, if $z^* \gtrsim z$, then

$$\mathcal{S}^{\omega,X,X^*} \models \forall y \leqslant_\sigma s(x) \left(\forall u \leqslant_0 \Phi(x, z^*) B_\forall(x, y, z, u) \rightarrow \exists v \leqslant_0 \Phi(x, z^*) C_\exists(x, y, z, v)\right)$$

holds whenever $\mathcal{S}^{\omega,X,X^*} \models \Delta$ for $\mathcal{S}^{\omega,X,X^*}$ defined via any (nontrivial) Banach space $(X, \|\cdot\|)$ with its dual X^* (and with suitable interpretations of the additional constants). Further:

1. *If $\widehat{\tau}$ is of degree 1, then Φ is a total computable functional.*
2. *We may have tuples instead of single variables x, y, z, u, v and a finite conjunction instead of a single premise $\forall u^0 B_\forall(x, y, z, u)$.*
3. *If the claim is proved without DC, then τ may be arbitrary and Φ will be a total functional on $S_\delta \times S_{\widehat{\tau}}$ which is primitive recursive in the sense of Gödel. In that case, also plain majorization can be used instead of strong majorization.*
4. *The claim of the theorem as well as the items (1) - (3) from above hold similarly for*
   (a) *$\mathcal{D}^\omega[X^{**}, \|\cdot\|_{X^{**}}]$ or its extension with the reflexivity axiom where the model and the majorizability relation, etc., have to be suitably extended,*
   (b) *$\mathcal{D}^\omega[f, \nabla f]$ and $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$, assuming a convex and Fréchet differentiable function $f : X \to \mathbb{R}$ where ∇f is uniformly continuous on bounded subsets for the former or where f is additionally supercoercive and ∇f^* is uniformly continuous on bounded subsets for the latter. Then the result holds for the additional constants suitably interpreted and the resulting bound will*

depend additionally on some $\omega \in \mathbb{N}^{\mathbb{N}}$ and some $n \in \mathbb{N}$ such that $\omega \gtrsim \omega^{\nabla f}$ and $n \geqslant_{\mathbb{R}} |f(0)|, \|\nabla f(0)\|_{X^*}$ for the former and where additionally $\omega \gtrsim \omega^{\nabla f^*}, \alpha^f$ and $n \geqslant_{\mathbb{R}} |f^*(0)|, \|\nabla f^*(0)\|_X$ for the latter.

Proof. The structure of the proof is very much standard and follows that of the proof of Theorem 3.7.9 from Chapter 3 and so we just briefly sketch the key parts. For this, we focus on $\mathcal{D}^\omega + \Delta$ and so we just assume for simplicity now that

$$\mathcal{D}^\omega + \Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x,y,z,u) \to \exists v^0 C_\exists(x,y,z,v)\right).$$

Clearly, it then also holds that

$$\overline{\mathcal{D}}^\omega_\Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x,y,z,u) \to \exists v^0 C_\exists(x,y,z,v)\right)$$

where $\overline{\mathcal{D}}^\omega_\Delta$ is defined as above. To this theory, Lemma 3.7.3 still applies and we can extract witnesses of the quantifiers which by Lemma 8.6.3 have majorants in the model defined over $\mathcal{M}^{\omega,X,X^*}$. We then can recover to the truth in $\mathcal{S}^{\omega,X,X^*}$ as the types are low enough as before. □

Further, following the methodology for the semi-constructive metatheorems laid out in Chapter 3, we obtain the following semi-constructive version for the system $\mathcal{D}^\omega_i$ defined similar to $\mathcal{D}^\omega$ but over $\mathcal{A}^\omega_i[X, \|\cdot\|]$ instead of $\mathcal{A}^\omega[X, \|\cdot\|]$ and similar for the respective extensions. For this, the additional axioms of type Δ as well as the linearity rules again have to be eliminated as above but as the constructions and proofs are completely analogous, we omit them here and just state the result:

Theorem 8.6.5. *Let δ be of the form $0(0)\dots(0)$ and σ, τ be arbitrary, s be a closed term of suitable type. Let $\Gamma_\neg$ be a set of sentences of the form $\forall \underline{u}^{\underline{\zeta}}(C(\underline{u}) \to \exists \underline{v} \leqslant_{\underline{\beta}} \underline{t}\underline{u} \neg D(\underline{u}, \underline{v}))$ with $\underline{\zeta}, \underline{\beta}$ and C, D arbitrary types and formulas respectively and where $\underline{t}$ is a tuple of closed terms. Let $B(x,y,z)/C(x,y,z,u)$ be arbitrary formulas of $\mathcal{D}^\omega_i$ with only $x,y,z/x,y,z,u$ free. If*

$$\mathcal{D}^\omega_i + \mathrm{IP}_\neg + \mathrm{CA}_\neg + \Gamma_\neg \vdash \forall x^\delta\, \forall y \leqslant_\sigma (x)\, \forall z^\tau\, (\neg B(x,y,z) \to \exists u^0 C(x,y,z,u)),$$

one can extract a $\Phi : S_\delta \times S_{\hat{\tau}} \to \mathbb{N}$ with is primitive recursive in the sense of Gödel such that for any $x \in S_\delta$, any $y \in S_\sigma$ with $y \leqslant_\sigma s(x)$, any $z \in S_\tau$ and $z^* \in S_{\hat{\tau}}$ with $z^* \gtrsim z$, we have that

$$\mathcal{S}^{\omega,X,X^*} \models \exists u \leqslant_0 \Phi(x,z^*)\, (\neg B(x,y,z) \to C(x,y,z,u))$$

holds whenever $\mathcal{S}^{\omega,X,X^*} \models \Gamma_\neg$ *where* $\mathcal{S}^{\omega,X,X^*}$ *is defined via any (nontrivial) Banach space* $(X, \|\cdot\|)$ *with dual* X^* *and with the constants interpreted as in Theorem 8.6.4.*

Further, the results also hold for the analogously defined theories $\mathcal{D}_i^\omega[X^{**}, \|\cdot\|_{X^{**}}]$, $\mathcal{D}_i^\omega[f, \nabla f]$ *and* $\mathcal{D}_i^\omega[f, \nabla f, f^*, \nabla f^*]$ *with similar modifications as in Theorem 8.6.4.*

9 Effective rates for iterations involving Bregman strongly nonexpansive operators

9.1 Introduction

In this chapter, we provide applications of the metatheorems established in the previous Chapter 8 to Picard- and Halpern-style iterations of Bregman strongly nonexpansive mappings. These types of mappings were first considered in [175], extending the influential notion of strongly nonexpansive maps [31] (or, more precisely, that of quasi strongly nonexpansive maps) to a notion involving Bregman distances.

The class of strongly nonexpansive maps is of vital importance for many influential developments in modern nonlinear optimization and analysis and consequently also has been at the focus of many recent developments in proof mining, first having been studied in [99] in the context of Picard-iterations involving such mappings. Subsequently, these mappings and their quantitative properties have in particular played a crucial role in the analysis given by Kohlenbach in [101] of Bauschke's proof [6] of the zero displacement conjecture [10].

The results presented here are partly in that same vein as the work [99] is situated in as we provide quantitative versions of the respective asymptotic regularity results for Picard iterations of these Bregman strongly nonexpansive maps contained in [143, 144]. In the context of Bregman distances and monotone operators on Banach spaces in the sense of Browder [28, 30], there also exists a notion of resolvent relative to the convex function f (as defined in [9, 58])[1] and as discussed in [9], such resolvents and thus in particular also the so-called Bregman projections as defined already in Bregman's foundational work [22] are Bregman strongly nonexpansive. In that way, the results

[1] These types of operators, while not covered by the formal discussions from the previous Chapter 8, will be discussed in the upcoming Chapter 10 which in particular will provide full formal justification for all extractions presented here.

presented here in particular also cover the influential proximal point algorithm (as introduced by Rockafellar [183] and Martinet [145]) extended to Bregman distances (as first considered in [58]) as well as the method of cyclic Bregman projections (see [175]).

Besides Picard-type iterations, we are also concerned here with Halpern-type iterations of Bregman strongly nonexpansive maps. The original method of Halpern, which relies on convex combinations of the iterations with an anchor to induce strong convergence (see [77] where the iteration was introduced for the anchor 0 by Halpern and [207] for the consequent seminal extension by Wittmann), is one of the most influential methods studied in nonlinear analysis in the recent decades. Consequently, also this method has attracted extensive attention from the research program of proof mining and the original iterations as well as a wide breadth of extension were analyzed (see e.g. [64, 98, 110, 124, 186] among many others).

Here, we are initially concerned with the work [199] where the authors extend the usual strong convergence results for Halpern-type iterations to Bregman strongly nonexpansive maps. We analyze this result in the similar spirit as in [104] and obtain a quantitative version providing a rate of metastability for the strong convergence. Further, we are able to also incorporate families of Bregman strongly nonexpansive maps which relate to an anchor map via a uniform quantitative version of the influential NST condition (see e.g. [2]). From this, by forgetting about the quantitative aspects, we are able to derive a new "ordinary" (that is non-quantitative) strong convergence result for this specific iteration involving a family of maps.

At last, we exploit this new generality and discuss what old and in particular new results can be derived from it. In that vein, we in particular obtain (quantitative) strong convergence results for Halpern-type variants of the method of cyclic Bregman projections, of the proximal point algorithm, of a special case of a method solving operator equations due to Butnariu and Resmerita [36] as well as of a special case of the forward-backward Bregman splitting method discussed by Búi and Combettes [32] (see also Van Nguyen [155]), of a method for finding common zeros of maximally monotone operators as discussed by Naraghirad [152] and of a Halpern-Mann type iteration of Bregman strongly nonexpansive maps [214] where we obtain a qualitative improvement on the conditions presented in [214].

Further, inspired by the recent considerations [41] on the relationship between modified Halpern methods in the sense of [53, 85] and Tikhonov-Mann type methods as developed in [20, 42, 210], we even provide a new strong convergence result for a

Tikhonov-Mann type iteration of Bregman strongly nonexpansive maps which we newly define in this chapter.

This chapter relies on various notions from convex analysis, in particular surrounding convex functions, their gradients and their corresponding Bregman distances. Although already discussed in the context of the logical investigations of Chapter 8, we will sometimes (re-)introduce these notions as needed throughout the chapter. For further expositions about convex analysis in Banach or Hilbert spaces, we again refer to the standard works [11, 182, 184, 212]. In this section, we just at first collect the essential notions regarding Bregman distances.

Throughout, if not specified otherwise, let X be a Banach space with norm $\|\cdot\|$ and let $f : X \to (-\infty, +\infty]$ be a given function with extended real values. In the following, we will assume that f is proper, lower-semicontinuous and convex (compare the definitions in Chapter 8). Similarly, we also rely on the other notions discussed in Chapter 8 regarding the differentiability of convex functions.

The fundamental notion of distance in this chapter is that of the influential Bregman distance already briefly discussed in Chapter 8: Let f be Gâteaux differentiable. The Bregman distance associated with f is the function $D_f : \mathrm{dom} f \times \mathrm{intdom} f \to [0, +\infty)$ which is defined as follows:

$$D_f(x, y) := f(x) - f(y) - \langle x - y, \nabla f(y)\rangle.$$

For this Bregman distance, in particular recall the so-called three and four point identities for D_f:

Lemma 9.1.1 (folklore, see e.g. [9])**.** *The following equalities are true for all $x, y, z, w \in \mathrm{intdom} f$:*

1. $D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle x - y, \nabla f(z) - \nabla f(y)\rangle.$
2. $D_f(y, x) - D_f(y, z) - D_f(w, x) + D_f(w, z) = \langle y - w, \nabla f(z) - \nabla f(x)\rangle.$

Recall also the following dual function $W_f : \mathrm{dom} f \times \mathrm{dom} f^* \to [0, +\infty)$ defined by

$$W_f(x, x^*) = f(x) - \langle x, x^*\rangle + f^*(x^*).$$

For this function, we will rely here on a few further properties: If $f : X \to \mathbb{R}$ is Legendre and supercoercive and if X is reflexive, one in particular has that

$$W_f(x, \nabla f(y)) = D_f(x, y)$$

for all $x, y \in X$ (see also Chapter 10) as well as that W_f is convex in its right argument and satisfies the inequality

$$W_f(x, x^*) \leqslant W_f(x, x^* + y^*) - \langle \nabla f^*(x^*) - x, y^* \rangle$$

for any $x \in X$ and any $x^*, y^* \in X^*$ (see [122]).

9.2 Gradients, Bregman distances and their quantitative properties

Throughout most of this chapter, if not indicated otherwise, we will now assume that f and f^* are total (i.e. $\mathrm{dom} f = X$ and $\mathrm{dom} f^* = X^*$, respectively) and that both are Fréchet differentiable everywhere with gradients ∇f and ∇f^*. This section now introduces the main quantitative notions related to the core objects like the gradients and distances. For this, and in this chapter in general, we use ε's to represent errors.

9.2.1 Quantitative properties of gradients

Definition 9.2.1. We say that a function $\omega^{\nabla f} : (0, \infty)^2 \to (0, \infty)$ is a modulus of uniform continuity (on bounded sets) for ∇f if for any $\varepsilon, b > 0$ and any $x, y \in \overline{B}_b(0)$:

$$\|x - y\| < \omega^{\nabla f}(\varepsilon, b) \to \|\nabla f(x) - \nabla f(y)\| < \varepsilon.$$

Using such a modulus, we can immediately derive quantitative witnesses for various central properties of ∇f and f. In that vein, the following lemma, giving such witnesses, is essentially just a reformulation of Lemma 8.5.5 written using ε's instead of 2^{-k} and as such, the proof is essentially the same and thus omitted.

Lemma 9.2.2. *Assume that ∇f is uniformly continuous on bounded subsets with a modulus $\omega^{\nabla f}$. Then:*

1. *f is uniformly Fréchet differentiable on bounded subsets with modulus*

$$\Delta(\varepsilon, b) = \min\{\omega^{\nabla f}(\varepsilon, b + 1), 1\},$$

i.e. for all $b, \varepsilon > 0$ and all $x \in \overline{B}_b(0), y \in X$:

$$0 < \|y\| < \Delta(\varepsilon, b) \to \frac{|f(x + y) - f(x) - \langle y, \nabla f(x) \rangle|}{\|y\|} < \varepsilon.$$

2. ∇f *is bounded on bounded subsets with modulus*

$$C(b) = \lceil b/\omega^{\nabla f}(1,b) \rceil + \|\nabla f(0)\| + 1,$$

i.e. for all $b > 0$ *and all* $x \in \overline{B}_b(0)$*:*

$$\|\nabla f(x)\| \leqslant C(b).$$

3. f *uniformly continuous on bounded subsets with modulus*

$$\omega^f(\varepsilon, b) = \frac{\varepsilon}{C(b)},$$

i.e. for all $\varepsilon, b > 0$ *and all* $x, y \in \overline{B}_b(0)$*:*

$$\|x - y\| < \omega^f(\varepsilon, b) \to |f(x) - f(y)| < \varepsilon.$$

4. f *is bounded on bounded sets with modulus*

$$D(b) = \lceil b/\omega^f(1,b) \rceil + |f(0)| + 1,$$

i.e. for all $b > 0$ *and all* $x \in \overline{B}_b(0)$*:*

$$|f(x)| \leqslant D(b).$$

Similar results of course also hold for the conjugate f^* if we assume a modulus of uniform continuity on bounded sets for the respective gradient ∇f^*.

If f is Fréchet differentiable, then the associated Bregman distance is continuous in both arguments and by analyzing the corresponding proof, we can extract a transformation that turns a modulus for the uniform continuity of the gradient of f into a modulus for the uniform continuity of the associated Bregman distance. This is collected in the following lemma:

Lemma 9.2.3. *Assume that* ∇f *is uniformly continuous on bounded subsets with a modulus* $\omega^{\nabla f}$*. Let* C *be a modulus for* ∇f *being bounded on bounded sets.*[2]

[2]As shown in the previous Lemma 9.2.2, such a C can actually be constructed from $\omega^{\nabla f}$. We however throughout work with a given C as a black box so that the contributions of the different types of moduli are highlighted.

1. *For any* $\varepsilon, b > 0$ *and any* $x, y, y' \in \overline{B}_b(0)$*:*

$$\|y - y'\| < \xi(\varepsilon, b) \to |D_f(x, y) - D_f(x, y')| < \varepsilon$$

where $\xi : (0, \infty)^2 \to (0, \infty)$ *can be explicitly given by*

$$\xi(\varepsilon, b) := \min\left\{\frac{\varepsilon}{4C(b)}, \omega^{\nabla f}\left(\frac{\varepsilon}{4b}, b\right)\right\}.$$

2. *For any* $\varepsilon, b > 0$ *and any* $x, x', y \in \overline{B}_b(0)$*:*

$$\|x - x'\| < \xi'(\varepsilon, b) \to |D_f(x, y) - D_f(x', y)| < \varepsilon$$

where $\xi' : (0, \infty)^2 \to (0, \infty)$ *can be explicitly given by*

$$\xi'(\varepsilon, b) := \frac{\varepsilon}{2C(b)}.$$

Proof. For item (1), note that we have

$$\begin{aligned}
|\langle y, \nabla f y\rangle - \langle y', \nabla f y'\rangle| &= |\langle y, \nabla f y\rangle - \langle y', \nabla f y\rangle + \langle y', \nabla f y\rangle - \langle y', \nabla f y'\rangle| \\
&\leqslant |\langle y - y', \nabla f y\rangle| + |\langle y', \nabla f y - \nabla f y'\rangle| \\
&\leqslant \|\nabla f y\| \, \|y - y'\| + \|y'\| \, \|\nabla f y - \nabla f y'\| .
\end{aligned}$$

Using that, we derive

$$\begin{aligned}
|D_f(x, y) - D_f(x, y')| &\leqslant |f(y) - f(y')| + |\langle x - y', \nabla f y'\rangle - \langle x - y, \nabla f y\rangle| \\
&\leqslant |f(y) - f(y')| + |\langle x, \nabla f y' - \nabla f y\rangle| \\
&\qquad + |\langle y, \nabla f y\rangle - \langle y', \nabla f y'\rangle| \\
&\leqslant |f(y) - f(y')| + \|x\| \, \|\nabla f y - \nabla f y'\| \\
&\qquad + \|\nabla f y\| \, \|y - y'\| + \|y'\| \, \|\nabla f y - \nabla f y'\| .
\end{aligned}$$

This yields the claim by the definition of ξ as by Lemma 9.2.2, we have that $\varepsilon/4C(b) = \omega^f(\varepsilon/4, b)$ for a suitably defined modulus of uniform continuity ω^f for f.

For item (2), note that

$$\begin{aligned}
|D_f(x, y) - D_f(x', y)| &\leqslant |f(x) - f(x')| + |\langle x - x', \nabla f(y)\rangle| \\
&\leqslant |f(x) - f(x')| + \|x - x'\| \, \|\nabla f(y)\|
\end{aligned}$$

and this yields the claim by the definition of ξ' as by Lemma 9.2.2, we have that $\varepsilon/2C(b) = \omega^f(\varepsilon/2, b)$ for a suitably defined modulus of uniform continuity ω^f for f. □

An assumption that is later used in the context of Halpern-type iterations is that f is uniformly strictly convex on bounded subsets in the sense of [36], i.e.

$$\forall \varepsilon, b > 0 \exists \delta > 0 \forall x, y \in X\Big(\|x\|, \|y\| \leqslant b \wedge \|x - y\| \geqslant \varepsilon \\ \to \left(f\left(\frac{x+y}{2}\right) \leqslant \frac{1}{2}f(x) + \frac{1}{2}f(y) - \delta \right)\Big).$$

In the following, we will occasionally assume a modulus of uniform strict convexity $\eta : (0,\infty)^2 \to (0,\infty)$ for f, i.e. an η witnessing the above quantifier $\exists\delta > 0$ in terms of ε and b. By the equivalent characterization of strictly convex functions f as those where ∇f is strictly monotone, we can translate such a modulus of uniform strict convexity into a modulus witnessing the "uniform strict monotonicity" of ∇f, i.e. an $\widehat{\eta} : (0,\infty)^2 \to (0,\infty)$ witnessing δ in terms of ε, b in the following condition:

$$\forall \varepsilon, b > 0 \exists \delta > 0 \forall x, y \in X(\|x\|, \|y\| \leqslant b \wedge \|x - y\| \geqslant \varepsilon \to (\langle x - y, \nabla f x - \nabla f y\rangle \geqslant \delta)).$$

This is collected in the following lemma.

Lemma 9.2.4. *Let $\eta(\varepsilon, b)$ be a modulus of uniform strict convexity for f. Then $\widehat{\eta}(\varepsilon, b) = 4\eta(\varepsilon, b)$ is a modulus of uniform strict monotonicity for ∇f.*

Proof. Note that we have

$$\begin{aligned} f\left(\frac{x+y}{2}\right) &\leqslant 1/2f(y) + 1/2f(x) - \eta(\varepsilon, b) \\ &= f(x) + 1/2(f(y) - f(x)) - \eta(\varepsilon, b) \end{aligned}$$

if $\|x - y\| \geqslant \varepsilon$ as η is a modulus of uniform strict convexity of f. As $\nabla f w$ is a subgradient of f at w, we have

$$\langle z, \nabla f w\rangle \leqslant \inf_{\alpha > 0} \frac{f(w + \alpha z) - f(w)}{\alpha},$$

for all w, z and from this we get

$$\langle y - x, \nabla f x\rangle \leqslant f(y) - f(x) - 2\eta(\varepsilon, b).$$

Similarly, we get

$$\langle x - y, \nabla f y\rangle \leqslant f(x) - f(y) - 2\eta(\varepsilon, b)$$

and this implies

$$\langle x - y, \nabla f y - \nabla f x\rangle \leqslant -4\eta(\varepsilon, b)$$

which gives that $\widehat{\eta}(\varepsilon, b) = 4\eta(\varepsilon, b)$ is a modulus of uniform strict monotonicity of ∇f. □

Conversely, also from a modulus $\widehat{\eta}$ for the uniform strict monotonicity we can construct a modulus η for the uniform strict convexity but we omit this other direction as, for one, this construction is rather messy and, for another, the one direction presented above suffices to justify that such an $\widehat{\eta}$ exists in the context of the central assumptions featured in the convergence results later on.

9.2.2 Sequential consistency and total convexity

Another central assumption featuring in the convergence results later on is that of the total convexity of f which we want to discuss in the following. For this, we briefly only assume that $f : X \to (-\infty, +\infty]$ is proper, lower-semicontinuous and convex.

Definition 9.2.5 (see e.g. [34])**.** Given a function f, define its modulus of total convexity $v_f : \mathrm{intdom} f \times [0, +\infty) \to [0, +\infty]$ by

$$v_f(x, t) := \inf\{D_f(y, x) \mid y \in \mathrm{dom} f, \|y - x\| = t\}.$$

The function f is called totally convex at a point $x \in \mathrm{intdom} f$ if $v_f(x, t) > 0$ whenever $t > 0$. It is called totally convex if it is totally convex at every point. Lastly, we call f totally convex on bounded sets if

$$v_f(B, t) := \inf\{v_f(x, t) \mid x \in B \cap \mathrm{intdom} f\} > 0$$

for any $t > 0$ and for any non-empty bounded set $B \subseteq X$.

This notion is intimately connected with the so-called sequential consistency for the function f:

Definition 9.2.6 ([36])**.** A function f is called sequentially consistent if for all bounded sequences (x_n) and (y_n) in $\mathrm{intdom} f$:

$$D_f(x_n, y_n) \to 0 \ (n \to \infty) \text{ implies } \|x_n - y_n\| \to 0 \ (n \to \infty).$$

Concretely, the main result connecting total convexity and sequential consistency is now the following:

Lemma 9.2.7 ([34])**.** *A proper, lower-semicontinuous and convex function $f : X \to (-\infty, +\infty]$ whose domain contains at least two points is totally convex on bounded sets if, and only if, it is sequentially consistent.*

In the following, we will rely on a modulus witnessing the sequential consistency of a function quantitatively. To motivate this, we move to another equivalent way of formulating sequential consistency (which is somewhat in spirit of e.g. Proposition 2.5 of [36], see also [179]). For the following, let f now again be total and Fréchet differentiable everywhere like in the previous standing assumptions.

Lemma 9.2.8. *A function f is sequentially consistent if, and only if, for all $b > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\forall x, y \in X \left(\|x\|, \|y\| \leqslant b \wedge D_f(x, y) < \delta \to \|x - y\| < \varepsilon\right). \tag{+}$$

Proof. For sufficiency, consider arbitrary sequences $(x_n), (y_n)$ with $\|x_n\|, \|y_n\| \leqslant b$ for some $b > 0$ and assume that $\lim D_f(x_n, y_n) = 0$. Let $\varepsilon > 0$ be given. By (+), there is a δ such that

$$\forall m \in \mathbb{N} \left(D_f(x_m, y_m) < \delta \to \|x_m - y_m\| < \varepsilon\right). \tag{++}$$

Then, by $\lim D_f(x_n, y_n) = 0$ there exists $N \in \mathbb{N}$ such that

$$\forall m \geqslant N \left(D_f(x_m, y_m) < \delta\right),$$

which by (++) entails that $\|x_m - y_m\| < \varepsilon$, for all $m \geqslant N$. This means that

$$\|x_n - y_n\| \to 0,$$

and we conclude the sequential consistency of f.

For necessity, suppose that (+) fails. Then for some $\varepsilon > 0$ and $b > 0$, we have

$$\forall n \in \mathbb{N} \exists x_n, y_n \in X \left(\|x_n\|, \|y_n\| \leqslant b \wedge D_f(x_n, y_n) < \frac{1}{n+1} \wedge \|x_n - y_n\| \geqslant \varepsilon\right).$$

Then in particular $D_f(x_n, y_n) < \frac{1}{n+1}$ for all $n \in \mathbb{N}$ which entails that

$$D_f(x_n, y_n) \to 0.$$

However $\|x_n - y_n\|$ is bounded away from zero by ε, and so f can not be sequentially consistent as x_n and y_n are bounded. □

Definition 9.2.9. Let f be sequentially consistent. A modulus of consistency for f is a function $\rho : (0, \infty)^2 \to (0, \infty)$ such that for all $b \in \mathbb{N}$ and $\varepsilon > 0$:

$$\forall x, y \in X \left(\|x\|, \|y\| \leqslant b \wedge D_f(x, y) < \rho(\varepsilon, b) \to \|x - y\| < \varepsilon\right).$$

By the above result, a function f is sequentially consistent if, and only if, it has a modulus of consistency.

We call a modulus of this type but for the converse implication, i.e. translating errors for the metric distance into errors for the Bregman distance, a *modulus of reverse consistency.* Further, such a modulus can actually be computed from a modulus of ∇f being bounded on bounded sets.

Lemma 9.2.10. *Let ∇f be bounded on bounded sets with a modulus C. Then for all $\varepsilon > 0$ and $b > 0$:*

$$\forall x, y \in X \left(\|x\|, \|y\| \leqslant b \wedge \|x - y\| < P(\varepsilon, b) \rightarrow D_f(x, y) < \varepsilon \right)$$

where $P(\varepsilon, b)$ can be given in terms of C via

$$P(\varepsilon, b) = \frac{\varepsilon}{2C(b)}.$$

Proof. By Lemma 9.2.2, we have that $\omega^f(\varepsilon, b) = \varepsilon/C(b)$ is a modulus of uniform continuity for f on bounded sets. So for $\|x - y\| < P(\varepsilon, b) = \omega^f(\varepsilon/2, b)$, we have $f(x) - f(y) < \varepsilon/2$ and thus

$$\begin{aligned} D_f(x, y) &= f(x) - f(y) - \langle x - y, \nabla f(y) \rangle \\ &< \varepsilon/2 + \|x - y\| \|\nabla f(y)\| \\ &\leqslant \varepsilon/2 + \|x - y\| C(b) \\ &< \varepsilon \end{aligned}$$

which is the claim. □

We want to note that the collection of such a modulus P together with a modulus of consistency ρ are called moduli of consistency in [164]. In particular, as discussed in [164], these moduli can be used to derive a so-called modulus of weak triangularity for D_f, i.e. a function $\theta : (0, \infty)^2 \rightarrow (0, \infty)$ such that

$$\forall \varepsilon, b > 0 \forall x, y, z \in X \left(\|x\|, \|y\|, \|z\| \leqslant b \wedge D_f(y, x), D_f(y, z) < \theta(\varepsilon, b) \rightarrow D_f(x, z) < \varepsilon \right).$$

In other words, θ witnesses that although the triangle inequality is not valid for D_f, it locally behaves similar to a distance function with a triangle inequality. To derive such a θ from a given ρ and P as above, set

$$\theta(\varepsilon, b) = \rho(P(\varepsilon, b)/2, b).$$

Then, if $D_f(y,x), D(y,z) < \theta(\varepsilon, b)$ for $\|x\|, \|y\|, \|z\| \leqslant b$, we have

$$\|x-y\|, \|z-y\| < P(\varepsilon, b)/2$$

using the properties of ρ. This implies

$$\|x-z\| < P(\varepsilon, b)$$

by triangle inequality of $\|\cdot\|$. So, using the properties of P, this yields $D_f(x,z) < \varepsilon$.

Remark 9.2.11. Note that in the presence of such moduli ρ and P, all moduli introduced later that depend on measuring a distance $\|x-y\|$ in the premise or conclusion could be translated into moduli that depend on measuring the distance $D_f(x,y)$.

Besides sequential consistency, being totally convex on bounded sets can be further recognized to be equivalent to another well-known convexity property for f already mentioned before, at least in the context of the standing assumptions of this chapter.

Lemma 9.2.12 (essentially [36, Theorem 2.10]). *Let $f : X \to \mathbb{R}$ be Fréchet differentiable and let ∇f be uniformly continuous on bounded sets. Then f is totally convex on bounded sets if, and only if, f is uniformly strictly convex on bounded sets.*

In that vein, the following remark shortly discusses the relationship between the modulus of consistency and the previous modulus of uniform strict convexity together with other convexity moduli from the literature.

Remark 9.2.13. Note that it can be easily shown that ρ is a modulus of consistency if

$$v_f(\overline{B}_b(0), t) \geqslant \rho(t, b)$$

for any $t, b > 0$ (using e.g. Proposition 2.1 from [36]) and conversely, if ρ is a modulus of consistency, then $v_f(\overline{B}_b(0), t) \geqslant \rho(t, b+t)$ for any $t, b > 0$. In that way, moduli of consistency as defined in this chapter actually immediately witness the total convexity of the function f.

Further, define the modulus of uniform convexity $\mu_f(x,t)$ as in [206] (see also [35, 211]), i.e.

$$\mu_f(x,t) := \inf\left\{\frac{\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y)}{\lambda(1-\lambda)} \mid y \in X, \|y-x\| = t, \lambda \in (0,1)\right\}$$

and write

$$\mu_f(B,t) := \inf\{\mu_f(x,t) \mid x \in B\}$$

for a given set $B \subseteq X$ similar as with v_f. Similarly define

$$\overline{\mu}_f(x,t) := \inf\left\{ f(x) + f(y) - 2f\left(\frac{x+y}{2}\right) \mid y \in X, \|y-x\| = t \right\}$$

as in [34] (see also [35]). Then as shown in [35], we have

$$\mu_f(x,t) \geqslant \overline{\mu}_f(x,t) \geqslant \frac{1}{2}\mu_f(x,t)$$

for any $x \in X$ and $t \geqslant 0$ as well as $v_f(x,t) \geqslant \mu_f(x,t)$ for any $x \in X$ and $t > 0$ as shown in [34, Proposition 1.2.5]. Now, it is also immediate that η is a modulus of uniform strict convexity of f as defined above if

$$\frac{1}{2}\overline{\mu}_f(\overline{B}_b(0),t) \geqslant \eta(t,b)$$

for any $t, b > 0$ where $\overline{\mu}_f(B,t)$, given a set $B \subseteq X$, is defined similarly as $\mu_f(B,t)$. Conversely, if η is a modulus of uniform strict convexity, then $\overline{\mu}_f(\overline{B}_b(0),t) \geqslant 2\eta(t,b+t)$ for any $t, b > 0$. Thus any modulus η of uniform strict convexity of f induces a modulus of consistency and thus witnesses the total convexity of f.

Conversely, as follows from the above Lemmas 9.2.7 and 9.2.12, if f is Fréchet differentiable with a gradient that is uniformly continuous on bounded sets, then f being sequentially consistent implies f being uniformly strictly convex on bounded sets. As shown in [35], both of these items are further equivalent to f^* being uniformly Fréchet differentiable (and thus to ∇f^* being uniformly continuous on bounded sets if f is also supercoercive by Propositions 8.5.4 and 8.5.7).

9.2.3 Boundedness properties of the Bregman distance

As is well-known, the distances D_f in general have very weak properties. In particular, a sequence (x_n) such that $D_f(x_n, y)$ is bounded for some y is not necessarily bounded itself. In that way, it is thus a common requirement in the context of Bregman distances to require that the level sets

$$\begin{aligned} L_1(y,\alpha) &= \{x \in X \mid D_f(x,y) \leqslant \alpha\}, \\ L_2(x,\alpha) &= \{y \in X \mid D_f(x,y) \leqslant \alpha\}, \end{aligned}$$

are bounded for every $\alpha > 0$ and $x, y \in X$. In particular, this condition features in the list of conditions exhibited by Eckstein in [58] and by Butnariu and Iusem in [34] regarding Bregman functions and a stronger requirement of these sets being compact already featured in Bregman's seminal work [22] for the conditions imposed on his general distances D.

As shown in [34], in the case that $L_2(x,\alpha)$ is bounded for all x and α and if f is additionally sequentially consistent, then $L_1(y,\alpha)$ is likewise bounded.

Further, as shown in [7] (Theorem 3.7),[3] if f is essentially strictly convex and $\mathrm{dom} f^*$ is open, then $D_f(x,\cdot)$ is coercive for any $x \in \mathrm{intdom} f$ and thus, in that case, $L_2(x,\alpha)$ is bounded for any α. This is in particular the case for supercoercive f and thus guaranteed in essentially all situations in this chapter.

In the following, we will rely on so-called moduli of boundedness for D_f that witness a uniform quantitative version of the boundedness of L_2. Concretely, by a modulus of boundedness for D_f we will mean a function $o : (0,\infty)^2 \to (0,\infty)$ such that

$$\forall x, y \in X \forall \alpha, b > 0 \left(\|x\| \leqslant b \wedge D_f(x,y) \leqslant \alpha \to \|y\| \leqslant o(\alpha, b) \right).$$

We call D_f uniformly bounded if such a modulus exists.

Remark 9.2.14. Such a modulus of boundedness for D_f in particular exists if $f, \nabla f^*$ are bounded on bounded sets and f^* is supercoercive (which actually follows from f being bounded on bounded sets by Proposition 8.5.7 since $f = f^{**}$ holds by the Fenchel-Moreau theorem) and it can be explicitly constructed from corresponding moduli witnessing these properties. This will be discussed in Chapter 10 in more detail.

9.3 Bregman strongly nonexpansive mappings and related notions

The main notion of mapping considered in this chapter will be that of a Bregman strongly nonexpansive mapping as introduced in [39, 175].

Let $T : X \to X$ be a mapping. We say that a point $p \in X$ is an asymptotic fixed point of T if there is a sequence (x_n) which converges weakly to p and satisfies $\lim_{n\to\infty} \|x_n - Tx_n\| = 0$. We write $\widehat{F}(T)$ for the set of all such asymptotic fixed points and $F(T)$ for the set of ordinary fixed points of T.

[3] While [7] is set in finite-dimensional spaces, the proof given there for Theorem 3.7 can be easily seen to be valid in general Banach spaces.

Definition 9.3.1. A map $T : X \to X$ is called

1. Bregman nonexpansive (see [144]) if
$$D_f(Tx, Ty) \leqslant D_f(x, y)$$
for any $x, y \in X$,

2. Bregman quasi-nonexpansive (see [142, 144]) if
$$D_f(p, Tx) \leqslant D_f(p, x)$$
for any $x \in X$ and $p \in \widehat{F}(T)$,

3. Bregman strongly nonexpansive (see [39, 175]) if
$$D_f(p, Tx) \leqslant D_f(p, x)$$
for any $x \in X$, $p \in \widehat{F}(T)$ and if additionally
$$\lim_{n\to\infty} (D_f(p, x_n) - D_f(p, Tx_n)) = 0 \to \lim_{n\to\infty} D_f(Tx_n, x_n) = 0$$
for any bounded sequence $(x_n) \subseteq X$ and any $p \in \widehat{F}(T)$,

4. Bregman firmly nonexpansive (see e.g. [9]) if
$$\langle Tx - Ty, \nabla fTx - \nabla fTy\rangle \leqslant \langle Tx - Ty, \nabla fx - \nabla fy\rangle$$
for all $x, y \in X$.

It is rather immediate to see that being Bregman firmly nonexpansive implies being Bregman strongly nonexpansive (see also Lemma 9.3.8 later) and it is clear that any Bregman strongly nonexpansive mapping is Bregman quasi-nonexpansive.

We want to note that the above notion of Bregman strongly nonexpansive operators is called strictly left Bregman strongly nonexpansive in other parts of the literature (see in particular [143]) since the fixed points occur in the left argument of the Bregman distance and since we used $\widehat{F}(T)$. If $F(T) = \widehat{F}(T)$ is further assumed, then the resulting notion is called fully left Bregman strongly nonexpansive in these parts of the literature. Note also that Bregman firmly nonexpansive maps are called D-firm in [9] and ∇f firmly nonexpansive in [15].

Fundamental for the quantitative results discussed later for iterations involving such mappings are moduli which quantitatively witness the defining properties of Bregman strongly nonexpansive mappings. The whole approach taken here in regard to quantitative moduli witnessing the Bregman strongly nonexpansiveness is modeled after the work [99] for "ordinary" quasi-nonexpansive functions. In these quantitative moduli, it will always be $F(T)$ that we use when deriving the moduli which results e.g. in the fact that instead of full fixed points, these moduli will concern approximate fixed points. If it is presumed that $\widehat{F}(T) = F(T)$ and if this assumption features crucially in a given proof, then a uniform quantitative version of this fact will feature necessarily in its analysis (see p. 224 for this uniform quantitative version).

Definition 9.3.2. A function $\omega : (0,\infty)^2 \to (0,\infty)$ such that

$$\forall \varepsilon, b > 0 \forall p \in F(T) \cap \overline{B}_b(0) \forall x \in \overline{B}_b(0)$$
$$(D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b) \to D_f(Tx,x) < \varepsilon)$$

is called a BSNE-modulus of T.

Conceptually, BSNE-moduli are similar to the (uniform) SQNE-moduli introduced in [99].

If we are given a specific element $p \in F(T)$, we will later say that a function $\omega : (0,\infty)^2 \to (0,\infty)$ is a BSNE-modulus w.r.t. p if

$$\forall \varepsilon, b > 0 \forall x \in \overline{B}_b(0) \, (D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b) \to D_f(Tx,x) < \varepsilon)$$

holds for that specific p.

We will later be concerned with a stronger type of modulus which only requires p to be a sufficiently good approximate fixed point.

Definition 9.3.3. A function $\omega : (0,\infty)^2 \to (0,\infty)$ is called a strong BSNE-modulus of T if

$$\forall \varepsilon, b > 0 \forall x, p \in X((\|p\|, \|x\| \leqslant b \wedge \|Tp - p\| < \omega(\varepsilon,b)$$
$$\wedge D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon,b)) \to D_f(Tx,x) < \varepsilon).$$

We say that T is uniformly Bregman strongly nonexpansive if it has such a modulus.

Clearly, a strong BSNE-modulus is also an ordinary BSNE-modulus.

From the following lemma, we get that a uniformly Bregman strongly nonexpansive map T is in particular Bregman strongly nonexpansive whenever $\widehat{F}(T) = F(T)$.

Lemma 9.3.4. *Let f be such that D_f satisfies $D_f(x,y) = 0 \leftrightarrow \|x - y\| = 0$ for any $x, y \in X$.*[4] *Let $T : X \to X$ be given. If T satisfies that for any $\varepsilon, b > 0$ there exists a $\delta > 0$ such that for any $p \in \overline{B}_b(0)$ with $\|p - Tp\| < \delta$ and any $x \in \overline{B}_b(0)$:*

$$D_f(p,x) - D_f(p,Tx) < \delta \to D_f(Tx,x) < \varepsilon$$

and if $\widehat{F}(T) = F(T)$, then T is Bregman strongly nonexpansive.

Proof. The existence of such a $\delta > 0$ for any $\varepsilon, b > 0$ clearly implies

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

for all bounded sequences (x_n) and fixed points p. As $\widehat{F}(T) = F(T)$, this also holds for all $p \in \widehat{F}(T)$. Further, note that it also implies that T is Bregman quasi-nonexpansive as either $D_f(Tx,x) = 0$ which yields $x = Tx$ and thus $D_f(p,Tx) = D_f(p,x)$, or $D_f(Tx,x) > 0$ which yields $D_f(p,x) - D_f(p,Tx) > 0$, i.e. $D_f(p,Tx) \leqslant D_f(p,x)$ as well. Thus T is Bregman strongly nonexpansive. □

Assuming that a given mapping even satisfies this strengthened notion of being uniformly Bregman strongly nonexpansive has practically often very little impact as in most concrete applications, a corresponding strong BSNE-modulus can actually be obtained (as is e.g. the case for Bregman firmly nonexpansive maps as Lemma 9.3.8 shows).

If D_f is uniformly bounded with a modulus of boundedness o as introduced in Section 9.2.3, then any Bregman quasi-nonexpansive map T with a non-empty fixed point set is bounded on bounded sets and we can also construct a witness for that in the following sense:

Lemma 9.3.5. *Let T be Bregman quasi-nonexpansive and let $p_0 \in F(T) \neq \emptyset$. Let ∇f, f be bounded on bounded sets with moduli C, D, respectively. Let o be a modulus of boundedness for D_f.*

[4] Naturally, this is the case if f is strictly convex.

Then T is bounded on bounded sets with

$$\|Tx\| \leqslant E(b) := o(2D(b) + 2bC(b), b)$$

for $b \geqslant \|x\|, \|p_0\|$.

Proof. Note that $D_f(p_0, Tx) \leqslant D_f(p_0, x)$ as T is Bregman quasi-nonexpansive and thus

$$\begin{aligned} D_f(p_0, Tx) &\leqslant |f(p_0)| + |f(x)| + |\langle p_0 - x, \nabla f(x)\rangle| \\ &\leqslant 2D(b) + 2bC(b) \end{aligned}$$

from which the claim follows using the properties of o. □

Conceptually, the strong BSNE-moduli are related to the notion of "quantitative quasiness" as discussed in [194] and from such a strong BSNE-modulus, one can in particular derive a modulus $\omega' : (0, \infty)^2 \to (0, \infty)$ which satisfies

$$\forall \varepsilon, b > 0 \forall x, p \in X(\|p\|, \|x\| \leqslant b \wedge \|Tp - p\| < \omega'(\varepsilon, b) \to D_f(p, Tx) - D_f(p, x) < \varepsilon).$$

This is collected in the following lemma:

Lemma 9.3.6. *Let ξ be a modulus of uniform continuity on bounded sets for D_f in its second argument and let ρ be a modulus of consistency for f. Let E be a modulus for T being bounded on bounded sets and let ω be a strong BSNE-modulus for T.*

Then there exists an ω' such that

$$\begin{aligned} \forall \varepsilon, b > 0 \forall x, p \in X(&\|p\|, \|x\| \leqslant b \\ &\wedge \|Tp - p\| < \omega'(\varepsilon, b) \to D_f(p, Tx) - D_f(p, x) < \varepsilon). \end{aligned}$$

which can be moreover constructed as

$$\omega'(\varepsilon, b) := \omega(\rho(\xi(\varepsilon, \widehat{b}), \widehat{b}), b)$$

where $\widehat{b} = \max\{b, E(b)\}$.

Proof. If $D_f(p, Tx) - D_f(p, x) \leqslant 0$, then the claim holds trivially. So suppose $D_f(p, Tx) - D_f(p, x) > 0$. Then trivially $D_f(p, x) - D_f(p, Tx) < 0 < \omega'(\varepsilon, b)$ which implies $D_f(Tx, x) < \rho(\xi(\varepsilon, \widehat{b}), \widehat{b})$. This yields $\|Tx - x\| < \xi(\varepsilon, \widehat{b})$. Thus, we in particular have $D_f(p, Tx) - D_f(p, x) < \varepsilon$. □

In the following, we will call such an ω' a derived modulus of ω.

As mentioned above, any Bregman firmly nonexpansive map is Bregman strongly nonexpansive. From the proof of this fact, we can immediately extract a (strong) BSNE-modulus for any Bregman firmly nonexpansive map T. Crucial for this is the following equivalent characterization of Bregman firmly nonexpansive mappings:

Lemma 9.3.7 ([9]). *A map $T : X \to X$ is Bregman firmly nonexpansive if, and only if,*

$$D_f(Tx, Ty) + D_f(Ty, Tx) \leqslant D_f(Tx, y) + D_f(Ty, x) - D_f(Tx, x) - D_f(Ty, y).$$

for all $x, y \in X$.

Lemma 9.3.8. *Let T be a Bregman firmly nonexpansive map which is bounded on bounded sets with a modulus E and let ξ, ξ' be moduli that D_f is uniformly continuous on bounded sets in its right and left argument, respectively.*

Then T is uniformly Bregman strongly nonexpansive with a strong BSNE-modulus ω defined by

$$\omega(\varepsilon, b) = \min\{\xi(\varepsilon/4, \widehat{b}), \xi'(\varepsilon/4, \widehat{b}), \varepsilon/4\}$$

where $\widehat{b} = \max\{b, E(b)\}$.

Further, one can choose $\omega(\varepsilon, b) = \varepsilon$ as a BSNE-modulus for any Bregman firmly nonexpansive T.

Proof. For the strong modulus, let x, p be given. Using Lemma 9.3.7 with $y = p$, we get

$$\begin{aligned} D_f(Tx, Tp) + D_f(Tp, Tx) &\leqslant D_f(Tx, p) + D_f(Tp, x) - D_f(Tx, x) - D_f(Tp, p) \\ &\leqslant D_f(Tx, p) + D_f(Tp, x) - D_f(Tx, x). \end{aligned}$$

Rearranging yields

$$\begin{aligned} D_f(Tx, x) &\leqslant D_f(Tx, p) - D_f(Tx, Tp) + D_f(Tp, x) - D_f(Tp, Tx) \\ &\leqslant (D_f(Tx, p) - D_f(Tx, Tp)) + (D_f(Tp, x) - D_f(p, x)) \\ &\qquad + (D_f(p, x) - D_f(p, Tx)) + (D_f(p, Tx) - D_f(Tp, Tx)). \end{aligned}$$

Thus if $\|p\|, \|x\| \leqslant b$ and $\|Tp - p\| < \omega(\varepsilon, b)$ as well as $D_f(p, x) - D_f(p, Tx) < \omega(\varepsilon, b) \leqslant \varepsilon/4$, then we get $D_f(Tx, x) < \varepsilon$.

For the ordinary BSNE-modulus, note that if $p = Tp$, then Lemma 9.3.7 with $y = p$ even yields

$$D_f(p, Tx) \leqslant D_f(p, x) - D_f(Tx, x)$$

which is equivalent to

$$D_f(Tx, x) \leqslant D_f(p, x) - D_f(p, Tx)$$

which yields the given modulus. □

Compare this BSNE-modulus in particular to the modulus extracted in [99] for ordinary (meaning in the usual metric sense) strongly (quasi-)nonexpansive maps which even in the simple case of Hilbert spaces (where the notions of firmly nonexpansive and Bregman firmly nonexpansive for $f = \|\cdot\|^2/2$ coincide) is quadratic in ε. By taking a look at the above proof, this seems due to the fact even in the Hilbert case with the specific choice $f = \|\cdot\|^2/2$, the distance D_f fits closer to the notion of firmly nonexpansive maps and the quadratic increase comes from converting from D_f to the usual norm.

A concrete example for Bregman firmly nonexpansive mappings are the resolvents Res^f_A relative to f for a given monotone operator A in Banach spaces. For this, we first recall the notion of monotone operators.

Definition 9.3.9 ([28, 30])**.** Let $A : X \to 2^{X^*}$ be a set-valued operator. The operator A is called monotone if

$$\langle x - y, x^* - y^* \rangle \geqslant 0$$

for all $(x, x^*), (y, y^*) \in A$.

Further, A is called maximally monotone if its graph is not strictly contained in the graph of another monotone operator.

The f-resolvents of A are then defined using ∇f:[5]

Definition 9.3.10 ([9, 58])**.** Let $A : X \to 2^{X^*}$ be a set-valued operator. Given f, we define the resolvent of A relative to f as the operator $\mathrm{Res}^f_A : X \to 2^X$ with

$$\mathrm{Res}^f_A(x) := \left((\nabla f + A)^{-1} \circ \nabla f\right)(x).$$

The following properties are essential for the resolvent relative to f:

[5] The idea of considering the above notion in general Banach spaces is due to [9] (where it was introduced under the name of D-resolvents) but this notion of a resolvent relative to f was already considered by Eckstein in [58] in the context of finite-dimensional spaces.

Proposition 9.3.11 ([9]). *Let $f : X \to (-\infty, +\infty]$ be a function which is proper, convex, lower semicontinuous, Gâteaux differentiable and strictly convex on* $\mathrm{intdom} f$ *and let A be a monotone operator such that* $\mathrm{intdom} f \cap \mathrm{dom} A \neq \emptyset$. *Then following statements hold:*

1. $\mathrm{dom}\ \mathrm{Res}^f_A \subseteq \mathrm{intdom} f$ *and* $\mathrm{ran}\ \mathrm{Res}^f_A \subseteq \mathrm{intdom} f$,
2. Res^f_A *is single-valued on its domain,*
3. $F(\mathrm{Res}^f_A) = \mathrm{intdom} f \cap A^{-1}0$,
4. Res^f_A *is Bregman firmly nonexpansive on its domain.*

Further, the classical result for monotone operators in Hilbert spaces established by Minty [147] that maximal monotonicity is equivalent to the totality of the resolvents extends to these resolvents relative to f under suitable assumptions on f:

Proposition 9.3.12 ([15]). *Let X be reflexive.*[6] *Let A be monotone and assume that $f : X \to \mathbb{R}$ is Gâteaux differentiable, strictly convex and cofinite (i.e.* $\mathrm{dom}\ f^* = X^*$*). Then A is maximal monotone if and only if* $\mathrm{ran}(A + \nabla f) = X^*$.

As we will mostly consider a fixed operator A in the following, we introduce a more compact notation for resolvents with real parameters in such a case: given $\gamma > 0$, we simply write Res^f_γ for $\mathrm{Res}^f_{\gamma A}$.

Important for the study of resolvents are their corresponding Yosida approximates defined by

$$A^f_\gamma(x) = \frac{1}{\gamma}\left(\nabla f(x) - \nabla f \mathrm{Res}^f_\gamma(x)\right)$$

for a given $\gamma > 0$.

It follows essentially by the definitions of Res^f_γ and A^f_γ (see e.g. [177]) that

$$(\mathrm{Res}^f_\gamma x, A^f_\gamma x) \in A$$

for any $\gamma > 0$ and any $x \in \mathrm{dom}\ \mathrm{Res}^f_\gamma$.

By the above results, as any Res^f_γ is Bregman firmly nonexpansive, all such resolvents for a maximal monotone A have the same BSNE-modulus (and also the same

[6] Recall Remark 8.5.13 by which this is true in the standing assumptions of this chapter.

strong BSNE-modulus if they are bounded on bounded sets with a common modulus).

These resolvents relative to f also include Bregman projections (see [22]) as these can be considered to be special resolvents: If C is a non-empty, closed and convex subset, we may define the indicator function

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{if } x \notin C. \end{cases}$$

It is straightforward to see that this function is proper, lower-semicontinuous and convex. Therefore, the subgradient $\partial\iota_C$ (recall Chapter 8) is maximally monotone [180, 182]. The Bregman projection P^f_C is then defined as the resolvent $\mathrm{Res}^f_{\partial\iota_C}$ and in particular is Bregman firmly nonexpansive. Thus also here the above moduli apply.

In general, already for Bregman firmly nonexpansive mappings, it is not immediately clear which (if any) form of ordinary metric continuity such mappings inherit. However, if one assumes that ∇f is uniformly continuous on bounded subsets as well as uniformly strictly monotone, then at least every Bregman firmly nonexpansive map that is bounded on bounded sets (i.e., by Lemma 9.3.5, in particular any such map with a fixed point) is indeed uniformly continuous on bounded subsets.

Lemma 9.3.13. *Let T be Bregman firmly nonexpansive and assume that T is bounded on bounded sets with a modulus E. Assume that ∇f is uniformly continuous on bounded sets with a modulus $\omega^{\nabla f}$ and that it is uniformly strictly monotone with a modulus η, i.e.*

$$\forall \varepsilon, b > 0 \forall x, y \in X \left(\|x\|, \|y\| \leqslant b \land \langle x - y, \nabla f x - \nabla f y \rangle < \eta(\varepsilon, b) \to \|x - y\| < \varepsilon \right).$$

Then T is uniformly continuous on bounded sets with

$$\forall \varepsilon, b > 0 \forall x, y \in \overline{B}_b(0) \left(\|x - y\| < \omega^{\nabla f}(\eta(\varepsilon, E(b))/2E(b), b) \to \|Tx - Ty\| < \varepsilon \right).$$

Proof. Let x, y be given with $\|x\|, \|y\| \leqslant b$. As T is Bregman firmly nonexpansive, we get by definition that

$$\begin{aligned} \langle Tx - Ty, \nabla f Tx - \nabla f Ty \rangle &\leqslant \langle Tx - Ty, \nabla f x - \nabla f y \rangle \\ &\leqslant \|Tx - Ty\| \|\nabla f x - \nabla f y\| \\ &\leqslant 2E(b) \|\nabla f x - \nabla f y\|. \end{aligned}$$

In particular, if $\|x-y\| < \omega^{\nabla f}(\varepsilon/2E(b), b)$, we have

$$\langle Tx - Ty, \nabla f Tx - \nabla f Ty \rangle < \varepsilon$$

and thus, if $\|x-y\| < \omega^{\nabla f}(\eta(\varepsilon, E(b))/2E(b), b)$, we get $\|Tx - Ty\| < \varepsilon$. □

A crucial feature of strongly nonexpansive maps (in the usual sense) as compared to e.g. firmly nonexpansive maps is that they are closed under composition. A similar result holds for Bregman strongly nonexpansive maps as established in [143]. We now derive a quantitative variant that allows one to combine (strong) BSNE-moduli for the factors into a (strong) BSNE-modulus for the composition. This result is similar to the corresponding results for "ordinary" (quasi)-strongly nonexpansive maps given in [99] (see Theorem 2.10 and Theorem 4.6 therein).

However, before we move to this result on moduli for compositions, we first consider a quantitative treatment of the fact that fixed points of compositions of Bregman strongly nonexpansive operators are fixed points of the factors (see e.g. Proposition 3.4 in [143]). This result, however, crucially relies on the fact that $\widehat{F}(T) \subseteq F(T)$ and so here, we will have to rely on a quantitative treatment of this aspect. The inclusion $\widehat{F}(T) \subseteq F(T)$ concretely expresses the closure property

$$\forall x \in X, (x_n) \subseteq X \, (\|x_n - Tx_n\| \to 0 \text{ and } x_n \to x \text{ (weakly)} \to x = Tx)$$

of which the underlying logical methods used in this chapter suggest the following uniform quantitative version to be necessary in the analysis:

$$\begin{aligned}\forall \varepsilon, b > 0 \exists \kappa > 0 \forall x, y \in X (&\|x\|, \|y\| \leqslant b \\ &\wedge \|y - Ty\|, \|y - x\| < \kappa \to \|x - Tx\| < \varepsilon).\end{aligned}$$

We call a function $\kappa(\varepsilon, b)$ that provides witness for such a κ in terms of ε, b a modulus of uniform closedness for $F(T)$ as this kind of modulus is essentially just a concrete instantiation of the moduli of uniform closedness considered in an abstract context in [112]. In particular, we want to note that this modulus can from a logical perspective be recognized as a quantitative form of a weak extensionality principle for T, namely

$$\forall x, y \, (y = Ty \wedge x = y \to x = Tx)$$

which has previously received attention in proof mining, in particular due to the fact that there are meaningful classes of maps that posses such moduli of uniform closedness but fail to be uniformly continuous (as e.g. maps satisfying Suzuki's (E) condition

[67, 201], see also the discussions in [100, 112]).

In the presence of such a modulus, we can now turn to the following quantitative result (which is anyhow analogous to Proposition 4.15 from [99]):

Theorem 9.3.14. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for f and let P be a modulus for reverse consistency for f. Let $T_1, \dots, T_N : X \to X$ be Bregman strongly nonexpansive with a (not necessarily strong) BSNE-modulus ω w.r.t. some common fixed point $p \in \bigcap_{i=1}^N F(T_i)$. Let κ be a common modulus of uniform closedness of $F(T_1), \dots, F(T_N)$.*

Then for all $\varepsilon > 0$:

$$\|T_N \circ \cdots \circ T_1 x - x\| < P(\varphi(\varepsilon, b, N), b) \to \bigwedge_{i=1}^{N} \|x - T_i x\| < \varepsilon$$

whenever $b \geqslant \|x\|, \|p\|$ and $b \geqslant \|T_k \circ \cdots \circ T_1 x\|$ for $1 \leqslant k \leqslant N$ where $\varphi(\varepsilon, b, N) = \chi_b(N-1, \varepsilon)$ and, given b, $\chi_b : \mathbb{N} \times (0, \infty) \to (0, \infty)$ is defined by

$$\begin{cases} \chi_b(0, \varepsilon) := \min\{\rho(\kappa(\varepsilon, b), b), \rho(\varepsilon, b)\}, \\ \chi_b(n+1, \varepsilon) := \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(n, \varepsilon), b), \rho(\kappa(\varepsilon, b), b)\}, b), b), b), \\ \qquad\qquad \chi_b(n, \varepsilon), \theta(\chi_b(n, \varepsilon), b)\}. \end{cases}$$

In particular, if E is a common modulus for $T_1, \dots, T_N$ being bounded on bounded sets, then above claim holds for $b \geqslant \|x\|, \|p\|$ and $P(\varphi(\varepsilon, \widehat{b}, N), \widehat{b})$ with $\varphi(\varepsilon, b, N) = \chi_{\widehat{b}}(N-1, \varepsilon)$ and where $\widehat{b} = \max\{b, E(b), \dots, E^{(N)}(b)\}$.

Proof (compare also [99]). Note first that

$$\chi_b(n, \varepsilon) \leqslant \min\{\rho(\kappa(\varepsilon, b), b), \rho(\varepsilon, b)\}. \tag{0}$$

Also note that every T_k is in particular Bregman quasi-nonexpansive w.r.t. p. We show by induction on $1 \leqslant k \leqslant N$ that $D_f(T_k \circ \cdots \circ T_1 x, x) < \chi_b(k-1, \varepsilon)$ implies $\|x - T_i x\| < \varepsilon$ for $1 \leqslant i \leqslant k$. For $k = 1$, the statement trivially holds since $\chi_b(0, \varepsilon) \leqslant \rho(\varepsilon, b)$. So let $1 < k \leqslant N$ and assume that the claim holds for $k-1$ and that

$$\begin{aligned} D_f(T_k \circ \cdots \circ T_1 x, x) &< \chi_b(k-1, \varepsilon) \\ &= \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(k-2, \varepsilon), b), \rho(\kappa(\varepsilon, b), b)\}, b), b), b), \\ &\qquad \chi_b(k-2, \varepsilon), \theta(\chi_b(k-2, \varepsilon), b)\}. \end{aligned} \tag{1}$$

For $y = T_{k-1} \circ \cdots \circ T_1 x$, we have

$$\|x - T_k y\| < \xi(\omega(\min\{\theta(\chi_b(k-2,\varepsilon),b), \rho(\kappa(\varepsilon,b),b)\},b),b). \tag{2}$$

Hence by (2), the assumption on ξ and $p \in \bigcap_{i=1}^k F(T_i)$, we derive

$$\begin{aligned} D_f(p,y) - \omega(\min\{\theta(\chi_b(k-2,\varepsilon),b), \rho(\kappa(\varepsilon,b),b)\},b) \\ \leqslant D_f(p,x) - \omega(\min\{\theta(\chi_b(k-2,\varepsilon),b), \rho(\kappa(\varepsilon,b),b)\},b) \\ < D_f(p,T_k y) \end{aligned}$$

where we in particular used that $D_f(p,y) \leqslant D_f(p,x)$. Thus, since ω is a BSNE-modulus for T_k:

$$D_f(T_k y, y) < \min\{\theta(\chi_b(k-2,\varepsilon),b), \rho(\kappa(\varepsilon,b),b)\}. \tag{3}$$

By (1) and (3) together with the assumption on θ, we thus obtain

$$D_f(T_{k-1} \circ \cdots \circ T_1 x, x) = D_f(y,x) < \chi_b(k-2,\varepsilon) \tag{4}$$

from which we derive

$$\bigwedge_{i=1}^{k-1} \|x - T_i x\| < \varepsilon$$

using the induction hypothesis. From (0) and (4) together with the definition of ρ, we also get

$$\|x - T_{k-1} \circ \cdots \circ T_1 x\| < \kappa(\varepsilon,b)$$

and so by (3), we obtain $\|x - T_k x\| < \varepsilon$. □

We now turn to the following result on moduli for compositions of Bregman strongly nonexpansive maps (which is modeled after Theorem 2.10 and Theorem 4.6 from [99]):

Theorem 9.3.15. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for D_f and let P be a modulus of reverse consistency. Let $T_1, \ldots, T_n : X \to X$ be uniformly Bregman strongly nonexpansive maps with strong BSNE-moduli $\omega_1, \ldots, \omega_n$ and derived moduli $\omega'_1, \ldots, \omega'_n$ and assume that the T_i's have a common fixed point. Let κ be a common modulus of uniform closedness of $F(T_1), \ldots, F(T_N)$.*

Then $T = T_n \circ \cdots \circ T_1$ is uniformly Bregman strongly nonexpansive with modulus

$$\omega(\varepsilon,b) := \min\left\{\widehat{\omega}(\varepsilon,b)/2, P(\varphi(\min\{\widehat{\omega}'(\varepsilon,b), \widehat{\omega}(\varepsilon,b)\}, \widehat{b}, n), \widehat{b})\right\}$$

where

$$\widehat{\omega}(\varepsilon,b) := \min\left\{\omega_1(\rho(P(\varepsilon,\widehat{b})/n,\widehat{b}),\widehat{b}),\dots,\omega_n(\rho(P(\varepsilon,\widehat{b})/n,\widehat{b}),\widehat{b})\right\},$$

$$\widehat{\omega}'(\varepsilon,b) := \min\left\{\omega'_1(\widehat{\omega}(\varepsilon,b)/2(n-1),\widehat{b}),\dots,\omega'_n(\widehat{\omega}(\varepsilon,b)/2(n-1),\widehat{b})\right\},$$

and where φ is defined as in Theorem 9.3.14, where $\widehat{b} = \max\{b, E(b),\dots,E^{(n)}(b)\}$ for b satisfying $b \geqslant \|q\|$ for a common fixed point q of the T_i's and where E is a common modulus for $T_1,\dots,T_n$ being bounded on bounded sets.

If the ω_i are ordinary BSNE-moduli, then ω defined by

$$\omega(\varepsilon,b) := \min\left\{\omega_1(\rho(P(\varepsilon,\widehat{b})/n,\widehat{b}),\widehat{b}),\omega_2(\rho(P(\varepsilon,\widehat{b})/n,\widehat{b}),\widehat{b}),\dots,\omega_n(\rho(P(\varepsilon,\widehat{b})/n,\widehat{b}),\widehat{b})\right\}$$

is a BSNE-modulus for $T = T_n \circ \dots \circ T_1$ where $\widehat{b}$ is defined as before.

Proof (compare also [99]). Define

$$\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b) := \min\left\{\omega_1(\rho(\varepsilon_1,\widehat{b}),\widehat{b}),\dots,\omega_n(\rho(\varepsilon_n,\widehat{b}),\widehat{b})\right\},$$

and

$$\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b) := \min\Big\{\omega'_1(\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1),\widehat{b}), \dots,\omega'_n(\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1),\widehat{b})\Big\}$$

as well as

$$\omega(\varepsilon_1,\dots,\varepsilon_n,b) := \min\Big\{\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2, P(\varphi(\min\{\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b),\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)\},\widehat{b},n),\widehat{b})\Big\}.$$

Now, suppose

$$\|p-Tp\|, D_f(p,x) - D_f(p,Tx) < \omega(\varepsilon_1,\dots,\varepsilon_n,b),$$

for points x,p with $\|x\|,\|p\| \leqslant b$. Then Theorem 9.3.14 yields that

$$\|p-T_ip\| < \min\{\widehat{\omega}'(\varepsilon_1,\dots,\varepsilon_n,b),\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)\}.$$

Therefore, we get

$$\begin{aligned}
D_f(p,Tx) &= D_f(p,T_n\circ T_{n-1}\circ\dots\circ T_1x)\\
&\leqslant D_f(p,T_{n-1}\circ\dots\circ T_1x) + \widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)\\
&\leqslant \dots\\
&\leqslant D_f(p,T_1x) + (n-1)\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)\\
&\leqslant D_f(p,x) + n\widehat{\omega}(\varepsilon_1,\dots,\varepsilon_n,b)/2(n-1)
\end{aligned}$$

and therefore

$$
\begin{aligned}
D_f(p, T_{i-1} \circ \cdots \circ T_1 x) &- D_f(p, T_i \circ T_{i-1} \circ \cdots \circ T_1 x) \\
&\leqslant D_f(p,x) - D_f(p,Tx) \\
&\qquad + (n-1)\widehat{\omega}(\varepsilon_1, \ldots, \varepsilon_n, b)/2(n-1) \\
&< \widehat{\omega}(\varepsilon_1, \ldots, \varepsilon_n, b) \\
&\leqslant \omega_i(\rho(\varepsilon_i, \widehat{b}), \widehat{b})
\end{aligned}
$$

for any $i = 1, \ldots, n$. This, together with $\|p - T_i p\| < \omega_i(\rho(\varepsilon_i, \widehat{b}), \widehat{b})$, yields

$$D_f(T_i \circ T_{i-1} \circ \cdots \circ T_1 x, T_{i-1} \circ \cdots \circ T_1 x) < \rho(\varepsilon_i, \widehat{b})$$

as ω_i is a strong BSNE-modulus for T_i. In particular, we have

$$\|T_i \circ T_{i-1} \circ \cdots \circ T_1 x - T_{i-1} \circ \cdots \circ T_1 x\| < \varepsilon_i$$

so that we get $\|x - Tx\| < \varepsilon_1 + \cdots + \varepsilon_n$. Now, for $\varepsilon_i = P(\varepsilon, \widehat{b})/n$, we then get $\|x - Tx\| < P(\varepsilon, \widehat{b})$ so that $D_f(Tx, x) < \varepsilon$.

If the ω_i's are BSNE-moduli and if p is a real fixed point of T (and thus a common fixed point of the T_i's as $\widehat{F}(T) \subseteq \bigcap_{i=1}^n \widehat{F}(T_i)$, see [143]), then it is clear that the second term involving φ can be dropped. □

The last type of operation on Bregman strongly nonexpansive operators that we consider here is that of the block operator introduced in [143, 144]:

Definition 9.3.16 ([143, 144])**.** Let T_i, $i = 1, \ldots, N$, be finitely many operators and let $w_i \in [0,1]$, $i = 1, \ldots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Then the associated block operator is defined as

$$Tx = \nabla f^* \left(\sum_{i=1}^N w_i \nabla f T_i x \right).$$

In particular, as shown in [143, 144], such block operators, if composed of Bregman strongly nonexpansive maps, are again Bregman strongly nonexpansive. For a quantitative version of the said result, we consider the following lemmas.

At first, we note that a block operator is bounded on bounded sets if its summands are.

Lemma 9.3.17. *Let ∇f, ∇f^* be bounded on bounded sets with moduli C, F, respectively. Let T_i, $i = 1, \ldots, N$, be finitely many operators which are bounded on bounded sets with a common modulus E and let $w_i \in [0,1]$, $i = 1, \ldots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$.*

Then the associated block operator T is bounded on bounded sets with a modulus $E'(b) := F(C(E(b)))$.

Proof. For $\|x\| \leqslant b$, we clearly have

$$\left\| \sum_{i=1}^N w_i \nabla f T_i x \right\| \leqslant \sum_{i=1}^N w_i \|\nabla f T_i x\| \leqslant C(E(b))$$

and thus $\|Tx\| = \left\| \nabla f^* \sum_{i=1}^N w_i \nabla f T_i x \right\| \leqslant F(C(E(b)))$. □

As shown in [144], one has $F(T) \subseteq F(T_i)$ for a block operator T and a summand T_i. The following lemma gives a quantitative version of this, translating bounds for approximate fixed points.

Theorem 9.3.18. *Let ξ be a modulus of uniform continuity of D_f in its second argument. Let T_i, $i = 1, \ldots, N$, be finitely many Bregman strongly nonexpansive operators with a (not necessarily strong) BSNE-modulus ω and let $w_i \in [0,1]$, $i = 1, \ldots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Let T be the associated block operator. Assume that T and all T_i's are bounded on bounded sets with a common modulus E. Let p_0 be a common fixed point of all T_i's and let $b \geqslant \|p_0\|$.*

Then for any x with $\|x\| \leqslant b$ and any $k = 1, \ldots, N$:

$$w_k \geqslant w > 0 \wedge \|x - Tx\| < \xi\left(w\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right) \rightarrow \|x - T_k x\| < \varepsilon$$

where $\widehat{b} = \max\{b, E(b)\}$.

Proof. If

$$\|x - Tx\| < \xi\left(w\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right),$$

then we get

$$D_f(p_0, x) - D_f(p_0, Tx) < w\omega(\rho(\varepsilon, \widehat{b}), b).$$

Fix $k = 1, \ldots, N$. Then

$$\begin{aligned} D_f(p_0, Tx) &\leqslant w_k D_f(p_0, T_k x) + \sum_{i \neq k} w_i D_f(p_0, T_i x) \\ &\leqslant w_k D_f(p_0, T_k x) + \sum_{i \neq k} w_i D_f(p_0, x) \\ &\leqslant w_k D_f(p_0, T_k x) + (1 - w_k) D_f(p_0, x) \\ &\leqslant w_k (D_f(p_0, T_k x) - D_f(p_0, x)) + D_f(p_0, x) \end{aligned}$$

and thus in particular

$$w_k(D_f(p_0, x) - D_f(p_0, T_k x)) \leqslant D_f(p_0, x) - D_f(p_0, Tx) < w\omega(\rho(\varepsilon, \widehat{b}), b)$$

which implies $D_f(p_0, x) - D_f(p_0, T_k x) < \omega(\rho(\varepsilon, \widehat{b}), b)$. As ω is a BSNE-modulus for T_k, we get $D_f(T_k x, x) < \rho(\varepsilon, \widehat{b})$ which yields $\|x - T_k x\| < \varepsilon$. □

The following lemma now provides a map that translates strong BSNE-moduli for the summands into strong BSNE-moduli for the block operator and in that sense is a quantitative version of Proposition 14 in [144].

Theorem 9.3.19. *Let ξ be a modulus of uniform continuity of D_f in its second argument. Let $\omega^{\nabla f}$ be a modulus of uniform continuity of ∇f on bounded sets and C be a modulus witnessing that ∇f is bounded on bounded sets. Let T_i, $i = 1, \ldots, N$, be finitely many uniformly Bregman strongly nonexpansive operators with a common strong BSNE-modulus ω and derived modulus ω' and let $w_i \in [0, 1]$, $i = 1, \ldots, N$, be finitely many weights with $\sum_{i=1}^N w_i = 1$. Let T be the associated block operator. Assume that T and all T_i's are bounded on bounded sets with a common modulus E. Let $p_0 \in F(T)$ be a common fixed point of all T_i's and let $b \geqslant \|p_0\|$.*

Then T is uniformly Bregman strongly nonexpansive with a strong BSNE-modulus $\widehat{\omega}$ which can be defined by

$$\widehat{\omega}(\varepsilon, b) := \min\{w^2\omega(\varepsilon', b), \xi(w\omega(\rho(\min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}, \widehat{b}), b), \widehat{b})\}$$

where $\widehat{b} = \max\{b, E(b)\}$ and $\varepsilon' = \rho(\omega^{\nabla f}(\varepsilon/4\widehat{b}, \widehat{b}), \widehat{b})$ and $w = \min\{\varepsilon/8N\widehat{b}C(\widehat{b}), 1\}$.

If ω is only a (not necessarily strong) BSNE-modulus, then we can chose $\widehat{\omega}(\varepsilon, b) = w\omega(\varepsilon', b)$ as a BSNE-modulus for T.

Proof. Let x, p be given with $\|x\|, \|p\| \leqslant b$, $\|p - Tp\| < \widehat{\omega}(\varepsilon, b)$ as well as

$$D_f(p, x) - D_f(p, Tx) < \widehat{\omega}(\varepsilon, b).$$

Then in particular

$$\|p - Tp\| < \xi(w\omega(\rho(\min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}, \widehat{b}), b), \widehat{b})$$

and by Theorem 9.3.18, we have

$$\|p - T_k p\| < \min\{\omega(\varepsilon', b), \omega'(w\omega(\varepsilon', b), b)\}$$

for any k with $w_k \geqslant w$.

We further have

$$D_f(p, Tx) \leqslant \sum_{i=1}^{N} w_i D_f(p, T_i x)$$

and, therefore,

$$\sum_{i=1}^{N} w_i \left(D_f(p, x) - D_f(p, T_i x)\right) \leqslant D_f(p, x) - D_f(p, Tx) < \widehat{\omega}(\varepsilon, b)$$

which implies

$$\begin{aligned} w_k \left(D_f(p, x) - D_f(p, T_k x)\right) &< \widehat{\omega}(\varepsilon, b) + \sum_{i \neq k} w_i \left(D_f(p, T_i x) - D_f(p, x)\right) \\ &< \widehat{\omega}(\varepsilon, b) + (1 - w_k) w\omega(\varepsilon', b) \\ &\leqslant w^2 \omega(\varepsilon', b) + (1 - w) w\omega(\varepsilon', b) \\ &= w\omega(\varepsilon', b) \end{aligned}$$

and thus $D_f(p, x) - D_f(p, T_k x) < \omega(\varepsilon', b)$ for any k with $w_k \geqslant w$. As ω is a strong BSNE-modulus for T_k, this gives $D_f(T_k x, x) < \varepsilon'$ for any such k. Thus in particular $\|x - T_k x\| < \omega^{\nabla f}(\varepsilon/4\widehat{b}, \widehat{b})$ which yields

$$\|\nabla f x - \nabla f T_k x\| < \varepsilon/4\widehat{b}.$$

As we have

$$\nabla f T x - \nabla f x = \sum_{i=1}^{N} w_i (\nabla f T_i x - \nabla f x)$$

the above yields

$$\begin{aligned} \|\nabla f T x - \nabla f x\| &\leqslant \sum_{i=1}^{N} w_i \|\nabla f x - \nabla f T_i x\| \\ &= \sum_{i: w_i \geqslant w} w_i \|\nabla f x - \nabla f T_i x\| + \sum_{i: w_i < w} w_i \|\nabla f x - \nabla f T_i x\| \\ &< \sum_{i: w_i \geqslant w} w_i \varepsilon/4\widehat{b} + \sum_{i: w_i < w} w_i 2C(\widehat{b}) \\ &< \varepsilon/4\widehat{b} \sum_{i: w_i \geqslant w} w_i + \sum_{i: w_i < w} w 2C(\widehat{b}) \\ &\leqslant \varepsilon/2\widehat{b}. \end{aligned}$$

Now using the three point identity, we have

$$\begin{aligned} D_f(Tx,x) + D_f(x,Tx) &= \langle Tx - x, \nabla f(Tx) - \nabla f(x)\rangle \\ &\leqslant \|\nabla f Tx - \nabla f x\| \, \|Tx - x\| \\ &\leqslant \|\nabla f Tx - \nabla f x\| \, 2\widehat{b} \\ &< \varepsilon \end{aligned}$$

which in particular yields $D_f(Tx,x) < \varepsilon$.

It is immediate to see that if ω is just a (not necessarily strong) BSNE-modulus and p is a fixed point of T, that $w\omega(\varepsilon', b)$ suffices. □

9.4 Picard iterations

We now consider the first type of iteration of Bregman strongly nonexpansive mappings: as shown in [142], a Bregman strongly nonexpansive map $T : X \to X$ (in the context of some surrounding assumptions) is asymptotically regular, i.e. it holds that $\|x_n - Tx_n\| \to 0$ where $x_n := T^n x$ is the Picard iteration of T. In this section, we now derive quantitative rates for the above limit. In fact, we will actually first establish a corresponding quantitative result for a more general iteration involving a family of Bregman strongly nonexpansive operators of which the above Picard iteration will be a special case.

For this, we now fix the following moduli abstractly:[7]

(a) Let $\theta : (0,\infty)^2 \to (0,\infty)$ be a modulus of weak triangularity for D_f, i.e.

$$\begin{aligned} &\forall \varepsilon, b > 0 \forall x,y,z \in X(\|x\|, \|y\|, \|z\| \leqslant b \\ &\qquad \wedge D_f(x,y), D_f(z,y) < \theta(\varepsilon,b) \to D_f(x,z) < \varepsilon). \end{aligned}$$

(b) Let $\xi : (0,\infty)^2 \to (0,\infty)$ be a modulus for $D_f(x,y)$ being uniformly continuous in y on bounded sets, i.e.

$$\begin{aligned} &\forall \varepsilon, b > 0 \forall x,y_1,y_2 \in X(\|x\|, \|y_1\|, \|y_2\| \leqslant b \\ &\qquad \wedge \|y_1 - y_2\| < \xi(\varepsilon,b) \to |D_f(x,y_1) - D_f(x,y_2)| < \varepsilon). \end{aligned}$$

[7] Note the previous sections for how such moduli can be derived from respective moduli for the uniform continuity of ∇f, etc.

(c) Let $\rho : (0, \infty)^2 \to (0, \infty)$ be a modulus of consistency for f, i.e.

$$\forall \varepsilon, b > 0 \forall x, y \in X \left(\|x\|, \|y\| \leqslant b \wedge D_f(x, y) < \rho(\varepsilon, b) \to \|x - y\| < \varepsilon \right).$$

We then obtain the following result on rates of metastability and rates of convergence for iterations of families of Bregman strongly nonexpansive mappings. In that vein, the result provides a quantitative version of the respective asymptotic regularity results contained in [143, 144]. Further, the theorem is an adaptation of a similar result (see Theorem 4.7 in [99]) on strongly quasi-nonexpansive mappings in the ordinary sense.

Theorem 9.4.1. *Let $(T_n)_{n \in \mathbb{N}}$ be a sequence of functions $T_n : X \to X$ which are Bregman strongly nonexpansive w.r.t. some $p \in \bigcap_{n \in \mathbb{N}} F(T_n)$ with a common BSNE-modulus $\omega(\varepsilon, b)$. Let $x_0 \in X$, $x_{n+1} = T_n x_n$ and $b \geqslant D_f(p, x_0), \|p\|, \|x_n\|$.*

Then[8]

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \psi_{b,\omega}(\varepsilon, g) \forall k \in [n; n + g(n)] \left(D_f(x_{k+1}, x_k) < \varepsilon \right)$$

where

$$\psi_{b,\omega}(\varepsilon, g) := \widetilde{g}^{\left(\left\lceil \frac{b}{\omega(\varepsilon, b)} \right\rceil\right)}(0)$$

and $\widetilde{g}(n) := n + g(n) + 1$.

In particular, if o is a modulus of boundedness of D_f, then the above results holds true for $\psi_{\widehat{b},\omega}(\varepsilon, g)$ where $b \geqslant D_f(p, x_0), \|p\|$ and $\widehat{b} = \max\{o(b, b), b\}$.

Further, if $T_n = T$ for all $n \in \mathbb{N}$ and T, additionally, is also Bregman nonexpansive, then we even have

$$\forall \varepsilon > 0 \forall k \geqslant \left\lceil \frac{\widehat{b}}{\omega(\varepsilon, \widehat{b})} \right\rceil \left(D_f(x_{k+1}, x_k) < \varepsilon \right).$$

Proof. Since T_n in particular is Bregman quasi-nonexpansive w.r.t. p, we get that

$$0 \leqslant D_f(p, x_n) \leqslant D_f(p, x_0) \leqslant b.$$

Hence by Corollary 2.28 and Remark 2.29 from [96], we get that the function

$$\varphi(\varepsilon, g) := \widetilde{g}^{\left(\left\lceil \frac{b}{\varepsilon} \right\rceil\right)}(0)$$

satisfies

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \varphi(\varepsilon, g) \forall i, j \in [n; n + g(n) + 1] \left(|D_f(p, x_i) - D_f(p, x_j)| < \varepsilon \right)$$

[8] Here, and in the following, we write $[n; m] = [n, m] \cap \mathbb{N}$.

and so, in particular, we have

$$\forall\varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \varphi(\varepsilon, g) \forall k \in [n; n + g(n)] \left(|D_f(p, x_k) - D_f(p, T_k x_k)| < \varepsilon \right).$$

Hence for $\omega(\varepsilon, b)$ in place of ε in the above and using the fact that T_k is Bregman strongly nonexpansive with modulus ω, we get that

$$\forall\varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \varphi(\omega(\varepsilon, b), g) \forall k \in [n; n + g(n)] \left(D_f(T_k x_k, x_k) < \varepsilon \right)$$

which proves the first claim. For $g(n) = 0$ for all n, we thus in particular have

$$\forall\varepsilon > 0 \exists n \leqslant \widetilde{g}^{(\lceil \frac{b}{\omega(\varepsilon,b)} \rceil)}(0) = \left\lceil \frac{b}{\omega(\varepsilon, b)} \right\rceil \left(D_f(x_{n+1}, x_n) < \varepsilon \right).$$

If now $T_k = T$ for all k and T is additionally Bregman nonexpansive, then

$$D_f(x_{k+1}, x_k) = D_f(T^{k+1}x, T^k x) \leqslant D_f(T^{n+1}x, T^n x) = D_f(x_{n+1}, x_n)$$

for all $k \geqslant n$ and so the second claim follows. □

From this, we get the following corollary to derive convergence of the norm distance:

Corollary 9.4.2. In addition to the assumptions in Theorem 9.4.1, let ρ be a modulus of consistency for f. Then

$$\forall\varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \psi_{b,\omega}(\rho(\varepsilon, d), g) \forall k \in [n; n + g(n)] \left(\|x_k - x_{k+1}\| < \varepsilon \right).$$

If again $T_k = T$ for all k and T is additionally Bregman nonexpansive *or* nonexpansive (w.r.t. $\|\cdot\|$), then

$$\forall\varepsilon > 0 \forall k \geqslant \left\lceil \frac{b}{\omega(\rho(\varepsilon, d), b)} \right\rceil \left(\|x_k - x_{k+1}\| < \varepsilon \right).$$

The main application of this Picard process now follows if the iterated map is a composition. Together with Theorem 9.3.15, we can then obtain the following result giving that the Picard iteration $x_{n+1} = Tx_n$ of a composition $T = T_k \circ \cdots \circ T_1$ is asymptotically regular w.r.t. each T_j (which in particular provides a quantitative perspective on the method of cyclic Bregman projections [175]):

Theorem 9.4.3. *Let ξ be a modulus of uniform continuity on bounded subsets for D_f in its second argument. Let θ be a modulus of weak triangularity for D_f. Let ρ be a modulus of consistency for f and let P be a modulus for reverse consistency for f. Let o be a modulus of boundedness of D_f. Let ∇f and f be bounded on bounded* *sets with moduli C, D. Let $T_1, \ldots, T_k : X \to X$ be Bregman strongly nonexpansive w.r.t. some $p \in F(T_1) \cap \cdots \cap F(T_k)$ with a (not necessarily strong) BSNE-modulus ω. Let κ be a common modulus of uniform closedness of $F(T_1), \ldots, F(T_N)$. Define $T = T_k \circ \cdots \circ T_1$ as well as $x_n = T^n x_0$ for some $x_0 \in X$. Let $b \geqslant D_f(p, x_0), \|p\|$ and define $\widetilde{b} = \max\{o(b,b), b\}$ as well as $\widehat{b} = \max\{\widetilde{b}, E(\widetilde{b}), \ldots, E^{(k)}(\widetilde{b})\}$ for $E(b) := o(2D(b) + 2bC(b), b)$.*

Then

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Phi(\varepsilon, g) \forall i \in [n; n+g(n)] \forall j \in [1;k] \left(\|T_j x_i - x_i\| < \varepsilon\right)$$

where Φ is defined by

$$\Phi(\varepsilon, g) := \widetilde{g}^{\left(\left\lceil \frac{\widehat{b}}{\widehat{\omega}(P(\varphi(\varepsilon,\widehat{b},k),\widehat{b}),\widehat{b})} \right\rceil\right)}(0)$$

where $\varphi(\varepsilon, b, k) = \chi_b(k-1, \varepsilon)$ with χ defined by

$$\begin{cases} \chi_b(0,\varepsilon) := \min\{\rho(\kappa(\varepsilon,b),b), \rho(\varepsilon,b)\}, \\ \chi_b(n+1,\varepsilon) := \min\{\rho(\xi(\omega(\min\{\theta(\chi_b(n,\varepsilon),b), \rho(\kappa(\varepsilon,b),b)\},b),b),b), \\ \qquad\qquad \chi_b(n,\varepsilon), \theta(\chi_b(n,\varepsilon),b)\}. \end{cases}$$

and where

$$\widehat{\omega}(\varepsilon, b) := \omega(\rho(P(\varepsilon,\widehat{b})/k, \widehat{b}), \widehat{b}).$$

Proof. The theorem is a straightforward combination of Corollary 9.4.2, Theorem 9.3.15, Theorem 9.3.14 and Lemma 9.3.5. □

Corollary 9.4.4. Let Ω_j, $j = 1, \ldots, k$, be non-empty, closed and convex sets with Bregman projections $P^f_{\Omega_j}$ and assume in addition to the assumptions in Theorem 9.4.3 that ∇f is uniformly continuous on bounded sets with a modulus $\omega^{\nabla f}$ and that it is uniformly strictly monotone with a modulus η, i.e.

$$\forall \varepsilon, b > 0 \forall x, y \in X \left(\|x\|, \|y\| \leqslant b \wedge \langle x - y, \nabla f x - \nabla f y\rangle < \eta(\varepsilon, b) \to \|x - y\| < \varepsilon\right).$$

Then for $T = P^f_{\Omega_k} \circ \cdots \circ P^f_{\Omega_1}$ and $x_n = T^n x_0$ for some $x_0 \in X$, we have

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \Phi(\varepsilon, g) \forall i \in [n; n+g(n)] \forall j \in [1;k] \left(\left\|P^f_{\Omega_j} x_i - x_i\right\| < \varepsilon\right)$$

with Φ defined by

$$\Phi(\varepsilon, g) := \widetilde{g}^{\left(\left\lceil \frac{\widehat{b}}{P(\varphi(\varepsilon,\widehat{b},k),\widehat{b})} \right\rceil\right)}(0)$$

with φ and χ defined as in Theorem 9.4.3, now using

$$\kappa(\varepsilon, b) = \min\{\varepsilon/3, \omega^{\nabla f}(\eta(\varepsilon/3, E(b))/2E(b), b)\}.$$

Proof. The corollary immediately follows from the above Theorem 9.4.3 where, for the particular case of Bregman projections, one additionally invokes Lemma 9.3.13 as well as Lemma 9.3.8 (by which we can use $\omega(\varepsilon, b) = \varepsilon$ as the common BSNE-modulus). □

The following proposition now provides an analogous result in the case that x_{n+1} is not exactly given by $T_n x_n$ but actually is allowed to differ from that point up to some summable error (compare this now to Theorem 4.9 from [99]). For that, we use the following result from [99]:

Lemma 9.4.5 (Lemma 4.8, [99]). *Let (a_n), (δ_n) be sequences of nonnegative reals with*

$$a_{n+1} \leqslant a_n + \delta_n,$$

where $\sum \delta_n < \infty$. Let $A, D \in \mathbb{N}$ with $A \geqslant a_0$ and $D \geqslant \sum \delta_n$. Define

$$\widetilde{\varphi}_{A,D}(\varepsilon, g) := \widetilde{g}^{(K)}(0), \text{ where } K = \left\lceil \frac{4(A+5D)}{\varepsilon} \right\rceil \text{ and } \widetilde{g}(n) := n + g(n).$$

Then $\widetilde{\varphi}_{A,D}$ is a rate of metastability for (a_n).

Proposition 9.4.6. *Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of functions $T_n : X \to X$ which are Bregman strongly nonexpansive w.r.t. some $p \in \bigcap_{n\in\mathbb{N}} F(T_n)$ with a common BSNE-modulus $\omega(\varepsilon, b)$. Let ξ be a modulus of uniform continuity of $D_f(p, u)$ in the argument u. Let $(x_n) \subseteq X$ be such that $\|x_{n+1} - T_n x_n\| < \xi(\delta_n, b)$ where $b \geqslant \|p\|, \|x_k\|, \|T_k x_k\|, D_f(p, x_0)$ for all k and where $(\delta_n) \subseteq [0, \infty)$ with $\sum \delta_n \leqslant D$. Let α be a rate of convergence for $\delta_n \to 0$, i.e.*

$$\forall \varepsilon > 0 \forall n \geqslant \alpha(\varepsilon) \left(\delta_n < \varepsilon\right).$$

Then

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \widetilde{\psi}_{b,\omega}(\varepsilon, g) \forall k \in [n; n + g(n)] \left(D_f(T_k x_k, x_k) < \varepsilon\right)$$

with

$$\widetilde{\psi}_{b,\omega}(\varepsilon, g) := \widetilde{\varphi}_{b,D}(\omega(\varepsilon, b)/2, g_{\alpha(\omega(\varepsilon,b)/2)} + 1)) + \alpha(\omega(\varepsilon, b)/2)$$

where $g_l(n) := g(n + l) + l$ and

$$\widetilde{\varphi}_{b,D}(\varepsilon, g) := \widetilde{g}^{(K)}(0) \text{ with } K = \left\lceil \frac{4(b+5D)}{\varepsilon} \right\rceil \text{ and } \widetilde{g}(n) := n + g(n).$$

In particular, if o is a modulus of boundedness of D_f, then the above results holds true for $\widetilde{\psi}_{\widehat{b},\omega}(\varepsilon, g)$ where $b \geqslant D_f(p, x_0), \|p\|$ and $\widehat{b} = \max\{o(b + D, b), b\}$.

Proof. Using the definition of ξ, we get for all $n \in \mathbb{N}$:

$$0 \leqslant D_f(p, x_{n+1}) \leqslant D_f(p, T_n x_n) + \delta_n \leqslant D_f(p, x_n) + \delta_n.$$

Hence by Lemma 9.4.5 applied to $a_n := D_f(p, x_n)$ (note that $b \geqslant a_0$), we get that

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \widetilde{\varphi}_{b,D}(\varepsilon, g+1) \forall i, j \in [n; n + g(n) + 1] \\ (|D_f(p, x_i) - D_f(p, x_j)| < \varepsilon)$$

and so, in particular, we have

$$\forall \varepsilon > 0 \forall g : \mathbb{N} \to \mathbb{N} \exists n \leqslant \widetilde{\varphi}_{b,D}(\varepsilon, g+1) \forall k \in [n; n + g(n)] \\ (|D_f(p, x_k) - D_f(p, x_{k+1})| < \varepsilon) .$$

Applied to $g_{\alpha(\varepsilon)}$ for a given ε, this yields

$$\exists n \leqslant \widetilde{\varphi}_{b,D}(\varepsilon, g_{\alpha(\varepsilon)} + 1) \forall k \in [n; n + g(n + \alpha(\varepsilon)) + \alpha(\varepsilon)] \, (|D_f(p, x_k) - D_f(p, x_{k+1})| < \varepsilon)$$

and so (by considering $n + \alpha(\varepsilon)$ instead of n), we get

$$\exists n \in [\alpha(\varepsilon); \widetilde{\varphi}_{b,D}(\varepsilon, g_{\alpha(\varepsilon)} + 1) + \alpha(\varepsilon)] \forall k \in [n; n + g(n)] \, (|D_f(p, x_k) - D_f(p, x_{k+1})| < \varepsilon) .$$

In turn, this then yields that

$$\exists n \leqslant \widetilde{\varphi}_{b,D}(\varepsilon, g_{\alpha(\varepsilon)} + 1) + \alpha(\varepsilon) \forall k \in [n; n + g(n)] \, (|D_f(p, x_k) - D_f(p, T_k x_k)| < 2\varepsilon)$$

since for $k \geqslant n \geqslant \alpha(\varepsilon)$, we have

$$\begin{aligned} |D_f(p, x_k) - D_f(p, T_k x_k)| & \\ &\leqslant |D_f(p, x_k) - D_f(p, x_{k+1})| + |D_f(p, x_{k+1}) - D_f(p, T_k x_k)| \\ &< \varepsilon + \delta_k \\ &\leqslant 2\varepsilon. \end{aligned}$$

Hence we lastly get that

$$\exists n \leqslant \big(\widetilde{\varphi}_{b,D}(\omega(\varepsilon, b)/2, g_{\alpha(\omega(\varepsilon,b)/2)} + 1)) + \alpha(\omega(\varepsilon, b)/2)\big) \, \forall k \in [n; n + g(n)] \\ (D_f(T_k x_k, x_k) < \varepsilon) .$$

□

9.5 A rate of metastability for a Halpern-type iteration of a family of maps

To obtain a strong convergence result, in [199], the authors defined a suitable Halpern-type iteration of a given Bregman strongly nonexpansive mapping. Concretely, the following result was established:

Theorem 9.5.1 ([199]). *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be a Bregman strongly nonexpansive mapping such that $F(T) = \widehat{F}(T) \neq \varnothing$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f u + (1 - \alpha_n)\nabla f T x_n\right)$$

where $(\alpha_n) \subseteq (0, 1)$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{F(T)}(u)$.

The aim of this section is to provide a quantitative analysis of this result as well as its extension to a family of mappings (T_n) as considered in [199], i.e. given u and x_0, we will consider the sequence

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f u + (1 - \alpha_n)\nabla f T_n x_n). \tag{$*$}$$

The proof of convergence for Theorem 9.5.1 as well as its extension to families of maps relies on a Lemma by Xu [209] as well as a subsequence construction due to Maingé [141], both of which have been treated quantitatively before in [124] as well as [104], respectively[9], and we present the quantitative versions of these crucial lemmas below.

Lemma 9.5.2 ([104], essentially [124]). *Let $b > 0$ and $(a_n) \subseteq [0, b]$ with*

$$a_{n+1} \leqslant (1 - \alpha_n)a_n + \alpha_n\beta_n + \gamma_n$$

for all n where $(\alpha_n) \subseteq (0, 1]$ with $\sum_{n=0}^{\infty} \alpha_n = +\infty$ (i.e. $\prod_{n=m}^{\infty}(1-\alpha_n) = 0$ for all $m \in \mathbb{N}$) and $(\beta_n) \subseteq \mathbb{R}$ as well as $(\gamma_n) \subseteq [0, \infty)$. Let $S : (0, \infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in m such that

$$\forall m \in \mathbb{N}, \varepsilon > 0 \left(\prod_{k=m}^{S(\varepsilon,m)} (1 - \alpha_k) \leqslant \varepsilon \right).$$

[9] The quantitative version of Xu's lemma presented in [124] works with slightly stronger assumptions than that presented in [104].

For $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, define

$$\widehat{g}(n) := g^M\left(n + S(\varepsilon/4b, n) + 1\right) + S(\varepsilon/4b, n).$$

Suppose that N satisfies

$$\exists m \leqslant N \forall i \in [m; m + \widehat{g}(m)]\left(\beta_i \leqslant \varepsilon/4\right).$$

Then for

$$\Phi(\varepsilon, S, N, b) := N + S(\varepsilon/4b, N) + 1,$$

we get that

$$\sum_{i=0}^{\Phi(\varepsilon,S,N,b)+g^M(\Phi(\varepsilon,S,N,b))} \gamma_i \leqslant \varepsilon/2 \to \exists n \leqslant \Phi(\varepsilon, S, N, b) \forall i \in [n; n + g(n)]\left(a_i \leqslant \varepsilon\right).$$

Lemma 9.5.3 ([104])**.** *Let $b > 0$ and $(a_n) \subseteq [0, b]$.*

1. *Let $\tau : \mathbb{N} \to \mathbb{N}$ be such that*

$$\forall n, k \in \mathbb{N}\left(k \leqslant n \wedge a_k < a_{k+1} \to k \leqslant \tau(n)\right). \tag{+}$$

For $K \in \mathbb{N}$, $g : \mathbb{N} \to \mathbb{N}$, $\varepsilon > 0$ and $\widetilde{g}(n) := n + g(n)$, define

$$\Psi(\varepsilon, g, K, b) := \widetilde{g}^{(\lceil b/\varepsilon \rceil)}(K).$$

Then

$$\begin{aligned}&\tau(\Psi(\varepsilon, g, K, b)) < K \\ &\qquad \to \exists n \leqslant \Psi(\varepsilon, g, K, b)\left(n \geqslant K \wedge \forall i, j \in [n; n + g(n)]\left(|a_i - a_j| \leqslant \varepsilon\right)\right).\end{aligned}$$

2. *Let $n_0 \in \mathbb{N}$ be such that $\exists n \leqslant n_0\left(a_n < a_{n+1}\right)$. Define*

$$\tau(n) := \max\{k \leqslant \max\{n_0, n\} \mid a_k < a_{k+1}\}.$$

Then τ is well-defined and satisfies (+). Moreover,

   (a) *$\forall n \in \mathbb{N}\left(a_{\tau(n)} \leqslant a_{\tau(n)+1}\right)$,*

   (b) *$\forall n \in \mathbb{N}\left(\tau(n) \leqslant \tau(n+1)\right)$,*

   (c) *$\forall n \geqslant n_0\left(a_n \leqslant a_{\tau(n)+1}\right)$.*

Before we move to quantitative results on the iteration considered above, we are first concerned with providing a quantitative account for Bregman projections onto fixed point sets of Bregman strongly nonexpansive maps. For this, the following lemma initially provides a quantitative version of the convexity of $F(T)$ as (essentially) shown in [178].

Lemma 9.5.4. *Let ρ be a modulus of consistency for f. Let T be uniformly Bregman strongly nonexpansive with strong BSNE-modulus ω and derived modulus ω'. Let T be bounded on bounded sets with a modulus E. Let $\varepsilon, b > 0$ be given and let x, y be such that $\|x\|, \|y\| \leqslant b$ and let $z = tx + (1-t)y$ for some $t \in [0,1]$.*

If

$$\|Tx - x\|, \|Ty - y\| < \omega'(\rho(\varepsilon, \max\{b, E(b)\}), b),$$

then we have

$$\|Tz - z\| < \varepsilon.$$

Proof. Note that $\|z\| \leqslant t\|x\| + (1-t)\|y\| \leqslant b$. As in [178], we get

$$D_f(z, Tz) = f(z) + tD_f(x, Tz) + (1-t)D_f(y, Tz) - tf(x) - (1-t)f(y).$$

Using ω', we get

$$D_f(x, Tz) - D_f(x, z), D_f(y, Tz) - D_f(y, z) < \rho(\varepsilon, \max\{b, E(b)\})$$

and thus, using the above and the definition of D_f, we get

$$\begin{aligned} D_f(z, Tz) &< f(z) + tD_f(x, z) + (1-t)D_f(y, z) - tf(x) - (1-t)f(y) \\ &\quad + \rho(\varepsilon, \max\{b, E(b)\}) \\ &= \rho(\varepsilon, \max\{b, E(b)\}). \end{aligned}$$

As $\|Tz\| \leqslant E(b)$, we get $\|z - Tz\| < \varepsilon$. □

Now, the following lemma provides a quantitative result on the existence of approximative projections onto fixed point sets of Bregman strongly nonexpansive maps. While the first part is concerned with the definition of said projections in terms of an infimum over Bregman distances, the second part is concerned with the characterization of Bregman projections in terms of a generalized type of variational inequality provided in [36] by which for a non-empty, closed and convex subset C and for a Gâteaux differentiable and totally convex function $f : X \to \mathbb{R}$, it holds that $z = P^f_C(x)$ if, and only if $z \in C$

$$\langle y - z, \nabla fx - \nabla fz \rangle \leqslant 0 \text{ for all } y \in C.$$

Note for both results that for a Bregman quasi-nonexpansive map T, the set of fixed points $F(T)$ is closed and convex (see e.g. [178][10]) and so $P^f_{F(T)}$ is defined for such a map whenever $F(T) \neq \varnothing$.

Lemma 9.5.5. *Let ρ be a modulus of consistency for f. Let T be uniformly Bregman strongly nonexpansive with strong BSNE-modulus ω and derived modulus ω'. Let T be bounded on bounded sets with a modulus E. Let $p_0 \in X$ be a fixed point of T with $D_f(p_0, u), \|p_0\| \leqslant b$.*

1. *For any $\varepsilon > 0$ and $\psi : (0, \infty) \to (0, \infty)$, let*

$$\varphi(\varepsilon, \psi) = \min\{\psi^{(r)}(1) \mid r \leqslant \lceil (b+1)/\varepsilon \rceil\}.$$

   Then there exists a $p \in X$ and a $\delta \geqslant \varphi(\varepsilon, \psi)$ with $\|p\| \leqslant b$ and $\|Tp - p\| < \psi(\delta)$ and

$$\forall q \in X \left(\|q\| \leqslant b \wedge \|Tq - q\| < \delta \to D_f(p, u) < D_f(q, u) + \varepsilon\right).$$

2. *Let further Δ be a modulus witnessing that $D_f(\cdot, u)$ is uniformly Fréchet differentiable on bounded subsets with derivative $x \mapsto \nabla f x - \nabla f u$, i.e. for any $b, \varepsilon > 0$ and any $x \in \overline{B}_b(0)$, $y \in X$:*

$$0 < \|y\| < \Delta(\varepsilon, b) \to \frac{|D_f(x+y, u) - D_f(x, u) - \langle y, \nabla f x - \nabla f u\rangle|}{\|y\|} < \varepsilon.$$

   For any $\varepsilon > 0$ and $\psi : (0, \infty) \to (0, \infty)$, let

$$\varphi'(\varepsilon, \psi) = \min\{\omega'(\rho({\psi'}^{(r)}(1), \max\{b, E(b)\}), b) \mid r \leqslant \lceil (b+1)/\varepsilon' \rceil\}$$

   with $\varepsilon' = \frac{\varepsilon}{2} \min\left\{\frac{\Delta(\varepsilon/4b, b)}{4b}, 1/2\right\}$ and with

$$\psi'(\delta) = \min\{\psi(\omega'(\rho(\delta, \max\{b, E(b)\}), b)), \omega'(\rho(\delta, \max\{b, E(b)\}), b)\}.$$

   Then there exists a $p \in X$ with $\|p\| \leqslant b$ and a $\delta' \geqslant \varphi'(\varepsilon, \psi)$ such that $\|Tp - p\| < \psi(\delta')$

$$\forall q \in X \left(\|q\| \leqslant b \text{ and } \|Tq - q\| < \delta' \to \langle q - p, \nabla f(u) - \nabla f(p)\rangle < \varepsilon\right).$$

[10]Note here, as well as already in the context of Lemma 9.5.4, that the results from [178], while phrased for Bregman firmly nonexpansive maps, clearly already hold for Bregman quasi-nonexpansive maps.

Proof. 1. Assume the contrary, i.e. that there are ε and ψ such that for any $p \in X$ and any $\delta \geqslant \varphi(\varepsilon, \psi)$ with $\|p\| \leqslant b$ and $\|Tp - p\| < \psi(\delta)$:

$$\exists q \in X \left(\|q\| \leqslant b \wedge \|Tq - q\| < \delta \wedge D_f(p, u) - \varepsilon \geqslant D_f(q, u)\right).$$

Let $r = \lceil (b+1)/\varepsilon \rceil$ and pick $q_0 = p_0$. Then clearly $\|q_0\| \leqslant b$ and

$$\|Tq_0 - q_0\| < \psi^{(r+1)}(1) = \psi(\psi^{(r)}(1)).$$

By definition, we have $\psi^{(r)}(1) \geqslant \varphi(\varepsilon, \psi)$ so that there exits a q_1 with $\|q_1\| \leqslant b$ and $\|Tq_1 - q_1\| < \psi^{(r)}(1)$ as well as

$$D_f(q_0, u) - \varepsilon \geqslant D_f(q_1, u).$$

Iterating this up to r yields a q_r such that

$$0 > D_f(q_0, u) - (b+1) \geqslant D_f(q_0, u) - \lceil (b+1)/\varepsilon \rceil \varepsilon = D_f(q_0, u) - r\varepsilon \geqslant D_f(q_r, u)$$

which is a contradiction.

2. Using (the proof of) (1), let $p \in X$ and $\delta = \psi'^{(i)}(1)$ for $i \leqslant \lceil (b+1)/\varepsilon' \rceil$ be such that $\|p\| \leqslant b$, $\|Tp - p\| < \psi'(\delta)$ and

$$\forall q \in X \left(\|q\| \leqslant b \wedge \|Tq - q\| < \delta \rightarrow D_f(p, u) < D_f(q, u) + \varepsilon'\right).$$

Let $\delta' = \omega'(\rho(\delta, \max\{b, E(b)\}), b)$. Then at first

$$\|Tp - p\| < \psi'(\delta) = \min\{\psi(\delta'), \delta'\} \leqslant \psi(\delta').$$

Now let q be such that $\|q\| \leqslant b$ and $\|Tq - q\| < \delta'$. If $q = p$, the claim is trivial. So suppose $q \neq p$. Then we can now reason along the lines of [34]: write $p(\alpha)$ for $p + \alpha(q - p)$. Using Lemma 9.5.4, as $\|Tp - p\| < \delta'$, we have

$$\|Tp(\alpha) - p(\alpha)\| < \delta.$$

Therefore, for any $\alpha \in [0, 1]$:

$$D_f(p, u) < D_f(p(\alpha), u) + \varepsilon'$$

Now, using the fact that D_f is convex and differentiable in its left argument with

$$[D_f(\cdot, x)]'(y) = \nabla f(y) - \nabla f(x),$$

we get

$$\frac{|D_f(p(\alpha),u) - D_f(p,u) - \langle \alpha(q-p), \nabla fp - \nabla fu\rangle|}{\|\alpha(q-p)\|} < \varepsilon/4b$$

if $\|\alpha(q-p)\| < \Delta(\varepsilon/4b, b)$, i.e. in particular if

$$\alpha < \frac{\Delta(\varepsilon/4b, b)}{2b}.$$

Thus in particular

$$\begin{aligned}\frac{\langle q-p, \nabla fu - \nabla fp\rangle}{\|q-p\|} &= \frac{-\langle \alpha(q-p), \nabla fp - \nabla fu\rangle}{\|\alpha(q-p)\|}\\ &< \frac{D_f(p,u) - D_f(p(\alpha),u)}{\|\alpha(q-p)\|} + \varepsilon/4b\end{aligned}$$

which implies

$$\langle q-p, \nabla fu - \nabla fp\rangle < \frac{D_f(p,u) - D_f(p(\alpha),u)}{\alpha} + \varepsilon/4b\,\|q-p\| \leqslant \frac{\varepsilon'}{\alpha} + \varepsilon/2$$

for any $\alpha < \min\left\{\frac{\Delta(\varepsilon/4b,b)}{2b}, 1\right\}$. In particular, for $\alpha = \min\left\{\frac{\Delta(\varepsilon/4b,b)}{4b}, 1/2\right\}$, we get

$$\langle q-p, \nabla fu - \nabla fp\rangle < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

Remark 9.5.6. Such a modulus Δ witnessing that $D_f(\cdot,u)$ is uniformly Fréchet differentiable on bounded subsets with derivative $x \mapsto \nabla fx - \nabla fu$ can be computed from $\omega^{\nabla f}$: we have that

$$\|[D_f(\cdot,u)]'(x) - [D_f(\cdot,u)]'(y)\| = \|\nabla fx - \nabla fy\|$$

so that $\omega^{\nabla f}$ is a modulus for $[D_f(\cdot,u)]'$ being uniformly continuous on bounded subsets. Therefore, we can apply Lemma 9.2.2, (1) to derive that $\Delta(\varepsilon,b) = \min\{\omega^{\nabla f}(\varepsilon,b+1),1\}$ is a suitable such modulus.

For the rest of this section, we are now concerned with quantitative results on the extension of the iteration from Theorem 9.5.1 to families of mappings discussed before. For the following quantitative results, we again fix some moduli abstractly:

(a) Let (T_n) be a family of uniformly Bregman strongly nonexpansive maps with a common strong BSNE-modulus ω and a common derived modulus ω', i.e.

$$\begin{aligned}&\forall \varepsilon, b > 0 \forall x,p \in X(\|p\|, \|x\| \leqslant b \land \|T_np - p\| < \omega(\varepsilon,b)\\ &\qquad \land D_f(p,x) - D_f(p,T_nx) < \omega(\varepsilon,b) \to D_f(T_nx,x) < \varepsilon)\end{aligned}$$

as well as

$$\forall \varepsilon > 0, b > 0, x, p \in X(\|p\|, \|x\| \leqslant b \\ \wedge \|T_n p - p\| < \omega'(\varepsilon, b) \to D_f(p, T_n x) - D_f(p, x) < \varepsilon)$$

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

(b) Let $(\alpha_n) \subseteq (0, 1]$ converge to zero with a rate $\sigma : (0, \infty) \to \mathbb{N}$, i.e.

$$\forall n \geqslant \sigma(\varepsilon) \left(\alpha_n < \varepsilon\right).$$

(c) Let f be sequentially consistent with a modulus of consistency ρ.

(d) Let $b \in \mathbb{N}^*$ be given and let x_n be defined by $(*)$ such that

$$b \geqslant \|x_n\|, \|T_n x_n\|, \|\nabla f(T_n x_n)\|, \|\nabla f(x_n)\|, \|u\|, \|\nabla f(u)\|, \\ \|p_0\|, \|\nabla f(p_0)\|, D_f(p_0, x_n), D_f(p_0, T_n x_n), D_f(p_0, u)$$

for all $n \in \mathbb{N}$ where p_0 is some given element of $F(T)$.

(e) Let $\omega^{\nabla f^*} : (0, \infty) \to (0, \infty)$ be a modulus of uniform continuity for ∇f^* on b-bounded sets.

(f) Let $\omega^f : (0, \infty) \to (0, \infty)$ be a modulus of uniform continuity for f on b-bounded sets.

(g) Let $S : (0, \infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in m such that

$$\forall m \in \mathbb{N}, \varepsilon > 0 \left(\prod_{k=m}^{S(\varepsilon,m)} (1 - \alpha_k) \leqslant \varepsilon \right).$$

(h) For each n, let $\overline{\alpha}_n$ be such that $0 < \overline{\alpha}_n \leqslant \alpha_n$ and define $\widetilde{\alpha}_n = \min\{\overline{\alpha}_i \mid i \leqslant n\}$.

Lemma 9.5.7. *Let $\varepsilon > 0$ be given and let x_n be defined by $(*)$. Define*

$$N := \sigma \left(\min \left\{ \frac{\widetilde{\varepsilon}}{8b}, \frac{\widetilde{\varepsilon}}{16b^2}, \frac{1}{2b} \omega^{\nabla f^*} \left(\min \left\{ \frac{\widetilde{\varepsilon}}{4b}, \omega^f \left(\frac{\widetilde{\varepsilon}}{4} \right) \right\} \right) \right\} \right)$$

where $\widetilde{\varepsilon} = \omega(\rho(\varepsilon, b), b)$.

For any $n \geqslant N$ and $p \in X$ with

$$\|p\|, \|\nabla f p\|, D_f(p, x_n), D_f(p, T_n x_n), D_f(p, u) \leqslant b$$

for the above b and where $\|T_n p - p\| < \min\{\widetilde{\varepsilon}, \omega'(\widetilde{\varepsilon}/8, b)\}$ as well as

$$D_f(p, x_n) \leqslant D_f(p, x_{n+1}) \text{ or } |D_f(p, x_{n+1}) - D_f(p, x_n)| < \widetilde{\varepsilon}/4,$$

it holds that

$$\|x_n - T_n x_n\| < \varepsilon.$$

Proof. At first, given an $n \geqslant N$ with $D_f(p, x_n) \leqslant D_f(p, x_{n+1})$, we have

$$\begin{aligned}
0 &\leqslant D_f(p, x_{n+1}) - D_f(p, x_n) \\
&\leqslant \alpha_n D_f(p, u) + (1 - \alpha_n) D_f(p, T_n x_n) - D_f(p, x_n) \\
&= \alpha_n \left(D_f(p, u) - D_f(p, T_n x_n)\right) + D_f(p, T_n x_n) - D_f(p, x_n) \\
&\leqslant \alpha_n \left(D_f(p, u) - D_f(p, T_n x_n)\right) + \frac{\widetilde{\varepsilon}}{8} && \text{(using (a) and (d))} \\
&\leqslant b\alpha_n + \frac{\widetilde{\varepsilon}}{8} && \text{(using (d))} \\
&< b\frac{\widetilde{\varepsilon}}{8b} + \frac{\widetilde{\varepsilon}}{8} && \text{(using that } n \geqslant N) \\
&= \frac{\widetilde{\varepsilon}}{4}.
\end{aligned}$$

Therefore, the first disjunct of the premise implies the second disjunct. So assume $n \geqslant N$ and $|D_f(p, x_{n+1}) - D_f(p, x_n)| < \widetilde{\varepsilon}/4$. Now, we have

$$\begin{aligned}
&\|\nabla f(x_{n+1}) - \nabla f(T_n x_n)\| \\
&\quad = \alpha_n \|\nabla f(u) - \nabla f(T_n x_n)\| \\
&\quad \leqslant \alpha_n 2b && \text{(using (d))} \\
&\quad < \min\left\{\frac{\widetilde{\varepsilon}}{8b}, \omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varepsilon}}{4b}, \omega^f\left(\frac{\widetilde{\varepsilon}}{4}\right)\right\}\right)\right\} && \text{(using (b) and } n \geqslant N)
\end{aligned}$$

and so by (e) and (d), we obtain

$$\begin{aligned}
\|x_{n+1} - T_n x_n\| &= \|\nabla f^*(\nabla f(x_{n+1})) - \nabla f^*(\nabla f(T_n x_n))\| \\
&\leqslant \min\left\{\frac{\widetilde{\varepsilon}}{4b}, \omega^f\left(\frac{\widetilde{\varepsilon}}{4}\right)\right\}.
\end{aligned}$$

By (f), we get

$$|f(x_{n+1}) - f(T_n x_n)| < \frac{\widetilde{\varepsilon}}{4}$$

and hence we obtain (reasoning similarly to [199])

$$\begin{aligned}
&|D_f(p, T_n x_n) - D_f(p, x_n)| \\
&\quad = |f(p) - f(T_n x_n) - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&\quad = |f(p) - f(x_{n+1}) + f(x_{n+1}) - f(T_n x_n) - \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle \\
&\qquad\quad + \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&\quad = |D_f(p, x_{n+1}) + f(x_{n+1}) - f(T_n x_n) + \langle p - x_{n+1}, \nabla f(x_{n+1})\rangle \\
&\qquad\quad - \langle p - T_n x_n, \nabla f(T_n x_n)\rangle - D_f(p, x_n)| \\
&\quad = |D_f(p, x_{n+1}) - D_f(p, x_n) + f(x_{n+1}) - f(T_n x_n) \\
&\qquad\quad + \langle p - x_{n+1}, \nabla f(x_{n+1}) - \nabla f(T_n x_n)\rangle - \langle x_{n+1} - T_n x_n, \nabla f(T_n x_n)\rangle| \\
&\quad \leqslant |D_f(p, x_{n+1}) - D_f(p, x_n)| + |f(x_{n+1}) - f(T_n x_n)| \\
&\qquad\quad + \|\nabla f(x_{n+1}) - \nabla f(T_n x_n)\| \, \|p - x_{n+1}\| + \|\nabla f(T_n x_n)\| \, \|x_{n+1} - T_n x_n\| \\
&\quad < \frac{\widetilde{\varepsilon}}{4} + \frac{\widetilde{\varepsilon}}{4} + \frac{\widetilde{\varepsilon}}{8b} 2b + \frac{\widetilde{\varepsilon}}{4b} b \\
&\quad = \widetilde{\varepsilon}.
\end{aligned}$$

Hence by (a) and (d), we obtain $D_f(T_n x_n, x_n) < \rho(\varepsilon, b)$ and so, by (c) and (d), we get $\|x_n - T_n x_n\| < \varepsilon$. □

Lemma 9.5.8. *For $\varepsilon > 0$ and $g : \mathbb{N} \to \mathbb{N}$, assume that we have a value φ and a $p \in X$ such that additionally*

$$\|p\|, \|\nabla f p\|, D_f(p, x_n), D_f(p, T_n x_n), D_f(p, u) \leqslant b$$

for the above b and

$$\begin{aligned}
\|T_n p - p\| < \min \Bigg\{ &\widetilde{\varphi}, \omega'\left(\frac{\widetilde{\varphi}}{8}, b\right), \\
&\omega'\left(\frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g) + g'^M(\Phi_\varphi(\varepsilon,g)))}}{2}, b\right), \\
&\omega'\left(\frac{\rho(\varepsilon, b)}{4(\Phi_\varphi(\varepsilon, g) + g^M(\Phi_\varphi(\varepsilon, g)) + 1)}, b\right) \Bigg\}
\end{aligned}$$

as well as

$$\forall y \in X \, (\|y\| \leqslant b \wedge \|T_n y - y\| < \varphi \to \langle y - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon, b)/8)$$

for any $n \leqslant \Phi_\varphi(\varepsilon, g) + {g'}^M(\Phi_\varphi(\varepsilon, g))$ *where we define*

$$N := \max\left\{\sigma\left(\frac{\omega^{\nabla f^*}(\rho(\varepsilon,b)/16b, b)}{2b}\right),\right.$$
$$\left.\sigma\left(\min\left\{\frac{\widetilde{\varphi}}{8b}, \frac{\widetilde{\varphi}}{16b^2}, \frac{1}{2b}\omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varphi}}{4b}, \omega^f\left(\frac{\widetilde{\varphi}}{4}\right)\right\}\right)\right\}\right), 1\right\}$$

with $\widetilde{\varphi} = \omega(\rho(\varphi', b), b)$ *and* $\varphi' = \min\{\varphi, \rho(\varepsilon, b)/16b\}$ *as well as*

$$\Phi_\varphi(\varepsilon, g) := K_1 + S(\rho(\varepsilon, b)/8b, K_1) + 1$$

with

$$K_0 = \widetilde{g'}^{(\lceil 4(b+1)/\widetilde{\varphi}\rceil)}(N),\ K_1 = \widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0),$$

and $\widetilde{g'}(n) = g'(n) + n$ *where* $g'(n) = \widehat{g}(n) + 2$ *for*

$$\widehat{g}(n) := g^M\left(n + S(\rho(\varepsilon, b)/8b, n) + 1\right) + S(\rho(\varepsilon, b)/8b, n).$$

Then it holds that

$$\exists n \leqslant \Phi_\varphi(\varepsilon, g)\forall i \in [n; n + g(n)]\left(\|p - x_i\| < \varepsilon\right).$$

Proof (compare also [104, 194]). We write $a_i := D_f(p, x_i)$. To establish the claim, we divide between two cases:

Case 1: $\forall i \leqslant K_0\,(a_{i+1} \leqslant a_i)$.
Suppose first that

$$\forall i < \lceil 4(b+1)/\widetilde{\varphi}\rceil\left(a_{\widetilde{g'}^{(i+1)}(N)} \leqslant a_{\widetilde{g'}^{(i)}(N)} - \widetilde{\varphi}/4\right).$$

Then we would get

$$a_{\widetilde{g'}^{(0)}(N)} \geqslant a_{\widetilde{g'}^{(1)}(N)} + \widetilde{\varphi}/4 \geqslant \dots \geqslant a_{\widetilde{g'}^{\lceil 4(b+1)/\widetilde{\varphi}\rceil}(N)} + \lceil 4(b+1)/\widetilde{\varphi}\rceil\widetilde{\varphi}/4 > b$$

which is a contradiction. Thus, we have

$$\exists i_0 < \lceil 4(b+1)/\widetilde{\varphi}\rceil\left(a_{\widetilde{g'}^{(i_0+1)}(N)} > a_{\widetilde{g'}^{(i_0)}(N)} - \widetilde{\varphi}/4\right).$$

and in particular for $n = \widetilde{g'}^{(i_0)}(N)$, we have

$$\forall i, j \in [n; n + \widehat{g}(n) + 2]\left(|a_i - a_j| \leqslant a_n - a_{n+\widehat{g}(n)+2} < \widetilde{\varphi}/4\right).$$

Therefore, we in particular have

$$\forall i \in [n; n + \widehat{g}(n) + 1] \, (|D_f(p, x_{i+1}) - D_f(p, x_i)| < \widetilde{\varphi}/4)$$

Using Lemma 9.5.7, we get

$$\forall i \in [n; n + \widehat{g}(n) + 1] \, (\|x_i - T_i x_i\| < \varphi' \leqslant \varphi) \, .$$

Using the assumption on φ, we in particular get

$$\forall i \in [n; n + \widehat{g}(n) + 1] \, (\langle x_i - p, \nabla f(u) - \nabla f(p) \rangle < \rho(\varepsilon, b)/8)$$

and thus

$$\forall i \in [n; n + \widehat{g}(n)] \, (\langle x_{i+1} - p, \nabla f(u) - \nabla f(p) \rangle < \rho(\varepsilon, b)/8) \, .$$

As in [199] (p. 495), we can derive

$$D_f(p, x_{i+1}) \leqslant (1 - \alpha_i) D_f(p, T_i x_i) + \alpha_i \langle x_{i+1} - p, \nabla f(u) - \nabla f(p) \rangle$$

for any i which implies

$$D_f(p, x_{i+1}) \leqslant \\ (1 - \alpha_i) D_f(p, x_i) + \alpha_i \langle x_{i+1} - p, \nabla f(u) - \nabla f(p) \rangle + \frac{\rho(\varepsilon, b)}{4(\Phi_\varphi(\varepsilon, g) + g^M(\Phi_\varphi(\varepsilon, g)) + 1)}$$

for any $i \leqslant \Phi_\varphi(\varepsilon, g) + g'^M(\Phi_\varphi(\varepsilon, g))$ using the assumption on p and the assumption on ω'. Using Lemma 9.5.2,[11] we get

$$\exists n \leqslant K_0 + S(\rho(\varepsilon, b)/8b, K_0) + 1 \leqslant \Phi_\varphi(\varepsilon, g) \forall i \in [n; n + g(n)] \\ \left(D_f(p, x_i) \leqslant \frac{\rho(\varepsilon, b)}{2} < \rho(\varepsilon, b) \right)$$

which implies $\|p - x_i\| < \varepsilon$ for all such i by (c).

Case 2: $\exists i \leqslant K_0 \, (a_{i+1} > a_i)$.
Then, we define τ as in Lemma 9.5.3.(2), i.e.

$$\tau(n) := \max\{k \leqslant \max\{K_0, n\} \mid a_k < a_{k+1}\}.$$

In particular, we have

[11] To apply Lemma 9.5.2, we set $a_n = D_f(p, x_n)$ on $[0; \Phi_\varphi(\varepsilon, g) + g'^M(\Phi_\varphi(\varepsilon, g)) + 1]$ and 0 otherwise. Further, we set $\beta_n = \langle x_{n+1} - p, \nabla f(u) - \nabla f(p) \rangle$ and $\gamma_n = \rho(\varepsilon, b)/4(\Phi_\varphi(\varepsilon, g) + g^M(\Phi_\varphi(\varepsilon, g)) + 1)$ and use the same α_n as in (b). Note that we then indeed have $a_i = D_f(p, x_i)$ for any a_i in the conclusion of Lemma 9.5.2.

1. $\forall n \left(a_{\tau(n)} \leqslant a_{\tau(n)+1}, \tau(n) \leqslant \tau(n+1)\right)$,

2. $\forall n \geqslant K_0 \left(a_n \leqslant a_{\tau(n)+1}\right)$.

Case 2.1: $\exists m \in [K_1; K_1 + g'(K_1)] \, (\tau(m) < K_0)$.
As $m \geqslant K_1$, we have

$$\tau\left(\widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0)\right) = \tau(K_1) \leqslant \tau(m) < K_0$$

and thus using Lemma 9.5.3.(1), we get

$$\exists n \leqslant \widetilde{g'}^{(\lceil 8(b+1)/\widetilde{\varphi}\rceil)}(K_0) \, (n \geqslant K_0 \wedge \forall i, j \in [n; n + \widehat{g}(n) + 2] \, (|a_i - a_j| \leqslant \widetilde{\varphi}/8 < \widetilde{\varphi}/4))$$

from which we can deduce

$$\exists n \leqslant \Phi_\varphi(\varepsilon, g) \forall i \in [n; n + g(n)] \, (\|p - x_i\| < \varepsilon)$$

as in Case 1.

Case 2.2: $\forall m \in [K_1; K_1 + g'(K_1)] \, (\tau(m) \geqslant K_0)$.
Using the properties of τ, we in particular have

$$D_f(p, x_{\tau(m)}) = a_{\tau(m)} \leqslant a_{\tau(m)+1} = D_f(p, x_{\tau(m)+1})$$

for all m. Therefore

$$\begin{aligned}
0 &\leqslant D_f(p, x_{\tau(m)+1}) - D_f(p, x_{\tau(m)}) \\
&\leqslant \alpha_{\tau(m)} D_f(p, u) + (1 - \alpha_{\tau(m)}) D_f(p, T_{\tau(m)} x_{\tau(m)}) - D_f(p, x_{\tau(m)}) \\
&\leqslant \alpha_{\tau(m)} \left(D_f(p, u) - D_f(p, T_{\tau(m)} x_{\tau(m)})\right) + \left(D_f(p, T_{\tau(m)} x_{\tau(m)}) - D_f(p, x_{\tau(m)})\right) \\
&< b\alpha_{\tau(m)} + \widetilde{\varphi}/8
\end{aligned}$$

using the assumption on p (as $\tau(m) \leqslant m \leqslant K_1 + g'(K_1)$). As for $m \in [K_1; K_1 + g'(K_1)]$, we have $\tau(m) \geqslant K_0 \geqslant N$, we get

$$0 \leqslant D_f(p, x_{\tau(m)+1}) - D_f(p, x_{\tau(m)}) < \widetilde{\varphi}/4$$

for all such m. Using Lemma 9.5.7, we get

$$\left\|x_{\tau(m)} - T_{\tau(m)} x_{\tau(m)}\right\| < \varphi' = \min\{\varphi, \rho(\varepsilon, b)/16b\}.$$

Thus, using the assumption on φ, we get

$$\langle x_{\tau(m)} - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon, b)/8 < \rho(\varepsilon, b)/4.$$

As in the proof Lemma 9.5.7, we get

$$\left\|\nabla f(x_{\tau(m)+1}) - \nabla f(T_{\tau(m)}x_{\tau(m)})\right\| \leqslant \alpha_{\tau(m)}2b.$$

As $\tau(m) \geqslant K_0 \geqslant N$, we in particular have $\left\|x_{\tau(m)+1} - T_{\tau(m)}x_{\tau(m)}\right\| < \rho(\varepsilon, b)/16b$. Further, from above we also have $\left\|x_{\tau(m)} - T_{\tau(m)}x_{\tau(m)}\right\| < \rho(\varepsilon, b)/16b$ such that this combined yields $\left\|x_{\tau(m)+1} - x_{\tau(m)}\right\| < \rho(\varepsilon, b)/8b$. Therefore:

$$\begin{aligned}
&\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle \\
&\quad = \langle x_{\tau(m)+1} - x_{\tau(m)}, \nabla f(u) - \nabla f(p)\rangle + \langle x_{\tau(m)} - p, \nabla f(u) - \nabla f(p)\rangle \\
&\quad < \left\|x_{\tau(m)+1} - x_{\tau(m)}\right\| 2b + \rho(\varepsilon, b)/4 \\
&\quad < \rho(\varepsilon, b)/2.
\end{aligned}$$

Similar to before, we can derive

$$\begin{aligned}
D_f(p, x_{\tau(m)+1}) &\leqslant (1-\alpha_{\tau(m)})D_f(p, T_{\tau(m)}x_{\tau(m)}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle \\
&\leqslant (1-\alpha_{\tau(m)})D_f(p, x_{\tau(m)}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle \\
&\qquad + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2} \\
&\leqslant (1-\alpha_{\tau(m)})D_f(p, x_{\tau(m)+1}) + \alpha_{\tau(m)}\langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle \\
&\qquad + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2}
\end{aligned}$$

for all $m \in [K_1; K_1 + g'(K_1)]$ (since $\tau(m) \leqslant m \leqslant K_1 + g'(K_1)$). From this, we get

$$D_f(p, x_{\tau(m)+1}) \leqslant \langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2\alpha_{\tau(m)}}$$

for all such m. Again as $\tau(m) \leqslant m \leqslant K_1 + g'(K_1)$, we get

$$\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))} \leqslant \overline{\alpha}_{\tau(m)} \leqslant \alpha_{\tau(m)}$$

for all such m and thus we have

$$D_f(p, x_{\tau(m)+1}) \leqslant \langle x_{\tau(m)+1} - p, \nabla f(u) - \nabla f(p)\rangle + \frac{\rho(\varepsilon, b)}{2} < \rho(\varepsilon, b).$$

Lastly, as we thus have

$$D_f(p, x_m) \leqslant D_f(p, x_{\tau(m)+1}) < \rho(\varepsilon, b)$$

for all such m by using the properties of τ, we can now deduce the claim of the theorem as before. □

Together with Lemma 9.5.5, we thus obtain the following combined result for sequences of uniformly Bregman strongly nonexpansive maps. One crucial property that features therein is a uniform version of the NST condition as e.g. considered in [2] for sequences of strongly nonexpansive maps in the ordinary sense: given a sequence (T_n) of strongly nonexpansive maps and an additional such map T, these are said to satisfy the NST condition if any fixed point of T is a common fixed point for all T_n and if $\|x_n - T_n x_n\| \to 0$ implies $\|x_n - Tx_n\| \to 0$ for any bounded sequence (x_n).

Concretely, the following uniform quantitative variant of this condition will feature crucially in the following combined result: we assume a modulus $\mu : (0,\infty)^2 \times \mathbb{N} \to (0,\infty)$ such that

$$\begin{aligned}\forall \varepsilon, b > 0 \forall K \in \mathbb{N} \forall p \in X (\|p\| \leqslant b \\ \wedge \|p - Tp\| < \mu(\varepsilon, b, K) \to \forall n \leqslant K (\|p - T_n p\| < \varepsilon))\end{aligned} \tag{$\dagger$}_1$$

as well as a modulus $\nu : (0,\infty)^2 \to (0,\infty)$ such that

$$\forall \varepsilon, b > 0 \forall n \in \mathbb{N} \forall p \in X \left(\|p\| \leqslant b \wedge \|p - T_n p\| < \nu(\varepsilon, b) \to \|p - Tp\| < \varepsilon\right). \tag{$\dagger$}_2$$

If such moduli exist, we say that (T_n) and T satisfy the uniform NST condition.

As we will discuss later, such moduli can in particular be explicitly computed for the resolvents relative to f, thereby allowing applications to a Halpern-type proximal point algorithm.

Theorem 9.5.9. *Let $(\alpha_n) \subseteq (0,1]$ converge to zero with a rate σ and, for any n, let $\overline{\alpha}_n$ be such that $0 < \overline{\alpha}_n \leqslant \alpha_n$ and define $\widetilde{\alpha}_n = \min\{\overline{\alpha}_i \mid i \leqslant n\}$. Let f be sequentially consistent with a modulus of consistency ρ. Let $S : (0,\infty) \times \mathbb{N} \to \mathbb{N}$ be nondecreasing in the right argument such that*

$$\forall m \in \mathbb{N}, \varepsilon > 0 \left(\prod_{k=m}^{S(\varepsilon,m)} (1 - \alpha_k) \leqslant \varepsilon \right).$$

Let (T_n) be a sequence of uniformly Bregman strongly nonexpansive maps and T be another uniformly Bregman strongly nonexpansive map with a common strong BSNE-modulus ω and a common derived modulus ω'. Let each T_n and T be bounded on bounded sets with a common modulus E and let $p_0 \in X$ be a common fixed point of all T_n and T. Let o be a modulus of boundedness for D_f. Let ∇f and f be bounded on bounded sets with moduli C, D, respectively. Let $b \in \mathbb{N}^*$ with

$$b \geqslant \|p_0\|, D_f(p_0, u), \|u\|, D_f(p_0, x_0)$$

and define

$$\begin{aligned}\widehat{b} = \max\{&b, C(b), o(b,b), E(o(b,b)), C(E(o(b,b))), C(o(b,b)),\\ &D(b) + D(E(o(b,b))) + (b + E(o(b,b)))C(E(o(b,b))),\\ &D(b) + D(o(b,b)) + (b + o(b,b))C(o(b,b)), 2D(b) + 2bC(b)\}.\end{aligned}$$

Let $\omega^{\nabla f^*}$, ω^f *be moduli of uniform continuity of* ∇f^*, f, *respectively. Let further* Δ *be a modulus witnessing that* $D_f(\cdot, u)$ *is uniformly Fréchet differentiable on bounded subsets with derivative* $x \mapsto \nabla f x - \nabla f u$ *as in Lemma 9.5.5. Assume that we have a modulus* $\mu : (0,\infty)^2 \times \mathbb{N} \to (0,\infty)$ *such that*

$$\forall \varepsilon, b > 0, K \in \mathbb{N}, p \in X \left(\|p\| \leqslant b \wedge \|p - Tp\| < \mu(\varepsilon, b, K) \to \forall n \leqslant K\left(\|p - T_n p\| < \varepsilon\right)\right)$$

as well as a modulus $\nu : (0,\infty)^2 \to (0,\infty)$ *such that*

$$\forall \varepsilon, b > 0, n \in \mathbb{N}, p \in X \left(\|p\| \leqslant b \wedge \|p - T_n p\| < \nu(\varepsilon, b) \to \|p - Tp\| < \varepsilon\right).$$

For any $\varepsilon > 0$ *and* $g : \mathbb{N} \to \mathbb{N}$ *as well as* $\varphi > 0$, *we define*

$$\begin{aligned}\psi(\varphi) := \min\Bigg\{&\widetilde{\varphi}, \omega'\left(\frac{\widetilde{\varphi}}{8}, \widehat{b}\right),\\ &\omega'\left(\frac{\rho(\varepsilon,\widehat{b})\widetilde{\alpha}_{(\Phi_\varphi(\varepsilon,g)+g'^M(\Phi_\varphi(\varepsilon,g)))}}{2}, \widehat{b}\right),\\ &\omega'\left(\frac{\rho(\varepsilon,\widehat{b})}{4(\Phi_\varphi(\varepsilon,g) + g^M(\Phi_\varphi(\varepsilon,g)) + 1)}, \widehat{b}\right)\Bigg\},\end{aligned}$$

and

$$\begin{aligned}\widehat{\psi}(\varphi) &= \mu(\psi(\nu(\varphi,\widehat{b})), \widehat{b}, \Phi_{\nu(\varphi,\widehat{b})}(\varepsilon,g) + g'^M(\Phi_{\nu(\varphi,\widehat{b})}(\varepsilon,g))),\\ \psi'(\varphi) &= \min\{\widehat{\psi}(\omega'(\rho(\varphi,\max\{\widehat{b},E(\widehat{b})\}),\widehat{b})), \omega'(\rho(\varphi,\max\{\widehat{b},E(\widehat{b})\}),\widehat{b})\},\end{aligned}$$

with

$$\begin{aligned}N := \max\Bigg\{&\sigma\left(\frac{\omega^{\nabla f^*}(\rho(\varepsilon,\widehat{b})/16\widehat{b},\widehat{b})}{2\widehat{b}}\right),\\ &\sigma\left(\min\left\{\frac{\widetilde{\varphi}}{8\widehat{b}}, \frac{\widetilde{\varphi}}{16\widehat{b}^2}, \frac{1}{2\widehat{b}}\omega^{\nabla f^*}\left(\min\left\{\frac{\widetilde{\varphi}}{4\widehat{b}}, \omega^f\left(\frac{\widetilde{\varphi}}{4},\widehat{b}\right)\right\},\widehat{b}\right)\right\}\right), 1\Bigg\}\end{aligned}$$

where $\widetilde{\varphi} = \omega(\rho(\varphi',\widehat{b}),\widehat{b})$ *and* $\varphi' = \min\{\varphi, \rho(\varepsilon,\widehat{b})/16\widehat{b}\}$ *as well as*

$$\Phi_\varphi(\varepsilon, g) := K_1 + S(\rho(\varepsilon,\widehat{b})/8\widehat{b}, K_1) + 1$$

with

$$K_0 = \widetilde{g'}^{(\lceil 4(\widehat{b}+1)/\widetilde{\varphi}\rceil)}(N),\ K_1 = \widetilde{g'}^{(\lceil 8(\widehat{b}+1)/\widetilde{\varphi}\rceil)}(K_0),$$

and $\widetilde{g'}(n) = g'(n) + n$ *where* $g'(n) = \widehat{g}(n) + 2$ *for*

$$\widehat{g}(n) := g^M\left(n + S(\rho(\varepsilon, \widehat{b})/8\widehat{b}, n) + 1\right) + S(\rho(\varepsilon, \widehat{b})/8\widehat{b}, n).$$

Then it holds that

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

where

$$\widehat{\Phi}(\varepsilon, g) := \max\left\{\Phi_{\nu(\omega'(\rho(\psi'^{(r)}(1), \max\{\widehat{b}, E(\widehat{b})\}), \widehat{b}), \widehat{b})}(\varepsilon/2, g) \mid r \leqslant \left\lceil(\widehat{b}+1)/\varepsilon'\right\rceil\right\}$$

with

$$\varepsilon' = \frac{\rho(\varepsilon/2, \widehat{b})}{16} \min\left\{\frac{\Delta(\rho(\varepsilon/2, \widehat{b})/32\widehat{b}, \widehat{b})}{4\widehat{b}}, 1/2\right\}.$$

Proof. Let ε and g be given. Using (the proof of) Lemma 9.5.5, (2), we get that for the above $\widehat{\psi}$, there exists a $p \in X$ with $\|p\| \leqslant b \leqslant \widehat{b}$ and an $r \leqslant \left\lceil(\widehat{b}+1)/\varepsilon'\right\rceil$ such that for $\delta = \omega'(\rho(\psi'^{(r)}(1), \max\{\widehat{b}, E(\widehat{b})\}), \widehat{b})$ we have $\|Tp - p\| < \widehat{\psi}(\delta)$

$$\forall q \in X \left(\|q\| \leqslant \widehat{b} \text{ and } \|Tq - q\| < \delta \to \langle q - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon/2, \widehat{b})/8\right).$$

Then, as

$$\|Tp - p\| < \widehat{\psi}(\delta) = \mu(\psi(\nu(\delta, \widehat{b})), \widehat{b}, \Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2, g) + g'^M(\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2, g))),$$

we get

$$\|T_n p - p\| < \psi(\nu(\delta, \widehat{b}))$$

for all $n \leqslant \Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2, g) + g'^M(\Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2, g))$. Further, if $\|q - T_n q\| < \nu(\delta, \widehat{b})$, we have $\|Tq - q\| < \delta$ and thus also

$$\forall q \in X \left(\|q\| \leqslant \widehat{b} \text{ and } \|T_n q - q\| < \nu(\delta, \widehat{b}) \to \langle q - p, \nabla f(u) - \nabla f(p)\rangle < \rho(\varepsilon/2, \widehat{b})/8\right)$$

for any n. Lemma 9.5.8 then yields that

$$\exists n \leqslant \Phi_{\nu(\delta,\widehat{b})}(\varepsilon/2, g) \forall i \in [n; n + g(n)] \left(\|p - x_i\| < \varepsilon/2\right)$$

as $\widehat{b}$ bounds all the objects involved. After using the triangle inequality, we get the desired claim. $\square$

In particular, since having a rate of metastability is equivalent to being convergent, the above quantitative result implies the following (non-quantitative) convergence result. For that, we say that (T_n) and T are commonly uniformly Bregman strongly nonexpansive if all T_n and T are uniformly Bregman strongly nonexpansive with a common strong BSNE-modulus and we say that they are commonly bounded on bounded sets if there exists a common modulus witnessing that all T_n and T are bounded on bounded sets.

Theorem 9.5.10. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let (T_n) be a sequence of selfmaps and T be a selfmap such that they are commonly uniformly Bregman strongly nonexpansive and commonly bounded on bounded sets. Assume that (T_n) and T satisfy the uniform NST condition and that they posses a common fixed point. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f u + (1-\alpha_n)\nabla f T_n x_n\right)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) is Cauchy.

Further, if we have $\overline{F}(T) \subseteq F(T)$ where $\overline{F}(T)$ is the set of all strong asymptotic fixed points (i.e. of all p such that there is a sequence (p_n) with $p_n \to p$ and $\|p_n - Tp_n\| \to 0$ for $n \to \infty$), then (x_n) converges strongly to $P^f_{F(T)}(u)$.

Proof. First, note that under the assumptions presented above, all moduli featured in Theorem 9.5.9 exist and we shortly discuss this for the assumptions not explicitly covered already: A modulus of consistency ρ exists for f as f is totally convex on bounded sets using Lemmas 9.2.7 and 9.2.8. As f is uniformly Fréchet differentiable and bounded on bounded sets, ∇f is uniformly continuous on bounded sets by Proposition 8.5.4 and thus a corresponding modulus $\omega^{\nabla f}$ exists which allows us to construct a corresponding modulus ω^f for the uniform continuity of f as well as moduli for ∇f, f being bounded on bounded sets using Lemma 9.2.2. Also, as discussed in Remark 9.5.6, $\omega^{\nabla f}$ can be used to construct the modulus Δ featured in Theorem 9.5.9. Now, as discussed in Remark 9.2.13, f being totally convex on bounded sets implies f^* being uniformly Fréchet differentiable and thus ∇f^* being uniformly continuous as f is supercoercive (again using Proposition 8.5.4). Thus a corresponding modulus $\omega^{\nabla f^*}$ exists. Lastly, a modulus of boundedness for D_f exists as well and can be constructed as discussed in Remark 9.2.14.

So Theorem 9.5.9 applies and we therefore get

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

Thus x_n is Cauchy as if not, there exists an $\varepsilon > 0$ such that for any n, there exists an m such that $\|x_n - x_{n+m}\| \geqslant \varepsilon$. Pick $g(n) = m$ for such an m. Then this ε and g refute the above property. Now, as x_n is Cauchy, it converges to a limit x.

To see that this limit is indeed the projection $P^f_{F(T)}(u)$, let $\Omega(\varepsilon, b)$ be a modulus of uniform continuity on bounded sets for the function $p \mapsto \langle y - p, \nabla f u - \nabla f p\rangle$ uniform in $\|u\|, \|y\| \leqslant b$.[12] Now, let $\varepsilon > 0$ be given and let K be so large that

$$\forall m \geqslant K \left(\|x_m - x\| < \frac{1}{2}\Omega\left(\frac{\varepsilon}{2}, \widehat{b}\right)\right).$$

Now, for $\varepsilon' := 1/2\Omega(\varepsilon/2, \widehat{b})$, we can use Lemma 9.5.5 to choose a $p \in X$ and a δ with $\|p\| \leqslant \widehat{b}$ and $\|p - Tp\| < \widehat{\psi}(\delta)$ as well as

$$\forall q \in X \left(\|q\| \leqslant \widehat{b} \text{ and } \|q - Tq\| < \delta \to \langle q - p, \nabla f u - \nabla f p\rangle < \rho(\varepsilon', \widehat{b})/8\right).$$

Then, using this p and reasoning as in the proof of Theorem 9.5.9, we can apply Lemma 9.5.8 to $g(n) := K$ and ε' which yields an $n \geqslant K$ such that $\|p - x_n\| < \varepsilon' = 1/2\Omega(\varepsilon/2, \widehat{b})$. That $n \geqslant K$ holds in particular yields $\|p - x\| < \Omega(\varepsilon/2, \widehat{b})$. Let w.l.o.g. $\rho(\varepsilon, b) \leqslant \varepsilon$ and $\Omega(\varepsilon, b) \leqslant \varepsilon$. Then we in particular have

$$\langle q - p, \nabla f u - \nabla f p\rangle < \varepsilon/2$$

for any q with $\|q\| \leqslant \widehat{b}$ and $\|q - Tq\| < \delta$. Thus

$$\langle q - x, \nabla f u - \nabla f x\rangle < \varepsilon$$

for all such q.

If now $q = Tq$, then we get $\langle q - x, \nabla f u - \nabla f x\rangle < \varepsilon$ for all $\varepsilon > 0$, i.e.

$$\forall q \in F(T) \left(\langle q - x, \nabla f u - \nabla f x\rangle \leqslant 0\right).$$

Further, if we assume that $\overline{F}(T) \subseteq F(T)$, then x is also fixed point of T. For this, note that as in Lemma 9.5.7, we have

$$\|\nabla f x_{n+1} - \nabla f T_n x_n\| = \alpha_n \|\nabla f u - \nabla f T x_n\| \to 0$$

[12] It can be easily seen that such a modulus Ω can actually be constructed from $\omega^{\nabla f}$

as $\alpha_n \to 0$ and as $\nabla f T x_n$ is bounded since x_n is bounded and since T and ∇f are bounded on bounded sets. Thus $\|x_{n+1} - T_n x_n\| \to 0$ and therefore also $\|x_n - T_n x_n\| \to 0$. As (T_n) and T satisfy the uniform NST condition, we get $\|x_n - T x_n\| \to 0$. As $\|x_n - x\| \to 0$, this yields $x \in \overline{F}(T) \subseteq F(T)$. Combined, this yields that $x = P^f_{F(T)}(u)$ (recall the discussion before Lemma 9.5.5). □

Remark 9.5.11. The above result in particular contains the previous Theorem 9.5.1 for uniformly Bregman strongly nonexpansive maps T by picking $T_n = T$. Naturally T is bounded on bounded sets as $F(T) \neq \emptyset$ and as T is Bregman quasi-nonexpansive. However, note that in the context of uniformly Bregman strongly nonexpansive maps T, the assumption that $\widehat{F}(T) \subseteq F(T)$ was properly weakened through the analysis to $\overline{F}(T) \subseteq F(T)$. As discussed before in Remark 8.5.13, reflexivity is already an inherent property from assuming that f is supercoercive, bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded sets as this implies that f^* is uniformly Fréchet differentiable. For that same reason, also being a Legendre function is an inherent property of any such function. So these assumptions could have been omitted in Theorem 9.5.1 already.

Using this theorem, we will in particular be able to derive the strong convergence of the Halpern-type proximal point algorithm in all Banach spaces together with other interesting instantiations that will be discussed in the following section.

9.6 Special cases and instantiations

We are now concerned with the range of the above results. For that, this section discusses how the above (quantitative) results can be instantiated in various ways so that they apply to many other well-known methods in the context of Bregman distance. In particular, we obtain quantitative strong convergence results for Halpern-type variants of the method of cyclic Bregman projections, of the proximal point algorithm, of a special case of a method solving operator equations due to Butnariu and Resmerita [36] as well as of a special case of the forward-backward Bregman splitting method discussed by Búi and Combettes [32] (see also Van Nguyen [155]), of a method for finding common zeros of maximally monotone operators as discussed by Naraghirad [152] and of a Halpern-Mann type iteration of Bregman strongly nonexpansive maps [214].

In particular, we show how the Halpern-Mann type iteration presented in [214] can be recognized as an instantiation of the Halpern-iteration considered before for a family of uniformly Bregman strongly nonexpansive maps. Further, inspired by the recent considerations [41] on the relationship between modified Halpern methods in the sense of [53, 85] and Tikhonov-Mann type methods as developed by [20, 42, 210], we use this instantiation to even provide a strong convergence result for a new Tikhonov-Mann type iteration of uniformly Bregman strongly nonexpansive maps which provides a suitable lift of such iterations to this Bregman context. Lastly, we discuss another new strongly convergent method for two uniformly Bregman strongly nonexpansive maps inspired by the recently introduced alternating Halpern-Mann type method introduced by Dinis and Pinto [57].

All these results in particular further show that the additional requirement in the previous theorems that the maps are even uniformly Bregman strongly nonexpansive is practically of lesser significance as most maps encountered in the literature that are Bregman strongly nonexpansive are already uniformly Bregman strongly nonexpansive.

9.6.1 Cyclic projections

A first readily defined instantiation of Theorem 9.5.10 on the Halpern-iteration is that obtained by using the cyclic projection operator

$$T = P^f_{\Omega_k} \circ \cdots \circ P^f_{\Omega_1}$$

where $P^f_{\Omega_j}$ is the Bregman projection onto a given non-empty closed convex set Ω_j for $j = 1, \ldots, k$. Assume that $\Omega_1 \cap \cdots \cap \Omega_k \neq \varnothing$. Then this operator T is uniformly Bregman strongly nonexpansive since every projection $P^f_{\Omega_j}$ is even Bregman firmly nonexpansive and moduli for the Bregman strong nonexpansivity of T can be calculated from the moduli of the factors by following Theorem 9.3.15 as well as Lemma 9.3.8. For this, note further that by $\Omega_1 \cap \cdots \cap \Omega_k \neq \varnothing$, using Lemma 9.3.5, each $P^f_{\Omega_j}$ and thus T is bounded on bounded sets. Further, note that any Bregman firmly nonexpansive map that is bounded on bounded sets actually possesses a modulus of uniform closedness if ∇f is uniformly continuous on bounded subsets as well as uniformly strictly monotone (the latter of which, recalling the discussion from Remark 9.2.13, follows from the assumption that f is totally convex on bounded sets) as by Lemma 9.3.13, each such map is then uniformly continuous on bounded subsets. Thus $P^f_{\Omega_j}$ is uniformly continuous on bounded subsets. In particular, from a corresponding (common) modulus of uniform continuity, a (common) modulus κ of uniform closedness can be immediately defined. Note that through the uniform continuity of each $P^f_{\Omega_j}$, also T is uniformly continuous on bounded sets and thus also T possesses a modulus of uniform closedness which in particular yields that $\overline{F}(T) \subseteq F(T)$.

Combining this with Theorem 9.5.10, we get the following corollary on a Halpern-type variant of the method of cyclic projections (where we can identify the limit as the corresponding projection as we have previously established $\overline{F}(T) \subseteq F(T)$).

Theorem 9.6.1. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a super-coercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let $\Omega_1, \ldots, \Omega_k$ be non-empty closed convex sets and assume that $\Omega_1 \cap \cdots \cap \Omega_k \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f u + (1-\alpha_n)\nabla f P^f_{\Omega_k} \circ \cdots \circ P^f_{\Omega_1} x_n\right)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{F(T)}(u)$ for $T = P^f_{\Omega_k} \circ \cdots \circ P^f_{\Omega_1}$.

In particular, a rate of metastability can be calculated using Theorem 9.5.9 together with Lemmas 9.3.5 and 9.3.8 as well as Theorems 9.3.15 and 9.3.14.

9.6.2 The proximal point algorithm

We are now concerned with a Halpern-type variant of the proximal point algorithm for a maximally monotone operator A with resolvents Res^f_γ as before. Concretely, for a given u and x_0, we consider the sequence

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f \mathrm{Res}^f_{r_n} x_n) \qquad (**)$$

for a given additional sequence r_n that satisfies

$$0 < \bar{r} = \inf\{r_n \mid n \in \mathbb{N}\}.$$

To show that the previous results contained in Theorems 9.5.9 and 9.5.10 apply here, we will in the following provide concrete instantiations for the moduli μ and ν for the concrete choices of

$$T_n = \mathrm{Res}^f_{r_n} \text{ and } T = \mathrm{Res}^f_{\bar{r}}.$$

For this, we will however need some further facts about the resolvent relative to f. It is straightforward to show that the set of fixed points of any Res^f_γ equals to the set of zeros $A^{-1}0$ of the operator A. The following lemma provides a quantitative result for one of the directions of the equivalence.

Lemma 9.6.2. *Let $\widehat{\eta}$ be a modulus of uniform strict monotonicity of ∇f on bounded sets. Given $\gamma > 0$ and $\varepsilon > 0$, let $(x,y) \in A$ with $b > 0$ such that $b \geqslant \|x\|, \left\|\mathrm{Res}^f_\gamma x\right\|, \gamma$. If we have*

$$\|y\| < \frac{\widehat{\eta}(\varepsilon,b)}{2b^2},$$

then $\left\|x - \mathrm{Res}^f_\gamma x\right\| < \varepsilon$.

Proof. By monotonicity of A, we have $\langle \mathrm{Res}^f_\gamma x - x, A^f_\gamma x - y\rangle \geqslant 0$ and thus

$$\begin{aligned}\langle x - \mathrm{Res}^f_\gamma x, \nabla f x - \nabla f \mathrm{Res}^f_\gamma x\rangle &\leqslant \gamma \langle x - \mathrm{Res}^f_\gamma x, y\rangle \\ &\leqslant \gamma(\|x\| + \left\|\mathrm{Res}^f_\gamma x\right\|)\|y\| \\ &\leqslant 2b^2 \|y\|.\end{aligned}$$

Thus $\|y\| < \widehat{\eta}(\varepsilon,b)/2b^2$ implies $\left\|x - \mathrm{Res}^f_\gamma x\right\| < \varepsilon$ by the assumptions on $\widehat{\eta}$. □

The following lemma due to Reich and Sabach provides a crucial relation between the resolvent relative to f and the Bregman distance associated with f.

Lemma 9.6.3 ([177, 178])**.** *Let A be maximally monotone and assume that $A^{-1}0 \neq \varnothing$. Then*

$$D_f(u, \mathrm{Res}^f_\gamma x) + D_f(\mathrm{Res}^f_\gamma x, x) \leqslant D_f(u,x)$$

for all $\gamma > 0$, $u \in A^{-1}0$ and $x \in X$.

In particular, we will in the following rely on a quantitative version of this result as given in the next lemma.

Lemma 9.6.4. *Let $\omega^{\nabla f}(\varepsilon,b) \leqslant \varepsilon$ be a modulus of uniform continuity of ∇f on bounded subsets. Let $x,y \in X$ and $r,s > 0$ be given such that*

$$b \geqslant \|x\|, \left\|\mathrm{Res}^f_s x\right\|, \|y\|, \left\|\mathrm{Res}^f_r y\right\|.$$

Then for any $\varepsilon > 0$, if

$$\left\|x - Res^f_s x\right\| < \omega^{\nabla f}\left(\frac{\varepsilon}{2E}, b\right) \text{ for } E \geqslant \max\{2b, rs^{-1}2b\},$$

then we have

$$D_f(x, \mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) < D_f(x,y) + \varepsilon.$$

Proof. Using the three-point-identity for D_f, we get

$$\begin{aligned} D_f(x,y) &= D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) + \langle x - \mathrm{Res}^f_r y, \nabla f \mathrm{Res}^f_r y - \nabla f y\rangle \\ &= D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) + r\langle x - \mathrm{Res}^f_r y, -A^f_r y\rangle. \end{aligned}$$

Using the monotonicity of A, we further derive that

$$\begin{aligned} &\langle x - \mathrm{Res}^f_r y, -A^f_r y\rangle \\ &\quad = \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, -A^f_r y\rangle \\ &\quad = \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, s^{-1}(\nabla f x - \nabla f \mathrm{Res}^f_s x) - A^f_r y\rangle \\ &\qquad + \langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, -s^{-1}(\nabla f x - \nabla f \mathrm{Res}^f_s x)\rangle \\ &\quad \geqslant \langle x - \mathrm{Res}^f_s x, -A^f_r y\rangle + s^{-1}\langle \mathrm{Res}^f_s x - \mathrm{Res}^f_r y, \nabla f \mathrm{Res}^f_s x - \nabla f x\rangle \\ &\quad \geqslant - \left\| x - \mathrm{Res}^f_s x\right\| \left\| A^f_r y\right\| - s^{-1}\left\|\mathrm{Res}^f_s x - \mathrm{Res}^f_r y\right\| \left\|\nabla f \mathrm{Res}^f_s x - \nabla f x\right\| \\ &\quad \geqslant - \left\| x - \mathrm{Res}^f_s x\right\| r^{-1}\left(\left\|\mathrm{Res}^f_r y\right\| + \|y\|\right) \\ &\qquad - s^{-1}\left(\left\|\mathrm{Res}^f_s x\right\| + \left\|\mathrm{Res}^f_r y\right\|\right) \left\|\nabla f \mathrm{Res}^f_s x - \nabla f x\right\|. \end{aligned}$$

Combined with the above, this yields

$$\begin{aligned} D_f(x,y) &\geqslant D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - \left\| x - \mathrm{Res}^f_s x\right\| \left(\left\|\mathrm{Res}^f_r y\right\| + \|y\|\right) \\ &\qquad - rs^{-1}\left(\left\|\mathrm{Res}^f_s x\right\| + \left\|\mathrm{Res}^f_r y\right\|\right) \left\|\nabla f \mathrm{Res}^f_s x - \nabla f x\right\| \\ &\geqslant D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - 2b\left\| x - \mathrm{Res}^f_s x\right\| \\ &\qquad - rs^{-1}2b\left\|\nabla f \mathrm{Res}^f_s x - \nabla f x\right\| \\ &\geqslant D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) \\ &\qquad - E\left(\left\| x - \mathrm{Res}^f_s x\right\| + \left\|\nabla f \mathrm{Res}^f_s x - \nabla f x\right\|\right) \end{aligned}$$

and therefore, for x such that

$$\left\| x - \mathrm{Res}^f_s x\right\| < \omega^{\nabla f}\left(\frac{\varepsilon}{2E}, b\right),$$

we get that

$$D_f(x,y) > D_f(x,\mathrm{Res}^f_r y) + D_f(\mathrm{Res}^f_r y, y) - \varepsilon$$

which is the claim. □

As a concrete instantiation of Theorem 9.5.10, we now obtain the following

Theorem 9.6.5. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet* *differentiable and totally convex on bounded subsets. Let A be a maximally monotone operator with resolvents Res^f_γ and assume that $A^{-1}0 \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f \mathrm{Res}^f_{r_n} x_n)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(r_n) \in (0,\infty)$ satisfies $0 < \bar{r} = \inf\{r_n \mid n \in \mathbb{N}\}$. Then (x_n) converges strongly to $P^f_{A^{-1}0}(u)$.

In particular, a rate of metastability can be calculated using Theorem 9.5.9 together with Lemmas 9.3.5 and 9.3.8 and with moduli

$$\mu(\varepsilon, b, K) = \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/2E'(K), \widehat{b}) \text{ and } \nu(\varepsilon, b) = \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/4\widehat{b}, \widehat{b})$$

for the uniform NST condition where $E'(K) = \max\{2\widehat{b}, R(K)\bar{r}^{-1}2\widehat{b}\}$ and $\widehat{b} = \max\{b, E(b)\}$ as well as $R(n) = \max\{r_k \mid k \leqslant n\}$ and where E is a modulus for Res^f_γ being bounded on bounded sets.

Proof. Note that using Lemmas 9.3.5 and 9.3.8 as well as $A^{-1}0 = \mathrm{F}(\mathrm{Res}^f_r)$ for any $r > 0$, it is immediate that the $\mathrm{Res}^f_{r_n}$ and $\mathrm{Res}^f_{\bar{r}}$ are commonly uniformly Bregman strongly nonexpansive and commonly bounded on bounded sets and corresponding moduli can be calculated. This also yields that a modulus of uniform closedness exists for $F(\mathrm{Res}^f_{\bar{r}})$.

The only thing left to prove is that the constructed μ and ν witness the uniform NST condition for $T_n = \mathrm{Res}^f_{r_n}$ and $T = \mathrm{Res}^f_{\bar{r}}$. By Lemma 9.6.4, we get that

$$\left\|x - \mathrm{Res}^f_s x\right\| < \omega^{\nabla f}(\varepsilon/2E', \widehat{b})$$

for $\|x\| \leqslant b$ implies that

$$D_f(x, \mathrm{Res}^f_r x) < \varepsilon$$

for $E' \geqslant \max\{2\widehat{b}, rs^{-1}2\widehat{b}\}$ and $\widehat{b} = \max\{b, E(b)\}$. In particular, we have

$$\left\|x - \mathrm{Res}^f_r x\right\| < \varepsilon$$

for any x with $\|x\| \leqslant b$ and

$$\left\|x - \mathrm{Res}^f_s x\right\| < \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/2E', \widehat{b}).$$

So, for $s \geqslant r$ we get for $\|x\| \leqslant b$ and

$$\left\|x - \mathrm{Res}^f_s x\right\| < \omega^{\nabla f}(\rho(\varepsilon, \widehat{b})/4\widehat{b}, \widehat{b})$$

that $\left\|x - \mathrm{Res}^f_{\bar{r}} x\right\| < \varepsilon$. Therefore, as $\bar{r} \leqslant r_n$ for all n, we get that ν indeed satisfies $(\dagger)_2$ for the given T_n and T.

Further, assuming that $\left\|x - \mathrm{Res}^f_{\bar{r}} x\right\| < \mu(\varepsilon, b, K)$, we get by the above that

$$\left\|x - \mathrm{Res}^f_{r_n} x\right\| < \varepsilon$$

as $E'(K) = \max\{2\widehat{b}, R(K)\bar{r}^{-1}2\widehat{b}\} \geqslant \max\{2\widehat{b}, r_n\bar{r}^{-1}2\widehat{b}\}$ for $n \leqslant K$. Thus μ satisfies $(\dagger)_1$ for the given T_n and T. □

9.6.3 A rate of convergence for the asymptotic regularity of the Halpern-type proximal point algorithm relative to resolvents in the case of $r_n \to \infty$

The convergence proof of the previous Halpern-type proximal point algorithm relies on an argument revolving around a case distinction and (essentially) because of this, we are not able to derive full rates of convergence for the asymptotic regularity relative to the resolvents, i.e. rates for the convergence

$$\left\|x_n - \mathrm{Res}^f_{\gamma} x_n\right\| \to 0 \quad (n \to \infty)$$

for $\gamma > 0$. In this section, we consider the previous Halpern-type proximal point algorithm under the additional condition $r_n \to \infty$ (conceptually similar to the work of Kohsaka and Takahashi [121]) for which we are able to derive full rates of convergence for the asymptotic regularity relative to the resolvents. In the case of Hilbert spaces with the ordinary Halpern-type proximal point algorithm induced by a maximally monotone operator, such a rate of convergence (in the context of the assumption of $r_n \to \infty$ similar to here) was first given by Pinto in [161].

Lemma 9.6.6. *Let $b \geqslant \|u\|, \|x_n\|, \left\|\mathrm{Res}^f_{r_n} x_n\right\|$ for all n with (x_n) defined as in $(**)$ and let σ be a rate of convergence for $\alpha_n \to 0$ as $n \to \infty$. Let further C be a modulus for ∇f being bounded on bounded sets. Then, for any $\varepsilon > 0$:*

$$\forall n \geqslant \sigma\left(\frac{\varepsilon}{2C(b)}\right)\left(\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| < \varepsilon\right).$$

In particular, if $\omega^{\nabla f^*}$ is a modulus of uniform continuity for ∇f^* on bounded subsets, then for any $\varepsilon > 0$:

$$\forall n \geqslant \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b))}{2C(b)}\right)\left(\left\|x_{n+1} - \mathrm{Res}^f_{r_n} x_n\right\| < \varepsilon\right).$$

Proof. As before, we have

$$\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| = \alpha_n \left\|\nabla f u - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| \leqslant \alpha_n 2C(b).$$

It immediately follows from the assumption on σ that

$$\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| \leqslant \alpha_n 2C(b) < \varepsilon$$

for all $n \geqslant \sigma(\varepsilon/2C(b))$. The second part of the lemma is immediate. □

Theorem 9.6.7. *Let $\gamma > 0$ be given. Assume that $b > 0$ is such that*

$$b \geqslant \|u\|, \|x_n\|, \left\|\mathrm{Res}^f_{r_n} x_n\right\|, \gamma, \left\|\mathrm{Res}^f_{\gamma}\mathrm{Res}^f_{r_n} x_n\right\|, \left\|\mathrm{Res}^f_{\gamma} x_n\right\|$$

for all n with (x_n) defined as in $(**)$. Let P be a modulus of reverse consistency. Assume that $\widehat{\eta}$ is a modulus of uniform strict monotonicity of ∇f on bounded sets and let ρ be a modulus of consistency for f. Let further C be a modulus for ∇f being bounded on bounded sets and let σ be a rate of convergence for $\alpha_n \to 0$ as $n \to \infty$. Let τ be a rate of divergence for $r_n \to \infty$ as $n \to \infty$, i.e.

$$\forall E > 0 \forall n \geqslant \tau(E)\,(r_n > E).$$

Let $\omega^{\nabla f^*}$ is a modulus of uniform continuity for ∇f^* on bounded subsets and let $\omega^{\nabla f}(\varepsilon, b) \leqslant \varepsilon$ be a modulus of uniform continuity of ∇f on bounded subsets. Then for any $\varepsilon > 0$:

$$\forall n \geqslant \Phi(\varepsilon)\left(\left\|x_{n+1} - \mathrm{Res}^f_{\gamma} x_{n+1}\right\| < \varepsilon\right)$$

where

$$\Phi(\varepsilon) := \max\left\{\tau\left(\frac{2C(b)}{\chi\left(\omega^{\nabla f}\left(\frac{\rho(\omega^{\nabla f}(\omega^{\nabla f^*}(\varepsilon, C(b))/2, b), b)}{8b}, b\right)\right)}\right), \varphi\left(P\left(\frac{\rho(\omega^{\nabla f}(\omega^{\nabla f^*}(\varepsilon, C(b))/2, b), b)}{2}, b\right)\right), \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b)/2}{2C(b)}\right)\right\}$$

and

$$\chi(\varepsilon) := \frac{\widehat{\eta}(\varepsilon, b)}{2b^2},\ \varphi(\varepsilon) := \sigma\left(\frac{\omega^{\nabla f^*}(\varepsilon, C(b))}{2C(b)}\right).$$

Proof. Note that

$$\left\|A^f_{r_n} x_n\right\| = \frac{1}{r_n}\left\|\nabla f x_n - \nabla f \mathrm{Res}^f_{r_n} x_n\right\| \leqslant \frac{2C(b)}{r_n}$$

and thus

$$\forall \varepsilon > 0 \forall n \geqslant \tau\left(\frac{2C(b)}{\varepsilon}\right)\left(\left\|A^f_{r_n}x_n\right\| < \varepsilon\right).$$

Therefore, since $A^f_{r_n}x_n \in A(\mathrm{Res}^f_{r_n}x_n)$, we have that

$$n \geqslant \tau\left(\frac{2C(b)}{\chi(\varepsilon)}\right)$$

implies $\left\|\mathrm{Res}^f_{r_n}x_n - \mathrm{Res}^f_{\gamma}\mathrm{Res}^f_{r_n}x_n\right\| < \varepsilon$ by Lemma 9.6.2. Therefore, we have for

$$n \geqslant \tau\left(\frac{2C(b)}{\chi(\omega^{\nabla f}\left(\frac{\varepsilon}{8b}, b\right))}\right)$$

that

$$\left\|\mathrm{Res}^f_{r_n}x_n - \mathrm{Res}^f_{\gamma}\mathrm{Res}^f_{r_n}x_n\right\| < \omega^{\nabla f}\left(\frac{\varepsilon}{8b}, b\right),$$

and thus

$$D_f(\mathrm{Res}^f_{r_n}x_n, \mathrm{Res}^f_{\gamma}x_{n+1}) \leqslant D_f(\mathrm{Res}^f_{r_n}x_n, x_{n+1}) + \varepsilon/2$$

in that case by Lemma 9.6.4 (with $s := r := \gamma$ and using $E = 2b$). Now, for

$$n \geqslant \max\left\{\tau\left(\frac{2C(b)}{\chi(\omega^{\nabla f}\left(\frac{\varepsilon}{8b}, b\right))}\right), \varphi(P(\varepsilon/2, b))\right\}$$

we get $D_f(\mathrm{Res}^f_{r_n}x_n, \mathrm{Res}^f_{\gamma}x_{n+1}) \leqslant D_f(\mathrm{Res}^f_{r_n}x_n, x_{n+1}) + \varepsilon/2$ from before as well as that $D_f(\mathrm{Res}^f_{r_n}x_n, x_{n+1}) < \varepsilon/2$ by Lemma 9.6.6 and the assumption on P. Thus in that case, we also have

$$D_f(\mathrm{Res}^f_{r_n}x_n, \mathrm{Res}^f_{\gamma}x_{n+1}) < \varepsilon.$$

Thus for

$$n \geqslant \max\left\{\tau\left(\frac{2C(b)}{\chi\left(\omega^{\nabla f}\left(\frac{\rho(\varepsilon,b)}{8b}, b\right)\right)}\right), \varphi(P(\rho(\varepsilon,b)/2, b))\right\}$$

we get $\left\|\mathrm{Res}^f_{r_n}x_n - \mathrm{Res}^f_{\gamma}x_{n+1}\right\| < \varepsilon$ using the assumption on ρ. Now, note that

$$\begin{aligned}\|\nabla f x_{n+1} - \nabla f \mathrm{Res}_{\gamma}x_{n+1}\| &\leqslant \alpha_n \left\|\nabla f u - \nabla f \mathrm{Res}^f_{r_n}x_n\right\| + \left\|\nabla f \mathrm{Res}^f_{r_n}x_n - \nabla f \mathrm{Res}^f_{\gamma}x_{n+1}\right\| \\ &\leqslant \alpha_n 2C(b) + \left\|\nabla f \mathrm{Res}^f_{r_n}x_n - \nabla f \mathrm{Res}^f_{\gamma}x_{n+1}\right\|\end{aligned}$$

Thus, for

$$n \geqslant \max\left\{\tau\left(\frac{2C(b)}{\chi\left(\omega^{\nabla f}\left(\frac{\rho(\omega^{\nabla f}(\varepsilon/2,b),b)}{8b}, b\right)\right)}\right),\right.$$
$$\left.\varphi\left(P\left(\frac{\rho(\omega^{\nabla f}(\varepsilon/2,b),b)}{2}, b\right)\right), \sigma\left(\frac{\varepsilon/2}{2C(b)}\right)\right\}$$

we get $\left\|\nabla f x_{n+1} - \nabla f \mathrm{Res}^f_{\gamma}x_{n+1}\right\| < \varepsilon$. This gives the claim using $\omega^{\nabla f^*}$. □

As before, a b bounding all objects involved can be constructed using the range of moduli discussed before together with some simple initial bounds. We refrain from spelling this out in more detail.

9.6.4 Finding common zeros of maximally monotone operators

Another readily defined instantiation of Theorem 9.5.10 on the Halpern-iteration is that of finding common zeros of a finite collection $(A_i)_{i=1,\dots,N}$ of maximally monotone operators with $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$. Similar to the idea in [152], we in that context can consider a composite operator

$$Tx = \nabla f^* \sum_{i=1}^{N} w_i \nabla f \mathrm{Res}^{\mathrm{f}}_{\mathrm{A_i}}$$

for weights $w_i \in (0,1)$ such that $\sum_{i=1}^{N} w_i = 1$. Then T is a block operator in the sense of [143, 144] (as also discussed in the previous sections) and moduli for the uniform Bregman strong nonexpansivity for this operator can be calculated from the moduli of the summands following Theorem 9.3.19. From Lemma 9.3.17, also a modulus for T being bounded on bounded sets can be calculated from corresponding moduli for $\nabla f, \nabla f^*$ and $\mathrm{Res}^f_{A_i}$ being bounded on bounded sets (using Lemma 9.3.5, the latter of which in particular exists as $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$ as any $\mathrm{Res}^f_{A_i}$ is Bregman firmly nonexpansive and thus Bregman quasi-nonexpansive). Lastly, note that by Lemma 9.3.13, each $\mathrm{Res}^f_{A_i}$ is uniformly continuous on bounded sets and it is easy to see that, since ∇f, ∇f^* are also uniformly continuous, this extends to T as well. Therefore, a corresponding modulus of uniform closedness exists for $F(T)$.

Combining this with Theorem 9.5.10, we get the following corollary on the approximation of common zeros:

Theorem 9.6.8. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let $A_1, \dots, A_N$ be maximally monotone operators with resolvents $\mathrm{Res}^f_{A_i}$ at parameter 1. Assume that $A_1^{-1}0 \cap \dots \cap A_N^{-1}0 \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^* \left(\alpha_n \nabla f u + (1-\alpha_n) \sum_{i=1}^{N} w_i \nabla f \mathrm{Res}^{\mathrm{f}}_{\mathrm{A_i}} x_n \right)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where the $w_i \in (0,1)$ are such that $\sum_{i=1}^{N} w_i = 1$. Then (x_n) converges strongly to $P^f_{F(T)}(u)$ for T defined as above.

In particular, a rate of metastability can be calculated using Theorem 9.5.9 together with Lemmas 9.3.5 and 9.3.8 as well as Theorems 9.3.19 and 9.3.18.

9.6.5 Bregman forward-backward splitting

We may also consider a forward-backward type of iteration in conjunction with the Halpern-type algorithm considered before. By abstracting the general approach taken in [155] for combining Bregman distances and forward-backward splitting methods, Bùi and Combettes in [32] considered the following iteration under suitable conditions on the scalars and f_n's:

$$x_{n+1} = (\nabla f_n + \gamma_n A)^{-1}(\nabla f_n x_n - \gamma_n B x_n)$$

for given $A : X \to 2^{X^*}$ and $B : X \to X^*$. At least in the context of the special case where $f_n = f$ for all n for a specific f as considered throughout this chapter and where the operator B is of the form

$$Bx = \nabla f x - \nabla f C x$$

for a given uniformly Bregman strongly nonexpansive map $C : X \to X$, we can now provide a strong convergence result for a Halpern-type variant of this method. For this, similar to [36], consider the map

$$B^f_\lambda x = \nabla f^*(\nabla f x - \lambda B x).$$

It is straightforward to verify that $B^f_\lambda x = \nabla f^*(\lambda \nabla f C x + (1-\lambda)\nabla f x)$. In particular, B^f_λ is a block operator for any $\lambda \in [0,1]$ and thus is uniformly Bregman strongly nonexpansive (where a corresponding strong BSNE-modulus can be computed from a modulus for C according to Theorem 9.3.19).

In particular, we have that

$$\mathrm{Res}^f_\lambda \circ B^f_\lambda = (\nabla f + \lambda A)^{-1}(\nabla f - \lambda B)$$

which is exactly of the form considered in [32]. We assume that $F(\mathrm{Res}^f_\lambda \circ B^f_\lambda) \neq \emptyset$. It is rather immediate to show that $(A+B)^{-1}0 = F(\mathrm{Res}^f_\lambda \circ B^f_\lambda)$. Thus, as a composition of uniformly Bregman strongly nonexpansive maps, it is itself uniformly Bregman strongly nonexpansive if we require a modulus of uniform closedness for $F(C)$. Note for this that $F(\mathrm{Res}^f_\lambda)$ is naturally uniformly closed since Res^f_λ is even Bregman firmly nonexpansive and that using Lemma 9.3.18, it can be easily seen that from the uniform closedness of $F(C)$, we get the uniform closedness of $F(B^f_\lambda)$. Further, note that using Lemma 9.3.14, we thus also get that $F(\mathrm{Res}^f_\lambda \circ B^f_\lambda)$ is uniformly closed. Also, the composition is therefore also bounded on bounded sets by Lemma 9.3.5 as we assume that $F(\mathrm{Res}^f_\lambda \circ B^f_\lambda) \neq \emptyset$.

We thus get the following corollary on a Halpern-type forward-backward splitting method using Bregman distances:

Theorem 9.6.9. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let C be uniformly Bregman strongly nonexpansive such that $F(C)$ is uniformly closed and let $Bx = \nabla fx - \nabla fCx$. Assume that $(A + B)^{-1}0 \neq \emptyset$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^*(\alpha_n \nabla fu + (1 - \alpha_n)\nabla f(\nabla f + \lambda A)^{-1}(\nabla fx_n - \lambda Bx_n))$$

for $\lambda > 0$ where $(\alpha_n) \subseteq (0, 1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{(A+B)^{-1}0}(u)$.

In particular, a rate of metastability can be calculated using Theorem 9.5.9 together with Theorems 9.3.19 and 9.3.18.

As a further special case, we want to note that for a B of this form, this also covers a Halpern-type variant of the iteration studied in [36] regarding the solution of operator equations of the form $Bx = 0$ for an operator $B : X \to X^*$ over a closed and convex set Ω. Concretely, in [36], the authors considered the iteration

$$x_{n+1} = \Pi^f_\Omega(\nabla fx_n - \lambda Bx_n)$$

where $\Pi^f_\Omega = P^f_\Omega \circ \nabla f^*$ for the Bregman projection P^f_Ω. As discussed in the previous section, we have that

$$P^f_\Omega = \mathrm{Res}^f_{\partial\iota_\Omega}$$

where ι_Ω is the characteristic function of Ω. In particular, as we then have

$$\Pi^f_\Omega(\nabla fx - \lambda Bx) = P^f_\Omega B^f_\lambda x$$

with $B^f_\lambda x = \nabla f^*(\nabla fx - \lambda Bx)$ as before, we find that the above iteration is a special case (if we would relax the parameter λ of Res^f_λ to a separate parameter) of the previous forward-backward method and we thus also obtain the following corollary:

Theorem 9.6.10. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let C be uniformly Bregman strongly nonexpansive such that $F(C)$ is uniformly closed and let $Bx = \nabla fx - \nabla fCx$. Assume that $B^{-1}0 \cap \Omega \neq \varnothing$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)\nabla f \Pi^f_{\Omega}(\nabla f x_n - \lambda B x_n))$$

for $\lambda > 0$ where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$. Then (x_n) converges strongly to $P^f_{B^{-1}0\cap\Omega}(u)$.

In particular, a rate of metastability can be calculated using Theorem 9.5.9 together with Theorems 9.3.19 and 9.3.18.

9.6.6 Modified-Halpern, Tikhonov-Mann and Halpern-Mann type methods

In this last subsection, we are concerned with a few generalizations of Halpern-type iterations that incorporate elements from Mann-type iterations. The first such generalization that we consider is the modified Halpern iteration as introduced in [85] (see also [53])

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

where (γ_n) and (α_n) are sequences in $[0,1]$ and $T : X \to X$ is a given mapping. Such a type of iteration has been considered for Bregman strongly nonexpansive maps in [214] under the name of *Halpern-Mann iterations.* Concretely, in [214] the authors proved the strong convergence of the iteration

$$x_{n+1} = \nabla f^*(\alpha_n \nabla fu + (1-\alpha_n)(\beta_n \nabla f x_n + (1-\beta_n)\nabla f T x_n))$$

under the scalar conditions that (α_n) and (β_n) are sequences in $(0,1)$ satisfying

1. $\alpha_n \to 0$ for $n \to \infty$,
2. $\sum \alpha_n = +\infty$,
3. $0 < \liminf \beta_n \leqslant \limsup \beta_n < 1$.

We begin by showing that for uniformly Bregman strongly nonexpansive maps, the convergence of this iteration can be derived by our previous result for families of maps.

For this, note first that the above iteration is nothing else but a usual Halpern-type iteration of the family of operators

$$T_n x = \nabla f^*(\beta_n \nabla f x + (1-\beta_n)\nabla f T x)$$

Assume $F(T) \neq \varnothing$. Then these operators, being block operators, are uniformly Bregman strongly nonexpansive and using Theorem 9.3.19, we can construct even a common strong BSNE-modulus from a strong BSNE-modulus of T using the assumption of a fixed point for T. Note however that for this, condition (3) is not needed at all and $\beta_n \in [0,1]$ can be permitted. Also note that the T_n together with T are commonly bounded on bounded sets by using Lemmas 9.3.17 and 9.3.5 together with the assumption of a fixed point for T.

To see that this sequence is permissible for our Halpern-type iteration for families of maps, we need to again provide concrete instantiations of the moduli μ and ν witnessing the uniform NST condition for the choice of these T_n together with the map T. For this, it is rather immediately clear that given moduli E, C, F for T, ∇f, ∇f^* being bounded on bounded sets as well as a modulus of consistency ρ and a modulus of reverse consistency P, one has that

$$\mu(\varepsilon, b, K) = P(\rho(\varepsilon, \max\{b, F(C(E(b)))\}), \max\{b, E(b)\})$$

suffices as we immediately have for given $\varepsilon, b > 0$ and $p \in X$ with $\|p - Tp\| < \mu(\varepsilon, b, K)$ that

$$\begin{aligned} D_f(p, T_n p) &\leqslant (1-\beta_n) D_f(p, Tp) \\ &\leqslant D_f(p, Tp) \\ &< \rho(\varepsilon, \max\{b, F(C(E(b)))\}) \end{aligned}$$

so that $\|p - T_n p\| < \varepsilon$.

For ν, assume that we have an $N_{\overline{\beta}}$ and a $\overline{\beta} < 1$ with $\beta_n \leqslant \overline{\beta}$ for all $n \geqslant N_{\overline{\beta}}$ (witnessing $\limsup_n \beta_n < 1$), a modulus of consistency ρ, a modulus of uniform continuity of D_f in its second argument ξ, a BSNE-modulus ω for T, moduli E, C, F for T, ∇f, ∇f^* being bounded on bounded sets, and a fixed point of T named p_0 with $b \geqslant \|p_0\|$. Then by Theorem 9.3.18: for any x with $\|x\| \leqslant b$, we have

$$\|x - T_n x\| < \xi\left((1-\overline{\beta})\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right) \to \|x - Tx\| < \varepsilon$$

for $n \geqslant N_{\overline{\beta}}$ where $\widehat{b} = \max\{b, E(b), F(C(E(b)))\}$ so that

$$\nu(\varepsilon, b) = \xi\left((1-\overline{\beta})\omega(\rho(\varepsilon, \widehat{b}), b), \widehat{b}\right)$$

suffices (after suitably shifting the sequence with $N_{\overline{\beta}}$). Combined, we thus derive the following result from Theorem 9.5.10.

Theorem 9.6.11. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be uniformly Bregman strongly nonexpansive with $F(T) \neq \varnothing$. Given a $u \in X$, define a sequence x_n by $x_0 = x \in X$ and*

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)(\beta_n \nabla f x_n + (1-\beta_n)\nabla f T x_n))$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(\beta_n) \subseteq [0,1)$ satisfies $\limsup \beta_n < 1$. If $\overline{F}(T) \subseteq F(T)$, then (x_n) converges strongly to $P^f_{F(T)}(u)$.

In particular, a rate of metastability can be obtained by suitably instantiating the rate given in Theorem 9.5.9 using Theorems 9.3.19 and 9.3.18 as well as the above moduli for μ and ν.

In particular, with this theorem we reobtain the strong convergence result for this iteration established in [214] (recall for this Remark 9.5.11) for uniformly Bregman strongly nonexpansive maps. However, the assumption (3) presented above which features in [214] could be substantially weakened to $\limsup \beta_n < 1$ which in particular allows $\beta_n = 0$ for all n. Thus, in the above iteration, the Mann-part can be "deactivated" and the original Halpern-type result can be reobtained, contrary to [214].

The other generalization of Halpern's method which we consider is an iteration of Tikhonov–Mann type. In the usual metric case, this type of iteration takes the form of

$$y_{n+1} = (1-\lambda_n)((1-\beta_n)u + \beta_n x_n) + \lambda_n T((1-\beta_n)u + \beta_n x_n)$$

as defined in [42] where (λ_n), (β_n) are sequences in $[0,1]$ and $T : X \to X$ is a again a given mapping. In particular, for $u = 0$, this iteration becomes the modified Mann iteration as studied in [210] and rediscovered in the seminal work by Boţ, Csetnek and Meier [20]. For these types of iterations, we can now prove a (new) strong convergence result for the following natural analog in the context of Bregman strongly nonexpansive maps:

$$y_{n+1} = \nabla f^*(\beta_n \nabla f u_n + (1-\beta_n)\nabla f T u_n) \text{ with } u_n = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f y_n).$$

As discussed in [41], methods of a modified Halpern type as well as methods of a Tikhonov-Mann type in both a normed and a hyperbolic context are closely related and in fact can be translated into each other.

By suitably adapting the arguments from [41] to this Bregman case, we arrive at the following result (which is similar to Proposition 3.2 in [41]):

Lemma 9.6.12. *Define the iterations*

$$x_{n+1} = \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fv_n) \text{ with } v_n = \nabla f^*(\beta_n\nabla fx_n + (1-\beta_n)\nabla fTx_n))$$

as well as

$$y_{n+1} = \nabla f^*(\beta_n\nabla fu_n + (1-\beta_n)\nabla fTu_n) \text{ with } u_n = \nabla f^*(\alpha_n\nabla fu + (1-\alpha_n)\nabla fy_n).$$

If $x_0 = \nabla f^*(\alpha_0\nabla fu + (1-\alpha_0)\nabla fy_0)$, *then for any* $n \in \mathbb{N}$*:*

$$u_n = x_n \text{ and } y_{n+1} = v_n.$$

Proof. The proof is by induction on n. For $n = 0$, it follows by the definition of u_0 as well as the assumption on x_0 that $x_0 = u_0$. From that, we get

$$\begin{aligned} y_1 &= \nabla f^*(\beta_0\nabla fu_0 + (1-\beta_0)\nabla fTu_0) \\ &= \nabla f^*(\beta_0\nabla fx_0 + (1-\beta_0)\nabla fTx_0) \\ &= v_0. \end{aligned}$$

For the induction step, suppose now that $u_n = x_n$ and $y_{n+1} = v_n$. Then

$$\begin{aligned} x_{n+1} &= \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fv_n) \\ &= \nabla f^*(\alpha_{n+1}\nabla fu + (1-\alpha_{n+1})\nabla fy_{n+1}) \\ &= u_{n+1} \end{aligned}$$

where the second equality follows by induction hypothesis. Further, we thus have

$$\begin{aligned} y_{n+2} &= \nabla f^*(\beta_{n+1}\nabla fu_{n+1} + (1-\beta_{n+1})\nabla fTu_{n+1}) \\ &= \nabla f^*(\beta_{n+1}\nabla fx_{n+1} + (1-\beta_{n+1})\nabla fTx_{n+1}) \\ &= v_{n+1}. \end{aligned}$$

□

Together with the above theorem, this allows us to derive the following new strong convergence result:

Theorem 9.6.13. *Let X be a real reflexive Banach space and $f : X \to \mathbb{R}$ be a supercoercive Legendre function which is bounded on bounded sets, uniformly Fréchet differentiable and totally convex on bounded subsets. Let T be uniformly Bregman strongly nonexpansive with $F(T) \neq \emptyset$. Given a $u \in X$, define a sequence y_n by $y_0 = y \in X$ and*

$$y_{n+1} = \nabla f^*(\beta_n \nabla f u_n + (1-\beta_n)\nabla f T u_n) \text{ with } u_n = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f y_n)$$

where $(\alpha_n) \subseteq (0,1]$ satisfies $\lim \alpha_n = 0$ and $\sum \alpha_n = \infty$ and where $(\beta_n) \subseteq [0,1)$ satisfies $\limsup \beta_n < 1$. If $\overline{F}(T) \subseteq F(T)$, then (y_n) converges strongly to $P^f_{F(T)}(u)$.

In particular, a rate of metastability can be obtained by suitably translating the rate from Theorem 9.6.11.

Proof. It suffices to show that given a rate of metastability Ω for the sequence x_n as defined in Theorem 9.6.11 (with α_{n+1} instead of α_n), i.e. Ω satisfying

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

we can construct a rate of metastability for y_n.

For this, note first that $\|y_n - u_n\| \to 0$ for $n \to \infty$ and we can witness this limit even by a rate of convergence. To see this, let $\bar{b}$ be such that $\bar{b} \geqslant D_f(y_n, u), \|y_n\|, \|u_n\|$ for all n.[13] Let σ be a rate of convergence for $\alpha_n \to 0$ as before. Then we get

$$\begin{aligned} D_f(y_n, u_n) &\leqslant \alpha_n D_f(y_n, u) + (1-\alpha_n) D_f(y_n, y_n) \\ &= \alpha_n D_f(y_n, u) \end{aligned}$$

so that for $n \geqslant \sigma(\varepsilon/\bar{b})$, we have $D_f(y_n, u_n) < \varepsilon$. In particular, for $n \geqslant \sigma(\rho(\varepsilon, \bar{b})/\bar{b})$ we get $\|y_n - u_n\| < \varepsilon$.

We can now construct a rate of metastability for y_n given one for x_n. At first, using Lemma 9.6.12, we get $u_n = x_n$ for all n so that Ω is also a rate of metastability for u_n. Then

$$\|y_i - y_j\| \leqslant \|y_i - u_i\| + \|u_i - u_j\| + \|u_j - y_j\|$$

and by reasoning similar to [41], it can be rather immediately seen that Ω' defined by

$$\begin{aligned} \Omega'(\varepsilon, g) &= \widetilde{\Omega}(\varepsilon/3, g, \sigma(\rho(\varepsilon/3, \bar{b})/\bar{b})), \\ \widetilde{\Omega}(\varepsilon, g, q) &= \Omega(\varepsilon, g_q) + q \text{ with } g_q(n) := g(n+q) + q, \end{aligned}$$

is therefore a rate of metastability for y_n. □

[13] Such a $\bar{b}$ can naturally be constructed from a $b \geqslant D_f(p, u), D_f(p, y_0)$ for a given fixed point p together with a modulus of boundedness for D_f and moduli for ∇f, f being bounded on bounded sets. We omit the details.

As the last application, we sketch how these constructions can be extended to define another new strongly convergent method for common fixed points of two uniformly Bregman strongly nonexpansive maps. Inspired by the recently introduced alternating Halpern-Mann type method by Dinis and Pinto [57], we consider the iteration

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f u + (1-\alpha_n)\nabla f S \nabla f^*(\beta_n \nabla f x_n + (1-\beta_n)\nabla f T x_n))$$

for two uniformly Bregman strongly nonexpansive maps $S, T : X \to X$ with similar conditions on the β_n and α_n.

Then, in similarity to before, the iteration is nothing else but the Halpern-type iteration of the family of maps $S \circ T_n$ for T_n as before. We can therefore derive a rate of metastability for (and therefore establish the convergence of) this sequence whenever the family $S \circ T_n$ possesses the quantitative NST-style moduli μ and ν relative to some other map R where we can exhibit a common strong BSNE-modulus. To find such an R, note that we can employ the previous Theorems 9.3.14 and 9.3.18 to derive that approximate fixed points of $S \circ T_n$ are approximate common fixed points of T and S. Using this, a modulus ν relating $S \circ T_n$ to $\nabla f^*((\nabla f S + \nabla f T)/2)$ (which has exactly as fixed points the common fixed points of S and T) can be constructed. This modulus will in particular depend on moduli of uniform closedness for $F(S)$ and $F(T_n)$ (the latter being definable from a corresponding modulus for $F(T)$).

Conversely, any approximate fixed point of $\nabla f^*((\nabla f S + \nabla f T)/2)$ can be shown to be a common fixed point of S and T by using Theorem 9.3.18 and thus of S and T_n. If we further have a modulus of uniform closedness for $F(S)$, then we can infer that an approximate fixed point of $\nabla f^*((\nabla f S + \nabla f T)/2)$ is also an approximate fixed point of $S \circ T_n$ from which we can extract a modulus μ.

A common strong BSNE-modulus can then be constructed using Theorems 9.3.15 and 9.3.19.

As these moduli will depend on moduli of uniform closedness, the resulting convergence theorem for the iteration above will in particular only hold for mappings where such a moduli exist. This in particular includes uniformly continuous mappings and thus in particular covers the case of Bregman firmly nonexpansive mappings which are bounded on bounded sets and a fortiori also the usual firmly nonexpansive maps in Hilbert spaces with which one in that case can re-obtain the convergence of the alternating Halpern-Mann Douglas-Rachford method from [57].

Presumably, a Tikhonov-Mann type variant of this iteration could be defined as well, relating to the above via a similar argument as in Lemma 9.6.12. We omit any concrete details on how all of this can be formally spelled out.

10 Monotone operators in Banach spaces and their resolvents

10.1 Introduction

In this chapter, we extend the considerations of Chapter 3 to monotone operators on Banach spaces which were already considered in the previous Chapter 9. To treat these operators (which also require the use of the dual of the underlying Banach space), we rely on the system $\mathcal{D}^\omega$ introduced in Chapter 8.

Even though the setting, being in the context of the dual of the space, is different in this chapter, the chosen approach to the set-valued operators is the same as in Chapter 3 and in that way, the present chapter further elucidates the naturalness and applicability of the methods developed therein to treat set-valued operators of various types.

Besides treating these operators, this chapter also provides a proof-theoretic treatment of the resolvents relative to a convex function f already discussed and used in Chapter 9 (and in that vein, we also rely on Chapter 9 for some analytical background). To that end, we show that the main properties of the operator and the resolvents relative to f are provable in the system that we define. Also, we show that the equivalence between maximality of the operator and extensionality discussed in Chapter 3 extends to these new objects.

At last, we extend the bound extraction results from the previous chapters to these systems. These new metatheorems in particular fully explain the applications given in Chapter 9 for these operators and their relativized resolvents.

10.2 Logical systems for operators and their resolvents

At first, we want to mention that all the considerations made here could be extended mutatis mutandis to the case where we only consider partial convex functions f with an intensional treatment of the domain as discussed in Chapter 8 but, for simplicity, we refrain from spelling this out in detail.

Also, while we previously have divided the treatment of resolvents on whether they are partial or total, we in the following will only consider systems for monotone operators on Banach spaces where the resolvents are all total. If one would want to treat operators with partial resolvents, then a similar approach as presented in Chapter 3 could also be followed here.

10.2.1 Further considerations on convex functions

The basic system for all extensions considered here will be $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ from Chapter 8, treating the dual of the abstract normed space together with a convex function, its Fenchel-conjugate and their uniformly continuous gradients. It will be convenient to slightly extend this system so that the theory of monotone operators can be developed smoothly. Concretely, it will be convenient to include a few more properties of the Fenchel conjugate f^* axiomatically into the previous systems: by the Fenchel–Moreau theorem (see e.g. [21]), we know that if f is proper, lower-semicontinuous and convex, then f^* is proper and $f = f^{**}$ where we define $f^{**} : X \to (-\infty, +\infty]$ by

$$f^{**}(x) := \sup_{x^* \in X^*} (\langle x, x^* \rangle - f^*(x^*)).$$

With this definition, we follow one particular approach to biconjugates as e.g. outlined in [21]. In other works, one finds f^{**} introduced as $(f^*)^*$ acting on X^{**} and thus on X by its embedding into X^{**} (which coincides with X in the context of reflexivity). As the spaces considered in the context of $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ are anyhow superreflexive by the results of [18], these different approaches yield the same object but the above formulation will also influence the types of objects considered later.

This fact that $f = f^{**}$ is crucial for the development of the theory of monotone operators and we need to deal with it formally. Naturally, a function f as axiomatized by the system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ satisfies the assumptions of the Fenchel-Moreau theorem but instead of analyzing the proof, we instead hardwire this fact into the system akin to how f^* was treated in Chapter 8.

In more detail, note that $f = f^{**} = (f^*)^*$ is bounded on bounded sets and therefore f^* is supercoercive by Proposition 8.5.7. So $f = f^{**}$ can be wired into the system by using a modulus of supercoercivity α^{f^*} for f^* together with the following axioms instantiating the schemes $(S)_1$, $(S)_2$ as in Chapter 8:

$(f^*)_3$ f^* is supercoercive with modulus α^{f^*}, i.e.

$$\forall K^0, x^{*X^*} \left(\|x^*\|_{X^*} >_{\mathbb{R}} \alpha^{f^*}(K) \to f^*(x^*)/\|x^*\|_{X^*} \geqslant_{\mathbb{R}} K \right).$$

Here, α^{f^*} is an additional constant of type 1.

$(f^{**})_1$ f is the pointwise upper bound for all affine functionals $g_{x^*}(x) = \langle x, x^* \rangle - f^*(x^*)$, i.e.

$$\forall x^X, x^{*X^*} \left(\langle x, x^* \rangle_{X^*} - f^*(x^*) \leqslant_{\mathbb{R}} f(x) \right).$$

$(f^{**})_2$ f is indeed the pointwise supremum of these affine functionals, i.e.

$$\forall x^X, b^0, k^0 \exists x^{*X^*} \leqslant_{X^*} \max\{\alpha^{f^*}(b+1)+1, [|f^*(0)|](0)+2\}1_{X^*}$$
$$\left(\|x\|_X <_{\mathbb{R}} b \to \left(f(x) - 2^{-k} \leqslant_{\mathbb{R}} \langle x, x^* \rangle_{X^*} - f^*(x^*) \right) \right).$$

With $\mathcal{D}^\omega_{f,f^*}[\mathrm{FM}]$ we abbreviate the system that arises from $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ by adding these constants and axioms.

Before we consider the monotone operators in Banach spaces, we first establish the following properties of f and f^* in $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ regarding the relationship between the continuity of ∇f and the convexity of f^* and vice versa (as already discussed at various points in Chapter 9, recall in particular Remark 9.2.13). For this, we also in particular establish a quantitative variant of the uniqueness of ∇f as a subgradient.

Lemma 10.2.1. *The system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ proves:*

1. *The "Fenchel-Young equality"*[1] *for any subgradient u^* of f at x, i.e.*

$$\forall x^X, u^{*X^*} (\forall y^X (f(y) \geqslant_{\mathbb{R}} f(x) + \langle y -_X x, u^* \rangle_{X^*}) \to f(x) + f^*(u^*) =_{\mathbb{R}} \langle x, u^* \rangle_{X^*}).$$

[1] By this expression, we mean in the following that the Fenchel-Young inequality is not strict, i.e. is satisfied with equality.

2. *The "Fenchel-Young equality" for $\nabla f x$, i.e.*

$$\forall x^X \left(f(x) + f^*(\nabla f x) =_{\mathbb{R}} \langle x, \nabla f x\rangle_{X^*}\right).$$

3. *Approximate subgradients of f are close to the gradient of f, i.e.*

$$\forall b^0, k^0 \exists j^0 \forall x^X, x^{*X^*} (\|x\|_X <_{\mathbb{R}} b \wedge \forall y^X \left(\langle y -_X x, x^*\rangle_{X^*} + f(x) \leqslant_{\mathbb{R}} f(y) + 2^{-j}\right) \\ \to \|x^* -_{X^*} \nabla f x\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k}),$$

where in fact we can take

$$j = k + 4 + \omega^{\nabla f}(k+3, b+1).$$

4. *The "Fenchel-Young equality" characterizes gradients of f, i.e.*

$$\forall x^X, x^{*X^*} \left(f(x) + f^*(x^*) =_{\mathbb{R}} \langle x, x^*\rangle_{X^*} \to x^* =_{X^*} \nabla f x\right),$$

where in fact it moreover holds that

$$\forall b^0, k^0 \exists j^0 \forall x^X, x^{*X^*} (\|x\|_X <_{\mathbb{R}} b \wedge f(x) + f^*(x^*) - \langle x, x^*\rangle_{X^*} \leqslant_{\mathbb{R}} 2^{-j} \\ \to \|x^* -_{X^*} \nabla f x\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k})$$

where we can take

$$j = k + 4 + \omega^{\nabla f}(k+3, b+1).$$

5. *$\nabla f x$ is the unique subgradient of f at x, i.e.*

$$\forall x^X, u^{*X^*} \left(\forall y^X \left(f(y) \geqslant_{\mathbb{R}} f(x) + \langle y -_X x, u^*\rangle_{X^*}\right) \to u^* =_{X^*} \nabla f x\right).$$

6. *f^* is uniformly strictly convex on bounded subsets, i.e.*

$$\forall k^0, i^0, b^0 \exists j^0 \forall x^{*X^*}, y^{*X^*}, t^1 (\|x^*\|_{X^*}, \|y^*\|_{X^*} <_{\mathbb{R}} b \wedge 2^{-i} \leqslant_{\mathbb{R}} t \leqslant_{\mathbb{R}} 1 - 2^{-i} \\ \wedge\, t f^*(x^*) + (1-t) f^*(y^*) - f^*(t x^* +_{X^*} (1-t) y^*) \leqslant_{\mathbb{R}} 2^{-j} \\ \to \|x^* -_{X^*} y^*\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k})$$

where we in fact can choose

$$j = (k + 4 + \omega^{\nabla f}(k+4, F(b)+2)) + i$$

where F is a modulus for ∇f^* being bounded on bounded sets (which can be constructed similar to Lemma 8.5.5).

7. *∇f^* is uniformly strictly monotone on bounded subsets, i.e.*

$$\forall k^0, b^0 \exists j^0 \forall x^{*X^*}, y^{*X^*} (\|x^*\|_{X^*}, \|y^*\|_{X^*} <_{\mathbb{R}} b \wedge \|x^* -_{X^*} y^*\|_{X^*} >_{\mathbb{R}} 2^{-k} \\ \to (\langle \nabla f^* x^* -_X \nabla f^* y^*, x^* -_{X^*} y^* \rangle_{X^*} \geqslant_{\mathbb{R}} 2^{-j})).$$

where we in fact can choose

$$j = k + 5 + \omega^{\nabla f}(k+4, F(b)+2).$$

with all other constants as in (6).

Proof. 1. Let u^* be such that

$$\forall y \left(f(y) \geqslant f(x) + \langle y - x, u^* \rangle\right),$$

i.e.

$$\langle y, u^* \rangle - f(y) \leqslant \langle x, u^* \rangle - f(x)$$

for all y. Using $(f^*)_2$, we get that for any j, there exists a y_j such that

$$f^*(u^*) - (\langle x, u^* \rangle - f(x)) \leqslant f^*(u^*) - (\langle y_j, u^* \rangle - f(y_j)) \leqslant 2^{-j}$$

and thus we have

$$f^*(u^*) \leqslant \langle x, u^* \rangle - f(x).$$

Using axiom $(f^*)_1$, we get $\langle x, u^* \rangle - f(x) \leqslant f^*(u^*)$ and combined this gives the result.

2. Follows immediately from (1) and $(\nabla f)_1$.

3. Let x^* be such that

$$\forall y \left(\langle y - x, x^* \rangle + f(x) \leqslant f(y) + 2^{-j}\right)$$

for j defined as above. This yields

$$\langle y - x, x^* - \nabla f x \rangle = \langle y - x, x^* \rangle - \langle y - x, \nabla f(x) \rangle \\ \leqslant f(y) - f(x) - \langle y - x, \nabla f(x) \rangle + 2^{-j}.$$

Using Lemma 8.5.5, (1), we get that for $\|y - x\| < 2^{-\omega^{\nabla f}(l, b+1)}$:

$$f(y) - f(x) - \langle y - x, \nabla f(x) \rangle \leqslant 2^{-l} \|y - x\|$$

and so $\langle y - x, x^* - \nabla fx \rangle \leqslant 2^{-j} + 2^{-l} \|y - x\|$ for all such y which in particular yields

$$\langle z, x^* - \nabla fx \rangle \leqslant 2^{-j} + 2^{-l} \|z\|$$

for all z with $\|z\| < 2^{-\omega^{\nabla f}(l,b+1)}$ given any l. For the given k, now use the same argument as in the proof of Lemma 8.3.4, (2).(b) to pick a z_k such that (w.l.o.g.) $0 < \|z_k\| \leqslant 1$ and

$$\langle z_k, x^* - \nabla fx \rangle \leqslant 2^{-(k+2)} \to \|x^* - \nabla fx\| \leqslant 2^{-k}.$$

Define

$$\widehat{z_k} = 2^{-(\omega^{\nabla f}(k+3,b+1)+1)} z_k.$$

Clearly $\|\widehat{z_k}\| < 2^{-\omega^{\nabla f}(k+3,b+1)}$ and thus

$$\langle \widehat{z_k}, x^* - \nabla fx \rangle \leqslant 2^{-j} + 2^{-(k+3)} \|\widehat{z_k}\|$$

which yields by definition of j that

$$\begin{aligned}
\langle z_k, x^* - \nabla fx \rangle &= 2^{(\omega^{\nabla f}(k+3,b+1)+1)} \langle \widehat{z_k}, x^* - \nabla fx \rangle \\
&\leqslant 2^{(\omega^{\nabla f}(k+3,b+1)+1)} (2^{-j} + 2^{-(k+3)} \|\widehat{z_k}\|) \\
&= 2^{(\omega^{\nabla f}(k+3,b+1)+1)} 2^{-j} + 2^{-(k+3)} 2^{(\omega^{\nabla f}(k+3,b+1)+1)} \|\widehat{z_k}\|) \\
&\leqslant 2^{-(k+4+\omega^{\nabla f}(k+3,b+1))} 2^{(\omega^{\nabla f}(k+3,b+1)+1)} + 2^{-(k+3)} \\
&= 2^{-(k+2)}
\end{aligned}$$

which implies $\|x^* - \nabla fx\| \leqslant 2^{-k}$ by the properties of z_k.

4. Let x^* be such that

$$f(x) + f^*(x^*) - \langle x, x^* \rangle_{X^*} \leqslant 2^{-j}$$

with j defined as above. Then we get

$$f^*(x^*) \leqslant 2^{-j} + \langle x, x^* \rangle - f(x)$$

which yields through $(f^*)_1$ that

$$\langle y, x^* \rangle - f(y) \leqslant 2^{-j} + \langle x, x^* \rangle - f(x)$$

for all y which is equivalent to

$$\langle y - x, x^* \rangle + f(x) \leqslant f(y) + 2^{-j}$$

for all y. Then item (3) yields the result.

5. This follows immediately from (3).

6. Suppose
$$|tf^*(x^*) + (1-t)f^*(y^*) - f^*(tx^* + (1-t)y^*)| \leqslant_{\mathbb{R}} 2^{-j}$$
for j as above. Then write $z^* = tx^* + (1-t)y^*$ and pick $x = \nabla f^* z^*$, i.e. $\nabla f x = z^*$ by (L). Then by item (2), the extensionality of f^* (recall Remark 8.5.15) and the extensionality of $\langle \cdot, \cdot \rangle$, we get
$$\begin{aligned} 0 &= f(x) + f^*(z^*) - \langle x, z^* \rangle \\ &\geqslant f(x) + tf^*(x^*) + (1-t)f^*(y^*) - 2^{-j} - \langle x, z^* \rangle, \end{aligned}$$
i.e. we have
$$2^{-j} \geqslant t(f(x) + f^*(x^*) - \langle x, x^* \rangle) + (1-t)(f(x) + f^*(y^*) - \langle x, y^* \rangle)$$
and thus, using $t, 1-t \geqslant 2^{-i}$ and that $f(x) + f^*(x^*) - \langle x, x^* \rangle \geqslant 0$ as well as $f(x) + f^*(y^*) - \langle x, y^* \rangle \geqslant 0$ by the Fenchel-Young inequality (which follows from axiom $(f^*)_1$), we get
$$2^{-j}2^i \geqslant f(x) + f^*(x^*) - \langle x, x^* \rangle, f(x) + f^*(y^*) - \langle x, y^* \rangle$$
By definition of j, we get
$$2^{-(k+4+\omega^{\nabla f}(k+4,F(b)+2))} \geqslant f(x) + f^*(x^*) - \langle x, x^* \rangle, f(x) + f^*(y^*) - \langle x, y^* \rangle.$$
Noting that $\|z^*\| \leqslant t\|x^*\| + (1-t)\|y^*\| < b$ and thus $\|x\| < F(b) + 1$, item (4) implies that
$$\|x^* - \nabla f x\|, \|y^* - \nabla f x\| \leqslant 2^{-(k+1)}$$
which yields $\|x^* - y^*\| \leqslant 2^{-k}$.

7. Using item (6), note that for $t = 1/2$, we have
$$\begin{aligned} f^*\left(\frac{x^* + y^*}{2}\right) &\leqslant 1/2 f^*(y^*) + 1/2 f^*(x^*) - 2^{-j} \\ &= f^*(x^*) + 1/2(f^*(y^*) - f^*(x^*)) - 2^{-j} \end{aligned}$$
if $\|x^* - y^*\| > 2^{-k}$. As
$$\langle \nabla f^* w^*, z^* \rangle \leqslant \frac{f^*(w^* + \alpha z^*) - f^*(w^*)}{\alpha},$$

for any $\alpha > 0$ (using $(\nabla f^*)_1$), we get

$$\langle \nabla f^* x^*, y^* - x^* \rangle \leqslant f^*(y^*) - f^*(x^*) - 2 \cdot 2^{-j}.$$

Similarly, we get

$$\langle \nabla f^* y^*, x^* - y^* \rangle \leqslant f^*(x^*) - f^*(y^*) - 2 \cdot 2^{-j}$$

and this implies

$$\langle \nabla f^* y^* - \nabla f^* x^*, x^* - y^* \rangle \leqslant -4 \cdot 2^{-j}$$

which gives the claim.

□

Now, the additional axioms in $\mathcal{D}^{\omega}_{f,f^*}[\mathrm{FM}]$ can be used to carry out the above proof with the roles of f and f^* exchanged. We collect this in the following lemma.

Lemma 10.2.2. *The system $\mathcal{D}^{\omega}_{f,f^*}[\mathrm{FM}]$ proves:*

1. *The "Fenchel-Young equality" for any subgradient u of f^* at x^*, i.e.*

$$\forall x^{*X^*}, u^X (\forall y^{*X^*} (f^*(y^*) \geqslant_{\mathbb{R}} f^*(x^*) + \langle u, y^* -_{X^*} x^* \rangle_{X^*}) \\ \to f^*(x^*) + f(u) =_{\mathbb{R}} \langle u, x^* \rangle_{X^*}).$$

2. *The "Fenchel-Young equality" for $\nabla f^* x$, i.e.*

$$\forall x^{*X^*} \left(f^*(x^*) + f(\nabla f^* x^*) =_{\mathbb{R}} \langle \nabla f^* x^*, x^* \rangle_{X^*} \right).$$

3. *Approximate subgradients of f^* are close to the gradient of f^*, i.e.*

$$\forall b^0, k^0 \exists j^0 \forall x^{*X^*}, x^X (\|x^*\|_{X^*} <_{\mathbb{R}} b \\ \wedge \forall y^{*X^*} \left(\langle x, y^* -_{X^*} x^* \rangle_{X^*} + f^*(x^*) \leqslant_{\mathbb{R}} f^*(y^*) + 2^{-j} \right) \\ \to \|x -_X \nabla f^* x^*\|_X \leqslant_{\mathbb{R}} 2^{-k}),$$

where in fact we can take

$$j = k + 4 + \omega^{\nabla f^*}(k + 3, b + 1).$$

4. *The "Fenchel-Young equality" characterizes gradients of f^*, i.e.*

$$\forall x^{*X^*}, x^X \left(f^*(x^*) + f(x) =_{\mathbb{R}} \langle x, x^* \rangle_{X^*} \to x =_X \nabla f^* x^* \right),$$

where in fact it moreover holds that

$$\forall b^0, k^0 \exists j^0 \forall x^{*X^*}, x^X (\|x^*\|_{X^*} <_{\mathbb{R}} b \wedge f^*(x^*) + f(x) - \langle x, x^* \rangle_{X^*} \leqslant_{\mathbb{R}} 2^{-j} \\ \to \|x -_X \nabla f^* x^*\|_X \leqslant_{\mathbb{R}} 2^{-k})$$

where we can take

$$j = k + 4 + \omega^{\nabla f^*}(k+3, b+1).$$

5. $\nabla f^* x^*$ *is the unique subgradient of* f^* *at* x^**, i.e.*

$$\forall x^{*X^*}, u^X \left(\forall y^{*X^*} (f^*(y^*) \geqslant_{\mathbb{R}} f^*(x^*) + \langle u, y^* -_{X^*} x^* \rangle_{X^*}) \to u =_X \nabla f^* x^* \right).$$

6. f *is uniformly strictly convex on bounded subsets, i.e.*

$$\forall k^0, i^0, b^0 \exists j^0 \forall x^X, y^X, t^1 (\|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge 2^{-i} \leqslant_{\mathbb{R}} t \leqslant_{\mathbb{R}} 1 - 2^{-i} \\ \wedge tf(x) + (1-t)f(y) - f(tx +_X (1-t)y) \leqslant_{\mathbb{R}} 2^{-j} \\ \to \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-k})$$

where we in fact can choose

$$j = (k + 4 + \omega^{\nabla f^*}(k+4, C(b)+2)) + i$$

where C *is a modulus for* ∇f *being bounded on bounded sets (which can be constructed as in Lemma 8.5.5).*

7. ∇f *is uniformly strictly monotone on bounded subsets, i.e.*

$$\forall k^0, b^0 \exists j^0 \forall x^X, y^X (\|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X >_{\mathbb{R}} 2^{-k} \\ \to (\langle x -_X y, \nabla f x -_{X^*} \nabla f y \rangle_{X^*} \geqslant_{\mathbb{R}} 2^{-j})).$$

where we in fact can choose

$$j = k + 5 + \omega^{\nabla f^*}(k+4, C(b)+2).$$

with all other constants as in (6).

In particular, in the system $\mathcal{D}^\omega_{f,f^*}$[FM] we can now formally establish some of the central properties of Bregman distances used extensively throughout Chapter 9. We begin with the fact that $W_f(x, \nabla f(y)) = D_f(x,y)$:

Lemma 10.2.3. *The system* $\mathcal{D}^\omega_{f,f^*}$[FM] *proves:*

$$\forall x^X, y^X (D_f(x,y) =_{\mathbb{R}} f(x) + f^*(\nabla f y) - \langle x, \nabla f y \rangle_{X^*}).$$

Proof. By Lemma 10.2.1, (2), we have

$$f^*(\nabla fy) = \langle y, \nabla fy\rangle - f(y)$$

and thus

$$\begin{aligned} f(x) + f^*(\nabla fy) - \langle x, \nabla fy\rangle &= f(x) - f(y) - \langle x - y, \nabla fy\rangle \\ &= D_f(x, y). \end{aligned}$$

□

Lemma 10.2.4. *The system $\mathcal{D}^{\omega}_{f,f^*}$[FM] proves that D_f is uniformly bounded in the sense of Chapter 9, i.e.*

$$\forall b^0, \alpha^0 \exists o^0 \forall x^X, y^X \left(\|x\|_X <_{\mathbb{R}} b \wedge D_f(x, y) <_{\mathbb{R}} \alpha \to \|y\|_X \leqslant_{\mathbb{R}} o\right)$$

and o can be realized by

$$o = o(\alpha, b) = F(\alpha^{f^*}(\alpha + D(b) + b) + 1).$$

where D, F are moduli of f, ∇f^* being bounded on bounded sets, respectively, and α^{f^*} is a modulus of supercoercivity for f^* as before.

Proof. First, note that $f^*(x^*) - \langle x, x^*\rangle$ is also supercoercive. For this, let $\|x\| < b$. If $\|x^*\| > \alpha^{f^*}(K + b)$, from axiom $(f^*)_3$ we derive

$$\frac{f^*(x^*) - \langle x, x^*\rangle}{\|x^*\|} \geqslant \frac{f^*(x^*)}{\|x^*\|} - \|x\| \geqslant K.$$

Now, we have

$$f^*(\nabla fy) - \langle x, \nabla fy\rangle = D_f(x, y) - f(x) < \alpha + D(b)$$

using the above Lemma 10.2.3. Therefore, we derive

$$\|\nabla fy\| \leqslant \alpha^{f^*}(\alpha + D(b) + b)$$

and thus we get $\|y\| = \|\nabla f^* \nabla fy\| \leqslant F(\alpha^{f^*}(\alpha + D(b) + b) + 1)$. □

10.2.2 Monotone operators and their relativized resolvents

Set-valued operators of the form $A : X \to 2^{X^*}$ are, in similarity to before in Chapter 3, modeled via a constant for their characteristic function. In the context of the system $\mathcal{D}^\omega$ for the dual of a Banach space, we in that way add a constant χ_A of type $0(X^*)(X)$ and write $x^* \in Ax$, $(x, x^*) \in A$ or $(x, x^*) \in \text{gra}A$ for $\chi_A x x^* =_0 0$. The first natural axiom is

$$\forall x^X, x^{*X^*} (\chi_A x x^* \leqslant_0 1) \qquad (I)^*$$

which witnesses that χ_A is a characteristic function as before.

Also, the treatment of the resolvent is conceptually similar to before. For this, let A be monotone (in the sense of Browder, recall Chapter 9) and recall Definition 9.3.10 for the resolvents of such monotone operators relative to f: $\text{Res}^f_\gamma : X \to 2^X$ is defined by

$$\text{Res}^f_\gamma x := \left((\nabla f + \gamma A)^{-1} \circ \nabla f\right)(x)$$

for any $x \in X$ and $\gamma > 0$ where, as before, since A remains fixed, we write Res^f_γ for $\text{Res}^f_{\gamma A}$. It follows by our assumptions on f and Proposition 9.3.11 that this map is single-valued, satisfies $F(\text{Res}^f_\gamma) = A^{-1}0$ (noting that $\text{dom}f = X$ in this chapter) and that it is Bregman firmly nonexpansive.

So, for treating an operator A with total relativized resolvents, we add a constant Res^f of type $X(X)(1)$ and write Res^f_γ for $\text{Res}^f\gamma$. The natural axiom for the resolvent now can be derived as before: If seen as a set-valued operator, the resolvent satisfies

$$\begin{aligned} p \in \text{Res}^f_\gamma x &\Leftrightarrow p \in (\nabla f + \gamma A)^{-1} \nabla f(x) \\ &\Leftrightarrow \nabla f(x) \in \nabla f(p) + \gamma Ap \\ &\Leftrightarrow \gamma^{-1} \left(\nabla f(x) - \nabla f(p)\right) \in Ap. \end{aligned}$$

This naturally leads us to consider the axiom scheme

$$\forall \gamma^1, x^X \left(\gamma >_{\mathbb{R}} 0 \to \gamma^{-1}(\nabla f x -_{X^*} \nabla f(\text{Res}^f_\gamma x)) \in A(\text{Res}^f_\gamma x)\right) \qquad (II)^*$$

in similarity to axiom (II) considered in Chapter 3 as an intensional version of the crucial direction of the above equivalence for total resolvents.

Remark 10.2.5. As in the context of the systems from Chapter 3, note that also here, the above axiom $(II)^*$ is actually an abbreviation for the following sentence where the dependence of γ^{-1} on a lower bound of γ is made explicit:

$$\forall \gamma^1, x^X, k^0 \left(\gamma >_{\mathbb{R}} 2^{-k} \to (\gamma)^{-1}_k(\nabla f x -_{X^*} \nabla f(\text{Res}^f_\gamma x)) \in A(\text{Res}^f_\gamma x)\right).$$

Also the monotonicity of A is easily specified by a universal axiom:

$$\forall x^X, y^X, x^{*X^*}, y^{*X^*} \left((x, x^*), (y, y^*) \in A \to \langle x -_X y, x^* -_{X^*} y^*\rangle_{X^*} \geqslant_{\mathbb{R}} 0\right). \qquad (III)^*$$

Lastly, all uses of the resolvent presented in Chapter 9 are made in the context of the assumption that $A^{-1}0 \neq \varnothing$ and we will also assume this here as it in particular will allow us to majorize the resolvent rather immediately. For this, we add a constant p_X of type X together with a corresponding axiom stating that p_X is a zero of A:

$$0 \in Ap_X. \qquad (IV)^*$$

This leads us to the following system:

Definition 10.2.6. The theory $\mathcal{B}^\omega$ is defined as the extension of the theory $\mathcal{D}^\omega_{f,f*}$[FM] with the above constants and corresponding axioms $(I)^*$ - $(IV)^*$.

Now, in similarity to the systems from Chapter 3, also $\mathcal{B}^\omega$ is sufficient for formalizing the first main aspects of the theory of monotone operators in Banach spaces and their resolvents relative to f as the following proposition shows.

Proposition 10.2.7. *The system $\mathcal{B}^\omega$ proves:*

1. Res^f_γ *is unique for any $\gamma > 0$, i.e.*

$$\forall \gamma^1, p^X, x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \gamma^{-1}(\nabla f x -_{X^*} \nabla f p) \in Ap \to p =_X \mathrm{Res}^f_\gamma x\right).$$

2. Res^f_γ *is Bregman firmly nonexpansive for any $\gamma > 0$, i.e.*

$$\begin{aligned}\forall \gamma^1, x^X, y^X (\gamma >_{\mathbb{R}} 0 \to \langle \mathrm{Res}^f_\gamma x -_X \mathrm{Res}^f_\gamma y, \nabla f \mathrm{Res}^f_\gamma x -_{X^*} \nabla f \mathrm{Res}^f_\gamma y\rangle_{X^*} \\ \leqslant_{\mathbb{R}} \langle \mathrm{Res}^f_\gamma x -_X \mathrm{Res}^f_\gamma y, \nabla f x -_{X^*} \nabla f y\rangle_{X^*}).\end{aligned}$$

3. Res^f_γ *satisfies the alternative notion of Bregman firm nonexpansivity for any $\gamma > 0$, i.e.*

$$\begin{aligned}\forall \gamma^1, x^X, y^X (\gamma >_{\mathbb{R}} 0 \to D_f(\mathrm{Res}^f_\gamma x, \mathrm{Res}^f_\gamma y) + D_f(\mathrm{Res}^f_\gamma y, \mathrm{Res}^f_\gamma x) \\ \leqslant_{\mathbb{R}} D_f(\mathrm{Res}^f_\gamma x, y) + D_f(\mathrm{Res}^f_\gamma y, x) - D_f(\mathrm{Res}^f_\gamma x, x) - D_f(\mathrm{Res}^f_\gamma y, y)).\end{aligned}$$

4. $A^{-1}0 \subseteq F(\mathrm{Res}^f_\gamma)$ *for any $\gamma > 0$, i.e.*

$$\forall p^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \wedge 0 \in Ap \to p =_X \mathrm{Res}^f_\gamma p\right).$$

Proof. 1. Suppose that $\gamma > 0$ and that $\gamma^{-1}(\nabla fx - \nabla fp) \in Ap$. Axiom $(II)^*$ gives $\gamma^{-1}(\nabla fx - \nabla f\mathrm{Res}^f_\gamma x) \in A(\mathrm{Res}^f_\gamma x)$. Axiom $(III)^*$ then implies that

$$\begin{aligned}0 &\leqslant \langle \mathrm{Res}^f_\gamma x - p, \gamma^{-1}(\nabla fx - \nabla f\mathrm{Res}^f_\gamma x) - \gamma^{-1}(\nabla fx - \nabla fp)\rangle \\ &= \langle \mathrm{Res}^f_\gamma x - p, \gamma^{-1}(\nabla fp - \nabla f\mathrm{Res}^f_\gamma x)\rangle\end{aligned}$$

where we have used extensionality of $\langle\cdot,\cdot\rangle$ and of the arithmetical operations in X^*. In particular, since $\gamma^{-1} > 0$ as $\gamma > 0$, we get that

$$\langle \mathrm{Res}^f_\gamma x - p, \nabla f\mathrm{Res}^f_\gamma x - \nabla fp\rangle \leqslant 0.$$

Thus, as ∇f is provably strictly monotone (Lemma 10.2.2), we get $\left\|\mathrm{Res}^f_\gamma x - p\right\| = 0$, i.e. $\mathrm{Res}^f_\gamma x = p$.

2. Let $\gamma > 0$. Axiom $(II)^*$ gives

$$\gamma^{-1}(\nabla fx - \nabla f\mathrm{Res}^f_\gamma x) \in A(\mathrm{Res}^f_\gamma x) \text{ and } \gamma^{-1}(\nabla fy - \nabla f\mathrm{Res}^f_\gamma y) \in A(\mathrm{Res}^f_\gamma y).$$

Axiom $(III)^*$ and $\gamma^{-1} > 0$ gives

$$\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla fx - \nabla fy - (\nabla f\mathrm{Res}^f_\gamma x - \nabla f\mathrm{Res}^f_\gamma y)\rangle \geqslant 0$$

which implies

$$\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla fx - \nabla fy\rangle \geqslant \langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla f\mathrm{Res}^f_\gamma x - \nabla f\mathrm{Res}^f_\gamma y\rangle.$$

3. By the provability of the three-point identity for D_f (Lemma 8.5.18), we get

$$\begin{aligned}&\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla f\mathrm{Res}^f_\gamma x - \nabla f\mathrm{Res}^f_\gamma y\rangle \\ &\quad= D_f(\mathrm{Res}^f_\gamma x, \mathrm{Res}^f_\gamma y) + D_f(\mathrm{Res}^f_\gamma y, \mathrm{Res}^f_\gamma x) - D_f(\mathrm{Res}^f_\gamma x, \mathrm{Res}^f_\gamma x) \\ &\quad= D_f(\mathrm{Res}^f_\gamma x, \mathrm{Res}^f_\gamma y) + D_f(\mathrm{Res}^f_\gamma y, \mathrm{Res}^f_\gamma x).\end{aligned}$$

Further, by the provability of the four-point identity for D_f (Lemma 8.5.18), we get

$$\begin{aligned}&\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla fx - \nabla fy\rangle \\ &\quad= D_f(\mathrm{Res}^f_\gamma x, y) - D_f(\mathrm{Res}^f_\gamma x, x) - D_f(\mathrm{Res}^f_\gamma y, y) + D_f(\mathrm{Res}^f_\gamma y, x).\end{aligned}$$

Thus, using item (2), we get the claimed inequality.

4. Let p be such that $0 \in Ap$. Then provably with the only assumption being $\gamma > 0$, we have $\gamma^{-1}(\nabla fp - \nabla fp) = 0$ and thus, using Σ_1-ER, we have that $0 \in Ap$ implies
$$\gamma^{-1}(\nabla fp - \nabla fp) \in Ap.$$
Using item (1), we get $p = \mathrm{Res}^f_\gamma p$.

□

Also the boundedness and continuity properties of maps that are Bregman firmly nonexpansive, as already discussed in Chapter 9, can now be formally replicated in the context of the system $\mathcal{B}^\omega$ (where we here formulate these properties just for the resolvents):

Proposition 10.2.8. $\mathcal{B}^\omega$ *proves:*

1. Res^f_γ *is bounded on bounded sets for any* $\gamma > 0$*, i.e.*
$$\forall \gamma^1 \forall b^0 \exists e^0 \forall x^X \left(\gamma >_{\mathbb{R}} 0 \wedge \|p_X\|_X, \|x\|_X <_{\mathbb{R}} b \to \left\|\mathrm{Res}^f_\gamma x\right\|_X \leqslant_{\mathbb{R}} e\right),$$
where in fact one can choose
$$e = E(b) = o(2D(b) + 2bC(b), b)$$
where C, D *are moduli witnessing that* $\nabla f, f$ *are bounded on bounded sets, respectively, and* o *is defined as in Lemma 10.2.4.*

2. Res^f_γ *is uniformly continuous on bounded sets for any* $\gamma > 0$*, i.e.*
$$\forall \gamma^1, k^0, b^0 \exists j^0 \forall x^X, y^X (\gamma >_{\mathbb{R}} 0 \wedge \|p_X\|_X, \|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-j} \to \left\|\mathrm{Res}^f_\gamma x -_X \mathrm{Res}^f_\gamma y\right\|_X \leqslant_{\mathbb{R}} 2^{-k})$$
where in fact one can choose
$$j = \varpi(k, b) = \omega^{\nabla f}(\widehat{k} + 1 + E(b), b)$$
for $\widehat{k} = k + 5 + \omega^{\nabla f^*}(k + 4, C(b) + 2)$ *with* C *being a modulus witnessing that* ∇f *is bounded on bounded sets and where* E *is defined as in (1).*

Proof. For item (1), note that by Lemma 10.2.7, (3) and (4), and with $p = p_X$ from axiom $(IV)^*$, we have (using the extensionality of D_f which follows from that of $f, \nabla f$ and $\langle \cdot, \cdot \rangle$):
$$\begin{aligned} D_f(\mathrm{Res}^f_\gamma x, p) + D_f(p, \mathrm{Res}^f_\gamma x) &\leqslant D_f(\mathrm{Res}^f_\gamma x, p) + D_f(p, x) - D_f(\mathrm{Res}^f_\gamma x, x) - D_f(p, p) \\ &\leqslant D_f(\mathrm{Res}^f_\gamma x, p) + D_f(p, x) \end{aligned}$$

and thus

$$D_f(p, \mathrm{Res}^f_\gamma x) \leqslant D_f(p,x) < 2D(b) + 2bC(b).$$

Thus, by Lemma 10.2.4, we get

$$\left\|\mathrm{Res}^f_\gamma x\right\| \leqslant o(2D(b) + 2bC(b), b).$$

For item (2), by Lemma 10.2.7, (2), we have

$$\begin{aligned}&\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla f \mathrm{Res}^f_\gamma x - \nabla f \mathrm{Res}^f_\gamma y\rangle\\ &\qquad \leqslant \langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla f x - \nabla f y\rangle\\ &\qquad \leqslant \left\|\mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y\right\| \left\|\nabla f x - \nabla f y\right\|\\ &\qquad \leqslant 2E(b) \left\|\nabla f x - \nabla f y\right\|.\end{aligned}$$

using the above item (1). So, for $\|x-y\| < 2^{-j}$, by the definition of j, we have

$$\|\nabla f x - \nabla f y\| \leqslant 2^{-(\widehat{k}+1+E(b))}$$

and thus

$$\langle \mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y, \nabla f \mathrm{Res}^f_\gamma x - \nabla f \mathrm{Res}^f_\gamma y\rangle \leqslant 2^{-\widehat{k}}.$$

Thus by Lemma 10.2.2, (7), we get

$$\left\|\mathrm{Res}^f_\gamma x - \mathrm{Res}^f_\gamma y\right\| \leqslant 2^{-k}.$$

□

Notice that therefore the system $\mathcal{B}^\omega$ proves that Res^f_γ is extensional.

10.3 Maximality and extensionality

As discussed in Chapter 3, a central theoretical result from [165] is the connection between the extensionality of A and the maximality statement for A. We can now extend this result to the monotone operators over Banach spaces.

Theorem 10.3.1. *Over $\mathcal{B}^\omega$, the following are equivalent:*

1. *Extensionality of A, i.e.*

$$\forall x^X, x^{*X^*}, y^X, y^{*X^*} \left(x =_X y \wedge x^* =_{X^*} y^* \rightarrow \chi_A x x^* =_0 \chi_A y y^*\right).$$

2. *The strong resolvent axiom, i.e.*

$$\forall x^X, p^X, \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \wedge p =_X \mathrm{Res}^f_\gamma x \to \gamma^{-1}(\nabla f x -_{X^*} \nabla f p) \in Ap\right).$$

3. *Maximal monotonicity of A, i.e.*

$$\forall x^X, x^{*X^*} \left(\forall y^X, y^{*X^*} \left(y^* \in Ay \to \langle x -_X y, x^* -_{X^*} y^* \rangle_{X^*} \geqslant_{\mathbb{R}} 0\right) \to x^* \in Ax\right).$$

Proof. For the direction (1) ⇒ (3), let x, x^* be such that

$$\forall y, y^* \left(y^* \in Ay \to \langle x - y, x^* - y^* \rangle \geqslant 0\right).$$

We consider $z = \nabla f^*(x^* + \nabla f x)$. Then

$$1^{-1}(\nabla f z - \nabla f \mathrm{Res}^f_1 z) \in A(\mathrm{Res}^f_1 z).$$

by axiom $(II)^*$. Thus by the assumption on x, x^*, axiom (L) and the extensionality of $\langle \cdot, \cdot \rangle$ we get

$$\begin{aligned} 0 &\leqslant \langle x - \mathrm{Res}^f_1 z, x^* - (\nabla f z - \nabla f \mathrm{Res}^f_1 z) \rangle \\ &= \langle x - \mathrm{Res}^f_1 z, \nabla f \mathrm{Res}^f_1 z - \nabla f x \rangle \end{aligned}$$

which is equivalent to

$$\langle x - \mathrm{Res}^f_1 z, \nabla f x - \nabla f \mathrm{Res}^f_1 z \rangle \leqslant 0$$

and this yields $x = \mathrm{Res}^f_1 z$ as ∇f is (provably) strictly monotone. Further, we have

$$1^{-1}\left(\nabla f z - \nabla f \mathrm{Res}^f_1 z\right) = x^* + \nabla f x - \nabla f x = x^*$$

using (L) and the extensionality of ∇f and thus the extensionality of A yields $x^* \in Ax$.

For the direction (3) ⇒ (2), assume that $\gamma > 0$ and $p = \mathrm{Res}^f_\gamma x$. Then at first

$$\gamma^{-1}(\nabla f x - \nabla f \mathrm{Res}^f_\gamma x) \in A(\mathrm{Res}^f_\gamma x)$$

by axiom $(II)^*$. By monotonicity (axiom $(III)^*$) together with the extensionality of $\langle \cdot, \cdot \rangle$ and ∇f, we get

$$\forall (y, y^*) \in A(\langle p - y, \gamma^{-1}(\nabla f x - \nabla f p) - y^* \rangle \geqslant 0).$$

By (3), we get

$$\gamma^{-1}(\nabla f x - \nabla f p) \in A(p).$$

For (2) $\Rightarrow$ (1), let $x = y$ and $x^* = y^*$ with $x^* \in Ax$. Define

$$z = \nabla f^*(y^* + \nabla f y).$$

By $(II)^*$, we get

$$1^{-1}(\nabla f z - \nabla f \mathrm{Res}^f_1 z) \in A(\mathrm{Res}^f_1 z).$$

Axiom $(III)^*$ together with the extensionality of $\langle \cdot, \cdot \rangle$ and ∇f^* as well as using (L) yields

$$\begin{aligned} 0 &\leqslant \langle x - \mathrm{Res}^f_1 z, x^* - (\nabla f z - \nabla f \mathrm{Res}^f_1 z)\rangle \\ &= \langle y - \mathrm{Res}^f_1 z, \nabla f \mathrm{Res}^f_1 z - \nabla f y\rangle \end{aligned}$$

and this is equivalent to

$$\langle y - \mathrm{Res}^f_1 z, \nabla f y - \nabla f \mathrm{Res}^f_1 z\rangle \leqslant 0$$

which yields $y = \mathrm{Res}^f_1 z$ by provable strict monotonicity of ∇f. Using (2), we have

$$1^{-1}(\nabla f z - \nabla f y) \in Ay$$

which yields by the quantifier-free extensionality rule that $y^* \in Ay$ as $1^{-1}(\nabla f z - \nabla f y) = y^*$ holds without any additional assumptions. □

As before, extensionality is not provable (as will be discussed in more detail later on in Chapter 11 as well). As all the results are considered in the context of a Legendre function where f and f^* are Fréchet differentiable with gradients that are uniformly continuous on bounded sets, we find by Proposition 9.3.12 that the totality of the resolvent implies that the operators A which are considered are maximally monotone. As before, this maximality can then not be provable due to the above equivalence. For now, we are content with the following replica of Theorem 3.4.2 which establishes that also here, the system $\mathcal{B}^\omega$ actually proves a weakened maximality principle.

Theorem 10.3.2. *The system $\mathcal{B}^\omega$ proves the following intensional maximality principle:*

$$\begin{aligned} \forall x^X, x^{*X^*} \Big(\forall y^X, y^{*X^*} \, (y^* \in Ay \to \langle x -_X y, x^* -_{X^*} y^* \rangle_{X^*} \geqslant_{\mathbb{R}} 0) \\ \to \exists x'^X, x'^{*X^*} \left(x =_X x' \wedge x^* =_{X^*} x'^* \wedge x'^* \in Ax' \right) \Big). \end{aligned}$$

Proof. As in the proof of the direction (1) $\Rightarrow$ (3) from the above Theorem 10.3.1, we get that

$$1^{-1}(\nabla f z - \nabla f \mathrm{Res}^f_1 z) \in A(\mathrm{Res}^f_1 z)$$

together with $x = \mathrm{Res}^f_1 z$ and $1^{-1}(\nabla f z - \nabla f \mathrm{Res}^f_1 z) = x^*$ for $z = \nabla f^*(x^* + \nabla f x)$ without any use of extensionality. This gives the claim. □

10.4 A bound extraction theorem

We now state the bound extraction theorems for the theory $\mathcal{B}^\omega$ which extends those for $\mathcal{D}^\omega$ and its extensions from Chapter 8. In that vein, we keep the proofs short and only briefly discuss the key ingredients. All other considerations regarding the dual space can be made similar to Chapter 8 and all considerations regarding the operator A can be made similar to Chapter 3.

In particular, the models of all set-theoretic and of all majorizable functionals can be defined for the theory $\mathcal{B}^\omega$ by combining the ideas from Chapter 8 and Chapter 3. We do not spell this out here any further.

We begin with the classical metatheorem:

Theorem 10.4.1. *Let τ be admissible, δ be of degree 1 and s be a closed term of $\mathcal{B}^\omega$ of type $\sigma(\delta)$ for admissible σ. Let Δ be a set of formulas of the form $\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} F_{qf}(\underline{a}, \underline{b}, \underline{c})$ where F_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and where $\underline{r}$ is a tuple of closed terms of appropriate type. Let $B_\forall(x,y,z,u)/C_\exists(x,y,z,v)$ be $\forall$-/$\exists$-formulas of $\mathcal{B}^\omega$ with only $x,y,z,u/x,y,z,v$ free. If*

$$\mathcal{B}^\omega + \Delta \vdash \forall x^\delta \forall y \leqslant_\sigma s(x) \forall z^\tau \left(\forall u^0 B_\forall(x,y,z,u) \to \exists v^0 C_\exists(x,y,z,v) \right),$$

then one can extract a partial functional $\Phi : \mathbb{N}^\mathbb{N} \times \mathbb{N} \times S_\delta \times S_{\widehat{\tau}} \rightharpoonup \mathbb{N}$ which is total and (bar-recursively) computable on $\mathbb{N}^\mathbb{N} \times \mathbb{N} \times M_\delta \times M_{\widehat{\tau}}$ and such that for all $x \in S_\delta$, $z \in S_\tau$, $z^* \in S_{\widehat{\tau}}$ with $z^* \gtrsim z$ and for all $\omega \in \mathbb{N}^\mathbb{N}$, $n \in \mathbb{N}$ with $\omega \gtrsim \omega^{\nabla f}, \omega^{\nabla f^*}, \alpha^f, \alpha^{f^*}$ and $n \geqslant_\mathbb{R} |f(0)|, \|\nabla f(0)\|_{X^*}, |f^*(0)|, \|\nabla f^*(0)\|_X, \|p_X\|_X$:

$$\mathcal{S}^{\omega,X,X^*} \models \forall y \leqslant_\sigma s(x)(\forall u \leqslant_0 \Phi(\omega,n,x,z^*) B_\forall(x,y,z,u) \\ \to \exists v \leqslant_0 \Phi(\omega,n,x,z^*) C_\exists(x,y,z,v))$$

holds whenever $\mathcal{S}^{\omega,X,X^*} \models \Delta$ for $\mathcal{S}^{\omega,X,X^*}$ defined via any (nontrivial) reflexive Banach space $(X, \|\cdot\|)$ with its dual X^* (and via a suitable interpretation of the additional constants similar to Chapters 8 and 3) and using a convex, supercoercive (with modulus α^f) and Fréchet differentiable function $f : X \to \mathbb{R}$ where ∇f, ∇f^* are uniformly continuous on bounded subsets with moduli $\omega^{\nabla f}, \omega^{\nabla f^*}$, respectively and where f^* is supercoercive (with modulus α^{f^*}). In particular, χ_A is interpreted by the characteristic function of a maximally monotone operator $A : X \to 2^{X^*}$ with $A^{-1}0 \neq \emptyset$ and Res^f by the corresponding resolvents $\mathrm{Res}^f_{\gamma A}$ for $\gamma > 0$.

Further:

1. *If $\widehat{\tau}$ is of degree* 1*, then* Φ *is a total computable functional.*

2. *We may have tuples instead of single variables* x, y, z, u, v *and a finite conjunction instead of a single premise* $\forall u^0 B_\forall(x, y, z, u)$*.*

3. *If the claim is proved without* DC*, then* τ *may be arbitrary and* Φ *will be a total functional on* $\mathbb{N}^{\mathbb{N}} \times \mathbb{N} \times S_\delta \times S_{\widehat{\tau}}$ *which is primitive recursive in the sense of Gödel. In that case, also plain majorization can be used instead of strong majorization.*

Proof. To define the models based on $\mathcal{M}^{\omega,X,X^*}$ and $\mathcal{S}^{\omega,X,X^*}$, the interpretation of the constants relating to the normed space and the dual can be achieved similar to Chapter 8 while the interpretation of the constants relating to the operators can be achieved similar to Chapter 3. Majorizability of the characteristic function is immediate as before and majorization of the resolvent follows from Proposition 10.2.8. Thus the previous Lemmas 3.7.7 and 8.6.3 extend to $\mathcal{B}^\omega$. The previous proofs therefore go through exactly as before since all the new axioms are purely universal or of type Δ (dealing with the linearity rule and any other axioms of type Δ in $\mathcal{B}^\omega$ as in the case of $\mathcal{D}^\omega$ in Chapter 8). □

Similar, we can also obtain a semi-constructive metatheorem by extending the ones presented in the previous chapters. For this, we can also define $\mathcal{B}^\omega_i$ similar to $\mathcal{B}^\omega$ but over $\mathcal{A}^\omega_i[X, \|\cdot\|]$ instead of $\mathcal{A}^\omega[X, \|\cdot\|]$. As before, since the constructions and proofs are completely analogous, we omit them here and just state the result:

Theorem 10.4.2. *Let* δ *be of the form* $0(0)\dots(0)$ *and* σ, τ *be arbitrary,* s *be a closed term of suitable type. Let* $\Gamma_\neg$ *be a set of sentences of the form* $\forall \underline{u}^{\underline{\zeta}}(C(\underline{u}) \to \exists \underline{v} \leqslant_{\underline{\beta}} \underline{t}\underline{u} \neg D(\underline{u}, \underline{v}))$ *with* $\underline{\zeta}, \underline{\beta}$ *and* C, D *arbitrary types and formulas respectively and where* $\underline{t}$ *is a tuple of closed terms. Let* $B(x, y, z)/C(x, y, z, u)$ *be arbitrary formulas of* $\mathcal{B}^\omega_i$ *with only* $x, y, z/x, y, z, u$ *free. If*

$$\mathcal{B}^\omega_i + \mathrm{IP}_\neg + \mathrm{CA}_\neg + \Gamma_\neg \vdash \forall x^\delta \, \forall y \leqslant_\sigma (x) \, \forall z^\tau \, (\neg B(x, y, z) \to \exists u^0 C(x, y, z, u)),$$

one can extract a $\Phi : S_1 \times S_0 \times S_\delta \times S_{\widehat{\tau}} \to \mathbb{N}$ *with is primitive recursive in the sense of Gödel such that for any* $x \in S_\delta$*, any* $y \in S_\sigma$ *with* $y \leqslant_\sigma s(x)$*, any* $z \in S_\tau$ *and* $z^* \in S_{\widehat{\tau}}$ *with* $z^* \gtrsim z$ *and for all* $\omega \in \mathbb{N}^{\mathbb{N}}$*,* $n \in \mathbb{N}$ *with* $\omega \gtrsim \omega^{\nabla f}, \omega^{\nabla f^*}, \alpha^f, \alpha^{f^*}$ *and* $n \geqslant_{\mathbb{R}} |f(0)|, \|\nabla f(0)\|_{X^*}, |f^*(0)|, \|\nabla f^*(0)\|_X, \|p_X\|_X$*:*

$$\mathcal{S}^{\omega,X,X^*} \models \exists u \leqslant_0 \Phi(\omega, n, x, z^*) \, (\neg B(x, y, z) \to C(x, y, z, u))$$

holds whenever $\mathcal{S}^{\omega,X,X^*} \models \Gamma_\neg$ *for* $\mathcal{S}^{\omega,X,X^*}$ *defined via any (nontrivial) reflexive Banach space* $(X, \|\cdot\|)$ *with its dual* X^* *(and via a suitable interpretation of the additional* *constants similar to Chapters 8 and 3) and using a convex, supercoercive (with modulus* α^f*) and Fréchet differentiable function* $f : X \to \mathbb{R}$ *where* ∇f, ∇f^* *are uniformly continuous on bounded subsets with moduli* $\omega^{\nabla f}, \omega^{\nabla f^*}$*, respectively and where* f^* *is supercoercive (with modulus* α^{f^*}*). In particular,* χ_A *is interpreted by the characteristic function of a maximally monotone* $A : X \to 2^{X^*}$ *with* $A^{-1}0 \neq \emptyset$ *and* Res^f *by the corresponding resolvents* $\mathrm{Res}^f_{\gamma A}$ *for* $\gamma > 0$*.*

11 On extensionality and uniform continuity for set-valued operators

11.1 Introduction

As discussed extensively in [165], the large applicability of the systems for accretive and monotone operators (and by extension also – presumably – the large applicability of the system $\mathcal{B}^\omega$ for monotone operators in Banach spaces from Chapter 10) is due to the empirical fact that in many situations from the mainstream literature of m-accretive or monotone operator theory, one does not require the full maximality of the operator but it actually suffices to have the intensional maximality principle (recall Theorems 3.4.2 and 10.3.2) together with an (intensionally) total resolvent. New examples for this are also discussed in [165] but many others can in particular be found throughout the previous case studies for set-valued operators in proof mining as most of them do not require any such quantitative treatment of extensionality.

If, however, the proof is not of that nature and really requires an extensionality (viz. maximality) principle, then a quantitative treatment of such will be necessary (as was e.g. the case in the recent application [166]). A short discussion of possible remedies and choices in that situation was given in [165], without indulging into too many details. In particular, a fragment of the extensionality statement corresponding to a certain continuity statement featured in [120, 166] was discussed, though only in brief.

The purpose of this chapter is now twofold:

1. We discuss the main issue with treating full extensionality in the context of the previous intensional approaches to accretivity and to the monotonicity notions for set-valued operators. Motivated by these problems, we discuss three fragments of the extensionality statement for set-valued operators which avoid these issues

and discuss their corresponding quantitative versions (under the guidance of the monotone functional interpretation). This provides a corresponding hierarchy of uniform continuity statements for set-valued operators with various strengths. To that end, we show that all of the considered uniform continuity statements can be added to the systems for proof mining in the context of set-valued operator theory while preserving the bound extraction theorems as discussed in the previous chapters. In particular, in the course of these discussions, we provide a proof-theoretic treatment of the Hausdorff-metric using the tame treatment of suprema over bounded sets developed in Chapter 8.

2. We further show that the correspondence of extensionality and maximality is a fundamental and robust phenomenon in the context of set-valued operators by extending it to the various weakenings of the full extensionality statement and corresponding natural weak forms of the maximality statement. In particular, we discuss similar equivalences in the context of the extensionality of the set of zeros of an operator.

11.2 Motivating considerations: full extensionality and issues with the intensional approach

As before, by (full) extensionality of A we mean the statement (E) defined as

$$\forall x^X, y^X, z^X, w^X \left(x =_X y \wedge z =_X w \wedge z \in Ax \to w \in Ay\right)$$

or, in the case of monotone operators on Banach spaces, defined as

$$\forall x^X, y^X, z^{*X^*}, w^{*X^*} \left(x =_X y \wedge z^* =_{X^*} w^* \wedge z^* \in Ax \to w^* \in Ay\right).$$

Note that through the bound extraction theorems established in Chapters 3 and 10, it is immediately clear that this version of full extensionality can not be provable in any of the theories $\mathcal{V}^\omega$, $\mathcal{T}^\omega$ as defined in Chapter 3 or $\mathcal{B}^\omega$ as defined in Chapter 10.

In fact, the situation regarding extensionality in this intensional approach to set-valued operators is much more dire: any extension of the systems $\mathcal{V}^\omega$, $\mathcal{T}^\omega$ or $\mathcal{B}^\omega$ which has a model based on $\mathcal{S}^{\omega,X}$ or $\mathcal{S}^{\omega,X,X^*}$, respectively, and which still allows for bound extraction theorems in the previous sense can not prove the extensionality of A. To see this, let $\mathcal{C}^\omega$ be any extension of $\mathcal{V}^\omega$ which has a model based on $\mathcal{S}^{\omega,X}$ and for which bound extraction theorems hold similar as for $\mathcal{V}^\omega$. If the extensionality statement (E) were to be provable in $\mathcal{C}^\omega$, we would be able to extract a functional ω such that

$$\forall x, y, z, w \in X \forall b \in \mathbb{N}(\|x\|, \|y\|, \|z\|, \|w\| \leqslant b \\ \wedge \|x - y\|, \|z - w\| \leqslant 2^{-\omega(b)} \wedge z \in Ax \rightarrow w \in Ay)$$

holds for any m-accretive operator A for which the model based on $\mathcal{S}^{\omega,X}$ which arises by interpreting χ_A by A actually is a model of $\mathcal{C}^\omega$. However, any such operator has to be open in $X \times X$: given $(x, z) \in A$ with $\|x\|, \|z\| \leqslant b$ and y, w such that

$$\|x - y\|, \|z - w\| \leqslant 2^{-\omega(b+1)},$$

we have $\|y\|, \|w\| \leqslant b + 1$ and so $(y, w) \in A$. But any m-accretive operator A is maximally accretive and thus closed in $X \times X$ (see e.g. [4]) so that the only two possible interpretations of A are $\varnothing$ or $X \times X$. The former is not allowed in $\mathcal{C}^\omega$ as the theory extends $\mathcal{V}^\omega$ by which axioms we have $\mathrm{dom}A \neq \varnothing$ and the latter is not accretive. So $\mathcal{C}^\omega$ does not have a model based on $\mathcal{S}^{\omega,X}$ after all. Similar considerations also hold for maximally monotone operators $A \subseteq X \times X^*$ and $\mathcal{B}^\omega$.

This is an inherent limitation that comes with the intensional approach chosen for set-valued operators (which however is essentially the only approach that allows for bound extraction theorems that do not outright distort the complexity of the extracted bounds due to analyzing maximality). In Section 11.6, we will later discuss a different approach for treating operators with full extensionality but for now, from the perspective of these intensional systems, there are two possible avenues to at least provide partial remedies to this situation.

The first might be to restrict the kind of x, y permissible in the extensionality statement: the above argument does not work if x and y are required to be contained in the domain of A before applying extensionality. As a second option, one might want to weaken the conclusion from $z \in Ax \leftrightarrow w \in Ay$ to a less "explicit" and more "analytic" version (we will later see concrete instantiations which illuminate what we mean by this).

For the former however, even if (E) is restricted to the domain of A, this still poses an exceedingly large limitation on the systems as if the principle

$$\forall x^X, y^X, z^X, w^X, v^X\, (x =_X y \wedge z =_X w \wedge z \in Ax \wedge v \in Ay \rightarrow w \in Ay) \qquad (E)^d$$

would be provable in the previously presumed extension $\mathcal{C}^\omega$ of $\mathcal{V}^\omega$, then the bound extraction results would yield the existence of a functional ω with

$$\forall x, y, z, w, v \in X \forall b \in \mathbb{N}(\|x\|, \|y\|, \|z\|, \|w\|, \|v\| \leqslant b \\ \wedge \|x - y\|, \|z - w\| \leqslant 2^{-\omega(b)} \wedge z \in Ax \wedge v \in Ay \to w \in Ay).$$

In particular, this implies still that Ax is open for any $x \in \mathrm{dom}A$ as if $z \in Ax$ with $\|z\|, \|x\| \leqslant b$, and w such that $\|z - w\| \leqslant 2^{-\omega(b+1)}$, then $w \in Ax$. As before, if A is m-accretive, then it is maximally accretive and so Ax is closed. Thus, in this case, Ax is clopen, i.e. equal to X or $\varnothing$. The latter is not possible as $x \in \mathrm{dom}A$. Therefore, the only m-accretive operators for which such an ω exists are of the form

$$A : x \mapsto \begin{cases} X & \text{if } x \in \mathrm{dom}A, \\ \varnothing & \text{otherwise.} \end{cases}$$

While such operators exist (take e.g. the normal cone N_x for the singleton $\{x\}$ in a Hilbert space, see [11]) this class is of course extremely restrictive. Similar considerations can also be made for monotone operators in Banach spaces.

So, even though the extension of $\mathcal{V}^\omega$ where the above quantitative variant is added as an axiom allows for bound extraction theorems and still has a model constructed over $\mathcal{S}^{\omega,X}$, the restrictions on the class of axiomatized operators are so strong that this is presumably of little practical relevance.

In the following, we will thus investigate the second option and discuss fragments of (E) and $(E)^d$ that arise by modifying the conclusion together with the quantitative notions that they induce on set-valued operators. Regarding these potential weakenings or reformulations of the conclusion, we are in particular interested in reformulations of the above full extensionality principle in the form of the general scheme

$$x =_X y \to E(Ax, Ay)$$

where E is a placeholder for some predicate expressing "$Ax = Ay$". If E is sufficiently of an "analytical nature" in the sense that it allows for quantifying the difference of Ax and Ay in the case that $Ax \neq Ay$, then such a formulation immediately gives rise to a meaningful associated uniform continuity principle and in this chapter, we will see three instantiations in that vein.

11.3 The Hausdorff-metric and its extensionality statement

A motivating example is the extensionality principle

$$\forall x, y\, (x, y \in \mathrm{dom}A \wedge x = y \rightarrow H(Ax, Ay) = 0)$$

where H is the Hausdorff-metric defined via

$$H(P, Q) := \max\left\{\sup_{p\in P}\inf_{q\in Q}\|p-q\|, \sup_{q\in Q}\inf_{p\in P}\|p-q\|\right\}$$

for closed non-empty sets P, Q in the space. This extensionality statement immediately induces a notion of uniform continuity for A (as commonly used in the analytic literature, see e.g. [151][1]), witnessed by an accompanying modulus of uniform continuity ω:

$$\forall x, y \in X \forall k, b \in \mathbb{N}\left(x, y \in \mathrm{dom}A \cap \overline{B}_b(0) \wedge \|x-y\| \leqslant 2^{-\omega(k,b)} \rightarrow H(Ax, Ay) \leqslant 2^{-k}\right).$$

Note that this restriction to $\mathrm{dom}A$ is necessary for $H(Ax, Ay)$ to be well-defined.

We now begin with showing that for certain sets P, Q, the Hausdorff distance $H(P, Q)$ can be treated in the context of the systems considered before. For this, we work over $\mathcal{V}^\omega$ for now. Let P be a set in a normed space X which is bounded, i.e. $\|p\| \leqslant c$ for all $p \in P$ where $c \in \mathbb{N}$. Then we can treat the real-valued distance function

$$d(x, P) = \inf_{p\in P}\|x-p\|$$

by adding an additional constant $d(\cdot, P)$ of type $1(X)$ with axioms determined similar to the schemes $(S)_1$, $(S)_2$ discussed in Chapter 8. Concretely, we consider the two axioms schemes

$$\forall x^X, p^X\, (P(p) \rightarrow d(x, P) \leqslant_{\mathbb{R}} \|x -_X p\|_X) \qquad (d_P)_1$$

as well as

$$\forall x^X, k^0 \exists p \leqslant_X c1_X \left(P(p) \wedge \|x -_X p\|_X \leqslant_{\mathbb{R}} d(x, P) + 2^{-k}\right) \qquad (d_P)_2$$

where $P(p)$ is a predicate describing $p \in P$. The motivation for $(d_P)_2$ is again the same as with $(S)_2$, just from the perspective of an infimum instead of a supremum: any

[1] While the following principle stipulates uniform continuity on bounded subsets, the literature often even considers situations where the continuity is uniform over the whole space.

interpretation of $d(x,P)$ by the real infimum will have to satisfy $(d_P)_2$ as there will be elements $p \in P$ such that $\|x-p\|$ is arbitrary close to $d(x,P)$ and this property of course also characterizes the infimum.

As before with the discussion in Chapter 8, these schemes become admissible if they are instantiated with an P such that the two axioms have a monotone functional interpretation which can again be guaranteed a priori if P is of a "$\Delta_1(\Delta^*)$"-form. However, later we will be only concerned with the case where $P(p)$ is quantifier-free (and potentially contains parameters).

Similarly, we can add a constant $d(\cdot,Q)$ of the same type for a second bounded set Q (w.l.o.g. also bounded by c) together with the following axioms determined as above over a predicate $Q(q)$ describing $q \in Q$:

$$\forall x^X, q^X \left(Q(q) \to d(x,Q) \leqslant_{\mathbb{R}} \|x -_X q\|_X\right), \qquad (d_Q)_1$$

$$\forall x^X, k^0 \exists q \leqslant_X c1_X \left(Q(q) \wedge \|x -_X q\|_X \leqslant_{\mathbb{R}} d(x,Q) + 2^{-k}\right). \qquad (d_Q)_2$$

In the context of both $d(x,P)$ and $d(x,Q)$, we can then introduce the quantities

$$d(P,Q) = \sup_{p \in P} d(p,Q) \text{ and } d(Q,P) = \sup_{q \in Q} d(q,P)$$

into the system by adding corresponding constants (for simplicity also denoted by) $d(P,Q)$ and $d(Q,P)$ of type 1 into the language together with another set of instantiations of the schemes $(S)_1$, $(S)_2$. Concretely, we consider the schemes

$$\forall p^X (P(p) \to d(P,Q) \geqslant_{\mathbb{R}} d(p,Q)) \qquad (d_{P,Q})_1$$

as well as

$$\forall k^0 \exists p \leqslant_X c1_X \left(P(p) \wedge d(p,Q) \geqslant_{\mathbb{R}} d(P,Q) - 2^{-k}\right). \qquad (d_{P,Q})_2$$

Similarly, for the quantity $d(Q,P)$, we consider the accompanying axiom schemes

$$\forall q^X (Q(q) \to d(Q,P) \geqslant_{\mathbb{R}} d(q,P)) \qquad (d_{Q,P})_1$$

as well as

$$\forall k^0 \exists q \leqslant_X c1_X \left(Q(q) \wedge d(q,P) \geqslant_{\mathbb{R}} d(Q,P) - 2^{-k}\right). \qquad (d_{Q,P})_2$$

Lastly, we move to the concrete Hausdorff-metric which can now just be introduced by a closed term involving $d(P,Q)$ and $d(Q,P)$:

$$H(P,Q) = \max\{d(P,Q), d(Q,P)\}.$$

Of course, this distance can also be introduced uniformly for a family of sets described by formulas $P(p, \underline{x}), Q(q, \underline{x})$ with parameters $\underline{x}$ of type $\underline{\sigma}$ if the sets described by $P(p, \underline{x}), Q(q, \underline{x})$ are bounded by a function $c(\underline{x})$ pointwise in the parameters. We could also introduce the Hausdorff-metric on X^* over the language of $\mathcal{D}^\omega$ (which we do not spell out here any further).

Note that the non-emptiness of the sets P, Q is not needed to define these formulas but the non-emptyness is required on a semantic level in order for these formulas to actually have a model as the objects, mapping to type 1, have to be interpreted by a real number (or by a function mapping into real numbers, respectively).

As mentioned before, this abstract treatment is fruitful at least in the context of sets describable by "$\Delta_1(\Delta^*)$"-formulas (in the sense of Chapter 8). Then these constants and axioms are suitable for extending the previous metatheorems where the interpretations of the constants $d(\cdot, P), d(\cdot, Q), d(P, Q)$ and $d(Q, P)$ in the respective models are naturally defined via $(\cdot)_\circ$. In particular, majorization of these constants can be easily achieved: For $d(\cdot, P)$, via the axiom $(d_P)_1$, we have

$$d(x, P) \leqslant \|x - p\| \leqslant \|x\| + \|p\| \leqslant \|x\| + c$$

where p is some point witnessing that P is non-empty (and thus the non-emptyness is also important for majorization). Further, we have

$$d(Q, P) \leqslant d(q, P) + 1 \leqslant \|q\| + c + 1 \leqslant 2c + 1$$

for a suitable q chosen with axiom $(d_{Q,P})_2$. From this, majorants for $d(\cdot, P)$ and $d(Q, P)$ are immediate.

By a similar reasoning, $d(\cdot, Q)$ as well as $d(P, Q)$ are majorizable and this extends to any variant using additional parameters if the sets are non-empty and bounded pointwise for all parameters. Naturally, also the resulting bounding function $c(\underline{x})$ then has to be majorizable as a function of type $0(\underline{\sigma}^t)$.

We are now in particular interested in using this way of formulating the Hausdorff-distance to talk about uniform continuity formulations for set-valued operators. Then the sets P and Q can be taken to be of the form Ax with a parameter x of type X for a given set-valued operator A which is represented in the system by an intensional description over its graph via χ_A as discussed in the preceding chapters. As the resulting formulation of the set Ax by $P(p,x) := \chi_A(x,p) =_0 0$ is quantifier-free, the above axioms in particular become admissible for bound extraction results if, as discussed before, the operator A is actually such that all Ax are bounded with a bounding function c of type $0(X)$ that is majorizable. In the language of Chapter 3, the existence of such a c is equivalent to the operator A being uniformly majorizable, i.e. bounded on bounded sets. Thus we consider an additional constant A^* of type 1 together with the axiom

$$\forall x^X, y^X, b^0\, (y \in Ax \wedge \|x\|_X <_{\mathbb{R}} b \to \|y\|_X \leqslant_{\mathbb{R}} A^*b) \tag{A^*}$$

which serves as a majorant of a witness to c. Then we can as above introduce constants $d(\cdot, Ax)$ and $d(Ax, Ay)$ for $x, y \in \mathrm{dom}A$ into the language using χ_A and A^* to form H such that the expression $H(Ax, Ay)$ is represented by a term for any x and y.

With this, the previous extensionality statement using the Hausdorff-metric now indeed can be written as a formal sentence in this extended language:

$$\forall x^X, y^X\, (x, y \in \mathrm{dom}A \wedge x =_X y \to H(Ax, Ay) =_{\mathbb{R}} 0)\,.$$

The monotone functional interpretation now suggests an associated uniform continuity principle as before together with a modulus ω of type $0(0)(0)$:

$$\begin{aligned}\forall x^X, y^X, u^X, v^X, k^0, b^0((x,u),(y,v) \in A \wedge \|x\|_X, \|y\|_X, \|u\|_X, \|v\|_X <_{\mathbb{R}} b \\ \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\omega(k,b)} \to H(Ax, Ay) \leqslant_{\mathbb{R}} 2^{-k}).\end{aligned} \tag{UC}$$

This statement is universal and can thus be added to the system together with a constant ω and, for this extension, one retains the bound extraction results.

A similar type of uniform continuity statement could of course also be defined for set-valued mappings $A : X \to 2^{X^*}$ but we do not spell this out in any detail here.

In Chapter 12, we will illustrate the applicability of this approach towards the Hausdorff-metric by analyzing iterative methods related to set-valued mappings which are uniformly continuous w.r.t. the Hausdorff-metric.

11.4 A Hausdorff-like predicate and approximate extensionality

In [120], Kohlenbach and Powell refrained from using H due to its use of infima and suprema and as a substitute, they introduced a so-called *Hausdorff-like predicate* H^* defined via

$$H^*[P,Q,\varepsilon] := \forall p \in P \exists q \in Q \left(\|p-q\| \leqslant \varepsilon\right)$$

and, corresponding to this, they considered a uniform continuity principle for A w.r.t. H^*:

$$\forall x,y \in X, k \in \mathbb{N} \left(x,y \in \mathrm{dom}A \wedge \|x-y\| \leqslant 2^{-\varpi(k)} \to H^*\left[Ax,Ay,2^{-k}\right]\right). \qquad (UC^*)$$

Regarding an associated extensionality statement, we follow the discussion laid out in [165]: we can immediately recognize (UC^*) as the uniform quantitative version, guided by the monotone functional interpretation, of the following *approximate extensionality principle*[2]

$$\forall x^X, y^X \left(x,y \in \mathrm{dom}A \wedge x =_X y \to \forall k^0 H^*\left[Ax,Ay,2^{-k}\right]\right) \qquad (AE)$$

as, by making the hidden quantifier in $=_X$ apparent, this is equivalent to

$$\forall x^X, y^X \left(x,y \in \mathrm{dom}A \wedge \forall j^0 \left(\|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-j}\right) \to \forall k^0 H^*\left[Ax,Ay,2^{-k}\right]\right)$$

and the monotone functional interpretation extracts from this statement a uniform bound (potentially depending on upper bounds on the norm of x and y) on j in terms of the k.[3]

Now, as also discussed in [165], this uniform continuity principle (UC^*) can be (rather immediately) phrased as an axiom of type Δ: considering the definition of H^*, the principle (UC^*) is equivalent to

$$\begin{aligned}\forall x^X, y^X, k^0(x,y \in \mathrm{dom}A \wedge \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-\varpi(k)} \\ \to \forall z \in Ax \exists w \in Ay \left(\|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k}\right)).\end{aligned}$$

[2]See the later parts of this section for a motivation of prefix "approximate" in the name.

[3]Note in the above that the formulation

$$H^*[Ax,Ay,\varepsilon] := \forall z^X \exists w^X \left(z \in Ax \to w \in Ay \wedge \|z -_X w\|_X \leqslant_{\mathbb{R}} \varepsilon\right)$$

is indeed a formula of the language of the underlying system.

Now, the quantifier "$\exists w \in Ay$" can be bounded in norm in terms of the other parameters as if $\|z - w\| \leqslant 2^{-k}$, then $\|w\| \leqslant \|z\| + \|z - w\| \leqslant \|z\| + 2^{-k} \leqslant \|z\| + 1$. Thus, (UC^*) is equivalent to

$$\forall k^0, x^X, y^X, z^X \exists w^X \leqslant_X (\|z\|_X + 1)\, 1_X \Big(x, y \in \mathrm{dom}A \wedge z \in Ax \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\varpi(k)} \\ \to w \in Ay \wedge \|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k} \Big).$$

This is of the form Δ as all variables have admissible types and the inner matrix is purely universal after making the hidden quantifiers in $<_{\mathbb{R}}, \leqslant_{\mathbb{R}}$ apparent and prenexing appropriately.

As (UC^*) can be transformed into a statement of the form Δ, we can add it (together with a corresponding constant ϖ of type 1) to the systems $\mathcal{V}^\omega, \mathcal{T}^\omega$ and still retain the bound extraction theorems established in the previous chapters.

Of course, we could also formulate H^* for X^* and then develop these principles over the language of $\mathcal{B}^\omega$ for operators $A : X \to 2^{X^*}$.

Remark 11.4.1. One can similarly show that the principle where ϖ additionally depends on norm upper bounds of x, y, z and v witnessing $y \in \mathrm{dom}A$ (i.e. $v \in Ay$) can be written as a formula of type Δ via

$$\forall k^0, b^0, x^X, y^X, z^X, v^X \exists w^X \leqslant_X (\|z\|_X + 1)\, 1_X \Big(\|x\|_X, \|y\|_X, \|z\|_X, \|v\|_X <_{\mathbb{R}} b \\ \wedge v \in Ay \wedge z \in Ax \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\varpi(k,b)} \to w \in Ay \wedge \|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k} \Big)$$

and thus can be added to the systems used in proof mining. However, we here wanted to focus on the principle introduced in [120].

11.5 A weak fragment of full extensionality

There is a case to be made regarding how natural the formulation of "$Ax = Ay$" via $\forall k^0 H^*[Ax, Ay, 2^{-k}]$ actually is, in particular compared to $H^*[Ax, Ay, 0]$ where the quantifier over k^0 is internalized in the bound. This leads us to consider the following *weak version of extensionality*

$$\forall x^X, y^X (x, y \in \mathrm{dom}A \wedge x =_X y \to H^*[Ax, Ay, 0]).$$

Unraveling the definition of the H^*-predicate, this can be rewritten as

$$\forall x^X, y^X(x, y \in \mathrm{dom}A \wedge x =_X y \rightarrow \forall z \in Ax \exists w \in Ay(z =_X w)). \qquad (WE)$$

Naturally, this fragment of the full extensionality statement is therefore *stronger* than the approximate extensionality statement from before.

Making the hidden quantifiers apparent, this statement is equivalent to

$$\begin{aligned}&\forall x^X, y^X, z^X, v^X \exists w^X \forall k^0 \exists j^0 (v \in Ay \wedge z \in Ax \\ &\qquad \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-j} \rightarrow \|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k} \wedge w \in Ay)\end{aligned}$$

and a quantitative version of this is therefore given by the existence of a modulus ϖ such that

$$\begin{aligned}&\forall x^X, y^X, z^X, v^X \exists w^X \forall k^0, b^0 (v \in Ay \wedge z \in Ax \wedge \|x\|_X, \|y\|_X, \|z\|_X, \|v\|_X <_{\mathbb{R}} b \\ &\qquad \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\varpi(k,b)} \rightarrow \|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k} \wedge w \in Ay). \qquad (WUC)\end{aligned}$$

Similar to before, (WUC) can be written in the form Δ as we can easily give the same norm bound on w as with (UC^*): for $\|z - w\| \leqslant 2^{-k} \leqslant 1$, we get $\|w\| \leqslant \|z\| + 1$. The statement (WUC) is thus equivalent to

$$\begin{aligned}&\forall x^X, y^X, z^X, v^X \exists w^X \leqslant_X (\|z\|_X + 1)\, 1_X \forall k^0, b^0 \\ &\Big(v \in Ay \wedge z \in Ax \wedge \|x\|_X, \|y\|_X, \|z\|_X, \|v\|_X <_{\mathbb{R}} b \\ &\qquad \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\varpi(k,b)} \rightarrow \|z -_X w\|_X \leqslant_{\mathbb{R}} 2^{-k} \wedge w \in Ay\Big).\end{aligned}$$

This is of the form Δ as the inner matrix is still (equivalent to a) universal formula and all quantifiers have admissible types as before.

11.6 The strength of not restricting to the domain

While the restriction $x, y \in \mathrm{dom}A$ is essential in uniform continuity statements formulated using the Hausdorff-metric, the principle (UC^*) could similarly well have been formulated without this restriction, i.e. we could consider the existence of a modulus ϖ with

$$\forall k \in \mathbb{N} \forall x, y, z \in X \exists w \in X \Big(z \in Ax \wedge \|x - y\| \leqslant 2^{-\varpi(k)} \rightarrow w \in Ay \wedge \|z - w\| \leqslant 2^{-k}\Big). \qquad (\dagger)$$

What we here want to discuss is that the existence of such a ϖ is already excessively strong. For simplicity, we here only focus on accretive operators. Then, for this, we rely on the following result of Chidume and Morales:[4]

Theorem 11.6.1 ([44]). *Let X, Y be topological spaces and call $A : X \to 2^Y$ lower semi-continuous if for every $x \in X$ and every neighborhood $V(y)$ of $y \in Ax$, there exists a neighborhood $U(x)$ of x such that for all $u \in U(x)$:*

$$A(u) \cap V(y) \neq \varnothing.$$

If X is a real normed space and $A : \mathrm{dom}A \subseteq X \to 2^X$ is a lower semi-continuous and accretive mapping, then A is a single-valued mapping on $\mathrm{intdom}A$.

Theorem 11.6.2. *Let $A \neq \varnothing$ be accretive and assume there exists a ϖ satisfying* (†). *Then* $\mathrm{dom}A = X$, *A is single-valued and uniformly continuous.*

Proof. We first show that $\mathrm{dom}A = X$. As $A \neq \varnothing$, let $(x, u) \in A$ as well as $y \in X$ be given. Then we can inductively construct points $y_1, \dots, y_k$ such that

$$\|x - y_1\|, \|y_i - y_{i+1}\|, \|y_k - y\| \leqslant 2^{-\varpi(0)}$$

for all i. Using (†), we pick $z_i \in Ay_i$ and $z' \in Ay$ with

$$\|z - z_1\|, \|z_i - z_{i+1}\|, \|z_k - z'\| \leqslant 1.$$

In particular $y \in \mathrm{dom}A$.

Now, if such a ϖ exists that satisfies (†), then we actually have that A is lower semi-continuous in the above sense: Let $x \in X(= \mathrm{dom}A)$ and $k \in \mathbb{N}$. Then for any $y \in \overline{B}_{2^{-\varpi(k)}}(x)$ and $z \in Ax$, there exists a $w \in \overline{B}_{2^{-k}}(z)$ such that $w \in Ay$, i.e. $A(y) \cap \overline{B}_{2^{-k}}(z) \neq \varnothing$.

The single-valuedness now follows from Theorem 11.6.1 and the uniform continuity is then immediate by using ϖ. □

Thus, to summarize, we seem to find a sort of dichotomy of proof-theoretic approaches to set-valued operators. Extensionality of A can only be treated up to a certain point in intensional systems if sensible bound extraction theorems shall be retained but if strong-enough fragments of extensionality are actually required by the proof, the methodology actually upgrades this assumption to A being total, single-valued and uniformly continuous, in which case we might switch the formal framework to a system where A is treated as an object of type $X(X)$ together with a modulus of uniform continuity.

[4] Actually, in [44] the authors work with locally accretive mappings but we do not care for this weakened assumption here.

11.7 Characterizations in terms of fragments of maximality

As discussed in Chapter 3, a crucial result from [165] is the equivalence between maximality as well as other analytic closure principles of the operators to extensionality, if the operators in questions are accretive or monotone (in Hilbert spaces). Also recall the extension of that result to monotone operators in Banach spaces from Chapter 10.

As mentioned in the introduction, we extend this correspondence by two new results (and, in a sense, another similar result given in the next Section 11.8), characterizing the previously discussed extensionality principles by respective fragments of the maximal accretiveness or maximal monotonicity as well as the closure of the graph which, for one, shows the robustness of this correspondence and, for another, provides equivalent forms of the fragments of full extensionality considered before in terms of principles which are more easily recognizable in actual applications to results from core mathematics.

Theorem 11.7.1. *Over the systems $\mathcal{V}^\omega$ or $\mathcal{T}^\omega$, the weak extensionality statement (WE) of A is equivalent to* weak closure of the graph of A, *i.e.*

$$\forall x^X, y^X, x_{(\cdot)}^{X(0)}, y_{(\cdot)}^{X(0)}\Big(x \in \mathrm{dom}A \wedge x_n \to_X x \wedge y_n \to_X y \\ \wedge \forall n^0(y_n \in Ax_n) \to \exists w^X(w =_X y \wedge w \in Ax)\Big) \qquad (WG)$$

with $x_n \to_X x$, $y_n \to_X y$ defined as in Theorem 3.4.1, as well as respectively to

1. weak maximal accretivity of A, *i.e.*

$$\forall x^X, u^X\Big(x \in \mathrm{dom}A \wedge \forall y^X, v^X\Big(v \in Ay \\ \to \|x -_X y +_X (u -_X v)\|_X \geqslant_{\mathbb{R}} \|x -_X y\|_X\Big) \to \exists w^X(w =_X u \wedge w \in Ax)\Big),$$

   in the case of $\mathcal{V}^\omega$,

2. weak maximal monotonicity of A, *i.e.*

$$\forall x^X, u^X(x \in \mathrm{dom}A \wedge \forall y^X, v^X\,(v \in Ay \to \langle x -_X y, u -_X v\rangle_X \geqslant_{\mathbb{R}} 0) \\ \to \exists w^X(w =_X u \wedge w \in Ax)),$$

   in the case of $\mathcal{T}^\omega$.

Similar equivalences also hold for $\mathcal{B}^\omega$ *if the statements are appropriately modified with* X^*.

Proof. We only show the equivalence for $\mathcal{V}^\omega$, the other systems can be treated similarly.

$(WE) \Rightarrow (1)$ Let x, u be such that $x \in \mathrm{dom}A$ and

$$\forall y, v(v \in Ay \to \|x - y + (u - v)\| \geqslant \|x - y\|).$$

By axiom (II) of $\mathcal{V}^\omega$, we have

$$1^{-1}((x+u) - J_1^A(x+u)) \in A(J_1^A(x+u))$$

We get by simple arithmetic together with the assumption on x, u that

$$\begin{aligned} 0 &= \left\|x - J_1^A(x+u) + (u - 1^{-1}((x+u) - J_1^A(x+u)))\right\| \\ &\geqslant \left\|x - J_1^A(x+u)\right\| \end{aligned}$$

using additionally the extensionality of the norm. Thus $x = J_1^A(x+u)$ and therefore $u = x + u - J_1^A(x+u) = 1^{-1}((x+u) - J_1^A(x+u))$. Thus, $1^{-1}((x+u) - J_1^A(x+u)) \in A(J_1^A(x+u))$ implies that there exists a $w \in Ax$ such that $w = u$ by weak extensionality of A.

$(1) \Rightarrow (WG)$ Let $x_n \to x$ and $y_n \to y$ as well as $y_n \in Ax_n$ for all n. Let v, w be arbitrary with $v \in Aw$. Then, by axiom (III)

$$\|x_n - w + y_n - v\| \geqslant \|x_n - w\|$$

for all n and thus by taking the limit $\|x - w + y - v\| \geqslant \|x - w\|$. By maximal accretivity, as v, w are arbitrary, we have $w \in Ax$ for some $w = y$.

$(WG) \Rightarrow (WE)$ Let $x = y$ and $z \in Ax$. Then $(x)_n \to y$ and $(z)_n \to z$ for the constant x- and z-sequences $(x)_n$ and $(z)_n$, respectively, and thus there exists a $w = z$ with $w \in Ay$.

□

Theorem 11.7.2. *Over the systems* $\mathcal{V}^\omega$ *or* $\mathcal{T}^\omega$, *the approximate extensionality statement* (AE) *of* A *is equivalent to* approximate closure of the graph of A, *i.e.*

$$\begin{aligned} \forall x^X, y^X, x_{(\cdot)}^{X(0)}, y_{(\cdot)}^{X(0)} \Big(& x \in \mathrm{dom}A \wedge x_n \to_X x \wedge y_n \to_X y \\ & \wedge \forall n^0 (y_n \in Ax_n) \to \forall k^0 \exists w^X \left(\|w -_X y\|_X \leqslant 2^{-k} \wedge w \in Ax \right) \Big) \quad (AG) \end{aligned}$$

as well as respectively to

1. approximate maximal accretivity of A, *i.e.*

$$\forall x^X, u^X \Big(\forall y^X, v^X \Big(x \in \mathrm{dom} A \wedge v \in Ay \rightarrow \|x -_X y +_X (u -_X v)\|_X \geqslant_{\mathbb{R}} \|x -_X y\|_X \Big) \rightarrow \forall k^0 \exists w^X \left(\|w -_X u\|_X \leqslant 2^{-k} \wedge w \in Ax \right) \Big).$$

in the case of $\mathcal{V}^\omega$*,*

2. approximate maximal monotonicity of A, *i.e.*

$$\forall x^X, u^X \Big(x \in \mathrm{dom} A \wedge \forall y^X, v^X \left(v \in Ay \rightarrow \langle x -_X y, u -_X v \rangle_X \geqslant_{\mathbb{R}} 0 \right) \rightarrow \forall k^0 \exists w^X \left(\|w -_X u\|_X \leqslant 2^{-k} \wedge w \in Ax \right) \Big),$$

in the case of $\mathcal{T}^\omega$*.*

Similar equivalences also hold for $\mathcal{B}^\omega$ *if the statements are appropriately modified with* X^**.*

Proof. As before, we only show the equivalence for $\mathcal{V}^\omega$, the other systems can be treated similarly. The cases (1) $\Rightarrow$ (AG) and (AG) $\Rightarrow$ (AE) are completely similar to the previous Theorem 11.7.1 and so we just sketch (AE) $\Rightarrow$ (1): As before with (WE) $\Rightarrow$ (1) in the proof of Theorem 11.7.1, we get $x =_X J^A_1(x+u)$ and $u =_X x+u-J^A_1(x+u) =_X 1^{-1}((x+u)-J^A_1(x+u))$ with $1^{-1}((x+u)-J^A_1(x+u)) \in A(J^A_1(x+u))$.

Approximate extensionality implies that for any k, there exists an $w \in Ax$ such that $\|w-u\| = \left\|w - 1^{-1}((x+u) - J^A_1(x+u))\right\| \leqslant 2^{-k}$ by extensionality of $\|\cdot\|$. □

These fragments can occur, or can be substituted for the full extensionality statement, in various situations in proofs from set-valued operator theory and we want to indicate on a high level what these situations could be: If, after an application of extensionality to points z, x with $z \in Ax$ to infer $w \in Ay$ from $w = z, x = y$, the rest of the formulas in the proof are not extensional in x/y but at least can be reformulated so that an approximation of z/w suffices for the rest to be carried out, then the approximate extensionality principle (AE) suffices. If the rest of the formulas in the proof are not extensional in x/y but at least are extensional in z/w, then the stronger weak extensionality principle (WE) suffices. As before, if after this application the rest of the proof is extensional in both x/y and z/w, then the previous intensional maximality principles already suffice which are provable in the underlying systems.

Of course, the above considerations on the quantitative versions of those principles then provide a guideline of what assumptions have to be placed on A in terms of moduli of uniform continuity to provide an analysis using the bound extraction theorems established in the previous chapters. Examples of these kind of scenarios in previous proof mining applications include [120, 166] where, in the latter case, the above considerations on different fragments of extensionality were crucial for obtaining the analysis.

11.8 Extensionality of the set of zeros

The last investigation regarding extensionality that we want to make here is on the set of zeros of the operator. If $0 \in Ax$, then $\gamma^{-1}(x - x) \in Ax$ or $\gamma^{-1}(\nabla f x - \nabla f x) \in Ax$ for $\gamma > 0$ by the quantifier-free extensionality rule and thus we get $J^A_\gamma x =_X x$ or $\mathrm{Res}^f_\gamma x =_X x$ in the systems $\mathcal{V}^\omega$, $\mathcal{T}^\omega$ or $\mathcal{B}^\omega$ as the uniqueness of the resolvents is provable.

Thus, these systems prove

$$\forall x^X \left(0 \in Ax \to \forall \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to J^A_\gamma x =_X x\right)\right)$$

or

$$\forall x^X \left(0 \in Ax \to \forall \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to \mathrm{Res}^f_\gamma x =_X x\right)\right)$$

respectively, as discussed already before.

Even further, we provably have

$$\forall x^X, z^X \left(z \in Ax \wedge z =_X 0 \to \forall \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to J^A_\gamma x =_X x\right)\right)$$

or

$$\forall x^X, z^X \left(z \in Ax \wedge z =_X 0 \to \forall \gamma^1 \left(\gamma >_{\mathbb{R}} 0 \to \mathrm{Res}^f_\gamma x =_X x\right)\right)$$

respectively. To see this, note that by the quantifier-free extensionality rule we have from $z \in Ax$ that $\gamma^{-1}(\gamma z + x - x) \in Ax$ or $\gamma^{-1}(\nabla f \nabla f^*(\gamma z + \nabla f x) - \nabla f x) \in Ax$ and thus

$$x =_X J^A_\gamma(\gamma z + x) =_X J^A_\gamma x$$

and similarly $\mathrm{Res}^f_\gamma x =_X x$, in both cases using the extensionality of the resolvent and $z = 0$.

As we will see now, the converse assertions are connected to the extensionality of the set of zeros of A.

Theorem 11.8.1. *Over $\mathcal{V}^\omega$, the following are equivalent:*

1. $\forall x^X, z^X \left(J^A_1 x =_X x \wedge z =_X 0 \to z \in Ax\right)$,
2. $\forall x^X, z^X \left(\forall \gamma^1 \left(\gamma >_\mathbb{R} 0 \to J^A_\gamma x =_X x\right) \wedge z =_X 0 \to z \in Ax\right)$,
3. $\forall x^X, y^X, z^X, z'^X \left(x =_X y \wedge z =_X z' =_X 0 \wedge z \in Ax \to z' \in Ay\right)$.

A fortiori, the same holds for $\mathcal{T}^\omega$. This statement also holds for $\mathcal{B}^\omega$ if J^A_γ is replaced by Res^f_γ and X by X^* at the appropriate places.

Proof. We only consider the case for $\mathcal{V}^\omega$. The implication from (1) to (2) is clear. For (2) $\Rightarrow$ (3), let $x = y$ and let $z = z' = 0$ with $z \in Ax$. Then using (2) and the provability of

$$\forall x^X, z^X \left(z \in Ax \wedge z =_X 0 \to \forall \gamma^1 \left(\gamma >_\mathbb{R} 0 \to J^A_\gamma x =_X x\right)\right),$$

we get

$$\begin{aligned} z \in Ax &\leftrightarrow \forall \gamma \left(\gamma > 0 \to J^A_\gamma x = x\right) \\ &\leftrightarrow \forall \gamma \left(\gamma > 0 \to J^A_\gamma y = y\right) \\ &\leftrightarrow z' \in Ay \end{aligned}$$

by the extensionality of J^A_γ. Lastly, for (3) $\Rightarrow$ (1), assume that $J^A_1 x = x$ and $z = 0$. Then by axiom (II), we get

$$1^{-1}(x - J^A_1 x) \in A(J^A_1 x).$$

By (3) and $1^{-1}(x - J^A_1 x) = 0 = z$, we get $z \in Ax$.

The case for $\mathcal{B}^\omega$ follows similarly as Res^f_γ is extensional (recall Proposition 10.2.8). □

Further, as we will see now, this form of the extensionality of the zero set of A is even equivalent to a corresponding fragment of the maximality principle:

Theorem 11.8.2. *Over $\mathcal{V}^\omega$, the following are equivalent:*

1. $\forall x^X, y^X, z^X, z'^X \left(x =_X y \wedge z =_X z' =_X 0 \wedge z \in Ax \to z' \in Ay\right)$,
2. $\forall x^X, z^X \left(\forall (y, v) \in A \left(\|x -_X y -_X v\|_X \geq_\mathbb{R} \|x -_X y\|_X\right) \wedge z =_X 0 \to z \in Ax\right)$.

A fortiori, the same holds for $\mathcal{T}^\omega$ where all are also equivalent to

3. $\forall x^X, z^X \left(\forall (y, v) \in A \left(\langle x -_X y, -_X v\rangle_X \geq_\mathbb{R} 0\right) \wedge z =_X 0 \to z \in Ax\right)$.

Lastly, the same holds for $\mathcal{B}^\omega$ *if items (1) and (2) are replaced by*

(1)' $\forall x^X, y^X, z^{*X^*}, z'^{*X^*} \, (x =_X y \wedge z^* =_{X^*} z'^* =_{X^*} 0 \wedge z^* \in Ax \to z'^* \in Ay)$,

(2)' $\forall x^X, z^{*X^*} \, (\forall (y, y^*) \in A \, (\langle x -_X y, -_{X^*} y^* \rangle_{X^*} \geqslant_{\mathbb{R}} 0) \wedge z^* =_{X^*} 0 \to z^* \in Ax)$.

Proof. Also here, we only consider the case for $\mathcal{V}^\omega$. For the direction (1) ⇒ (2), assume $\forall (y, v) \in A \, (\|x - y - v\| \geqslant \|x - y\|)$. Without any assumptions, we have $1^{-1}(x - J^A_1 x) \in A(J^A_1 x)$. Therefore, we have

$$0 = \left\|x - J^A_1 x - 1^{-1}(x - J^A_1 x)\right\| \geqslant \left\|x - J^A_1 x\right\|$$

and thus $x = J^A_1 x$ as well as $1^{-1}(x - J^A_1 x) = 0$. (1) yields $z \in Ax$ for any $z = 0$.

Conversely, for (2) ⇒ (1), assume that $x = y$ and $z = z' = 0$ as well as $z \in Ax$. Then by accretivity and the extensionality of the norm, we have

$$\forall (a, b) \in A \, (\|y - a - b\| \geqslant \|y - a\|) \, .$$

(2) yields that $z' \in Ay$. □

However, similar to the full extensionality statement, already the provability of the extensionality of the zero set of A as formulated by

$$\forall x^X, y^X \, (x =_X y \wedge 0 \in Ax \to 0 \in Ay)$$

will result in any standard model, i.e. any model based on $\mathcal{S}^{\omega,X}$ (or $\mathcal{S}^{\omega,X,X^*}$ in the case of $\mathcal{B}^\omega$), for extensions of $\mathcal{V}^\omega$, $\mathcal{T}^\omega$ (or $\mathcal{B}^\omega$) that allow for bound extractions to be rather degenerate: For suppose this extensionality principle would be provable in some extension, then we would be able to extract a functional ω such that

$$\forall x, y \in X \forall b \in \mathbb{N} \left(\|x\|, \|y\| \leqslant b \wedge \|x - y\| \leqslant 2^{-\omega(b)} \wedge 0 \in Ax \to 0 \in Ay\right).$$

Note that in the presence of such an ω, the set $A^{-1}0$ is open in X as for $x \in A^{-1}0$ with $\|x\| \leqslant b$, we have that any y with $\|x - y\| \leqslant 2^{-\omega(b+1)}$ also satisfies $y \in A^{-1}0$. However, for any maximal operator, as it is closed in $X \times X$, we get that also $A^{-1}0$ is closed and therefore $A^{-1}0$ would be clopen, i.e. would be equal to $\varnothing$ or X, which renders any discussion on zeros of the operator pointless. The same argument can also be made in the context of X^* and $\mathcal{B}^\omega$.

12 Quantitative results on Mann-iterations for set-valued mappings in Banach spaces

12.1 Set-valued nonexpansive maps and Mann-type iterations

As a last application, we provide quantitative results on a Mann-type iteration of set-valued mappings which are nonexpansive w.r.t. the Hausdorff-metric.

Concretely, let X be a Banach space and denote by $CB(X)$ the collection of non-empty, closed and bounded subsets of X. Then the Hausdorff-metric

$$H(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\right\}$$

is well-defined and real-valued for $A, B \in CB(X)$. We write

$$d(x,A) = \inf_{a\in A}\|x-a\|$$

for a given set $A \in CB(X)$ as before. A set-valued map $T : D \subseteq X \to CB(X)$ is called nonexpansive if

$$H(Tx,Ty) \leqslant \|x-y\|$$

for any $x, y \in D$. We say that a point x is a fixed point of T if $x \in Tx$ and we denote the set of fixed points of T by $F(T)$.

The following is a rather immediate consequence of the definition of the Hausdorff-metric:

Lemma 12.1.1 (see e.g. [82])**.** *Let $A, B \in CB(X)$. For any $a \in A$ and $\varepsilon > 0$, there exists some $b \in B$ with*

$$\|a-b\| \leqslant H(A,B) + \varepsilon.$$

Based on this lemma, it is immediately clear that given a non-empty convex set K and starting points $x_0 \in K$, $y_0 \in Tx_0$ together with scalars $\alpha_n \in [0,1]$ and $\gamma_n \in (0,\infty)$, one can inductively define an iteration

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n y_n \tag{†}$$

where $y_{n+1} \in Tx_{n+1}$ is chosen such that $\|y_{n+1} - y_n\| \leqslant H(Tx_{n+1}, Tx_n) + \gamma_n$. This iteration defined in that way was studied in [196] and in the case that the set K is additionally compact, the authors obtained the following convergence result:

Theorem 12.1.2 ([196]). *Let $K \subseteq X$ be non-empty, convex and compact. Let $T : K \to CB(K)$ be a set-valued map that is nonexpansive and suppose that $F(T) \neq \emptyset$ as well as $T(p) = \{p\}$ for each $p \in F(T)$. Let (x_n) be defined as in (†) with starting points $x_0 \in K$, $y_0 \in Tx_0$ and scalars $(\alpha_n) \subseteq [0,1]$ and $(\gamma_n) \subseteq (0,\infty)$ such that*

1. $\lim_{n\to\infty} \gamma_n \to 0$,
2. $0 < \liminf_{n\to\infty} \alpha_n \leqslant \limsup_{n\to\infty} \alpha_n < 1$.

Then (x_n) converges strongly to a fixed point of T.

The main feature of the sequence exploited in the proof is that it is Fejér monotone (see in particular [46, 47]). This well-studied class of sequences possesses very general convergence theorems which guarantee the weak convergence of such sequences under very mild asymptotic regularity assumptions. In compact (metric) spaces, like in the above result, the convergence is in particular strong.

These general convergence results for Fejér monotone sequences from compact sets were analyzed through the lens of proof mining in [112] where, under the assumption of the existence of moduli which witness quantitative reformulations of the central properties involved, a construction of a rate of metastability for the sequence in question is presented. Further, in [114], a general principle of metric regularity is studied (encompassing various forms of well-known regularity assumptions from nonlinear analysis and optimization like metric subregularity, weak sharp minima, error bounds, etc.) and under the assumption of such a metric regularity principle, the authors then provide a construction for a computable as well as highly uniform full rate of convergence for a given Fejér monotone iteration which moreover holds in the absence of any compactness assumptions.

These general but abstract proof mining results were previously successfully instantiated for many different situations in which Fejér monotone sequences occur to derive rates of metastability and rates of convergence. In particular, we want to mention the applications in the context of the asymptotic behavior of the composition of two mappings [113], the proximal point algorithm in uniformly convex Banach spaces [105] and in CAT(0)-spaces [135, 136], subgradient-methods for equilibrium problems [169] as well as algorithms for finding zeros of differences of monotone operators [166].

It is also here that we apply the results from [112, 114] to derive rates of metastability and rates of convergence (under a metric regularity assumption) for the above iteration which are, as before, not only computable in their parameters but also highly uniform. For that, we need to extract the previously mentioned moduli witnessing quantitative versions of the Fejér monotonicity and asymptotic regularity which themselves arise from an application of proof mining to the respective proofs of these properties given in the course of the proof of Theorem 12.1.2 in [196]. As these proofs in particular rely on the utilization of the Hausdorff-metric, this application given here is in particular to be seen as a case study to illustrate the applicability of the treatment of the Hausdorff-metric discussed in Chapter 11.

12.2 The central assumptions and their quantitative content

In this section, we now first discuss the central assumptions present in Theorem 12.1.2 and in particular discuss (using the underlying logical methodology) what kind of quantitative assumptions they entail to potentially feature in the analysis of the main theorem given later.

The first important assumption present in Theorem 12.1.2 is the compactness of the set K. This compactness assumption on K is witnessed in the following by a quantitative modulus of compactness introduced in [69] under the name of a *modulus of total boundedness*[1] which takes the form of a function $\gamma : \mathbb{N} \to \mathbb{N}$ such that for any

[1]In [112], the name *II-modulus of total boundedness* is used but we here follow the conventions from [69] where such a modulus is just called a *modulus of total boundedness*.

$k \in \mathbb{N}$ and for any $(x_n) \subseteq K$:

$$\exists 0 \leqslant i < j \leqslant \gamma(k) \left(\|x_i - x_j\| \leqslant \frac{1}{k+1} \right).$$

As discussed in [112], such a modulus exists if, and only if, K is compact and we refer to [112] for various discussions on the construction of such moduli for certain concrete classes of compact sets and spaces.

As a second assumption, we find the non-emptyness of the fixed point set $F(T)$ which will be represented by a concrete witness p_0 (i.e. $p_0 \in K$ and $p_0 \in Tp_0$) in the following. As follows by the metatheorems, the bounds extracted later will of course only depend on an upper bound on the norm of p_0, which by the compactness and therefore the boundedness of K, is in particular represented by any upper bound on the diameter of K.

One of the most crucial assumptions, in some sense, is the single-valuedness of T on actual fixed points, i.e. the assumption that $Tp = \{p\}$ if $p \in F(T)$. This implication is equivalent to

$$\forall p \in K \, (d(p, Tp) = 0 \to H(\{p\}, Tp) = 0) \tag{$*$}$$

which in turn unravels into

$$\forall p \in K \forall k \in \mathbb{N} \exists j \in \mathbb{N} \left(d(p, Tp) \leqslant \frac{1}{j+1} \to H(\{p\}, Tp) \leqslant \frac{1}{k+1} \right)$$

and in that way the logical methodology induces[2] a modulus $\theta : \mathbb{N} \to \mathbb{N}$ bounding (and thus witnessing) such a j in terms of k, i.e. such that[3]

$$\forall p \in K \forall k \in \mathbb{N} \left(d(p, Tp) \leqslant \frac{1}{\theta(k)+1} \to H(\{p\}, Tp) \leqslant \frac{1}{k+1} \right).$$

[2] To formalize the above statement in the language of the previous systems, we have to represent the set $\{p\}$ using an additional constant χ_s of type $0(X)(X)$ together with two axioms expressing that $\chi_s(p, \cdot)$ intensionally codes the singleton $\{p\}$ for all p:

$$\forall p^X \, (\chi_s(p,p) =_0 0),$$
$$\forall p^X, x^X \, (\chi_s(p,x) =_0 0 \to x =_X p).$$

In that way, the treatment of $\{p\}$ is intensional as we can not prove that for $x = p$, we also have $x \in \{p\}$ in the sense that $\chi_s(p,x) =_0 0$. Then $H(\{p\}, Tx)$ can be introduced using χ_s and some χ_T coding T as detailed in Chapter 11.

[3] Note that the (full) independence on p is suggested by the logical methodology as the set K is in particular bounded.

Note that by a simple compactness argument, possessing such a modulus is equivalent to the property $(*)$ in compact spaces:

Lemma 12.2.1. *Let K be compact and let $T : K \to CB(K)$ be a nonexpansive operator. Then T satisfies $(*)$ if, and only if,*

$$\forall k \in \mathbb{N} \exists j \in \mathbb{N} \forall p \in K \left(d(p, Tp) \leqslant \frac{1}{j+1} \to H(\{p\}, Tp) \leqslant \frac{1}{k+1} \right). \tag{$**$}$$

Proof. Clearly, $(**)$ implies $(*)$. Conversely, suppose that $(**)$ fails, i.e. suppose there exists a $k \in \mathbb{N}$ such that for any $j \in \mathbb{N}$:

$$\exists p_j \in K \left(d(p_j, Tp_j) \leqslant \frac{1}{j+1} \wedge H(\{p_j\}, Tp_j) > \frac{1}{k+1} \right).$$

Then $d(p_j, Tp_j) \leqslant \frac{1}{j+1}$ implies that for any $j \geqslant 1$, there exists a $q_j \in Tp_j$ such that $\|p_j - q_j\| \leqslant 1/j$. Further, $H(\{p_j\}, Tp_j) > \frac{1}{k+1}$ now implies that there exists a $q'_j \in Tp_j$ such that $\left\|p_j - q'_j\right\| > \frac{1}{k+1}$.

We now pick subsequences p_{j_i}, q_{j_i} and q'_{j_i} such that $p_{j_i} \to p$, $q_{j_i} \to q$ and $q'_{j_i} \to q'$ with $p, q, q' \in K$. Then $\|p - q\| = 0$ and $H(Tp_{j_i}, Tp) \to 0$ for $i \to \infty$ as T is nonexpansive. Thus in particular $d(q_{j_i}, Tp), d(q'_{j_i}, Tp) \to 0$ which yields

$$d(q, Tp) \leqslant \|q - q_{j_i}\| + d(q_{j_i}, Tp) \to 0$$

and thus $d(p, Tp) = d(q, Tp) = 0$. Similarly $d(q', Tp) = 0$ and thus $q' \in Tp$. However, we have $\|p - q'\| \geqslant \frac{1}{k+1}$ and so $H(\{p\}, Tp) \geqslant \|p - q'\| \geqslant \frac{1}{k+1}$. This is a contradiction to $(*)$. □

In that way, the existence of such a modulus is implied already by the assumptions in Theorem 12.1.2.

At last, we consider the assumptions on the auxiliary sequences γ_n and α_n. For γ_n, where it is assumed that

$$\lim_{n \to \infty} \gamma_n \to 0,$$

we will later rely on a rate of convergence τ witnessing this property, i.e. on a τ satisfying

$$\forall k \in \mathbb{N} \forall n \geqslant \tau(k) \left(\gamma_n \leqslant \frac{1}{k+1} \right).$$

For α_n, the assumption that

$$0 < \liminf_{n \to \infty} \alpha_n \leqslant \limsup_{n \to \infty} \alpha_n < 1$$

is witnessed by a value $a \in \mathbb{N}^*$ with the property

$$\forall n \geqslant a \left(\frac{1}{a} \leqslant \alpha_n \leqslant 1 - \frac{1}{a} \right)$$

in similarity to [55].

Remark 12.2.2. For the previous treatment of the Hausdorff-metric, it was crucial that the sets come equipped with a modulus witnessing their boundedness. Note that the existence of such a modulus is immediate for sets of the form Tx as $Tx \in CB(K)$ and thus $Tx \subseteq K$ which is bounded as K is compact. In that way, for the quantitative results, we will later rely on a bound on the diameter of K (as mentioned before). Note that such a bound can not be computed from the modulus of total boundedness γ for K as this modulus is only non-effectively equivalent to the total boundedness of K in the usual sense and thus only implies the boundedness of K non-effectively (see [112] for a further discussion of this).

12.3 Suzuki's lemma and its analysis

The main analytical ingredient into the convergence proof from [196] is a well-known lemma from Suzuki [200]:

Lemma 12.3.1 (Suzuki [200]). *Let $(x_n), (y_n)$ be bounded sequences in a Banach space X and let $(\alpha_n) \subseteq [0,1]$ be such that $0 < \liminf_{n\to\infty} \alpha_n \leqslant \limsup_{n\in\mathbb{N}} \alpha_n < 1$. Suppose that $x_{n+1} = \alpha_n x_n + (1-\alpha_n)y_n$ as well as*

$$\limsup_{n\to\infty} (\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\|) \leqslant 0.$$

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

This lemma was analyzed quantitatively in [55] and we will rely in the following on this analysis:

Lemma 12.3.2 (Dinis and Pinto [55]). *Let $(x_n), (y_n)$ be sequences in a Banach space X with $\|x_n\|, \|y_n\| \leqslant b$ for $b \in \mathbb{N}^*$ and let $(\alpha_n) \subseteq [0,1]$ be such that there exists a $a \in \mathbb{N}^*$ with the property*

$$\forall n \geqslant a \left(\frac{1}{a} \leqslant \alpha_n \leqslant 1 - \frac{1}{a} \right).$$

Suppose that $x_{n+1} = \alpha_n x_n + (1-\alpha_n)y_n$ as well as that there exists a monotone function $\tau : \mathbb{N} \to \mathbb{N}$ such that

$$\forall k \in \mathbb{N} \forall n \geqslant \tau(k) \left((\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\|) \leqslant \frac{1}{k+1} \right).$$

Then for any $k \in \mathbb{N}$ *and any* $g : \mathbb{N} \to \mathbb{N}$*:*

$$\exists n \leqslant \varphi_{a,\tau,b}(k,g) \forall m \in [n; n+g(n)] \left(\|x_m - y_m\| \leqslant \frac{1}{k+1} \right),$$

where $\varphi_{a,\tau,b}(k,g) = \max\{a, \tau(t(2t+1)a^t(k+1)-1)\} + (bt(2t+1)a^t(k+1)-1)t + r_0$ *for*

$$r_i := \begin{cases} 0 & \text{if } i = b(k+1), \\ t + r_{i+1} + \widehat{g}(\max\{a, \tau(t(2t+1)a^t(k+1)-1)\} + it + r_{i+1}) & \text{if } i < b(k+1). \end{cases}$$

where $\widehat{g}(m) = t + g(m)$ *and* $t = 2ba(k+1)$*.*

12.4 Fejér monotonicity and metastability

We now present the extractions of the quantitative versions of Fejér monotonicity and asymptotic regularity.

For this, we first need to define an appropriate notion of an approximate solution (i.e. of an approximate fixed point) as the results given in [112] rely on uniform reformulations of the respective properties in terms of such approximate solutions. For our concrete situation here, note that p is a fixed point of T if, and only if, $d(p, Tp) = 0$ (as Tp is closed since $Tp \in CB(K)$). In that vein, we call p a $\frac{1}{k+1}$-approximate fixed point of T if

$$d(p, Tp) \leqslant \frac{1}{k+1}$$

and define correspondingly

$$AF_k = \left\{ p \in K \mid d(p, Tp) \leqslant \frac{1}{k+1} \right\}$$

as the set of approximate solutions which extend the set of full solutions

$$F = \{p \in K \mid d(p, Tp) = 0\} = F(T).$$

Now, for the Fejér monotonicity of (x_n), we concretely strive to establish the existence of the following modulus relative to the chosen AF_k:

Definition 12.4.1 ([112])**.** A function $\chi : \mathbb{N}^3 \to \mathbb{N}$ is a modulus of uniform Fejér monotonicity for (x_n) w.r.t. AF_k if for any $n, m, r \in \mathbb{N}$, any $p \in AF_{\chi(k,m,r)}$ and any $l \leqslant m$:

$$\|x_{n+l} - p\| < \|x_n - p\| + \frac{1}{r+1}.$$

For this, we can now extract the following from the proof of Fejér monotonicity given in [196] for the sequence (x_n) defined as in (†).

Lemma 12.4.2. *Let θ be such that*

$$\forall p \in K \forall k \in \mathbb{N} \left(d(p, Tp) \leqslant \frac{1}{\theta(k)+1} \to H(\{p\}, Tp) \leqslant \frac{1}{k+1} \right).$$

Then the sequence (x_n) defined as in (†) is uniformly Fejér monotone w.r.t. AF_k with a modulus

$$\chi(n, m, r) = \theta(m(r+1)+1).$$

Proof. Let p be given with $d(p, Tp) \leqslant \frac{1}{\chi(n,m,r)+1}$. Then

$$\begin{aligned}
\|x_{n+1} - p\| &\leqslant (1-\alpha_n)\|x_n - p\| + \alpha_n \|y_n - p\| \\
&\leqslant (1-\alpha_n)\|x_n - p\| + \alpha_n d(y_n, Tp) + \alpha_n(\|y_n - p\| - d(y_n, Tp)) \\
&\leqslant (1-\alpha_n)\|x_n - p\| + \alpha_n H(Tx_n, Tp) + \alpha_n(\|y_n - p\| - d(y_n, Tp)) \\
&\leqslant \|x_n - p\| + (\|y_n - p\| - d(y_n, Tp))
\end{aligned}$$

and by induction we get

$$\|x_{n+l} - p\| \leqslant \|x_n - p\| + \sum_{i=0}^{l-1} (\|y_{n+i} - p\| - d(y_{n+i}, Tp))$$

for any $l \geqslant 1$. It is rather immediate to see that in general, for non-empty sets $Y, Z \subseteq X$ and a point x, we have $d(x, Y) \leqslant d(x, Z) + H(Y, Z)$ and instantiating this yields

$$\|y_{n+i} - p\| = d(y_{n+i}, \{p\}) \leqslant d(y_{n+i}, Tp) + H(\{p\}, Tp)$$

and thus $\|y_{n+i} - p\| - d(y_{n+i}, Tp) \leqslant H(\{p\}, Tp)$. As now $p \in AF_{\chi(n,m,r)}$, we get

$$H(\{p\}, Tp) < \frac{1}{m(r+1)}.$$

In particular, in that case we have

$$\begin{aligned}
\|x_{n+l} - p\| &\leqslant \|x_n - p\| + mH(\{p\}, Tp) \\
&< \|x_n - p\| + \frac{1}{r+1}
\end{aligned}$$

for $l \leqslant m$. □

Remark 12.4.3. Note that if T satisfies $(*)$, the sequence is Fejér monotone w.r.t. $F(T)$ in the usual sense as can be shown by following the proof of the above Lemma 12.4.2. In particular, this results holds without any compactness assumption for K.

For the asymptotic behavior, we are interested in the following type of quantitative information:

Definition 12.4.4 ([112])**.** A function Φ is an approximate F-point bound for (x_n) w.r.t. AF_k if for any $k \in \mathbb{N}$:

$$\exists n \leqslant \Phi(k)\, (x_n \in AF_k)\,.$$

The construction of such a Φ for the sequence studied here relies on analyzing the proof of the statement $d(x_n, Tx_n) \to 0$ from [196] which relies on Suzuki's lemma. Concretely, we get the following:

Lemma 12.4.5. *Let b be a bound on the diameter of K and let $(\alpha_n) \subseteq [0,1]$ be such that there exists an $a \in \mathbb{N}^*$ with the property*

$$\forall n \geqslant a \left(\frac{1}{a} \leqslant \alpha_n \leqslant 1 - \frac{1}{a}\right).$$

Let τ be a monotone rate of convergence for $\gamma_n \to 0$. Let $\varphi_{a,\tau,b}$ be defined as in Lemma 12.3.2. Then (x_n) defined as in (†) has approximate F-points with an approximate F-point bound

$$\Phi(k) = \varphi_{a,\tau,b}(k, 0).$$

Proof. As in [196], we can derive

$$\|y_{n+1} - y_n\| \leqslant H(Tx_{n+1}, Tx_n) + \gamma_n \leqslant \|x_{n+1} - x_n\| + \gamma_n$$

which yields that

$$\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\| \leqslant \gamma_n$$

and thus τ satisfies the assumption of Lemma 12.3.2. Applying Lemma 12.3.2, we get that for any $k \in \mathbb{N}$ and any $g : \mathbb{N} \to \mathbb{N}$:

$$\exists n \leqslant \varphi_{a,\tau,b}(k, g) \forall m \in [n; n + g(n)] \left(\|x_m - y_m\| \leqslant \frac{1}{k+1}\right).$$

In particular, we get for any $k \in \mathbb{N}$ that

$$\exists n \leqslant \varphi_{a,\tau,b}(k, 0) \left(\|x_n - y_n\| \leqslant \frac{1}{k+1}\right)$$

which yields that for this n, we have

$$d(x_n, Tx_n) \leqslant \|x_n - y_n\| \leqslant \frac{1}{k+1},$$

i.e. $x_n \in AF_k$. □

Remark 12.4.6. Note that

$$\varphi_{a,\tau,b}(k,0) = \max\{a, \tau(t(2t+1)a^t(k+1)-1)\} + (bt(2t+1)a^t(k+1)-1)t + 2b(k+1)t$$

for $t = 2ba(k+1)$.

Lastly, we show that $F(T)$ is not only closed but that it is even sufficiently uniformly closed respective to the approximations AF_k in a concrete way introduced in [112]:

Definition 12.4.7 ([112])**.** The solution set F is called uniformly closed w.r.t. AF_k with moduli δ, ω if for any $k \in \mathbb{N}$, any $q \in AF_{\delta(k)}$ and any p with $\|p - q\| \leqslant 1/(\omega(k)+1)$, we have $p \in AF_k$.

Lemma 12.4.8. *The set $F = F(T)$ is uniformly closed w.r.t. AF_k with moduli*

$$\begin{cases} \delta(k) = 2k+1, \\ \omega(k) = 4k+3. \end{cases}$$

Proof. Note that we have

$$\begin{aligned} d(p, Tp) &\leqslant d(p, Tq) + H(Tp, Tq) \\ &\leqslant \|p - q\| + d(q, Tq) + \|q - p\| \end{aligned}$$

and thus if $q \in AF_{2k+1}$ and $\|p - q\| \leqslant \frac{1}{4(k+1)}$, then $d(p, Tp) \leqslant \frac{1}{k+1}$, i.e. $p \in AF_k$. □

Combined, we can now apply the general result from [112] to get the following quantitative version of Theorem 12.1.2:

Theorem 12.4.9. *Let γ be a modulus of total boundedness for K. Let b be a bound on the diameter of K and let $(\alpha_n) \subseteq [0,1]$ be such that there exists an $a \in \mathbb{N}^*$ with the property*

$$\forall n \geqslant a \left(\frac{1}{a} \leqslant \alpha_n \leqslant 1 - \frac{1}{a} \right).$$

Let τ be a monotone rate of convergence for $\gamma_n \to 0$. Let θ be such that

$$\forall p \in K \forall k \in \mathbb{N} \left(d(p, Tp) \leqslant \frac{1}{\theta(k)+1} \to H(\{p\}, Tp) \leqslant \frac{1}{k+1} \right).$$

Let $\varphi_{a,\tau,b}(k,0)$ be defined as in Remark 12.4.6, i.e.

$$\varphi_{a,\tau,b}(k,0) = \max\{a, \tau(t(2t+1)a^t(k+1)-1)\} + (bt(2t+1)a^t(k+1)-1)t + 2b(k+1)t$$

for $t = 2ba(k+1)$. *Then* (x_n) *defined as in* (†) *is Cauchy and moreover, for all* $k \in \mathbb{N}$ *and all* $g : \mathbb{N} \to \mathbb{N}$,

$$\exists N \leqslant \Psi(k,g) \forall i,j \in [N; N+g(N)] \left(\|x_i - x_j\| \leqslant \frac{1}{k+1} \wedge x_i \in AF_k \right)$$

where $\Psi(k,g) = \Psi_0(P,k,g)$ *for* $P = \gamma(4k+3)$ *and with*

$$\begin{cases} \Psi_0(0,k,g) = 0, \\ \Psi_0(n+1,k,g) = \varphi_{a,\tau,b}(\chi^M_{k,g}(\Psi_0(n,k,g), 8k+7), 0), \end{cases}$$

and where

$$\chi(n,m,r) = \theta(m(r+1)+1),$$
$$\chi_k(n,m,r) = \max\{2k+1, \chi(n,m,r)\},$$
$$\chi^M_{k,g}(n,r) = \max\{\chi_k(i,g(i),r) \mid i \leqslant n\}.$$

Proof. The result rather immediately follows from Theorem 5.3 in [112] (which itself builds on Theorem 5.1 in [112]) by instantiating the bound given there with the moduli obtained in Lemmas 12.4.2, 12.4.5, 12.4.8 and where $G = H = \mathrm{id}$ and thus $\alpha_G(k) = \beta_H(k) = k$. □

Remark 12.4.10. Theorem 12.4.9 is a real finitization of Theorem 12.1.2 in the sense of Tao as it only references finite segments of the iteration (x_n) but it trivially implies back the original formulation of Theorem 12.1.2 as all the moduli naturally exist and since metastability is (non-effectively) equivalent to convergence (see also Remark 5.5 in [112]).

12.5 Moduli of regularity and rates of convergence

In this section, using the results from [114], we give constructions for rates of convergence based on the assumption of a (very general) kind of regularity notion as discussed in the introduction.

The central notion here is consequently the following instantiation of the abstract notion of a modulus of regularity from [114]:

Definition 12.5.1. Let $z \in F(T)$ and $r > 0$. A function $\phi : (0,\infty) \to (0,\infty)$ is called a modulus of regularity for T w.r.t $\overline{B}_r(z)$ if for all $\varepsilon > 0$ and all $x \in \overline{B}_r(z)$:

$$d(p,Tp) < \phi(\varepsilon) \to d(x,F(T)) < \varepsilon.$$

If there is a $z \in F(T)$ such that ϕ is a modulus of regularity w.r.t. $\overline{B}_r(z)$ for all $r > 0$, then ϕ is just called a modulus of regularity for T.

Remark 12.5.2. Note that the work [114] is written in the context of a formal setup where instead of using sets F/AF_k as above to formulate the solutions and approximative solutions, a function $F : X \to [0, +\infty]$ is employed and the roles of the sets F/AF_k are (conceptually) replaced by $\mathrm{zer}F/\{x \mid F(x) \leqslant \varepsilon\}$ for $\varepsilon > 0$. The above notion arises from the general definition given in [114] by using $F(x) := d(x, Tx)$ but we in the following suppress this whole setup from [114].

Note that the function $d(p, Tp)$ is continuous in p if T is nonexpansive as

$$\begin{aligned} d(p, Tp) &\leqslant d(p, Tq) + H(Tp, Tq) \\ &\leqslant \|p - q\| + d(q, Tq) + \|q - p\| \end{aligned}$$

and thus

$$|d(p, Tp) - d(q, Tq)| \leqslant 2\|p - q\|.$$

It follows from Proposition 3.3 of [114] that any such nonexpansive map T has a modulus of regularity (albeit in general being uncomputable) if K is compact.

Under the assumption of such a modulus, we now get the following result on rates of convergence by instantiating the corresponding abstract result from [114]:

Theorem 12.5.3. *Let $z \in F(T) \neq \varnothing$ and let b be a bound on the diameter of K. Assume that K is closed. Let (x_n) be defined as in (†). Assume that T satisfies (∗). Let $(\alpha_n) \subseteq [0, 1]$ be such that there exists an $a \in \mathbb{N}^*$ with the property*

$$\forall n \geqslant a \left(\frac{1}{a} \leqslant \alpha_n \leqslant 1 - \frac{1}{a} \right).$$

Let τ be a monotone rate of convergence for $\gamma_n \to 0$. Let $\varphi_{a,\tau,b}(k, 0)$ be defined as in Remark 12.4.6, i.e.

$$\varphi_{a,\tau,b}(k, 0) = \max\{a, \tau(t(2t+1)a^t(k+1) - 1)\} + (bt(2t+1)a^t(k+1) - 1)t + 2b(k+1)t$$

for $t = 2ba(k+1)$. Let ϕ be a modulus of regularity for T w.r.t. $\overline{B}_b(z)$. Then (x_n) is Cauchy with

$$\forall \varepsilon > 0 \forall i, j \geqslant \varphi_{a,\tau,b}\left(\left\lceil \frac{1}{\phi(\varepsilon/2)} \right\rceil, 0 \right) (d(x_i, x_j) < \varepsilon).$$

and further (x_n) converges to a fixed point of T with a rate of convergence

$$\varphi_{a,\tau,b}\left(\left\lceil \frac{1}{\phi(\varepsilon/2)} \right\rceil, 0 \right).$$

Proof. The result is a straightforward instantiation of the general abstract Theorem 4.1 from [114], using the previous Lemma 12.4.5 by which we have that

$$\forall \varepsilon > 0 \exists n \leqslant \varphi_{a,\tau,b}\left(\left\lceil\frac{1}{\varepsilon}\right\rceil, 0\right)(d(x_n, Tx_n) < \varepsilon).$$

Note for this that the sequence (x_n) is Fejér monotone w.r.t. $F(T)$ by Remark 12.4.3 since T satisfies $(*)$. That (x_n) converges to a fixed point of T with the given rate follows from Theorem 4.1, (i) in [114] for which we need that K is complete (which follows as X is a Banach space and as K is closed) and that $F(T)$ is closed which follows from the fact that $d(p, Tp)$ is uniformly continuous in p and $F(T) = (d(\cdot, T\cdot))^{-1}(0)$. □

Remark 12.5.4. Note that the above Theorem 12.5.3 holds without any compactness assumptions on K. Thus, in the presence of a modulus of regularity, the convergence result from Theorem 12.1.2 immediately holds for any closed, bounded and non-empty set K and any nonexpansive mapping T with $F(T) \neq \varnothing$ that satisfies $(*)$.

At last, we look at a notion for multi-valued mappings where simple instances of such moduli of regularity can be derived. Following Senter and Dotson [190], a multivalued mapping $T : K \to \mathrm{CB}(K)$ is said to satisfy Condition I if there is a nondecreasing function $f : [0, \infty) \to [0, \infty)$ with $f(0) = 0$, $f(r) > 0$ for $r \in (0, \infty)$ and

$$d(x, Tx) \geqslant f(d(x, F(T)))$$

for all $x \in K$. If the property that $f(r) > 0$ for $r \in (0, \infty)$ is witnessed in a uniform and quantitative way by a function $\phi : (0, \infty) \to (0, \infty)$ with

$$f(r) < \phi(\varepsilon) \to r < \varepsilon$$

for any $r, \varepsilon > 0$, then such a ϕ is clearly already a modulus of regularity for T. This in particular is true for mappings that satisfy Condition II of Senter and Dotson [190], i.e. where there exists a real $\alpha > 0$ such that

$$d(x, Tx) \geqslant \alpha d(x, F(T))$$

where then ϕ can be given by $\phi(\varepsilon) = \alpha\varepsilon$. Examples of mappings which satisfy Condition II are for instance discussed in [190] and for these, the above rates of convergence therefore instantiate immediately.

Bibliography

[1] Y.I. Alber. Metric and generalized projections in Banach spaces: properties and applications. In A.G. Kartsatos, editor, *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, pages 15–50. Marcel Dekker, New York, 1996.

[2] K. Aoyama and M. Toyoda. Approximation of common fixed points of strongly nonexpansive sequences in a Banach space. *Journal of Fixed Point Theory and Applications*, 21, 2019. Article no. 35.

[3] V. Barbu. A class of boundary problems for second order abstract differential equations. *Journal of the Faculty of Science, University of Tokyo*, 19:295–319, 1972.

[4] V. Barbu. *Nonlinear semigroups and differential equations in Banach spaces*. Springer Netherlands, 1976.

[5] V. Barbu. *Nonlinear Differential Equations of Monotone Types in Banach Spaces*. Springer Monographs in Mathematics. Springer New York, NY, 2010.

[6] H.H. Bauschke. The composition of projections onto closed convex sets in Hilbert space is asymptotically regular. *Proceedings of the American Mathematical Society*, 131:141–146, 2003.

[7] H.H. Bauschke and J.M. Borwein. Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis*, 4(1):27–67, 1997.

[8] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. *Communications in Contemporary Mathematics*, 3:615–647, 2001.

[9] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42:596–636, 2003.

[10] H.H. Bauschke, J.M. Borwein, and A.S. Lewis. The method of cyclic projections for closed convex sets in Hilbert space. In Y. Censor and S. Reich, editors, *Recent developments in optimization theory and nonlinear analysis (Jerusalem, 1995)*, pages 1–38. American Mathematical Society, Providence, RI, 1997.

[11] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, Cham, 2nd edition, 2017.

[12] H.H. Bauschke and A.S. Lewis. Dykstra's algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.

[13] H.H. Bauschke, S.M. Moffat, and X. Wang. Firmly Nonexpansive Mappings and Maximally Monotone Operators: Correspondence and Duality. *Set-Valued and Variational Analysis*, 20:131–153, 2012.

[14] H.H. Bauschke, W.M. Moursi, and X. Wang. Generalized monotone operators and their averaged resolvents. *Mathematical Programming*, 189:55–74, 2021.

[15] H.H. Bauschke, X. Wang, and L. Yao. General resolvents for monotone operators: characterization and extension. In Y. Censor, M. Jiang, and G. Wang, editors, *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, pages 57–74. Medical Physics Publishing, Madison, WI, USA, 2010.

[16] M. Bezem. Strongly majorizable functionals of finite type: a model for bar recursion containing discontinuous functionals. *The Journal of Symbolic Logic*, 50:652–660, 1985.

[17] E. Bishop. Mathematics as a numerical language. In A. Kino, J. Myhill, and R.E. Vesley, editors, *Intuitionism and proof theory*, pages 53–71. North Holland, Amsterdam, 1970.

[18] J.M. Borwein, A.J. Guirao, P. Hájek, and J. Vanderwerff. Uniformly convex functions on Banach spaces. *Proceedings of the American Mathematical Society*, 137(3):1081–1091, 2009.

[19] J.M. Borwein and J. Vanderwerff. Fréchet-Legendre functions and reflexive Banach spaces. *Journal of Convex Analysis*, 17(3–4):915–924, 2010.

[20] R.I. Boţ, E.R. Csetnek, and D. Meier. Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces. *Optimization Methods and Software*, 34:489–514, 2019.

[21] R.I. Boţ, S.-M. Grad, and G. Wanka, editors. *Duality in Vector Optimization.* Vector Optimization. Springer Berlin, Heidelberg, 2009.

[22] L.M. Bregman. The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[23] H. Brezis. On a problem of T. Kato. *Communications on Pure and Applied Mathematics*, 24:1–6, 1971.

[24] H. Brezis. Equations d'evolution du second ordre associees a des operateurs monotones. *Israel Journal of Mathematics*, 12:51–60, 1972.

[25] H. Brezis and A. Pazy. Accretive sets and differential equations in Banach spaces. *Israel Journal of Mathematics*, 8:367–383, 1970.

[26] A. Brøndsted and R.T. Rockafellar. On the subdifferentiability of convex functions. *Proceedings of the American Mathematical Society*, 16(4):605–611, 1965.

[27] F.E. Browder. Non-Linear Equations of Evolution. *Annals of Mathematics*, 80(3):485–523, 1964.

[28] F.E. Browder. Nonlinear monotone operators and convex sets in Banach spaces. *Bulletin of the American Mathematical Society*, 71(5):780–785, 1965.

[29] F.E. Browder. Nonlinear accretive operators in Banach spaces. *Bulletin of the American Mathematical Society*, 73:470–476, 1967.

[30] F.E. Browder. Nonlinear maximal monotone operators in Banach space. *Mathematische Annalen*, 175:89–113, 1968.

[31] R.E. Bruck and S. Reich. Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston Journal of Mathematics*, 3(4):459–470, 1977.

[32] M.N. Búi and P.L. Combettes. Bregman forward-backward operator splitting. *Set-Valued and Variational Analysis*, 29:583–603, 2021.

[33] R.S. Burachik, M.N. Dao, and S.B. Lindstrom. The generalized Bregman distance. *SIAM Journal on Optimization*, 31(1):404–424, 2021.

[34] D. Butnariu and A.N. Iusem. *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, volume 40 of *Applied Optimization*. Springer Dordrecht, 2000.

[35] D. Butnariu, A.N. Iusem, and C. Zălinescu. On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithm in Banach spaces. *Journal of Convex Analysis*, 10(1):35–62, 2003.

[36] D. Butnariu and E. Resmerita. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis*, pages 1–39, 2006. Art. ID 84919.

[37] B. Calvert. Maximal accretive is not m-accretive. *Bollettino dell'Unione Matematica Italiana*, 3:1042–1044, 1970.

[38] Y. Censor and T. Elfving. A multiprojection algorithm using Bregman projections in a product space. *Numerical Algorithms*, 8:221–239, 1994.

[39] Y. Censor and S. Reich. Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization. *Optimization*, 37:323–339, 1996.

[40] Y. Censor and S. Reich. The Dykstra algorithm with Bregman projections. *Communications in Applied Analysis*, 2(3):407–420, 1998.

[41] H. Cheval, U. Kohlenbach, and L. Leuştean. On modified Halpern and Tikhonov-Mann iterations. *Journal of Optimization Theory and Applications*, 197:233–251, 2023.

[42] H. Cheval and L. Leuştean. Quadratic rates of asymptotic regularity for the Tikhonov–Mann iteration. *Optimization Methods and Software*, 37:2225–2240, 2022.

[43] C. Chidume. *Geometric Properties of Banach Spaces and Nonlinear Iterations*, volume 1965 of *Lecture Notes in Mathematics*. Springer-Verlag London, 2009.

[44] C.E. Chidume and C.H. Morales. Accretive operators which are always single-valued in normed spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 67(12):3328–3334, 2007.

[45] J.A. Clarkson. Uniformly Convex Spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.

[46] P.L. Combettes. Quasi-Fejérian Analysis of Some Optimization Algorithms. *Studies in Computational Mathematics*, 8:115–152, 2001.

[47] P.L. Combettes. Fejér monotonicity in convex optimization. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1016–1024. Springer, New York, 2009.

[48] M.G. Crandall. A generalized domain for semigroup generators. *Proceedings of the American Mathematical Society*, 37:434–440, 1973.

[49] M.G. Crandall and T.M. Liggett. A Theorem and a Counterexample in the Theory of Semigroups of Nonlinear Transformations. *Transactions of the American Mathematical Society*, 160:263–278, 1971.

[50] M.G. Crandall and T.M. Liggett. Generation of semigroups of nonlinear transformations on general Banach spaces. *American Journal of Mathematics*, 93:265–298, 1971.

[51] M.G. Crandall and A. Pazy. Semi-groups of nonlinear contractions and dissipative sets. *Journal of Functional Analysis*, 3(3):376–418, 1969.

[52] M.G. Crandall and A. Pazy. Nonlinear evolution equations in Banach spaces. *Israel Journal of Mathematics*, 11:57–94, 1972.

[53] A. Cuntavenapit and B. Panyanak. Strong convergence of modified Halpern iterations in CAT(0) spaces. *Journal of Fixed Point Theory and Applications*, 16:Article no. 869458, 2011.

[54] J. Diestel. *Geometry of Banach Spaces - Selected Topics*, volume 485 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg, 1975.

[55] B. Dinis and P. Pinto. Metastability of the multi-parameters proximal point algorithm. *Portugaliae Mathematica*, 77(3):345–381, 2020.

[56] B. Dinis and P. Pinto. Quantitative results on the multi-parameters proximal point algorithm. *Journal of Convex Analysis*, 28(3), 2021. 23 pp.

[57] B. Dinis and P. Pinto. Strong convergence for the alternating Halpern-Mann iteration in CAT(0) spaces. *SIAM Journal on Optimization*, 33(2):785–815, 2023.

[58] J. Eckstein. Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.

[59] P. Engrácia. *Proof-theoretic studies on the bounded functional interpretation*. PhD thesis, Universidade de Lisboa, 2009.

[60] S. Feferman. Kreisel's "Unwinding" Program. In P. Odifreddi, editor, *Kreiseliana: About and Around Georg Kreisel*, pages 247–274. Wellesley, Mass.: A K Peters, 1996.

[61] W. Fenchel. On conjugate convex functions. *Canadian Journal of Mathematics*, 1:73–77, 1949.

[62] F. Ferreira. Injecting uniformities into Peano arithmetic. *Annals of Pure and Applied Logic*, 157(2-3):122–129, 2009.

[63] F. Ferreira and P. Engrácia. Bounded functional interpretation with an abstract type. In A. Rezuş, editor, *Contemporary Logic and Computing*, volume 1 of *Landscapes in Logic*, pages 87–112. College Publications, 2020.

[64] F. Ferreira, L. Leuştean, and P. Pinto. On the removal of weak compactness arguments in proof mining. *Advances in Mathematics*, 354, 2019. 106728.

[65] F. Ferreira and P. Oliva. Bounded functional interpretation. *Annals of Pure and Applied Logic*, 135:73–112, 2005.

[66] J. García-Falset. The asymptotic behavior of the solutions of the Cauchy problem generated by φ-accretive operators. *Journal of Mathematical Analysis and Applications*, 310:594–608, 2005.

[67] J. García-Falset, E. Llorens-Fuster, and T. Suzuki. Fixed point theory for a class of generalized nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 375:185–195, 2011.

[68] J. García-Falset and C.H. Morales. Existence theorems for m-accretive operators in Banach spaces. *Journal of Mathematical Analysis and Applications*, 309:453–461, 2005.

[69] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49:431–446, 2008.

[70] P. Gerhardy and U. Kohlenbach. Strongly uniform bounds from semi-constructive proofs. *Annals of Pure and Applied Logic*, 141:89–107, 2006.

[71] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360:2615–2660, 2008.

[72] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.

[73] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines Mathematischen Kolloquiums*, 4:34–38, 1933.

[74] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958.

[75] A. Grzegorczyk. *Some classes of recursive functions*. Rozprawny Matematyczne. Warsaw, 1953.

[76] D. Günzel and U. Kohlenbach. Logical metatheorems for abstract spaces axiomatized in positive bounded logic. *Advances in Mathematics*, 290:503–551, 2016.

[77] B. Halpern. Fixed points of nonexpanding maps. *Bulletin of the American Mathematical Society*, 73:957–961, 1967.

[78] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95(1):161–190, 1926.

[79] W.A. Howard. Hereditarily majorizable functionals of finite type. In A.S. Troelstra, editor, *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*, pages 454–461. Springer, New York, 1973.

[80] W.A. Howard. The formulae-as-types notion of construction. In J.R. Hindley and J.P. Seldin, editors, *To H.B. Curry*, pages 479–490. Academic, London, 1980.

[81] R.C. James. Reflexivity and the Supremum of Linear Functionals. *Annals of Mathematics*, 66(1):159–169, 1957.

[82] S.B. Nadler Jr. Multi-valued contraction mappings. *Pacific Journal of Mathematics*, 30:475–487, 1969.

[83] S. Kamimura, F. Kohsaka, and W. Takahashi. Weak and strong convergence theorems for maximal monotone operators in a Banach space. *Set-Valued Analysis*, 12:417–429, 2004.

[84] T. Kato. Nonlinear semigroups and evolution equations. *Journal of the Mathematical Society of Japan*, 19:508–520, 1967.

[85] T.-H. Kim and H.-K. Xu. Strong convergence of modified Mann iterations. *Nonlinear Analysis*, 61:51–60, 2005.

[86] U. Kohlenbach. *Theorie der majorisierbaren und stetigen Funktionale und ihre Anwendung bei der Extraktion von Schranken aus inkonstruktiven Beweisen: Effektive Eindeutigkeitsmodule bei besten Approximationen aus ineffektiven Eindeutigkeitsbeweisen.* PhD thesis, Goethe-Universität Frankfurt am Main, 1990.

[87] U. Kohlenbach. Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization. *The Journal of Symbolic Logic*, 57:1239–1273, 1992.

[88] U. Kohlenbach. Pointwise hereditary majorization and some applications. *Archive for Mathematical Logic*, 31:227–241, 1992.

[89] U. Kohlenbach. Effective moduli from ineffective uniqueness proofs. An unwinding of de La Vallee Poussin's proof for Chebycheff approximation. *Annals of Pure and Applied Logic*, 64:27–94, 1993.

[90] U. Kohlenbach. New effective moduli of uniqueness and uniform a-priori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory. *Numerical Functional Analysis and Optimization*, 14:581–606, 1993.

[91] U. Kohlenbach. Analysing proofs in analysis. In W. Hodges, M. Hyland, C. Steinhorn, and J. Truss, editors, *Logic: From Foundations to Applications. European Logic Colloquium*, pages 225–260. Oxford University Press, Oxford, 1996.

[92] U. Kohlenbach. Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. *Archive of Mathematical Logic*, 36:31–71, 1996.

[93] U. Kohlenbach. Arithmetizing proofs in analysis. In J.M. Larrazabal, D. Lascar, and G. Mints, editors, *Logic Colloquium '96*, volume 12 of *Lecture Notes in Logic*, pages 115–158. Springer, 1998.

[94] U. Kohlenbach. Relative constructivity. *The Journal of Symbolic Logic*, 63:1218–1238, 1998.

[95] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.

[96] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.

[97] U. Kohlenbach. Gödel's functional interpretation and its use in current mathematics. In M. Baaz, C.H. Papadimitriou, H.W. Putnam, D.S. Scott, and C.L. Harper Jr, editors, *Kurt Gödel and the Foundations of Mathematics. Horizons of Truth*, pages 361–398. Cambridge University Press, New York, 2011.

[98] U. Kohlenbach. On quantitative versions of theorems due to F.E. Browder and R. Wittmann. *Advances in Mathematics*, 226:2764–2795, 2011.

[99] U. Kohlenbach. On the quantitative asymptotic behavior of strongly nonexpansive mappings in Banach and geodesic spaces. *Israel Journal of Mathematics*, 216:215–246, 2016.

[100] U. Kohlenbach. Recent progress in proof mining in nonlinear analysis. *IFCoLog Journal of Logics and their Applications*, 10(4):3361–3410, 2017.

[101] U. Kohlenbach. A polynomial rate of asymptotic regularity for compositions of projections in Hilbert space. *Foundations of Computational Mathematics*, 19:83–99, 2019.

[102] U. Kohlenbach. Proof-theoretic Methods in Nonlinear Analysis. In B. Sirakov, P. Ney de Souza, and M. Viana, editors, *Proc. ICM 2018*, volume 2, pages 61–82. World Scientific, 2019.

[103] U. Kohlenbach. Local formalizations in nonlinear analysis and related areas and proof-theoretic tameness. In P. Weingartner and H.-P. Leeb, editors, *Kreisel's Interests. On the Foundations of Logic and Mathematics*, volume 41 of *Tributes*, pages 45–61. College Publications, 2020.

[104] U. Kohlenbach. Quantitative analysis of a Halpern-type Proximal Point Algorithm for accretive operators in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 21(9):2125–2138, 2020.

[105] U. Kohlenbach. Quantitative results on the Proximal Point Algorithm in uniformly convex Banach spaces. *Journal of Convex Analysis*, 28(1):11–18, 2021.

[106] U. Kohlenbach. On the Proximal Point Algorithm and its Halpern-type variant for generalized monotone operators in Hilbert space. *Optimization Letters*, 16:611–621, 2022.

[107] U. Kohlenbach. Kreisel's 'shift of emphasis' and contemporary proof mining. *Studies in Logic*, 16(3):1–35, 2023.

[108] U. Kohlenbach and A. Koutsoukou-Argyraki. Rates of convergence and metastability for abstract Cauchy problems generated by accretive operators. *Journal of Mathematical Analysis and Applications*, 423:1089–1112, 2015.

[109] U. Kohlenbach and A. Koutsoukou-Argyraki. Effective asymptotic regularity for one-parameter nonexpansive semigroups. *Journal of Mathematical Analysis and Applications*, 433(2):1883–1903, 2016.

[110] U. Kohlenbach and L. Leuştean. Effective metastability of Halpern iterates in CAT(0) spaces. *Advances in Mathematics*, 231:2526–2556, 2012.

[111] U. Kohlenbach and L. Leuştean. On the computational content of convergence proofs via Banach limits. *Philosophical Transactions of the Royal Society A*, 370:3449–3463, 2012.

[112] U. Kohlenbach, L. Leuştean, and A. Nicolae. Quantitative Results on Fejér Monotone Sequences. *Communications in Contemporary Mathematics*, 20(2):1750015, 42pp., 2018.

[113] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Quantitative Asymptotic Regularity Results for the Composition of Two Mappings. *Optimization*, 66:1291–1299, 2017.

[114] U. Kohlenbach, G. López-Acedo, and A. Nicolae. Moduli of regularity and rates of convergence for Fejér monotone sequenc es. *Israel Journal of Mathematics*, 232:261 – 297, 2019.

[115] U. Kohlenbach and A. Nicolae. A proof-theoretic bound extraction theorem for CAT(κ)-spaces. *Studia Logica*, 105:611–624, 2017.

[116] U. Kohlenbach and P. Oliva. Proof Mining: A Systematic Way of Analysing Proofs in Mathematics. *Proceedings of the Steklov Institute of Mathematics*, 242:136–164, 2003.

[117] U. Kohlenbach and P. Oliva. Proof mining in L_1-approximation. *Annals of Pure and Applied Logic*, 121:1–38, 2003.

[118] U. Kohlenbach and P. Pinto. Quantitative translations for viscosity approximation methods in hyperbolic spaces. *Journal of Mathematical Analysis and Applications*, 507:125823, 33 pages, 2022.

[119] U. Kohlenbach and N. Pischke. Proof Theory and Nonsmooth Analysis. *Philosophical Transactions of the Royal Society A*, 381(2248), 2023. 20220015, 21pp.

[120] U. Kohlenbach and T. Powell. Rates of convergence for iterative solutions of equations involving set-valued accretive operators. *Computers & Mathematics with Applications*, 80:490–503, 2020.

[121] F. Kohsaka and W. Takahashi. Strong convergence of an iterative sequence for maximal monotone operators in a Banach space. *Abstract and Applied Analysis*, 3:239–249, 2004.

[122] F. Kohsaka and W. Takahashi. Proximal point algorithms with Bregman functions in Banach spaces. *Journal of Nonlinear and Convex Analysis*, 6(3):505–523, 2005.

[123] Y. Komura. Nonlinear semi-groups in Hilbert space. *Journal of the Mathematical Society of Japan*, 19(4):493–507, 1967.

[124] D. Körnlein. Quantitative results for Halpern iterations of nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 428:1161–1172, 2015.

[125] G. Kreisel. On the Interpretation of Non-Finitist Proofs–Part I. *The Journal of Symbolic Logic*, 16(4):241–267, 1951.

[126] G. Kreisel. On the Interpretation of Non-Finitist Proofs–Part II. Interpretation of Number Theory. Applications. *The Journal of Symbolic Logic*, 17(1):43–58, 1952.

[127] G. Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting, editor, *Constructivity in Mathematics*, Lecture Notes in Mathematics, pages 101–128. North-Holland Publishing Company, Amsterdam, 1959.

[128] G. Kreisel. On weak completeness of intuitionistic predicate logic. *The Journal of Symbolic Logic*, 27:139–158, 1962.

[129] G. Kreisel. What have we learnt from Hilbert's second problem? In F.E. Browder, editor, *Mathematical developments arising from Hilbert problems*, volume 28 of *Proceedings of Symposia in Pure Mathematics*, pages 93–130. American Mathematical Society, 1976.

[130] S. Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Mathematical Journal*, 2:35–47, 1951.

[131] G. Leoni. *A First Course in Sobolev Spaces*, volume 105 of *Grad. Stud. Math.* Amer. Math. Soc., Providence, RI, 2009.

[132] L. Leuştean. Proof Mining in $\mathbb{R}$-trees and Hyperbolic Spaces. *Electronic Notes in Theoretical Computer Science*, 165:95–106, 2006.

[133] L. Leuştean. An application of proof mining to nonlinear iterations. *Annals of Pure and Applied Logic*, 165(9):1484–1500, 2014.

[134] L. Leuştean and P. Pinto. Quantitative results on the Halpern type proximal point algorithm. *Computational Optimization and Applications*, 79(1):101–125, 2021.

[135] L. Leuştean and A. Sipoş. An application of proof mining to the proximal point algorithm in CAT(0) spaces. In A. Bellow, C. Calude, and T. Zamfirescu, editors, *Mathematics Almost Everywhere. In Memory of Solomon Marcus*, pages 153–168. World Scientific, 2018.

[136] L. Leuştean and A. Sipoş. Effective strong convergence of the proximal point algorithm in CAT(0) spaces. *Journal of Nonlinear and Variational Analysis*, 2(2):219–228, 2018.

[137] H. Luckhardt. Herbrand-Analysen Zweier Beweise Des Satzes von Roth: Polynomiale Anzahlschranken. *Journal of Symbolic Logic*, 54(1):234–263, 1989.

[138] H. Luckhardt. Bounds Extracted by Kreisel from Ineffective Proofs. In P. Odifreddi, editor, *Kreiseliana: About and Around Georg Kreisel*, pages 289–300. Wellesley, Mass.: A K Peters, 1996.

[139] A. Macintyre. The mathematical significance of proof theory. *Philosophical Transactions of the Royal Society A*, 363:2419–2435, 2005.

[140] A. Macintyre. The impact of Gödel's incompleteness theorems on mathematics. In M. Baaz, C.H. Papadimitriou, H.W. Putnam, D.S. Scott, and C.L. Harper, editors, *Kurt Gödel and the foundations of mathematics: Horizons of truth*, pages 3–25. Cambridge University Press, 2011.

[141] P.-E. Maingé. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set-Valued Analysis*, 16:899–912, 2008.

[142] V. Martín-Márquez, S. Reich, and S. Sabach. Right Bregman nonexpansive operators in Banach spaces. *Nonlinear Analysis*, 75:5448–5465, 2012.

[143] V. Martín-Márquez, S. Reich, and S. Sabach. Bregman strongly nonexpansive operators in reflexive Banach spaces. *Journal of Mathematical Analysis and Applications*, 400(2):597–614, 2013.

[144] V. Martín-Márquez, S. Reich, and S. Sabach. Iterative methods for approximating fixed points of Bregman nonexpansive operators. *Discrete and Continuous Dynamical Systems-Series S*, 6(4):1043–1063, 2013.

[145] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Revue française d'informatique et de recherche opérationnelle*, 4:154–158, 1970.

[146] G.J. Minty. Monotone networks. *Proceedings of the Royal Society A*, 257:194–212, 1960.

[147] G.J. Minty. Monotone (nonlinear) operators in Hilbert spaces. *Duke Mathematical Journal*, 29:341–346, 1962.

[148] I. Miyadera. Some remarks on semigroups of nonlinear operators. *Tohoku Mathematical Journal*, 23:245–258, 1971.

[149] I. Miyadera. *Nonlinear Semigroups*. Translations of Mathematical Monographs. AMS, Providence, 1992.

[150] I. Miyadera and K. Kobayasi. On the asymptotic behaviour of almost-orbits of nonlinear contraction semigroups in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 6(4):349–365, 1982.

[151] C. Moore and B.V.C. Nnoli. Iterative solution of nonlinear equations involving set-valued uniformly accretive operators. *Computers and Mathematics with Applications*, 42(1–2):131–140, 2001.

[152] E. Naraghirad. Compositions and convex combinations of Bregman weakly relatively nonexpansive operators in reflexive Banach spaces. *Journal of Fixed Point Theory and Applications*, 22, 2020. Article no. 65.

[153] E. Neumann. Computational Problems in Metric Fixed Point Theory and their Weihrauch Degrees. *Logical Methods in Computer Science*, 11(4), 2015.

[154] O. Nevanlinna and S. Reich. Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces. *Israel Journal of Mathematics*, 32:44–58, 1979.

[155] Q. Van Nguyen. Forward-backward splitting with Bregman distances. *Vietnam Journal of Mathematics*, 45(3):519–539, 2017.

[156] P. Odifreddi, editor. *Kreiseliana: About and Around Georg Kreisel*. Wellesley, Mass.: A K Peters, 1996.

[157] L. Păunescu and A. Sipoş. A proof-theoretic metatheorem for tracial von Neumann algebras. *Mathematical Logic Quarterly*, 69(1):63–76, 2023.

[158] N.H. Pavel. *Nonlinear Evolution Operators and Semigroups: Applications to Partial Differential Equations*. Lecture Notes in Mathematics. Springer Berlin, Heidelberg, 1987.

[159] A. Pazy. Asymptotic behavior of contractions in Hilbert space. *Israel Journal of Mathematics*, 9(2):235–240, 1971.

[160] A. Pazy. Strong convergence of semigroups on nonlinear contractions in Hilbert space. *Journal of Mathematical Analysis and Applications*, 34:1–35, 1978.

[161] P. Pinto. A rate of metastability for the Halpern type Proximal Point Algorithm. *Numerical Functional Analysis and Optimization*, 42(3):320–343, 2021.

[162] P. Pinto and N. Pischke. On computational properties of Cauchy problems generated by accretive operators. *Documenta Mathematica*, 28(5):1235–1274, 2023.

[163] N. Pischke. A proof-theoretic metatheorem for nonlinear semigroups generated by an accretive operator and applications. *ArXiv e-prints*, 2023. arXiv, math.LO, 2304.01723.

[164] N. Pischke. Generalized Fejér monotone sequences and their finitary content. *ArXiv e-prints*, 2023. arXiv, math.FA, 2312.01852.

[165] N. Pischke. Logical Metatheorems for Accretive and (Generalized) Monotone Set-Valued Operators. *Journal of Mathematical Logic*, 2023. To appear, 59pp.

[166] N. Pischke. Quantitative Results on Algorithms for Zeros of Differences of Monotone Operators in Hilbert Space. *Journal of Convex Analysis*, 30(1):295–315, 2023.

[167] N. Pischke. Rates of convergence for the asymptotic behavior of second-order Cauchy problems. *Journal of Mathematical Analysis and Applications*, 533(2), 2024. 128078, 15pp.

[168] N. Pischke. Logical metatheorems for set-valued operators and their use in the analysis of Moudafis algorithm for the difference of two monotone operators in Hilbert space. Master's thesis, TU Darmstadt, January 2022. 127pp.

[169] N. Pischke and U. Kohlenbach. Quantitative Analysis of a Subgradient-Type Method for Equilibrium Problems. *Numerical Algorithms*, 90:197–219, 2022.

[170] A.T. Plant. The differentiability of nonlinear semigroups in uniformly convex spaces. *Israel Journal of Mathematics*, 38(3):257–268, 1981.

[171] E.I. Poffald and S. Reich. An incomplete Cauchy problem. *Journal of Mathematical Analysis and Applications*, 113(2):514–543, 1986.

[172] S. Reich. Product formulas, nonlinear semigroups, and accretive operators. *Journal of Functional Analysis*, 36(2):147–168, 1980.

[173] S. Reich. Strong Convergence Theorems for Resolvents of Accretive Operators in Banach Spaces. *Journal of Mathematical Analysis and Applications*, 75:287–292, 1980.

[174] S. Reich. On the Asymptotic Behavior of Nonlinear Semigroups and the Range of Accretive Operators. *Journal of Mathematical Analysis and Applications*, 79:113–126, 1981.

[175] S. Reich. A weak convergence theorem for the alternating method with Bregman distances. In A.G. Kartsatos, editor, *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, pages 313–318. Marcel Dekker, New York, 1996.

[176] S. Reich and S. Sabach. A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces. *Journal of Nonlinear and Convex Analysis*, 10:471–485, 2009.

[177] S. Reich and S. Sabach. Two strong convergence theorems for a proximal method in reflexive Banach spaces. *Numerical Functional Analysis and Optimization*, 31(1):22–44, 2010.

[178] S. Reich and S. Sabach. Existence and approximation of fixed points of Bregman firmly nonexpansive mappings in reflexive Banach spaces. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 301–316. Springer, New York, 2011.

[179] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *Journal of Convex Analysis*, 11:1–16, 2004.

[180] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17:497–510, 1966.

[181] R.T. Rockafellar. Level sets and continuity of conjugate convex functions. *Transactions of the American Mathematical Society*, 123(1):46–63, 1966.

[182] R.T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific Journal of Mathematics*, 33(1):209–216, 1970.

[183] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[184] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis.* Springer, New York, 1998.

[185] W. Rudin. *Functional Analysis*, volume 8 of *International Series in Pure and Applied Mathematics.* McGraw-Hill, New York, 1991.

[186] K. Schade and U. Kohlenbach. Effective metastability for modified Halpern iterations in CAT (0) spaces. *Fixed Point Theory and Applications*, 1:1–19, 2012.

[187] H.H. Schaefer and M.P. Wolff. *Topological Vector Spaces*, volume 8 of *Graduate Texts in Mathematics.* Springer, New York, 1999.

[188] M. Schönfinkel. Über die Bausteine der mathematischen Logik. *Mathematische Annalen*, 92:305–316, 1924.

[189] A. Schönhage. *Approximationstheorie.* De Gruyter Berlin, New York, 1971.

[190] H.F. Senter and W.G. Dotson. Approximating fixed points of nonexpansive mappings. *Proceedings of the American Mathematical Society*, 44(2):375–380, 1974.

[191] Y. Shehu. Convergence Results of Forward-Backward Algorithms for Sum of Monotone Operators in Banach Spaces. *Results in Mathematics*, 74:Article number 138, 24pp., 2019.

[192] A. Sipoş. Proof mining in L^p spaces. *The Journal of Symbolic Logic*, 84(4):1612–1629, 2019.

[193] A. Sipoş. Bounds on strong unicity for Chebyshev approximation with bounded coefficients. *Mathematische Nachrichten*, 294(12):2425–2440, 2021.

[194] A. Sipoş. Abstract strongly convergent variants of the proximal point algorithm. *Computational Optimization and Applications*, 83:349–380, 2022.

[195] C. Smorynski. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, volume 90 of *Studies in Logic and the Foundations of Mathematics*, pages 821–865. Elsevier, 1977.

[196] Y. Song and H. Wang. Convergence of iterative algorithms for multivalued mappings in Banach spaces. *Nonlinear Analysis: Theory, Methods & Applications*, 70(4):1547–1556, 2009.

[197] E. Specker. Nicht konstruktiv beweisbare Sätze der Analysis. *Journal of Symbolic Logic*, 14:145–158, 1949.

[198] C. Spector. Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics. In J.C.E. Dekker, editor, *Recursive Function Theory*, volume 5 of *Proceedings of Symposia in Pure Mathematics*, pages 1–27. American Mathematical Society, 1962.

[199] S. Suantai, Y.J. Cho, and P. Cholamjiak. Halpern's iteration for Bregman strongly nonexpansive mappings in reflexive Banach spaces. *Computers & Mathematics with Applications*, 64(4):489–499, 2012.

[200] T. Suzuki. Strong convergence theorems for infinite families of nonexpansive mappings in general Banach spaces. *Journal of Fixed Point Theory and Applications*, 2005(1):103–123, 2005.

[201] T. Suzuki. Fixed point theorems and convergence theorems for some generalized nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 340:1088–1095, 2008.

[202] W. Takahashi. *Nonlinear Functional Analysis*. Yokohama Publishers, Yokohama, Japan, 2000.

[203] T. Tao. Norm Convergence of Multiple Ergodic Averages for Commuting Transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.

[204] T. Tao. *Structure and Randomness: Pages from Year One of a Mathematical Blog*, chapter Soft analysis, hard analysis, and the finite convergence principle. American Mathematical Society, Providence, RI, 2008.

[205] A.S. Troelstra, editor. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer, New York, 1973.

[206] A.A. Vladimirov, Y.E. Nesterov, and Y.N. Cěkanov. Uniformly convex functionals. *Vestnik Moskovskogo Universiteta. Seriya XV. Vychislitel'naya Matematika i Kibernetika*, 3:12–23, 1978.

[207] R. Wittmann. Approximation of fixed points of nonexpansive mappings. *Archiv der Mathematik*, 58:486–491, 1992.

[208] H.-K. Xu. Strong asymptotic behavior of almost-orbits of nonlinear semigroups. *Nonlinear Analysis: Theory, Methods & Applications*, 46(1):135–151, 2001.

[209] H.-K. Xu. An iterative approach to quadratic optimization. *Journal of Optimization Theory and Applications*, 116:659–678, 2003.

[210] Y. Yao, H. Zhou, and Y.-C. Liou. Strong convergence of a modified Krasnoselski–Mann iterative algorithm for non-expansive mappings. *Journal of Applied Mathematics and Computing*, 29:383–389, 2009.

[211] C. Zălinescu. On uniformly convex functions. *Journal of Mathematical Analysis and Applications*, 95(2):344–374, 1983.

[212] C. Zălinescu. *Convex analysis in general vector spaces*. World scientific, 2002.

[213] H. Zegeye. Strong convergence theorems for maximal monotone mappings in Banach spaces. *Journal of Mathematical Analysis and Applications*, 343(2):663–671, 2008.

[214] J.H. Zhu and S.S. Chang. Halpern-Mann's iterations for Bregman strongly non-expansive mappings in reflexive Banach spaces with applications. *Journal of Inequalities and Applications*, page Article no. 146, 2013.

List of Theories

List of Axioms and Rules

(I) – (IV), Axioms for accretive and monotone operators with total resolvents. p. 25.

(II)$'$, (V), (Alternative) axioms for accretive and monotone operators with partial resolvent. p. 29 and p. 29.

(J), (A), Axioms for the dual formulation of accretivity. p. 55.

$(+)_1$ – $(+)_5$, Axioms for $\langle \cdot, \cdot \rangle_s$. p. 58.

$(RC)_{\lambda_0}$, Axiom for the range condition up to λ_0 using the closure of the domain. p. 63.

(RC), Axiom for the full range condition using the closure of the domain. p. 63.

(IV)$'$, (V)$'$, Alternative axioms for the majorization of the resolvent in the context of the dual formulation of accretivity. p. 64.

(VI), Axiom for the resolvent at zero. p. 65.

(S1), (S2), Axioms for the semigroup generated by an accretive operator. p. 66 and p. 68.

(NE), Axiom for the non-emptyness of the zero set of an operator. p. 105.

(P1), Axiom for a selection of the projection. p. 105.

(P2), Axiom for the uniform continuity of a selection of the projection. p. 106.

(j'), (M1), (M2), Axioms for the special selection j' of the duality map for the uniform formulation of accretivity. p. 107.

(L1), Axiom for the approximation of $|A \cdot|$ by the Yosida approximates. p. 151.

(L2), Axiom for the uniform continuity of $|A \cdot|$. p. 152.

(d), Axioms for the value $d = \inf\{\|y\| \mid y \in Ax \text{ for some } x\}$. p. 154.

$(*)_1$, $(*)_2$, Axioms for the dual norm. p. 165.

$(*)_3$, $(*)_4$, Axioms for the application functional of the dual space. p. 166.

$(*)_5$, The vector space axioms for the dual space. p. 166.

(QF-LR), The quantifier-free linearity rule. p. 166.

$(*)_6$, Axiom for the non-emptyness of the normalized duality map. p. 167.

$(**)_1 - (**)_6$ and (QF-LR**), Axioms and rules for the bidual. p. 176.

(R), Axiom for the reflexivity of a normed space. p. 177.

(JT), Axiom for the alternative characterization of reflexivity as in James' theorem. p. 178.

$(f)_1$, $(\nabla f)_1$, $(\nabla f)_2$, Axioms for a convex function f and its uniformly continuous Fréchet derivative. p. 182.

$(f)_2$, Axiom for the supercoercivity of f. p. 188.

$(f^*)_1$, $(f^*)_2$, $(\nabla f^*)_1$, $(\nabla f^*)_2$, Axioms for the conjugate of a convex function f and its uniformly continuous Fréchet derivative. p. 188 and p. 189.

(L), Additional axioms to formulate the property of Legendre functions that the gradients ∇f and ∇f^* are inverses of each other. p. 190.

$(f^*)_3$, $(f^{**})_1$, $(f^{**})_2$, Axioms for hardwiring the Fenchel-Moreau theorem, i.e. that $f = f^{**}$. p. 277.

$(\mathrm{I})^* - (\mathrm{IV})^*$, Axioms for monotone operators in Banach spaces and their relativized resolvents. p. 285 and p. 286.

Wissenschaftlicher Werdegang

1999	Geboren in Berlin am 5. Februar
2017	Abitur an der Johann-Philipp-Reis-Schule in Friedberg, Hessen
2017 - 2021	Technische Universität Darmstadt,
	BSc. Mathematik mit Nebenfach Informatik
2021 - 2022	Technische Universität Darmstadt,
	MSc. Mathematik mit Nebenfach Informatik
Seit Oktober 2022	Technische Universität Darmstadt,
	Wissenschaftlicher Mitarbeiter am Fachbereich Mathematik