PROOF MINING FOR THE DUAL OF A BANACH SPACE WITH EXTENSIONS FOR UNIFORMLY FRÉCHET DIFFERENTIABLE FUNCTIONS

NICHOLAS PISCHKE

Department of Mathematics, Technische Universität Darmstadt,
Schlossgartenstraße 7, 64289 Darmstadt, Germany,

E-mail: pischke@mathematik.tu-darmstadt.de

Abstract. We present a proof-theoretically tame approach for treating the dual space of an abstract Banach space in systems amenable to proof mining metatheorems which quantify and allow for the extraction of the computational content of large classes of theorems about the dual of a Banach space and its corresponding norm, unlocking a major branch of functional analysis as a new area of applications for these methods. The approach relies on using intensional methods to deal with the high quantifier complexity of the predicate defining the dual space as well as on a proof-theoretically tame treatment of suprema over (certain) bounded sets in normed spaces to deal with the norm of the dual. Beyond this, we discuss further possible extensions of this system to deal with convex functions and corresponding Fréchet derivatives and their duality theory through Fenchel conjugates, together with the associated Bregman distances, which provide the logical basis for a range of recent applications of proof mining methods to these branches of nonlinear analysis.

Keywords: Proof mining; Metatheorems; Banach spaces; Dual spaces; Fréchet derivatives; Fenchel conjugates.
MSC2020 Classification: 46G05; 47H04; 46B10; 03F10; 03F35

1. Introduction

One of the fundamental forces driving many developments in proof theory since its earliest days has been the question for what the computational content of a given mathematical theorem is. *Proof mining*[1] emerged as a new program in mathematical logic in the later 1990's through the works of U. Kohlenbach and his collaborators (going back conceptually to Kreisel's *unwinding of proofs* [45, 46]) that aims to extract said computational information from classical proofs as found in the mainstream literature. Since such proofs are, as common in ordinary mathematical practice, prima facie noneffective, this is a nontrivial task.

The aim of this paper is 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 [25, 36], 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 [10]) as well as the theory of von Neumann algebras.[2]

The fundamental logical 'substrate' of this discipline of proof mining are the aforementioned so-called *logical metatheorems* on bound extractions.[3] These use well-known proof interpretations like negative translations

Date: August 12, 2024.

[1]The development of proof mining is detailed comprehensively up to 2008 in the monograph [37] (see also [42] for a survey of the early stages of proof mining) and recent progress, with a focus on nonlinear analysis and optimization, is surveyed in [38].

[2]For the latter, an approach for extending proof mining methods to the context of tracial von Neumann algebras has recently been given in [55].

[3]Examples of such metatheorems may be found in [25, 27, 36, 37, 41, 49, 50, 55, 56, 58, 69], for the metatheorems obtained via (modifications of) Gödel's Dialectica interpretation, and [23] for subsequent metatheorems obtained via the bounded functional interpretation [24] due to F. Ferreira and P. Oliva.

(see e.g. [48]), Kreisel's *modified realizability* [47], Gödel's *functional (Dialectica) interpretation* [26] and their variants to provide general results that quantify and allow for the extraction of the computational content from large classes of theorems and proofs from the core mathematical literature (e.g. potentially involving a wide range of non-computational 'ideal principles' and using classical logic). As such, proof mining as substantiated by these metatheorems has lead to hundreds of new results in the respective areas of applications over the last three decades.

A crucial innovation introduced in [36] and which has since been adopted in most approaches to the logical foundations of proof mining is the incorporation of so-called abstract types. While the first logical metatheorems in proof mining relied on pure systems of arithmetic in all finite types and consequently only covered applications involving Polish metric spaces (as those can be represented in the underlying language), these additional abstract types allow for the treatment of spaces which are not separable and thus not representable in the (bare) language of finite type arithmetic.[4] In this way, for example general metric and normed spaces, so-called W-hyperbolic spaces and CAT(0)-spaces have been successfully studied in proof mining, among many others.

Now, the main results of this paper 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[5] $1(X)$ 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 [58] as will be discussed later). This intensional treatment of the dual then allows us to utilize a proof-theoretically tame[6] approach for treating suprema over (certain) bounded sets in abstract spaces, developed in the first part of this paper, to provide defining axioms for the norm of the dual.

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

[4]Whether a class of spaces or objects can be treated via this approach to such metatheorems ultimately depends on the complexity and uniformity of the defining axioms.

[5]Here, and throughout this paper for that matter, we follow the notation from [37] and consequently write 1 for the type of number-theoretic functions. The type $1(X)$ then signifies the type of functions $X \to \mathbb{R}$ by following a standard coding of real numbers as functions $\mathbb{N} \to \mathbb{N}$. We refer to Section 2 for further details.

[6]We understand 'proof-theoretically tame' here in the sense of [39], i.e. pertaining to the phenomenon that although these notions could be subject to well-known Gödelian phenomena, they nevertheless 'seem to be tame in the sense of allowing for the extraction of bounds of rather low complexity' (as phrased in [39], see also [51, 52] for further discussions of these types of phenomenas and their implications for logic and mathematics).

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 developed 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.

Practically, to ultimately decide whether a proof falls into the range of the systems presented here, one would have to try to fully formalize the proof using these restricted means (i.e. only using the intensional descriptions of the involved objects). While this seems to be a prevalent issue here where we do not fully axiomatize the objects in question but rather just intensionally specify them, it is in fact an issue that concerns all of proof mining to a certain degree as it is, in a way, never fully formally clear whether a proof does or does not use a property that transcends the means of a given system (in particular if abstract types are involved) until one has performed an analysis or has fully formalized the corresponding proof. However, proof mining as a whole benefits from the fact that the systems used therein are designed by taking the way mathematicians actually work in the respective area of application into account, the success of which being empirically substantiated by the large amount of case studies developed using these systems. In that way, gauging the possibility of whether a theorem can be formally established in a given system geared for proof mining is practically often much more direct and accessible. We therefore expect that these considerations also apply to the systems presented here, i.e. that the systems are rather easy to apply in many situations commonly occurring in the literature, in particular when, as a rule of thumb, the proof uses the dual of the underlying space only in an abstract manner, relying mainly on structural properties rather than on concrete objects. This is in particular confirmed from the practical perspective through the recent new case studies given in [57, 60], situated in the area of convex analysis on Banach spaces, where all the objects relating to the dual space are used in a rather abstract way so that formalizing these given proofs using the systems presented here is a most natural task. However, we expect that these metatheorems allow for many further new case studies to be carried out in the areas discussed above and for that, we want to also in particular mention the works [1, 2, 6, 10, 19, 20, 31, 44, 68, 74] as promising future applications as, by inspection of the proofs, they also seem to be formalizable in (suitable extensions of) the systems introduced here, being of a sufficiently abstract nature.

Lastly, we also strongly believe that the tame approach to suprema over bounded sets introduced here as well as the general 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. A first indication of this will be provided by two subsequent papers where, in one, the treatment of monotone operators in Hilbert spaces given in [58] is adapted to monotone operators on Banach spaces as introduced by Browder [13, 14] and, in the other, we provide the first proof-theoretic treatment of the Hausdorff-metric in systems amenable to proof mining metatheorems using an intensional approach to the sets measured by the Hausdorff-metric together with the tame approach to suprema over bounded sets presented here.[7]

2. Preliminaries: basic formal systems for analysis in all finite types

The base system for all the extensions introduced in this paper is the system $\mathcal{A}^\omega[X,\|\cdot\|]$ as introduced by Kohlenbach [36] (see also [25]). This system itself is an extension of the system $\mathcal{A}^\omega = \text{WE-PA}^\omega + \text{QF-AC} + \text{DC}$ for (a fragment of) classical analysis over all finite types T which extends weakly extensional Peano arithmetic in all finite types WE-PA$^\omega$ (as defined e.g. in [72]), for one, with the quantifier-free axiom of choice in finite types

$$\text{(QF-AC)} \qquad \forall \underline{x} \exists \underline{y} A_0(\underline{x},\underline{y}) \to \exists \underline{Y} \forall \underline{x} A_0(\underline{x}, \underline{Y}\underline{x})$$

where A_0 is quantifier-free but the types of the variable tuples $\underline{x}$, $\underline{y}$ are arbitrary and where we use the notation $\underline{Y}\underline{x}$ to abbreviate $Y_1\underline{x},\dots,Y_k\underline{x}$ if $\underline{Y} = Y_1,\dots,Y_k$, and, for another, with the schema of dependent choice $\text{DC} = \{\text{DC}^{\underline{\rho}} \mid \underline{\rho} \subseteq T\}$ with

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

[7]Besides these forthcoming works, we however want to mention that intensional methods together with the tame approach to suprema over bounded sets may in particular, further, be useful to treat the so-called generalized Bregman distances recently introduced by Burachik, Dao and Lindstrom [15] while the approach to the dual space may be adapted to treat function spaces between general vector spaces in order to treat associated operator algebras.

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

To form $\mathcal{A}^\omega[X, \|\cdot\|]$, this system $\mathcal{A}^\omega$ is extended with a new abstract type X and formulated over the extended set of types T^X defined by[8]

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

This new type can then be used with additional constants and axioms to represent a wide range of abstract classes of spaces and operations on them and the concrete extension $\mathcal{A}^\omega[X, \|\cdot\|]$ is obtained by adding 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 $1(X)$ together with 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 [25, 36, 37]). It should be noted that in the system $\mathcal{A}^\omega[X, \|\cdot\|]$, equality at type 0 ($=_0$) is the only primitive relation and equality at different types is introduced via abbreviations. Concretely, identity on X is treated as a defined predicate via[9]

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

and higher-type equality is then defined recursively via

$$x^{\tau(\xi)} =_{\tau(\xi)} y^{\tau(\xi)} := \forall z^\xi \left(xz =_\tau yz\right).$$

Further, we define a relation $\leqslant$ 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 $\underline{\sigma} = (\sigma_1, \dots, \sigma_k)$.

For the notions like $=_\mathbb{R}$ or the arithmetical operations $+_\mathbb{R}$ and $\cdot_\mathbb{R}$, we in that context rely on a chosen representation of the real numbers as a Polish space in the system $\mathcal{A}^\omega$ in which context we follow the definitions and conventions given in [37]. 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 [36]:

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

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 with a fixed Cauchy modulus 2^{-n} (see [37] 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 [35] 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 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 [37], 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 [37] for any further discussions of its properties.

[8] We largely follow the conventions for writing and abbreviating types established in [37].

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

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 [36], 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. However, later we will need an extension of this function $(\cdot)_\circ$ to all real numbers such that we retain these nice properties regarding majorizability. For this, if $r < 0$, we define[10]

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

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 (extending Lemma 2.10 from [36]):

Lemma 2.1. *Let $r \in \mathbb{R}$. Then:*

(1) $(r)_\circ$ is a representation of r in the sense of the above (see again e.g. [37]*).*
(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 is immediate and clearly $(r)_\circ$ is nondecreasing as a type 1 object as j is monotone. For item (2), let $|r| \leqslant s$. If $r \geqslant 0$, the result is contained in the Lemma 2.10 from [36]. 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. □

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.

To make everything more readable, we will omit the subscripts of the arithmetical operations for $\mathbb{R}$ in the following parts. Similarly, we will also omit types of variables whenever convenient and omit types in proofs almost always. Lastly, 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.

3. 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 $\underline{\sigma}^t = \sigma_k, \dots, \sigma_1$ similar to [37]. Then, stating for an additional term s of type $1(\underline{\sigma}^t)$

[10] Here, $\dot{-}$ is defined via $n \dot{-} m = \max\{n - m, 0\}$ for $n, m \in \mathbb{N}$.

that it represents the supremum of a function f, given via another term 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.

$$(S)_1 \qquad \forall \underline{p}^{\underline{\sigma}}, x^X \left(D(\underline{p}) \land C(x,\underline{p}) \to f(x,\underline{p}) \leqslant_{\mathbb{R}} s(\underline{p})\right),$$

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.

$$(S)_2' \qquad \forall \underline{p}^{\underline{\sigma}} \left(D(\underline{p}) \to \forall k^0 \exists x^X \left(C(x,\underline{p}) \land s(\underline{p}) - 2^{-k} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right).$$

Remark 3.1. 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.

$$(+) \qquad \forall \underline{p}^{\underline{\sigma}}, M^1 \left(D(\underline{p}) \land M <_{\mathbb{R}} s(\underline{p}) \to \exists x^X \left(C(x,\underline{p}) \land M <_{\mathbb{R}} f(x,\underline{p})\right)\right),$$

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

$$(++) \qquad \forall \underline{p}^{\underline{\sigma}}, M^1, k^0 \exists x^X, j^0 \left(D(\underline{p}) \land M + 2^{-k} <_{\mathbb{R}} s(\underline{p}) \to \left(C(x,\underline{p}) \land M + 2^{-j} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right),$$

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[11]

$$(S)_2 \qquad \forall \underline{p}^{\underline{\sigma}}, k^0 \exists x^X \leqslant_X b(\underline{p}) 1_X \left(D(\underline{p}) \to \left(C(x,\underline{p}) \land s(\underline{p}) - 2^{-k} \leqslant_{\mathbb{R}} f(x,\underline{p})\right)\right).$$

Now, in the case of a quantifier-free C and an existential D, the above statement is of the form Δ exhibited in [27, 37] (which will also be discussed in more detail later on) 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

$$\forall \underline{a}_1^{\underline{\delta}_1} \exists \underline{b}_1 \leqslant_{\underline{\sigma}_1} \underline{r}_1 \underline{a}_1 \dots \forall \underline{a}_n^{\underline{\delta}_n} \exists \underline{b}_n \leqslant_{\underline{\sigma}_n} \underline{r}_n \underline{a}_1 \dots \underline{a}_n \forall \underline{c}^{\underline{\gamma}} D_{qf}(x, \underline{p}, \underline{a}_1, \dots, \underline{a}_n, \underline{b}_1, \dots, \underline{b}_n, \underline{c})$$

which is a kind of generalized form Δ which we, following Remark 10.24 in [37], denote by Δ^* as well as equivalently as a formula of the form

$$\exists \underline{\widetilde{a}}_1^{\underline{\widetilde{\delta}}_1} \forall \underline{\widetilde{b}}_1 \leqslant_{\underline{\widetilde{\sigma}}_1} \underline{\widetilde{r}}_1 \underline{\widetilde{a}}_1 \dots \exists \underline{\widetilde{a}}_m^{\underline{\widetilde{\delta}}_m} \forall \underline{\widetilde{b}}_m \leqslant_{\underline{\widetilde{\sigma}}_m} \underline{\widetilde{r}}_m \underline{\widetilde{a}}_1 \dots \underline{\widetilde{a}}_m \exists \underline{\widetilde{c}}^{\underline{\widetilde{\gamma}}} \widetilde{D}_{qf}(x, \underline{p}, \underline{\widetilde{a}}_1, \dots, \underline{\widetilde{a}}_m, \underline{\widetilde{b}}_1, \dots, \underline{\widetilde{b}}_m, \underline{\widetilde{c}})$$

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 $(S)_2$ are a priori admissible in the context of bound extraction theorems (in the sense that they have a monotone functional interpretation, see the later Section 7 for further details).

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

[11] Here, 1_X is the constant of $\mathcal{A}^\omega[X, \|\cdot\|]$ representing a canonical unit vector as before.

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

$$\forall r^1, \underline{p}^{\underline{\sigma}} \forall x^X \left(D(\underline{p}, r) \rightarrow f(\widetilde{x}^r, \underline{p}, r) \leqslant_{\mathbb{R}} s(\underline{p}, r)\right)$$

where we make use of the functional[12]

$$\widetilde{x}^r = \frac{rx}{\max_{\mathbb{R}}\{\|x\|_X, r\}}$$

which allows for implicit quantification over elements from $\overline{B}_r(0)$.

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 paper 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. These examples then in particular also make crucial use of an extended language where new constants for the respective functions and their suprema are included. If a potential proof would make use of further suprema, then further such extensions of the language and axioms of the system would presumably be required.

4. 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 paper but we in general refer to [66, 71] 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 types $1(X)$ 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^*.

[12] This functional seems to have first been used for $r = 1$ in [40].

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

$$(*)_1 \qquad \forall x^{*X^*}, x^X \left(|\langle x, x^*\rangle_{X^*}| \leqslant_{\mathbb{R}} \|x^*\|_{X^*} \|x\|_X\right),$$

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

$$(*)_2 \qquad \forall x^{*X^*}, k^0 \exists x \leqslant_X 1_X \left(\|x^*\|_{X^*} - 2^{-k} \leqslant_{\mathbb{R}} |\langle x, x^*\rangle_{X^*}|\right),$$

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 4.1. Similar to Remark 3.1, 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 3.1, 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[13]

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

for x^{*X^*}, y^{*X^*}.

[13]Similar to before with $-_X$, we write $x^* -_{X^*} y^*$ for $x^* +_{X^*} (-_{X^*} y^*)$.

We now turn to the axioms for the application constant $\langle\cdot,\cdot\rangle_{X^*}$ which essentially just state that the map is bilinear:[14]

$$(*)_3 \qquad \begin{cases} \forall x^X, x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1 \left(\langle x, \alpha x^* +_{X^*} \beta y^*\rangle_{X^*} =_{\mathbb{R}} \alpha\langle x, x^*\rangle_{X^*} + \beta\langle x, y^*\rangle_{X^*}\right), \\ \forall x^X, x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1 \left(\langle x, \alpha x^* -_{X^*} \beta y^*\rangle_{X^*} =_{\mathbb{R}} \alpha\langle x, x^*\rangle_{X^*} - \beta\langle x, y^*\rangle_{X^*}\right), \end{cases}$$

$$(*)_4 \qquad \begin{cases} \forall x^X, y^X, x^{*X^*}, \alpha^1, \beta^1 \left(\langle \alpha x +_X \beta y, x^*\rangle_{X^*} =_{\mathbb{R}} \alpha\langle x, x^*\rangle_{X^*} + \beta\langle y, x^*\rangle_{X^*}\right), \\ \forall x^X, y^X, x^{*X^*}, \alpha^1, \beta^1 \left(\langle \alpha x -_X \beta y, x^*\rangle_{X^*} =_{\mathbb{R}} \alpha\langle x, x^*\rangle_{X^*} - \beta\langle y, x^*\rangle_{X^*}\right), \end{cases}$$

Lastly, we specify the vector space structure of X^* further, akin to [36]:[15]

$$(*)_5 \qquad \text{The vector space axioms for } +_{X^*},\ -_{X^*},\ \cdot_{X^*},\ 0_{X^*},\ 1_{X^*} \text{ w.r.t. } =_{X^*}.$$

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[16]

$$\text{(QF-LR)} \qquad \frac{A_0 \to \left(\forall x^X, y^X, \alpha^1, \beta^1 \left(t(\alpha x +_X \beta y) =_{\mathbb{R}} \alpha tx + \beta ty\right) \wedge \forall x^X \left(|tx| \leqslant_{\mathbb{R}} M \|x\|_X\right)\right)}{A_0 \to \exists x^* \leqslant_{X^*} M 1_{X^*} \forall x^X \left(tx =_{\mathbb{R}} \langle x, x^*\rangle_{X^*}\right)}$$

where A_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 4.5 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.

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

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:

$$(*)_6 \qquad \forall x^X \exists x^* \leqslant_{X^*} \|x\|_X 1_{X^*} \left(\langle x, x^*\rangle_{X^*} =_{\mathbb{R}} \|x\|_X^2 =_{\mathbb{R}} \|x^*\|_{X^*}^2\right).$$

Definition 4.2. 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 4.3. 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

[14]In the following, we omit the types from $\cdot_{X^*}$ or $\cdot_{X^*}$ altogether, similar to $\cdot_X$.

[15]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 $\|1_{X^*}\|_{X^*} =_{\mathbb{R}} 1$.

[16]Similar to before, given objects x^*, y^* of type X^*, we here write $x^* \leqslant_{X^*} y^*$ for $\|x^*\|_{X^*} \leqslant_{\mathbb{R}} \|y^*\|_{X^*}$.

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 [58]. 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 4.5) 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 4.5) and extensional:

Lemma 4.4. *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 4.1), 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 4.5. The above treatment of X^* ties to the notion of dual systems from the context of topological vector spaces (see e.g. [67]). 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 4.6. *The system $\mathcal{D}^\omega$ proves the norm axioms exhibited in* [36]*, 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^*.[17] Also, in the context of the use of axiom $(*)_2$, recall Remark 4.1 for the particular consequence of $(*)_2$ that formalizes the usual least upper bound property of the supremum for $\|\cdot\|_{X^*}$.

(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

[17] 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 4.4) 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.

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 4.7. A simple property of Banach spaces (see e.g. [66]) 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 [37]): 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 $fk:=\Phi\widehat{(k+3)}(k+3)$. 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[18]

$$\forall x^X\forall n^0\forall m,k\geqslant_0(n+[\|x\|_X](0)+1)\left(|\langle x,x^*k\rangle_{X^*}-\langle x,x^*m\rangle_{X^*}|\leqslant_{\mathbb{R}}2^{-n}\right).$$

By the above completeness of $\mathbb{R}$ we can define its limit by a term in x in the sense that provably

$$\forall x^X\forall n\geqslant_0([\|x\|_X](0)+1)\left(|\langle x,x^*n\rangle_{X^*}-fx|\leqslant_{\mathbb{R}}2^{-n}\right)$$

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 4.3 since also here, only a corresponding rule could be formulated which states the closure of χ_{X^*} under functions 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 [37]). 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 4.8. By formalizing a standard argument (see e.g. Chapter 2, §4, Theorem 1 in [21]), 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 [40]), i.e.[19]

$$\forall x^X,y^X,k^0\left(\|x\|_X>_{\mathbb{R}}1\wedge\|y\|_X<_{\mathbb{R}}2^{-\tau(k)}\to\left\|\widetilde{x}^1+_X y\right\|_X+\left\|\widetilde{x}^1-_X y\right\|_X\leqslant_{\mathbb{R}}2+2^{-k}\|y\|_X\right),$$

is equivalent to the uniform convexity of X^*, formulated using a modulus of uniform convexity η of type 1 (in similarity to e.g. [36]), i.e.

$$\forall x^{*X^*},y^{*X^*},k^0\left(\|x^*\|_{X^*},\|y^*\|_{X^*}<_{\mathbb{R}}1\wedge\left\|\frac{x^*+_{X^*}y^*}{2}\right\|_{X^*}>_{\mathbb{R}}1-2^{-\eta(k)}\to\|x^*-_{X^*}y^*\|_{X^*}\leqslant_{\mathbb{R}}2^{-k}\right).$$

We do not spell this out here any further.

[18]Here, we write $[a](n)$ for the n-th number in the type 1 representation of the real number a as before.

[19]Here, $\widetilde{x}^1$ is defined as in Section 3.

5. Reflexivity of Banach spaces

5.1. **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 5.1. 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 5.2 (James' theorem [30])**.** *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 $\langle x, x^* \rangle = 1$.*

5.2. **Treating reflexivity.** To treat reflexivity in its version given by Definition 5.1, 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.[20]

(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:[21]

$(**)_1$ $$\forall x^{**X^{**}}, x^{*X^*} \left(|\langle x^*, x^{**} \rangle_{X^{**}}| \leq_{\mathbb{R}} \|x^{**}\|_{X^{**}} \|x^*\|_{X^*} \right).$$

$(**)_2$ $$\forall x^{**X^{**}}, k^0 \exists x^* \leq_{X^*} 1_{X^*} \left(\|x^{**}\|_{X^{**}} - 2^{-k} \leq_{\mathbb{R}} |\langle x^*, x^{**} \rangle_{X^{**}}| \right).$$

$(**)_3$ $$\begin{cases} \forall x^{*X^*}, x^{**X^{**}}, y^{**X^{**}}, \alpha^1, \beta^1 \left(\langle x^*, \alpha x^{**} +_{X^{**}} \beta y^{**} \rangle_{X^{**}} =_{\mathbb{R}} \alpha \langle x^*, x^{**} \rangle_{X^{**}} + \beta \langle x^*, y^{**} \rangle_{X^{**}} \right), \\ \forall x^{*X^*}, x^{**X^{**}}, y^{**X^{**}}, \alpha^1, \beta^1 \left(\langle x^*, \alpha x^{**} -_{X^{**}} \beta y^{**} \rangle_{X^{**}} =_{\mathbb{R}} \alpha \langle x^*, x^{**} \rangle_{X^{**}} - \beta \langle x^*, y^{**} \rangle_{X^{**}} \right). \end{cases}$$

$(**)_4$ $$\begin{cases} \forall x^{*X^*}, y^{*X^*}, x^{**X^{**}}, \alpha^1, \beta^1 \left(\langle \alpha x^* +_{X^*} \beta y^*, x^{**} \rangle_{X^{**}} =_{\mathbb{R}} \alpha \langle x^*, x^{**} \rangle_{X^{**}} + \beta \langle y^*, x^{**} \rangle_{X^{**}} \right), \\ \forall x^{*X^*}, y^{*X^*}, x^{**X^{**}}, \alpha^1, \beta^1 \left(\langle \alpha x^* -_{X^*} \beta y^*, x^{**} \rangle_{X^{**}} =_{\mathbb{R}} \alpha \langle x^*, x^{**} \rangle_{X^{**}} - \beta \langle y^*, x^{**} \rangle_{X^{**}} \right). \end{cases}$$

$(**)_5$ The vector space axioms for $+_{X^{**}}$, $-_{X^{**}}$, $\cdot_{X^{**}}$, $0_{X^{**}}$, $1_{X^{**}}$ w.r.t. $=_{X^{**}}$.

$(**)_6$ $$\forall x^{*X^*} \exists x^{**} \leq_{X^{**}} \|x^*\|_{X^*} 1_{X^{**}} \left(\langle x^*, x^{**} \rangle_{X^{**}} =_{\mathbb{R}} \|x^*\|^2_{X^*} =_{\mathbb{R}} \|x^{**}\|^2_{X^{**}} \right).$$

For the rule, we opt for the formulation[22]

(QF-LR**) $$\frac{\begin{cases} A_0 \to (\forall x^{*X^*}, y^{*X^*}, \alpha^1, \beta^1 \left(t(\alpha x^* +_{X^*} \beta y^*) =_{\mathbb{R}} \alpha t x^* + \beta t y^* \right) \\ \qquad \wedge \forall x^{*X^*} \left(|t x^*| \leq_{\mathbb{R}} M \|x^*\|_{X^*} \right)) \end{cases}}{A_0 \to \exists x^{**} \leq_{X^{**}} M 1_{X^{**}} \forall x^{*X^*} \left(t x^* =_{\mathbb{R}} \langle x^*, x^{**} \rangle_{X^{**}} \right)},$$

[20] As before, in formulas, we often omit the types around the $\cdot_{X^{**}}$-operation or we omit the operation entirely.

[21] 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 $\|1_{X^{**}}\|_{X^{**}} =_{\mathbb{R}} 1$.

[22] Also here, given objects x^{**}, y^{**} of type X^{**}, we write $x^{**} \leq_{X^{**}} y^{**}$ for $\|x^{**}\|_{X^{**}} \leq_{\mathbb{R}} \|y^{**}\|_{X^{**}}$ similarly to before.

where A_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:

$$\text{(R)} \qquad \forall x^{**X^{**}} \exists x \leqslant_X \|x^{**}\|_{X^{**}} 1_X \forall x^{*X^*} \left(\|x\|_X =_{\mathbb{R}} \|x^{**}\|_{X^{**}} \wedge \langle x, x^* \rangle_{X^*} =_{\mathbb{R}} \langle x^*, x^{**} \rangle_{X^{**}}\right).$$

As a simple example for the use of the axiom (R), we now consider the formal provability of one direction of James' theorem.

Lemma 5.3. *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 $\langle x, x^* \rangle = \langle x^*, x^{**} \rangle = \|x^*\|^2$. $\square$

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

$$\text{(JT)} \qquad \forall x^{*X^*} \exists x^X \leqslant_X \|x^*\|_{X^*} 1_X \left(\langle x, x^* \rangle_{X^*} =_{\mathbb{R}} \|x\|_X^2 =_{\mathbb{R}} \|x^*\|_{X^*}^2\right).$$

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.[23] Clearly, the above formal proof of one direction of the equivalence in James' theorem only concerns the simpler of the two directions and as such just represents a small litmus test for the appropriateness of the phrasing of both (R) and (JT). We do not know whether the deep converse direction of James' theorem is formalizable in the above systems and it would be very interesting to further investigate this topic.

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 mainly 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. So, from a practical perspective, a good rule of thumb for gauging whether the above approach to reflexivity is useful in the context of an analysis of a given proof is whether the use of reflexivity is suitably abstract, concerning not concrete objects but just structural properties. In particular, this seems to be the case in the context of the current proof mining practice focused on modern convex analysis over Banach spaces where these issues seem to not feature at all. This is in particular illustrated by the applications given in [60] where reflexivity features prominently in the underlying theory but is only ever used in a rather abstract way to guarantee the existence and well-definedness of certain involved objects corresponding to the geometry of the space which can be easily hardwired into system.

Remark 5.4. By formalizing a standard argument (see e.g. Chapter 2, §4, Theorem 2 in [21]), 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).

[23]Note that this system is conservative over the base system by relativizing the quantifiers over elements of X^* accordingly.

6. 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 is that it 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 (beyond the previous abstract treatment of generalized gradients and monotone operators [43, 58]) 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 form the basis for the extraction of quantitative results on the asymptotic regularity and convergence of iterations involving Bregman strongly nonexpansive operators given in [60] as well as of rates of metastability for algorithms involving Bregman projections in [57]. We refer to the references given in the introduction for further examples from the vast array of potential future applications of these systems.

6.1. **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 [5, 64, 65, 73].

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. [12, 62]). For this, we write $\mathrm{intdom}f$ for the interior of $\mathrm{dom}f$.

Definition 6.1 (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 6.2 (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 on a set $C \subseteq X$ if the limit is also uniform in $x \in C$.

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. [73]). 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 [73] (or in [5] for the case of Hilbert spaces where the proofs are rather similar).

Proposition 6.3. *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, being uniformly Fréchet differentiable on bounded sets (essentially) implies being Fréchet differentiable with a gradient that is uniformly norm-to-norm continuous on bounded sets.

Proposition 6.4 (essentially [61])**.** *If $f : X \to \mathbb{R}$ is uniformly Fréchet differentiable on bounded sets 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 functions that are uniformly Fréchet differentiable on bounded sets where ∇f is bounded on bounded sets.

6.2. **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 6.6 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 [37] 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 the convexity of f requires the extensionality of f to work as expected (as extensionality of f is e.g. needed to prove that the convexity property really holds for all λ of type 1 with $0 \leqslant_{\mathbb{R}} \lambda \leqslant_{\mathbb{R}} 1$ as would be desired). However, Lemma 6.5 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, since these functions naturally satisfy $\mathrm{dom}\nabla f = \mathrm{intdom}\, f$ and since we assume $\mathrm{dom}\, f = X$, 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.

$$\forall x^X, y^X, b^0, k^0 \left(\|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X <_{\mathbb{R}} 2^{-\omega^{\nabla f}(k,b)} \to \|\nabla f(x) -_{X^*} \nabla f(y)\|_{X^*} \leqslant_{\mathbb{R}} 2^{-k} \right).$$

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 6.3, 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 6.5. *The theory $\mathcal{D}^\omega[f, \nabla f]$ proves:*

(1) *f is uniformly Fréchet differentiable on bounded subsets, i.e.*

$$\forall b^0, k^0 \exists j^0 \forall x^X, y^X \Bigg(\|x\|_X <_{\mathbb{R}} b \wedge 0 <_{\mathbb{R}} \|y\|_X <_{\mathbb{R}} 2^{-j} \to \frac{|f(x+y) - f(x) - \langle y, \nabla f(x)\rangle_{X*}|}{\|y\|_X} \leqslant_{\mathbb{R}} 2^{-k} \Bigg),$$

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.*

$$\forall k^0, b^0 \exists j^0 \forall x^X, y^X \left(\|x\|_X, \|y\|_X <_{\mathbb{R}} b \wedge \|x -_X y\|_X \leqslant_{\mathbb{R}} 2^{-j} \to |f(x) - f(y)| \leqslant_{\mathbb{R}} 2^{-k} \right),$$

where in fact one can choose

$$j = \omega^f(k, b) = k + C(b).$$

(4) *f is bounded on bounded sets, i.e.*

$$\forall b^0 \exists d^0 \forall x^X \left(\|x\|_X <_{\mathbb{R}} b \to |f(x)| \leqslant_{\mathbb{R}} d \right),$$

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

$$\begin{aligned} f(x+y) - f(x) &\geqslant \langle x + y - x, \nabla f(x)\rangle \\ &= \langle y, \nabla f(x)\rangle. \end{aligned}$$

Similarly we derive

$$f(x) - f(x+y) \geqslant \langle -y, \nabla f(x+y)\rangle.$$

Together, we get

$$\begin{aligned} 0 &\leqslant f(x+y) - f(x) - \langle y, \nabla f(x)\rangle \\ &\leqslant \langle y, \nabla f(x+y)\rangle - \langle y, \nabla f(x)\rangle \\ &\leqslant \|y\| \, \|\nabla f(x+y) - \nabla f(x)\|. \end{aligned}$$

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 6.6. 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[24]

$$\forall x^X, k^0 \left(\chi_{\nabla f} x k =_0 0 \to \forall y^k \left(\left(x -_X \widetilde{y}^{(2^{-k})} \right) \in \mathrm{dom} f \right) \right)$$

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

$$(x, k) \in \mathrm{dom}\nabla f := \chi_{\nabla f} x k =_0 0$$

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 $\widetilde{y}^{(2^{-k})}$.

---

[24]Here, we use the $\widetilde{\cdot}^{r}$ operation on elements of type X as defined in Section 3.

6.3. **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 [22], see also [11, 63]): 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 6.7 ([3])**.** *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 6.8. *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 of 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 6.5 which immediately allows us to compute a majorant for f^*.

The axioms for f^* are now readily presented by formalizing the supercoercivity of f using a corresponding modulus and then using the above properties of f^* to instantiate the previous schemes $(S)_1$ and $(S)_2$:

$(f)_2$ 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 6.9. Similar to Remark 3.1 (recall also Remark 4.1), 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 6.10. *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 6.9), 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.

$$\begin{aligned}\forall {x^*}^{X^*}, {y^*}^{X^*}, b^0, k^0 \Bigg(\bigg(\|x^*\|_{X^*}, \|y^*\|_{X^*} <_{\mathbb{R}} b \\ \wedge \|x^* -_{X^*} y^*\|_{X^*} <_{\mathbb{R}} 2^{-\omega^{\nabla f^*}(k,b)}\bigg) \to \|\nabla f^*(x^*) -_X \nabla f^*(y^*)\|_X \leqslant_{\mathbb{R}} 2^{-k}\Bigg).\end{aligned}$$

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 6.11 ([3]). 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 6.12 ([3]). *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 on bounded sets.

Remark 6.13. 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 [9], 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 [8] 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 6.14 ([3]). *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 6.15. Note that the previous Lemma 6.5, 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 6.16. 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. [16, 17, 18] 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 [60] for further (formal) investigations into the interrelations of these properties and their quantitative analogues as guided by the logical methodology introduced in this paper.

6.4. **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 [10] which features in many algorithmic approaches in that field (see in particular again the references in the introduction as well as the references in [4]).

These Bregman distances are defined relative to a convex function in terms of its gradient:

Definition 6.17 ([10]). 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 $W^f : \text{dom } f \times \text{dom } f^* \to [0, +\infty)$ 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. [54, 53]). Also this function can be represented by a closed term in the system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$. Similarly, the above system is strong enough to prove many of the common properties of D_f outright and we just mention two of these here:

Lemma 6.18. *The system $\mathcal{D}^\omega[f, \nabla f, f^*, \nabla f^*]$ proves the three and four point identities (see e.g. [4]):*

(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. [18]), 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 is also the case in the forthcoming works [57, 60].

7. 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 [25, 36, 37] 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.

The basis for the upcoming metatheorems, as well as for most of the previously established ones in the literature, is the utilization of *Gödel's functional interpretation* [26] in combination with a negative translation [48]. We recall the definitions of those interpretations here.

Definition 7.1 ([26, 72])**.** The Dialectica interpretation $A^D = \exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y})$ of a formula A in the language of $\mathcal{A}^\omega[X, \|\cdot\|]$ (and its extensions) is defined via the following recursion on the structure of the formula:

(1) $A^D := A_D := A$ for A being a prime formula.

If $A^D = \exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y})$ and $B^D = \exists \underline{u} \forall \underline{v} B_D(\underline{u}, \underline{v})$, we set

(2) $(A \wedge B)^D := \exists \underline{x}, \underline{u} \forall \underline{y}, \underline{v} (A \wedge B)_D$
where $(A \wedge B)_D(\underline{x}, \underline{u}, \underline{y}, \underline{v}) := A_D(\underline{x}, \underline{y}) \wedge B_D(\underline{u}, \underline{v})$,

(3) $(A \vee B)^D := \exists z^0, \underline{x}, \underline{u} \forall \underline{y}, \underline{v} (A \vee B)_D$
where $(A \vee B)_D(z^0, \underline{x}, \underline{u}, \underline{y}, \underline{v}) := (z = 0 \to A_D(\underline{x}, \underline{y})) \wedge (z \neq 0 \to B_D(\underline{u}, \underline{v}))$,

(4) $(A \to B)^D := \exists \underline{U}, \underline{Y} \forall \underline{x}, \underline{v} (A \to B)_D$
where $(A \to B)_D(\underline{U}, \underline{Y}, \underline{x}, \underline{v}) := A_D(\underline{x}, \underline{Yxv}) \to B_D(\underline{Ux}, \underline{v})$,

(5) $(\exists z^\tau A(z))^D := \exists z, \underline{x} \forall \underline{y} (\exists z^\tau A(z))_D$
where $(\exists z^\tau A(z))_D(z, \underline{x}, \underline{y}) := A_D(\underline{x}, \underline{y}, z)$,

(6) $(\forall z^\tau A(z))^D := \exists \underline{X} \forall z, \underline{y} (\forall z^\tau A(z))_D$
where $(\forall z^\tau A(z))_D(\underline{X}, z, \underline{y}) := A_D(\underline{X}z, \underline{y}, z)$.

Definition 7.2 ([48])**.** The negative translation of A is defined by $A' := \neg\neg A^*$ where A^* is defined by the following recursion on the structure of A:

(1) $A^* := A$ for prime A,
(2) $(A \circ B)^* := A^* \circ B^*$ for $\circ \in \{\wedge, \vee, \to\}$,
(3) $(\exists x^\tau A)^* := \exists x^\tau A^*$,

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

The following is then a soundness result for the combination of both interpretations which forms the basis for the upcoming metatheorems. In that context, we write $\mathcal{A}^\omega[X, \|\cdot\|]^-$ for the respective system *without* the axiom schemes QF-AC and DC.

Lemma 7.3 (essentially [36]). *Let $\mathcal{P}$ be a set of universal sentences and let $A(\underline{a})$ 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*

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

holds where $\underline{t}$ is a tuple of closed terms of $\mathcal{A}^\omega[X, \|\cdot\|]^- + (\mathrm{BR})$ which can be extracted from the respective proof and (BR) *is the schema of* simultaneous bar-recursion *of Spector* [70]*, here extended to all types from T^X (see e.g.* [37]*).*

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.

Besides Gödel's functional interpretation, the other central notion used in the bound extraction results is that of majorizability. Majorizability was first introduced by Howard [29] and subsequently extended by Bezem [7] to the notion of strong majorizability which Bezem used to provide a model of bar-recursion featuring discontinuous functionals. This model also forms the basis for the bound extraction theorems in proof mining as developed in [25, 36]. In that context, we however rely on a further extension of Bezem's strong majorizability to the new abstract types devised in [25, 36]. In this work, 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 [37], 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 7.4 (essentially [25]). 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 7.5 (essentially [25, 36]). 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 follows: 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 a model based on the usual full set-theoretic structure $\mathcal{S}^{\omega,X,X^*}$.

For the concrete implementation of 'low enough', we follow [36, 25] and in that way introduce the following specific classes of types: We call of type ξ *of degree n* if $\xi \in T$ and it has degree $\leqslant n$ in the usual sense (see e.g. [37]). Further 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 $\xi_0 \in \{0, X, X^*\}$.

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

We now define the previously only vaguely discussed class Δ more precisely. In similarity to [27, 37], we consider formulas of type Δ to be formulas of the form

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

where A_{qf} is quantifier-free, the types in $\underline{\delta}$, $\underline{\sigma}$ and $\underline{\gamma}$ are admissible and $\underline{r}$ are tuples of closed terms of appropriate type. Here, $\leqslant$ is defined as in the preliminaries by recursion on the type with the respective additional clause for X^*, i.e.

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

So the main property beyond the form of the sentences is that we now in particular also require that the occurring types are low enough (note that all formulas of type Δ considered before satisfy these restrictions).

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

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

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

In the bound extraction theorems, axioms of type Δ are treated 'in spirit' of the monotone functional interpretation (as introduced in [34] but conceptually already due to [32]). 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 similar as in e.g. Corollary 5.14 from [27]. 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

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

as new constants to the language and adding its 'instantiated Skolem normal form', i.e. the sentence

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

as a new axiom. Further, we do the same with all conclusions of the rule (QF-LR): for any provable premise

$$\mathcal{D}^\omega + \Delta \vdash A_0 \to \left(\forall x^X, y^X, \alpha^1, \beta^1 \left(t(\alpha x +_X \beta y) =_{\mathbb{R}} \alpha t x + \beta t y\right) \wedge \forall x^X \left(|tx| \leqslant_{\mathbb{R}} M \|x\|_X\right)\right)$$

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

$$\|x^*_t\|_{X^*} \leqslant_{\mathbb{R}} M \wedge \left(A_0 \to \forall x^X \left(tx =_{\mathbb{R}} \langle x, x^*_t \rangle_{X^*}\right)\right).$$

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 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 $\mathcal{D}^\omega$.

We now move to the central majorizability result (Lemma 7.7), 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 [25]. Before this, we just note that majorization behaves as expected for functionals with multiple arguments (represented by their 'curryied' variants):

Lemma 7.6 ([25, 36], see also Kohlenbach [37], Lemma 17.80)**.** *Let* $\xi = \tau(\xi_k)\dots(\xi_1)$*. For* $x^* : M_{\widehat{\xi_1}} \to (M_{\widehat{\xi_2}} \to \dots \to M_{\widehat{\tau}})\dots)$ *and* $x : M_{\xi_1} \to (M_{\xi_2} \to \dots \to M_\tau)\dots)$*, we have* $x^* \gtrsim_\xi x$ *iff*

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

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

Lemma 7.7. *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* $\omega \gtrsim \omega^{\nabla f^*}$ *and* $n \geqslant_{\mathbb{R}} |f^*(0)|, \|\nabla f^*(0)\|_X$ *together with a parameter* ${\omega'}^{0(0)}$ *such that* $\omega' \gtrsim \alpha^f$*. 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. The structure of the proof is very much standard and follows that of the proof of Lemma 17.85 in [37]. 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 [37] (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) $\lambda \alpha^1, x^0.((\alpha(0) + 1)x) \gtrsim \cdot_{X^*}$,

(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 [37]) and we thus omit the details for them (note that item (7), similar to item (6) discussed below, relies on Lemma 2.1). 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 implies $(n \cdot m)_\circ \gtrsim (\langle x, x^*\rangle)_\circ$. The latter follows immediately from Lemma 2.1 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 $\omega \in \mathbb{N}^{\mathbb{N}\times\mathbb{N}}$, $\omega' \in \mathbb{N}^\mathbb{N}$ and $n \in \mathbb{N}$ with $\omega \gtrsim \omega^{\nabla f}, \omega^{\nabla f^*}$, $\omega' \gtrsim \alpha^f$ 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 as before but also Lemma 6.5 and its variant for f^* and ∇f^* as in Remark 6.15) 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. [37]). 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 [27], 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 [27] 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 [33], one can show (similar to [33]) that $\mathcal{M}^{\omega,X,X^*} \models \text{b-AC}_{X,X^*}$ where

$$\text{b-AC}_{X,X^*} = \bigcup_{\delta,\rho\in T^{X,X^*}} \text{b-AC}^{\delta,\rho}$$

with

$$(\text{b-AC}^{\delta,\rho}) \qquad \forall Z^{\rho(\delta)} \left(\forall x^\delta \exists y \leqslant_\rho ZxA(x,y,Z) \to \exists Y \leqslant_{\rho(\delta)} Z \forall x^\delta A(x, Yx, Z)\right).$$

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 [25] carries over. Further, we need that it is provable in $\mathcal{D}^{\omega-}$ that

$$(\dagger) \qquad \forall x', x, y \left(x' \gtrsim_\rho x \wedge x \geqslant_\rho y \to x' \gtrsim_\rho y\right)$$

holds for all types ρ which can be shown similar as e.g. in [37].

Suppose now that

$$\mathcal{S}^{\omega,X,X^*} \models \forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} A_{qf}(\underline{a}, \underline{b}, \underline{c}).$$

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 $\underline{r}\underline{a}$ is majorizable. That we have $\underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a}$ 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 7.8. *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 $\forall \underline{a}^{\underline{\delta}} \exists \underline{b} \leqslant_{\underline{\sigma}} \underline{r}\underline{a} \forall \underline{c}^{\underline{\gamma}} A_{qf}(\underline{a}, \underline{b}, \underline{c})$ where A_{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 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*

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

then one can extract a partial functional $\Phi : S_\delta \times S_{\widehat{\tau}} \rightharpoonup \mathbb{N}$ 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

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

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* [26] *and Hilbert* [28]*. In that case, also plain majorization can be used instead of strong majorization (see e.g.* [37]*).*
(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 $\omega \in \mathbb{N}^{\mathbb{N}\times\mathbb{N}}$ 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 $\omega \gtrsim \omega^{\nabla f^*}$ and $n \geqslant_{\mathbb{R}} |f^*(0)|, \|\nabla f^*(0)\|_X$ together with a new parameter $\omega' \in \mathbb{N}^{\mathbb{N}}$ such that $\omega' \gtrsim \alpha^f$ for the latter.

Proof. We only treat the case of $\mathcal{D}^\omega + \Delta$. First, assume that

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

Clearly, the same statement is then also provable in $\overline{\mathcal{D}}^{\omega}_{\Delta}$. 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, by prenexiation, we get

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

Using Lemma 7.3 (which is applicable as $\overline{\mathcal{D}}^{\omega}_{\Delta}$ is an extension of $\mathcal{A}^\omega[X, \|\cdot\|]$ only by new constants and purely universal axioms) and disregarding the realizers for $\underline{a}, \underline{b}$, we get closed terms t_u, t_v of $\overline{\mathcal{D}}^{\omega-}_{\Delta}$ + (BR) such that

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

By Lemma 7.7 there are closed terms t^*_u, t^*_v of $\mathcal{A}^\omega$ + (BR) such that

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

for all nontrivial normed spaces $(X, \|\cdot\|)$ with their duals X^* and where the constants are interpreted as in Lemma 7.7. Define

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

Then

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

holds for all $z \in M_\tau$ and $z^* \in M_{\hat{\tau}}$ with $z^* \gtrsim z$. The conclusion that $\mathcal{S}^{\omega,X,X^*}$ satisfies the same sentence can be achieved as in the proof of Theorem 17.52 in [37] 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 $\mathcal{M}^{\omega,X,X^*}$ and $\mathcal{S}^{\omega,X,X^*}$ coincide (recall the proof of Lemma 7.7). Therefore, any such z, z^* from $\mathcal{S}^{\omega,X,X^*}$ also live in $\mathcal{M}^{\omega,X,X^*}$ so that $\Phi(z^*)$ is well-defined for z, z^* belonging to $\mathcal{S}^{\omega,X,X^*}$ with $z^* \gtrsim z$. In $B_\forall$, all types are admissible to that truth in $\mathcal{S}^{\omega,X,X^*}$ implies truth in $\mathcal{M}^{\omega,X,X^*}$ and similarly for $C_\exists$ where thus truth in $\mathcal{M}^{\omega,X,X^*}$ implies truth in $\mathcal{S}^{\omega,X,X^*}$. Lastly, as in Lemma 17.84 in [37], we can show that as Φ is of type $0(\hat{\tau})$, the interpretations of Φ in $\mathcal{S}^{\omega,X,X^*}$ and $\mathcal{M}^{\omega,X,X^*}$ coincide on majorizable elements. All in all we have that

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

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

The additional $\forall x^\delta \forall y \leqslant_\sigma s(x)$ can be treated as e.g. discussed in [58] and we thus omit any details. Similarly, item (1) can be shown as in the proof of Theorem 17.52 from [37] (see page 428 therein). Further, (2) is immediate and (3) follows from the fact that without DC, bar recursion becomes superfluous and the model $\mathcal{M}^{\omega,X,X^*}$ can be avoided. □

Acknowledgments: I want to thank Ulrich Kohlenbach for many helpful discussions on the topic of this paper. Also, I want to thank the referee for many very valuable comments which have improved the paper at various places. The results of this paper form the main part of Chapter 8 of my doctoral dissertation [59], submitted on October 10th, 2023 to the Department of Mathematics of TU Darmstadt. The author was supported by the 'Deutsche Forschungsgemeinschaft' Project DFG KO 1737/6-2.

References

[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] 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.

[3] 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.

[4] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[5] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, Cham, 2017.

[6] H.H. Bauschke and A.S. Lewis. Dykstra's algorithm with Bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.

[7] 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.

[8] 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.

[9] J.M. Borwein and J. Vanderwerff. Fréchet-Legendre functions and reflexive Banach spaces. *Journal of Convex Analysis*, 17(3-4):915–924, 2010.

[10] 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.
[11] A. Brøndsted. Conjugate convex functions in topological vector spaces. *Mat. Fys. Medd. Danske Vid. Selsk.*, 34, 1964.
[12] A. Brøndsted and R.T. Rockafellar. On the subdifferentiability of convex functions. *Proceedings of the American Mathematical Society*, 16(4):605–611, 1965.
[13] F.E. Browder. Nonlinear monotone operators and convex sets in Banach spaces. *Bulletin of the American Mathematical Society*, 71(5):780–785, 1965.
[14] F.E. Browder. Nonlinear maximal monotone operators in Banach space. *Mathematische Annalen*, 175:89–113, 1968.
[15] R.S. Burachik, M.N. Dao, and S.B. Lindstrom. The generalized Bregman distance. *SIAM Journal on Optimization*, 31(1):404–424, 2021.
[16] 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.
[17] 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.
[18] 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.
[19] Y. Censor and T. Elfving. A multiprojection algorithm using Bregman projections in a product space. *Numerical Algorithms*, 8:221–239, 1994.
[20] Y. Censor and S. Reich. The Dykstra algorithm with Bregman projections. *Communications in Applied Analysis*, 2(3):407–420, 1998.
[21] J. Diestel. *Geometry of Banach Spaces - Selected Topics*, volume 485 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin Heidelberg, 1975.
[22] W. Fenchel. On conjugate convex functions. *Canadian Journal of Mathematics*, 1:73–77, 1949.
[23] F. Ferreira, L. Leuştean, and P. Pinto. On the removal of weak compactness arguments in proof mining. *Advances in Mathematics*, 354, 2019. 106728.
[24] F. Ferreira and P. Oliva. Bounded functional interpretation. *Annals of Pure and Applied Logic*, 135:73–112, 2005.
[25] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Transactions of the American Mathematical Society*, 360:2615–2660, 2008.
[26] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958.
[27] D. Günzel and U. Kohlenbach. Logical metatheorems for abstract spaces axiomatized in positive bounded logic. *Advances in Mathematics*, 290:503–551, 2016.
[28] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95(1):161–190, 1926.
[29] 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.
[30] R.C. James. Reflexivity and the Supremum of Linear Functionals. *Annals of Mathematics*, 66(1):159–169, 1957.
[31] 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.
[32] 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.
[33] U. Kohlenbach. Pointwise hereditary majorization and some applications. *Archive for Mathematical Logic*, 31:227–241, 1992.
[34] 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 (Keele, 1993)*, pages 225–260. Oxford University Press, 1996.
[35] U. Kohlenbach. Arithmetizing proofs in analysis. In J.M. Larrazabal, D. Lascar, and G. Mints, editors, *Logic Colloquium '96*, volume 12 of *Springer Lecture Notes in Logic*, pages 115–158. Springer, 1998.
[36] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Transactions of the American Mathematical Society*, 357(1):89–128, 2005.
[37] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, 2008.
[38] 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.
[39] 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.
[40] 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.
[41] U. Kohlenbach and A. Nicolae. A proof-theoretic bound extraction theorem for CAT(κ)-spaces. *Studia Logica*, 105:611–624, 2017.
[42] 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.
[43] U. Kohlenbach and N. Pischke. Proof Theory and Nonsmooth Analysis. *Philosophical Transactions of the Royal Society A*, 381:20220015, 21pp., 2023.
[44] 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.
[45] G. Kreisel. On the Interpretation of Non-Finitist Proofs–Part I. *The Journal of Symbolic Logic*, 16(4):241–267, 1951.
[46] 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.

[47] G. Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting, editor, *Constructivity in Mathematics*, pages 101–128. North-Holland, Amsterdam, 1959.
[48] S. Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Mathematical Journal*, 2:35–47, 1951.
[49] L. Leuştean. Proof Mining in $\mathbb{R}$-trees and Hyperbolic Spaces. *Electronic Notes in Theoretical Computer Science*, 165:95–106, 2006.
[50] L. Leuştean. An application of proof mining to nonlinear iterations. *Annals of Pure and Applied Logic*, 165(9):1484–1500, 2014.
[51] A. Macintyre. The mathematical significance of proof theory. *Philosophical Transactions of the Royal Society A*, 363:2419–2435, 2005.
[52] A. Macintyre. The impact of Gödel's incompleteness theorems on mathematics. In M. Baaz, C.H. Papadimitriou, H.W. Putnam, D.S. Scott, and jr. C.L. Harper, editors, *Kurt Gödel and the foundations of mathematics: Horizons of truth*, pages 3–25. Cambridge University Press, 2011.
[53] 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.
[54] 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.
[55] L. Paunescu and A. Sipoş. A proof-theoretic metatheorem for tracial von Neumann algebras. *Mathematical Logic Quarterly*, 69(1):63–76, 2023.
[56] 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.
[57] N. Pischke. Generalized Fejér monotone sequences and their finitary content. *Optimization*, 2024. To appear, doi:10.1080/02331934.2024.2390114.
[58] N. Pischke. Logical Metatheorems for Accretive and (Generalized) Monotone Set-Valued Operators. *Journal of Mathematical Logic*, 24(2), 2024. 2350008, 59pp.
[59] N. Pischke. *Proof-Theoretical Aspects of Nonlinear and Set-Valued Analysis*. PhD thesis, TU Darmstadt, 2024. Thesis available at `https://tuprints.ulb.tu-darmstadt.de/26584/`.
[60] N. Pischke and U. Kohlenbach. Effective rates for iterations involving Bregman strongly nonexpansive operators. 2024. Submitted manuscript available at `https://sites.google.com/view/nicholaspischke/notes-and-papers`.
[61] 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.
[62] R.T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497–510, 1966.
[63] R.T. Rockafellar. Level sets and continuity of conjugate convex functions. *Transactions of the American Mathematical Society*, 123(1):46–63, 1966.
[64] R.T. Rockafellar. *Convex analysis*. Princeton university press, 1970.
[65] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Berlin, Heidelberg, 1998.
[66] W. Rudin. *Functional Analysis*, volume 8 of *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York, 1991.
[67] H.H. Schaefer and M.P. Wolff. *Topological Vector Spaces*, volume 8 of *Graduate Texts in Mathematics*. Springer New York, 1999.
[68] Y. Shehu. Convergence Results of Forward-Backward Algorithms for Sum of Monotone Operators in Banach Spaces. *Results in Mathematics*, 74:Article number 138, 2019.
[69] A. Sipoş. Proof mining in L^p spaces. *The Journal of Symbolic Logic*, 84(4):1612–1629, 2019.
[70] 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, Proceedings of Symposia in Pure Mathematics*, volume 5, pages 1–27. AMS, Providence, 1962.
[71] W. Takahashi. *Nonlinear Functional Analysis*. Yokohama Publishers, Yokohama, Japan, 2000.
[72] A.S. Troelstra, editor. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer, Berlin, 1973.
[73] C. Zălinescu. *Convex analysis in general vector spaces*. World scientific, 2002.
[74] H. Zegeye. Strong convergence theorems for maximal monotone mappings in Banach spaces. *Journal of Mathematical Analysis and Applications*, 343(2):663–671, 2008.