Iterative Regularization Methods for the Solution of the Split Feasibility Problem in Banach Spaces

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultäten
der Universität des Saarlandes

von

Frank Schöpfer

Saarbrücken
2007

Tag des Kolloquiums:	06. Juli 2007
Dekan:	Univ.-Prof. Dr. Th. Herfet
Mitglieder des Prüfungsausschusses:	
Vorsitzender:	Univ.-Prof. Dr. S. Rjasanow
Berichterstatter:	Univ.-Prof. Dr. A.K. Louis
	Univ.-Prof. Dr. Th. Schuster
akademischer Mitarbeiter:	Dr. A. Lakhal

Kurze Zusammenfassung

Gegenstand unserer Arbeit ist die Entwicklung von Iterationsverfahren zur Lösung des *split feasibility problem* (SFP) in Banachräumen und deren Untersuchung hinsichtlich Stabilität und regularisierender Eigenschaften. Das SFP besteht darin, einen gemeinsamen Punkt im Schnitt endlich vieler abgeschlossener konvexer Mengen zu finden, wobei einige der Mengen dadurch gegeben sind, dass Zwangsbedingungen im Bild eines linearen Operators auferlegt sind. Das SFP lässt sich prinzipiell dadurch lösen, dass zyklisch auf die einzelnen Mengen projeziert wird. In den Anwendungen sind solche Projektionsverfahren effizient, wenn die Projektionen auf die einzelnen Mengen relativ einfach zu berechnen sind. Wenn die Mengen jedoch durch Zwangsbedingungen im Bild eines linearen Operators gegeben sind, dann ist es i.A. zu schwierig oder zu aufwendig, in jedem Iterationsschritt auf diese Mengen zu projezieren. In endlichdimensionalen euklidischen Räumen schlug Byrne den CQ Algorithmus zur Lösung des SFP vor, bei dem nicht direkt auf solche Mengen projeziert wird, sondern Gradienten geeigneter Funktionale verwendet werden.

Zur Lösung des SFP in Banachräumen verallgemeinern wir diesen Algorithmus mittels Dualitätsabbildungen, metrischer und Bregman Projektionen. Dazu stellen wir die theoretischen Grundlagen zur Verfügung und ergänzen diese durch weitere Ergebnisse. Wir zeigen die Konvergenz der resultierenden Verfahren und untersuchen sie hinsichtlich der Verwendung approximativer Daten sowie ihre regularisierenden Eigenschaften mit Hilfe eines Diskrepanzprinzips. Insbesondere beschäftigen wir uns auch mit der Berechnung von Projektionen auf affine Unterräume, die durch den Nullraum oder das Bild eines linearen Operators gegeben sind. Dazu dient das gleiche Iterationsschema wie beim SFP und darauf aufbauend schlagen wir auch entsprechend verallgemeinerte sequentielle Unterraum und konjugierte Gradienten Verfahren vor.

Abstract

We develop iterative methods for the solution of the *split feasibility problem* (SFP) in Banach spaces and analyze stability and regularizing properties. The SFP consists in finding a common point in the intersection of finitely many closed convex sets, whereby some of the sets arise by imposing constraints in the range of a linear operator. In principle the SFP can be solved by cyclically projecting onto the individual sets. In applications such projection algorithms are efficient if the projections onto the individual sets are relatively simple to calculate. If the sets arise by imposing constraints in the range of a linear operator then it is in general too difficult or too costly to project onto these sets in each iterative step. In finite-dimensional euclidean spaces BYRNE suggested the CQ algorithm for the solution of the SFP, which avoids projecting directly onto such sets by using gradients of suitable functionals. To solve the SFP in Banach spaces we generalize this algorithm via duality mappings, metric projections and Bregman projections. We provide the necessary theoretical framework and extend it by some further contributions. We prove convergence of the resulting methods, show how approximate data may be used, and analyze their regularizing properties by applying a discrepancy principle. Especially we are also concerned with the computation of projections onto affine subspaces that are given via the nullspace or the range of a linear operator. To this end we can use the same iterative scheme as for the SFP and we also propose generalized sequential subspace and conjugate gradient methods.

Zusammenfassung

Gegenstand unserer Arbeit ist die Entwicklung von Iterationsverfahren zur Lösung des *split feasibility problem* (SFP) in Banachräumen und deren Untersuchung hinsichtlich Stabilität und regularisierender Eigenschaften. Das von Censor und Elfving [20] so gennante SFP ist ein Spezialfall des *convex feasibility problem* (CFP). Das CFP besteht darin, einen gemeinsamen Punkt im Schnitt endlich vieler abgeschlossener konvexer Mengen zu finden. Beim SFP werden dabei Mengen, die dadurch gegeben sind, dass Zwangsbedingungen im Bild eines linearen Operators auferlegt sind, gesondert behandelt. Ein klassisches Lösungsverfahren für das CFP in Hilberträumen ist die Methode der zyklischen Orthogonalprojektionen [30], bei der iterativ eine konvergente Folge durch zyklisches Projezieren auf die einzelnen Mengen erzeugt wird. Bregman [11] zeigte 1967, dass auch allgemeinerer Projektionen verwendet werden können, welche durch konvexe Funktionen erzeugt werden. Mit Hilfe solcher *Bregman Projektionen* konnten Alber und Butnariu [1] das CFP in Banachräumen lösen.

In den Anwendungen sind solche Projektionsverfahren effizient, wenn die Projektionen auf die einzelnen Mengen relativ einfach zu berechnen sind. Wenn die Mengen jedoch dadurch gegeben sind, dass Zwangsbedingungen im Bild eines linearen Operators auferlegt sind, dann ist es i.A. zu schwierig oder zu aufwendig, in jedem Iterationsschritt auf diese Mengen zu projezieren. Byrne [17] schlug 2002 den *CQ Algorithmus* vor, um einen Punkt x in einer abgeschlossenen konvexen Menge $C \subset \mathbb{R}^N$ zu finden, so dass $Ax \in Q$ liegt für eine abgeschlossene konvexe Menge $Q \subset \mathbb{R}^M$ und eine $M \times N$-matrix A. Der CQ Algorithmus hat die iterative Form

$$x_{n+1} = P_C\Big(x_n - \mu A^*\big(Ax_n - P_Q(Ax_n)\big)\Big),$$

wobei $\mu > 0$ ein Parameter ist und P_C, P_Q die Orthogonalprojektionen auf die entsprechenden Mengen bezeichnen. Der Vorteil liegt darin, dass man es vermeidet, direkt auf die Menge $\{x \in \mathbb{R}^N \mid Ax \in Q\}$ zu projezieren, indem man den Gradienten des Funktionals $f(x) = \frac{1}{2}\|Ax - P_Q(Ax)\|^2$ und damit auch nur die Projektion auf Q verwendet.

Wir verallgemeinern dieses Verfahren zur Lösung des SFP in Banachräumen mittels *Dualitätsabbildungen*, metrischer und Bregman Projektionen. Den dazu benötigten theoretischen Rahmen stellen wir im ersten Kapitel zur Verfügung. Dabei werden die Banachräume, in denen die Verfahren konvergieren, durch ihre geometrischen Eigenschaften charakterisiert. Wir geben einen kurzen Überblick über Dualitätsabbildungen, da sie das Hauptwerkzeug unserer Arbeit sind. Wir verwenden auch positive Dualitätsabbildungen in Banachverbänden, um lineare Ungleichungen „$Ax \leq y$" zu behandeln. Die Breg-

man Projektionen, die wir hier verwenden, werden durch Potenzen der Norm der zugrundeliegenden Räume induziert. Wir ergänzen die bestehende Theorie dieser speziellen Bregman Projektionen durch weitere nützliche Eigenschaften und klären den Zusammenhang mit den metrischen Projektionen. Dabei beweisen wir auch einen Zerlegungssatz der Form

$$X = P_U(X) \oplus J^*\Big(\Pi^*_{U^\perp}\big(J(x)\big)\Big),$$

wobei $U \subset X$ ein abgeschlossener Unterraum eines reflexiven, strikt konvexen und glatten Banachraums ist, P_U die metrische Projektion auf U, $\Pi^*_{U^\perp}$ eine Bregman Projektion auf den Annihilator $U^\perp$ von U bezeichnet und J, J^* Dualitätsabbildungen in X bzw. dem Dualraum X^* sind. Ein Resultat über gleichmäßige Stetigkeit der Projektionen erhalten wir bezüglich *beschränkter Hausdorff Konvergenz* konvexer Mengen (dieser Konvergenzbegriff basiert auf lokalen Versionen der Hausdorff Metrik und wurde auch von Penot [39] im Zusammenhang mit metrischen Projektionen verwendet).

Die Verfahren zur Lösung des SFP behandeln wir im zweiten Kapitel. Wir beweisen ihre (schwache) Konvergenz und untersuchen mit Hilfe eines *Diskrepanzprinzips* ihre regularisierenden Eigenschaften und wie auch Approximationen der Daten (rechte Seiten von Operatorgleichungen, die Operatoren selbst, die konvexen Mengen) verwendet werden können.

Insbesondere beschäftigen wir uns auch mit der Berechnung von Projektionen auf affine Unterräume, die durch den Nullraum oder das Bild eines linearen Operators gegeben sind. Dazu dient das gleiche Iterationsschema wie beim SFP. Hierbei verbessern und ergänzen wir unter Verwendung des oben erwähnten Zerlegungssatzes unsere Arbeit in [46], wo wir schon die starke Konvergenz dieses Verfahrens zeigen konnten.

Schließlich gehen wir noch auf Möglichkeiten ein, die Verfahren effizient zu implementieren. Dazu gehören passende *line search* Verfahren, um die beteiligten Parameter optimal zu bestimmen, sowie entsprechend verallgemeinerte *sequentielle Unterraum* und *konjugierte Gradienten* Verfahren, um im Fall exakter Daten Projektionen auf affine Unterräume zu berechnen.

Contents

Introduction

Many problems in mathematics, natural sciences and engineering can be formulated as the *convex feasibility problem* (CFP), which consists in finding a common point in the intersection of finitely many closed convex sets. A classical procedure for the solution of the CFP in Hilbert spaces is the method of cyclic orthogonal projections [30], where a convergent sequence is iteratively generated by projecting cyclically onto the individual sets. In 1967 BREGMAN [11] extended this method to non-orthogonal projections that are induced by convex functions. ALBER and BUTNARIU [1] used these nowadays called *Bregman projections* to solve the CFP in Banach spaces.
In applications such projection algorithms are efficient if the projections onto the individual sets are relatively simple to calculate. If the sets arise by imposing constraints in the range of a linear operator then it is in general too difficult or too costly to project onto these sets in each iterative step. In 2002 BYRNE [17] suggested the *CQ algorithm* to solve the problem of finding a point x in a closed convex set $C \subset \mathbb{R}^N$ such that $Ax \in Q$ for a closed convex set $Q \subset \mathbb{R}^M$ and an $M \times N$-matrix A. The CQ algorithm has the iterative form

$$x_{n+1} = P_C\Big(x_n - \mu A^*\big(Ax_n - P_Q(Ax_n)\big)\Big)\,,$$

whereby $\mu > 0$ is a parameter and P_C, P_Q denote the orthogonal projections onto the respective sets. The special case of $Q = \{y\}$ being a singleton is also known as the *projected Landweber method*. The advantage is that the difficulty of directly projecting onto the set $\{x \in \mathbb{R}^N \mid Ax \in Q\}$ is avoided by using the gradient of the functional $f(x) = \frac{1}{2}\|Ax - P_Q(Ax)\|^2$ and thus only the projection onto Q is involved. Recently CENSOR ET AL. [21] generalized this procedure to the case where several constraints are imposed in the domain as well as in the range of the linear operator. This special case of the CFP was also called the *split feasibility problem* (SFP) by CENSOR and ELFVING [20] and its solution in Banach spaces by means of a CQ algorithm has not been analyzed yet.
Often the available data (right hand sides of operator equations, the opera-

tors themselves, the convex sets) is given only approximately or contaminated with noise or it is even preferable to use approximate data. Therefore it is important not only to have solution methods but also to analyze their stability and regularizing properties and to modify them if necessary. Some results in this direction were given by Eicke [27], who analyzed the regularizing properties of the projected Landweber method in Hilbert spaces with respect to perturbed right hand side $\{y\}$, or Zhao and Yang [49], who studied a relaxed version of the CQ algorithm for the use of approximately given convex sets.

We are concerned with the solution of the SFP in Banach spaces and the special case of computing projections onto affine subspaces that are given via a linear operator. To this end we generalize the CQ algorithm via *duality mappings*, metric projections and Bregman projections induced by powers of the norm of the underlying Banach spaces (also called *generalized projections* by Alber [3]). We show how approximate data may be used in the resulting methods and analyze regularizing properties and stability with respect to all given data by applying a *discrepancy principle*.

In chapter 1 we provide the theoretical framework necessary to develop and discuss the methods in Banach spaces. The spaces in which the methods work are characterized by their geometrical properties dealt with in section 1.1, where we also give a survey of duality mappings since they are the main tool throughout this thesis. Positive duality mappings in Banach lattices (subsec. 1.1.8) are used to handle linear inequalities. In section 1.2 we are concerned with the Bregman distances and Bregman projections we use here. We make some contributions to the existing theory and clarify the relationship between these Bregman projections and the metric projection. Especially we prove a decomposition theorem of the form

$$X = P_U(X) \oplus J^*\Big(\Pi^*_{U^\perp}\big(J(x)\big)\Big) ,$$

whereby $U \subset X$ is a closed subspace of a reflexive, smooth and strictly convex Banach space, P_U is the metric projection onto U, $\Pi^*_{U^\perp}$ is a Bregman projection onto the annihilator $U^\perp$ of U and J, J^* denote duality mappings of X resp. the dual space X^*. In the last section we prove a uniform continuity result with respect to *bounded Hausdorff convergence* of convex sets (this notion of convergence is based on local versions of the Hausdorff metric and has also been used by Penot [39] in the context of metric projections).

In chapter 2 we are concerned with the iteration methods for the solution of the SFP and the computation of projections onto affine subspaces. In the first section we examine the operators that are used in the iterative process to handle different kinds of constraints. The operators related to constraints in the range of a linear operator depend on a positive parameter which in general has to be chosen a posteriori. The question of how to choose these parameters is settled in the following section. Thereby we make use of the characteristic inequalities of uniformly smooth Banach spaces [48] of subsection 1.1.7. The iteration methods for the SFP are analyzed in section 2.3. In section 2.4 we concentrate on the special case of computing projections onto affine subspaces, which can be used to compute minimum norm solutions of operator equations and best approximations in the range of a linear operator. Hereby we improve and complement our work in [46] by using the above mentioned decomposition theorem. The last two sections deal with possibilities to efficiently implement the methods. We show that the choice of parameters can be replaced by *line searches* and propose generalized *conjugate gradient* and *sequential subspace methods* to compute projections onto affine subspaces.

I am deeply grateful to Univ.-Prof. Dr. Alfred K. Louis for sparking my interest in this topic and for encouraging me whenever I did not know how to continue.
Special thanks go to Univ.-Prof. Dr. Thomas Schuster for advising and helping me in so many ways.
I would also like to thank my friends who accepted my frequent absence, and especially my parents for supporting me throughout the years.
And I thank my beloved wife Birgit for being with me.

Saarbrücken, im März 2007 Frank Schöpfer

1

Theoretical Framework

In this chapter we provide the theoretical framework in which the problems we deal with are formulated and which is necessary to develop and discuss the methods we propose for their solution.

1.1 Geometry of Banach Spaces and Duality Mappings

At first we recall some basic definitions and properties of Banach spaces which can be found in [25, 33, 47] or any other book about Banach space theory or functional analysis. We mainly concentrate on some convergence principles that we will use frequently. Then we will give a short survey of geometrical aspects of Banach spaces and (positive) duality mappings. A detailed introduction to this topic can be found in [22]. We only give some of the proofs, when we think they are not too involved and to make ourselves familiar with duality mappings and the techniques to analyze the behaviour of the iteration methods.

1.1.1 Preliminaries

Let us make some conventions: We shortly write "iff" for "if and only if". $p, q \in [1, \infty]$ are always supposed to be conjugate exponents so that

$$\frac{1}{p} + \frac{1}{q} = 1\,,$$

whereby we set $\frac{1}{\infty} := 0$. For convenience we also mention some equalities which we will use more frequently for $p \in (1, \infty)$:

$$q = \frac{p}{p-1} \quad , \quad p\,q = p + q \quad , \quad (p-1)(q-1) = 1\,.$$

Further for extended real valued a, b[1] we write

[1] i.e. scalars, sequences or functions with values in $[-\infty, +\infty]$

$$a \vee b = \max\{a, b\} \quad \text{and} \quad a \wedge b = \min\{a, b\}\,,$$

which is to be understood componentwise in case of sequences and pointwise in case of functions.

1.1.2 Basic Definitions and Properties of Banach Spaces

Throughout this thesis X and Y are *real Banach spaces*, i.e. normed vector spaces over the real field such that every Cauchy sequence is convergent. By $\|.\|_X$ and $\|.\|_Y$ we denote their respective norms and omit indices whenever it becomes clear from the context which space is meant. The *dual space* X^* of X is the space of continuous linear functionals on X. It becomes itself a Banach space if we endow it with the norm $\|.\|_{X^*}$ defined by

$$\|x^*\|_{X^*} := \sup\left\{ |\langle x^* \,|\, x\rangle| \,\middle|\, x \in X\,,\, \|x\|_X \le 1 \right\} \quad , \quad x^* \in X^*\,, \tag{1.1}$$

whereby we write $\langle x^* \,|\, x\rangle = x^*(x)$ for the application of an element $x^* \in X^*$ on an element $x \in X$ to emphasize that there are some similarities with scalar-products in real Hilbert spaces. Obvious are bilinearity and continuity in both arguments and by looking at

$$\frac{|\langle x^* \,|\, x\rangle|}{\|x\|_X} = \left| \left\langle x^* \,\middle|\, \frac{x}{\|x\|_X} \right\rangle \right| \le \|x^*\|_{X^*}$$

we see that the following generalization of the Cauchy-Schwarz inequality holds

$$|\langle x^* \,|\, x\rangle| \le \|x^*\|_{X^*}\, \|x\|_X \quad \text{for all} \quad x^* \in X^*, x \in X\,. \tag{1.2}$$

The canonical embedding ι_X of X in its bidual $X^{**} = (X^*)^*$ can then be written in the form

$$\iota_X : X \longrightarrow X^{**} \quad , \quad \langle x \,|\, x^*\rangle = \iota_X(x)(x^*) := \langle x^* \,|\, x\rangle \quad , \quad x \in X\,,\, x^* \in X^*\,.$$

This mapping is linear and isometric. If it is also surjective and thus an isometric isomorphism, the Banach space X is called *reflexive*. In this case we can identify X^{**} with X. It is a fact that X is reflexive iff X^* is reflexive.

Example 1.1.

(a) Every finite-dimensional normed vector space is a reflexive Banach space.
(b) Every Hilbert space X is a reflexive Banach space with the identification $X^* = X$ and $\langle . \,|\, . \rangle$ is just the scalar-product.
(c) The L_p-spaces are Banach spaces with

$$\|x\|_p := \begin{cases} \left(\sum_n |x_n|^p \right)^{\frac{1}{p}} & , \quad p < \infty \\ \sup_n \{|x_n|\} & , \quad p = \infty \end{cases}$$

and
$$\langle x \mid y \rangle = \sum_n x_n y_n \quad , \quad x = (x_n)_n \in L_p \, , \, y = (y_n)_n \in L_q$$
in case of sequence spaces, respectively
$$\|x\|_p := \begin{cases} \left(\int_{\omega \in \Omega} |x(\omega)|^p \, \mathrm{d}\omega \right)^{\frac{1}{p}} & , \quad p < \infty \\ \operatorname*{ess\,sup}_{\omega \in \Omega} \{|x(\omega)|\} & , \quad p = \infty \end{cases}$$
and
$$\langle x \mid y \rangle = \int_{\omega \in \Omega} x(\omega) y(\omega) \, \mathrm{d}\omega \quad , \quad x \in L_p \, , \, y \in L_q \, .$$
in case of function spaces. For $p \in (1, \infty)$ they are reflexive and the dual spaces are $L_p^* = L_q$. In finite dimensions the dual of $(\mathbb{R}^n, \|.\|_\infty)$ is $(\mathbb{R}^n, \|.\|_1)$ and vice versa.

(d) Of course there are many other classes of Banach spaces like spaces of continuous or differentiable functions, Sobolev spaces, Orlicz spaces [42],...

Besides convergence in norm, which is often referred to as *strong convergence*, we will make use of the concept of weak convergence. A sequence $(x_n)_n$ in a Banach space X is called *weakly convergent*, if there is an $x \in X$ such that
$$\lim_{n \to \infty} \langle x^* \mid x_n \rangle = \langle x^* \mid x \rangle \quad \text{for all} \quad x^* \in X^* \, . \tag{1.3}$$
Since in a reflexive space $X^{**} = X$, a sequence $(x_n^*)_n$ in the dual X^* is weakly convergent, if there is an $x^* \in X$ such that
$$\lim_{n \to \infty} \langle x_n^* \mid x \rangle = \langle x^* \mid x \rangle \quad \text{for all} \quad x \in X \, . \tag{1.4}$$
By $|\langle x^* \mid x - x_n \rangle| \leq \|x^*\|_{X^*} \, \|x - x_n\|_X$ we see that strong convergence implies weak convergence. The converse is true iff X is finite-dimensional. As a consequence of the Hahn-Banach theorem[2] the above weak limit points are unique and a convex subset $C \subset X$ is (norm-)closed iff it is weakly closed. Moreover a closed and convex subset $\emptyset \neq C \subsetneq X$ coincides with the intersection of all halfspaces containing C, i.e.
$$C = \bigcap_{C \subset H_{\leq}(u^*, \alpha)} H_{\leq}(u^*, \alpha) \, , \tag{1.5}$$
whereby for $0 \neq u^* \in X^*$ and $\alpha \in \mathbb{R}$ we define the *hyperplane*

---

[2] There are several versions of this theorem. We just mention the following two:
(a) For every $x \in X$ there exists an $x^* \in X^*$ with $\|x^*\| = 1$ and $\langle x^* \mid x \rangle = \|x\|$.
(b) If $C \subset X$ is a closed convex subset and $x_0 \notin C$ then there exist $x^* \in X^*$ and $\alpha \in \mathbb{R}$ such that $\langle x^* \mid x \rangle \leq \alpha < \langle x^* \mid x_0 \rangle$ for all $x \in C$.

$$H(u^*, \alpha) := \{x \in X \mid \langle u^* \mid x\rangle = \alpha\} \tag{1.6}$$

and the *halfspace*

$$H_{\leq}(u^*, \alpha) := \{x \in X \mid \langle u^* \mid x\rangle \leq \alpha\} \tag{1.7}$$

and analogously for $\geq$, $<$ and $>$. A hyperplane H is called a *supporting hyperplane* of C (at $x \in C$) if H has non-empty intersection with C (at x) and C is contained in one of the halfspaces $H_{\leq}$ or $H_{\geq}$. In this case $H_{\leq}$ (resp. $H_{\geq}$) is called a *supporting halfspace* of C (at $x \in C$). By $\mathcal{C}(X)$ we denote the set of all non-empty, closed and convex subsets of X. For a subspace $U \subset X$ the *annihilator* of U is the set

$$U^{\perp} := \{x^* \in X^* \mid \langle x^* \mid u\rangle = 0 \quad \text{for every} \quad u \in U\}\,.$$

It is a closed subspace of X^* and in a reflexive Banach space X we have $U^{\perp\perp} = \left(U^{\perp}\right)^{\perp} = \overline{U}$, the closure of U.

Another important characterization of reflexive spaces is the following.

Proposition 1.2. *A Banach space X is reflexive iff the unit ball of X is weakly compact iff every bounded sequence has a weakly convergent subsequence.*

The next proposition shows that in reflexive spaces we can solve a special kind of convex optimization problem, namely

$$\min f(x) \quad \text{s.t.} \quad x \in C \tag{1.8}$$

for a $C \in \mathcal{C}(X)$, whereby $f(x) = \|x-y\|$ for an arbitrary $y \in X$ (or equivalently $f(x) = \frac{1}{p}\|x-y\|^p$ for any $p > 1$).

Proposition 1.3. *In a reflexive Banach space problem (1.8) has at least one solution.*

Proof. There is a sequence $(x_n)_n \in C$ with

$$\lim_{n\to\infty} \|x_n - y\| = m := \inf_{x \in C} \|x - y\|\,.$$

In particular $(x_n)_n$ is bounded and by 1.2 it has a weakly convergent subsequence $(x_{n_k})_k$. Since C is convex and closed and therefore also weakly closed, the weak limit point x_0 of $(x_{n_k})_k$ lies again in C. Hence $\|x_0 - y\| \geq m$ and for all $x^* \in X^*$ with $\|x^*\| \leq 1$ we have

$$|\langle x_0 - y \mid x^*\rangle| = \lim_{k\to\infty} |\langle x_{n_k} - y \mid x^*\rangle| \leq \lim_{k\to\infty} \|x_{n_k} - y\| = m\,.$$

It follows that $\|x_0 - y\| = \sup\left\{|\langle x_0 - y \mid x^*\rangle| \,\middle|\, x^* \in X^*\,,\ \|x^*\| \leq 1\right\} \leq m.$ □

1.1.3 Geometry

We turn to some geometrical aspects of Banach spaces because we need them to characterize those spaces in which we can apply the iteration methods and prove their convergence. The results presented here are taken from [22, 24, 29, 34].

Definition 1.4. *The function* $\delta_X : [0,2] \longrightarrow [0,1]$ *defined by*

$$\delta_X(\epsilon) = \inf\left\{1 - \left\|\frac{1}{2}(x+y)\right\| \,:\, \|x\| = \|y\| = 1\,,\ \|x-y\| \geq \epsilon\right\}$$

is referred to as the modulus of convexity *of* X*.*

The function $\rho_X : [0,\infty) \longrightarrow [0,\infty)$ *defined by*

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

is referred to as the modulus of smoothness *of* X*.*

These functions can be seen as a measure of the degree of convexity resp. smoothness of the norm. They have the following basic properties.

Proposition 1.5.

(a) δ_X *is continuous and nondecreasing with* $\delta_X(0) = 0$*.*
(b) ρ_X *is continuous, convex and nondecreasing with* $\rho_X(0) = 0$ *and* $\rho_X(\tau) \leq \tau$*.*
(c) The function $\tau \mapsto \frac{\rho_X(\tau)}{\tau}$ *is nondecreasing and fulfills* $\frac{\rho_X(\tau)}{\tau} > 0$ *for all* $\tau > 0$*.*
(d) For every Hilbert space H *we have* $\delta_H(\epsilon) = 1 - \sqrt{1 - \left(\frac{\epsilon}{2}\right)^2}$ *and* $\rho_H(\tau) = \sqrt{1+\tau^2} - 1$ *and* $\delta_X(\epsilon) \leq \delta_H(\epsilon)$*,* $\rho_X(\tau) \geq \rho_H(\tau)$ *for arbitrary Banach spaces* X*.*

The classes of Banach spaces we will deal with are:

Definition 1.6. *A Banach space* X *is said to be*

(a) strictly convex*, if* $\|\lambda x + (1-\lambda)y\| < 1$ *for all* $\lambda \in (0,1)$ *and* $x, y \in X$ *with* $x \neq y$ *and* $\|x\| = \|y\| = 1$*, i.e. the boundary of the unit ball contains no line segment,*
(b) smooth*, if for every* $0 \neq x \in X$ *there is a unique* $x^* \in X^*$ *such that* $\|x^*\| = 1$ *and* $\langle x, x^*\rangle = \|x\|$*, i.e. there is a unique supporting hyperplane for the ball* $B_{\|x\|}$ *around the origin with radius* $\|x\|$ *at* x*,*
(c) uniformly convex*, if* $\delta_X(\epsilon) > 0$ *for any* $\epsilon \in (0,2]$*,*
(d) uniformly smooth*, if* $\lim_{\tau\to 0}\frac{\rho_X(\tau)}{\tau} = 0$*.*

Uniform convexity implies strict convexity and uniform smoothness implies smoothness. In finite dimensions the converse is also true. The next proposition shows that convexity and smoothness are dual concepts.

Proposition 1.7.

(a) Let X be reflexive. Then X is strictly convex (resp. smooth) iff X^* is smooth (resp. strictly convex).
(b) If X is uniformly convex then X is reflexive and strictly convex.
(c) If X is uniformly smooth then X is reflexive and smooth.
(d) X is uniformly convex (resp. uniformly smooth) iff X^* is uniformly smooth (resp. uniformly convex).

Example 1.8. L_p-spaces $(1 < p < \infty)$ are known to be both uniformly convex and uniformly smooth and

$$\delta_{L_p}(\epsilon) = \begin{cases} \frac{p-1}{8}\epsilon^2 + o(\epsilon^2) > \frac{p-1}{8}\epsilon^2 & , 1 < p < 2 \\ 1 - \left(1 - \left(\frac{\epsilon}{2}\right)^p\right)^{\frac{1}{p}} > \frac{1}{p}\left(\frac{\epsilon}{2}\right)^p & , p \geq 2 \end{cases} \tag{1.9}$$

$$\rho_{L_p}(\tau) = \begin{cases} (1+\tau^p)^{\frac{1}{p}} - 1 < \frac{1}{p}\tau^p & , 1 < p \leq 2 \\ \frac{p-1}{2}\tau^2 + o(\tau^2) < \frac{p-1}{2}\tau^2 & , p > 2\,, \end{cases} \tag{1.10}$$

whereas the spaces L_1 and L_∞ are neither smooth nor strictly convex.

1.1.4 Duality Mappings

Duality mappings are a very important tool in nonlinear functional analysis, in theory as well as in applications. One reason for this is that they serve as a suitable substitute for the isomorphism $H = H^*$ in Hilber spaces. For more information about duality mappings we refer the reader to the book of Ioana Cioranescu [22] and the references cited therein.

Definition 1.9. *Let $p \in (1, \infty)$ be given. The mapping $J_X^p : X \longrightarrow 2^{X^*}$ defined by*

$$J_X^p(x) = \left\{x^* \in X^* \,\middle|\, \langle x^*|x\rangle = \|x^*\|\,\|x\|\,,\ \|x^*\| = \|x\|^{p-1}\right\} \tag{1.11}$$

is called the duality mapping *of X with gauge function $t \mapsto t^{p-1}$. J_X^2 is also called the* normalized *duality mapping. By j_X^p we denote a single-valued selection of J_X^p, i.e $j_X^p(x) \in J_X^p(x)$ for every $x \in X$.*

These mappings have the following basic properties.

Proposition 1.10.

(a) For every $x \in X$ the set $J_X^p(x)$ is not empty and convex.
(b) J_X^p is homogenous *of degree $p-1$, i.e.*

$$J_X^p(\lambda x) = |\lambda|^{p-1}\,\mathrm{sgn}(\lambda) J_X^p(x) \quad \textit{for all} \quad x \in X\,,\ \lambda \in \mathbb{R}\,.$$

(c) J_X^p *is* monotone, *i.e.*

$$\langle x^* - y^* \,|\, x - y \rangle \geq 0 \quad \textit{for all} \quad x, y \in X\,,\; x^* \in J_X^p(x)\,,\; y^* \in J_X^p(y)\,.$$

(d) If $J_{X^*}^q$ *is the duality mapping of* X^* *with gauge function* $t \mapsto t^{q-1}$ *then* $x^* \in J_X^p(x)$ *whenever* $x \in J_{X^*}^q(x^*)$*; if* X *is reflexive then* $x^* \in J_X^p(x)$ *iff* $x \in J_{X^*}^q(x^*)$.

(e) If J_X^r *is the duality mapping of* X *with gauge function* $t \mapsto t^{r-1}$ *then*

$$J_X^r(x) = \|x\|^{r-p} J_X^p(x)\,.$$

(So it suffices to know J_X^p *for one value of* p*.)*

(f) The normalized duality mapping is linear iff X *is a Hilbert space and in this case it is just the identity mapping.*

Proof. Obviously $J_X^p(0) = 0$. For $x \neq 0$ by the Hahn-Banach theorem we can find $x^* \in X^*$ with $\|x^*\| = 1$ and $\langle x^* \,|\, x \rangle = \|x\|$. The element $\tilde{x}^* := \|x\|^{p-1} x^*$ then lies in $J_X^p(x)$ ant thus $J_X^p(x)$ is not empty.
For $x^*, y^* \in J_X^p(x)$, $\lambda \in (0,1)$ and $z^* = \lambda x^* + (1-\lambda) y^*$ we have

$$\langle z^* \,|\, x \rangle = \lambda \langle x^* \,|\, x \rangle + (1-\lambda)\langle y^* \,|\, x \rangle = \|x\|^p$$

and therefore by the triangle-inequality of the norm

$$\|x\|^{p-1} = \left\langle z^* \,\middle|\, \frac{x}{\|x\|} \right\rangle \leq \|z^*\| \leq \lambda \|x^*\| + (1-\lambda)\|y^*\| = \|x\|^{p-1}\,.$$

Hence $z^* \in J_X^p(x)$.
For $x^* \in J_X^p(x)$ and $\lambda > 0$ we see that

$$\langle -x^* \,|\, -x \rangle = \langle x^* \,|\, x \rangle = \|x\|^p = \|-x\|^p\,,$$

$$\|-x^*\| = \|x^*\| = \|x\|^{p-1} = \|-x\|^{p-1}\,,$$

$$\langle \lambda^{p-1} x^* \,|\, \lambda x \rangle = \lambda^p \langle x^* \,|\, x \rangle = \lambda^p \|x\|^p = \|\lambda x\|^p\,,$$

$$\|\lambda^{p-1} x^*\| = \lambda^{p-1}\|x^*\| = \lambda^{p-1}\|x\|^{p-1} = \|\lambda x\|^{p-1}\,.$$

and thus $-J_X^p(x) \subset J_X^p(-x)$ and $\lambda^{p-1} J_X^p(x) \subset J_X^p(\lambda x)$. The inverse inclusions can be proven analogously.
For all $x, y \in X$ and $x^* \in J_X^p(x)\,,\; y^* \in J_X^p(y)$ we have

$$\begin{aligned}\langle x^* - y^* \,|\, x - y \rangle &= \langle x^* \,|\, x \rangle + \langle y^* \,|\, y \rangle - \langle x^* \,|\, y \rangle - \langle y^* \,|\, x \rangle \\ &\geq \|x\|^p + \|y\|^p - \|x\|^{p-1}\|y\| - \|y\|^{p-1}\|x\| \\ &= \left(\|x\|^{p-1} - \|y\|^{p-1}\right)(\|x\| - \|y\|) \geq 0\,.\end{aligned}$$

Let $J_{X^*}^q$ be the duality mapping of X^* with gauge function $t \mapsto t^{q-1}$ and $x \in J_{X^*}^q(x^*)$ be given. By

$$\|x^*\|^q = \langle x \,|\, x^* \rangle = \langle x^* \,|\, x \rangle \leq \|x^*\| \, \|x\| = \|x^*\| \, \|x^*\|^{q-1}$$

it follows that $\|x\| = \|x^*\|^{q-1} \Leftrightarrow \|x\|^{p-1} = \|x^*\|$ and

$$\langle x^* \,|\, x \rangle = \|x^*\|^q = \|x\|^{(p-1)q} = \|x\|^p \,.$$

Hence $x^* \in J^p_X(x)$.
The relation in (e) is straightforward. Now let X be a Hilbert space. Then

$$\|J^2_X(x) - x\|^2 = \|J^2_X(x)\|^2 + \|x\|^2 - 2\langle J^2_X(x) \,|\, x \rangle = \|x\|^2 + \|x\|^2 - 2\|x\|^2 = 0$$

and thus J^2_X is the identity mapping. Conversely suppose that for a real Banach space X the normalized duality mapping is linear. We show that in this case the parallelogram equality holds which characterizes Hilbert spaces. If J^2_X is linear then $x^* \pm y^* \in J^2_X(x \pm y)$ for all $x, y \in X$ and $x^* \in J^p_X(x)$, $y^* \in J^p_X(y)$ and therefore

$$\|x \pm y\|^2 = \langle x^* \pm y^* \,|\, x \pm y \rangle = \|x\|^2 + \|y\|^2 \pm \langle x^* \,|\, y \rangle \pm \langle y^* \,|\, x \rangle \,.$$

Adding these two equalities yields

$$\|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) \,.$$

□

In the following we will frequently use J, J_X, J^p and J^p_X for the duality mappings in X and J^*, J^*_X, J^q and J_{X^*} for the duality mappings in the dual X^* depending on which index is to be emphasized or to facilitate notation. By checking (1.9) we see that the following mappings are duality mappings.

Example 1.11.

(a) In L_p-spaces $(1 < p < \infty)$ we have

$$J^p(x) = |x|^{p-1} \operatorname{sgn}(x) \,,$$

which is to be understood componentwise resp. pointwise ($\operatorname{sgn}(x) := \frac{x}{|x|}$ for $0 \neq x \in \mathbb{R}$ and $\operatorname{sgn}(0) := 0$).

(b) A single-valued selection for the normalized duality mapping in $(\mathbb{R}^n, \|.\|_\infty)$ is given by

$$j(x) = \|x\|_\infty \operatorname{sgn}(x_k)(\delta_{l,k})^n_{l=1} \,,$$

whereby k is an index with $|x_k| = \|x\|_\infty$ ($\delta_{l,k} = 1$ for $l = k$ and $\delta_{l,k} = 0$ for $l \neq k$).

(c) If we equip $\mathbb{R}^n$ with the L_1-norm we may choose

$$j(x) = \|x\|_1 \operatorname{sgn}(x) \,.$$

1.1.5 Relationship between Geometry, Duality Mappings und Convex Functionals

Duality mappings are in fact subdifferentials of convex functions. We recall that a function $f : X \longrightarrow \mathbb{R}$ is said to be *subdifferentiable* at a point $x \in X$, if there exists an $x^* \in X^*$, called *subgradient* of f at x, such that

$$f(y) - f(x) \geq \langle x^* \,|\, y - x \rangle \quad \text{for all} \quad y \in X\,. \tag{1.12}$$

By $\partial f(x)$ we denote the set of all subgradients of f at x and the mapping $\partial f : X \longrightarrow 2^{X^*}$ is called the *subdifferential* of f.

It is known that if $f, g : X \longrightarrow \mathbb{R}$ are continuous convex functions then they are subdifferentiable and

$$\partial(f+g)(x) = \partial f(x) + \partial g(x) \quad \text{for all} \quad x \in X\,. \tag{1.13}$$

Proposition 1.12. *Let J_X^p be the duality mapping of X with gauge function $t \mapsto t^{p-1}$ and let $f : X \longrightarrow \mathbb{R}$ be defined by*

$$f(x) = \frac{1}{p}\|x\|^p \quad , \quad x \in X\,.$$

Then $J_X^p = \partial f$.

Proof. For $x^* \in J_X^p(x)$ and all $y \in X$ we have by Young's inequality

$$\frac{1}{q}\|x\|^p + \frac{1}{p}\|y\|^p = \frac{1}{q}\left(\|x\|^{p-1}\right)^q + \frac{1}{p}\|y\|^p \geq \|x\|^{p-1}\|y\| = \|x^*\|\,\|y\| \geq \langle x^* \,|\, y \rangle\,.$$

Since $\frac{1}{q}\|x\|^p = \left(1 - \frac{1}{p}\right)\|x\|^p = \langle x^* \,|\, x \rangle - \frac{1}{p}\|x\|^p$ we conclude that

$$\frac{1}{p}\|y\|^p - \frac{1}{p}\|x\|^p \geq \langle x^* \,|\, y - x \rangle \tag{1.14}$$

and therefore $x^* \in \partial f(x)$. Now let $x^* \in \partial f(x)$ be given. By the above inequality (1.14) we get for all $1 \neq t > 0$ and $y = tx$

$$\frac{(t^p - 1)}{p}\|x\|^p \geq \langle x^* \,|\, (t-1)x \rangle\,.$$

We divide by $t - 1$ and obtain

$$\frac{\frac{1}{p}t^p - \frac{1}{p}}{t-1}\|x\|^p \begin{cases} \leq \langle x^* \,|\, x \rangle & , \quad t < 1 \\ \geq \langle x^* \,|\, x \rangle & , \quad t > 1 \end{cases}.$$

By letting $t \to 1$ we arrive at $\|x\|^p = \langle x^* \,|\, x \rangle$. Finally for $y \in X$ with $\|y\| = \|x\|$ inequality (1.14) yields $\langle x^* \,|\, y \rangle \leq \langle x^* \,|\, x \rangle$ and therefore

$$\|x^*\|\,\|x\| = \sup_{\|y\| = \|x\|} |\langle x^* \,|\, y \rangle| \leq \langle x^* \,|\, x \rangle\,.$$

Hence $\|x^*\|\,\|x\| = \langle x^* \,|\, x \rangle = \|x\|^p$ which also gives $\|x^*\| = \|x\|^{p-1}$. $\square$

In order to clarify even more the tight relationship between geometry, duality mappings and the above defined convex functionals, we need a little more notation.

Definition 1.13. *A function $f : X \longrightarrow \mathbb{R}$ is called*

(a) Gâteaux differentiable *at $x \in X$, if there exists an element $f'(x) \in X^*$ such that*

$$\lim_{t \to 0} \frac{f(x+ty) - f(x)}{t} = \langle f'(x) \,|\, y \rangle \quad \textit{for every} \quad y \in X \,,$$

(b) Fréchet differentiable *at $x \in X$, if it is Gâteaux differentiable at x and*

$$\lim_{t \to 0} \sup_{\|y\|=1} \left\| \frac{f(x+ty) - f(x)}{t} - \langle f'(x) \,|\, y \rangle \right\| = 0 \,,$$

(c) uniformly Fréchet differentiable on the unit sphere, *if it is Fréchet differentiable and*

$$\lim_{t \to 0} \sup_{\|y\|=\|x\|=1} \left\| \frac{f(x+ty) - f(x)}{t} - \langle f'(x) \,|\, y \rangle \right\| = 0 \,,$$

(d) uniformly convex, *if it is convex and*

$$\inf_{\substack{\|x\|=1 \\ \|y-x\| \geq \epsilon}} \left\{ f(y) + f(x) - 2f\left(\frac{x+y}{2}\right) \right\} > 0 \quad \textit{for all} \quad \epsilon > 0 \,.$$

It is not difficult to see that a continuous convex function $f : X \longrightarrow \mathbb{R}$ is Gâteaux differentiable at $x \in X$ iff it has a unique subgradient at x; and in this case $f'(x) = \partial f(x)$.

Proposition 1.14. *Let J be any duality mapping of X with gauge function $t \mapsto t^{p-1}$ and let $f : X \longrightarrow \mathbb{R}$ be the function $f(x) = \frac{1}{p}\|x\|^p$.*

(a) X is strictly convex iff f is strictly convex iff J is strictly monotone, *i.e.*

$$\langle x^* - y^* \,|\, x - y \rangle > 0 \quad \textit{for all} \quad x \neq y \in X \,,\, x^* \in J(x) \,,\, y^* \in J(y) \,.$$

(b) X is smooth iff f is Gâteaux differentiable iff J is single-valued. In this case $\partial f(x) = f'(x) = J(x)$.
(c) X is uniformly convex iff f is uniformly convex.
(d) X is uniformly smooth iff f is uniformly Fréchet differentiable on the unit sphere iff J is single-valued and uniformly continuous on bounded sets.
(e) X is reflexive, strictly convex and smooth iff J is bijective. In this case the inverse $J^{-1} : X^* \longrightarrow X$ is given by $J^{-1} = J^*$ with J^* being the duality mapping of X^* with gauge function $t \mapsto t^{q-1}$.
(f) If X is reflexive and smooth then J is norm-to-weak-continuous.

1.1.6 Metric Projections

In proposition 1.3 we proved that in reflexive spaces the optimization problem (1.8) has a solution. Now we are concerned with uniqueness of the solution and when we can formulate (1.8) as a variational problem. And we use the opportunity to define metric projections (see also [3, 5, 40]).

Proposition 1.15. *Let X be a reflexive, smooth and strictly convex Banach space and J^p be the duality mapping of X with gauge function $t \mapsto t^{p-1}$. Then for every $C \in \mathcal{C}(X)$ and every $x \in X$ there exists a unique element $P_C(x) \in C$ such that*

$$\|P_C(x) - x\| = \min_{y \in C} \|y - x\| \,. \tag{1.15}$$

$P_C(x)$ is called the metric projection *of x onto C. An element $x_0 \in C$ is the metric projection of x onto C iff*

$$\langle J^p(x_0 - x) \,|\, y - x_0 \rangle \geq 0 \quad \textit{for all} \quad y \in C \,. \tag{1.16}$$

Proof. The existence of an element $P_C(x) \in C$ with (1.15) follows by proposition 1.3. Suppose there are two different such solutions $P_C(x) \neq \tilde{P}_C(x) \in C$. Then $\|P_C(x) - x\| = \|\tilde{P}_C(x) - x\| = m := \min_{y \in C} \|y - x\|$ and the element $z := \frac{P_C(x) + \tilde{P}_C(x)}{2}$ lies in C since C is convex. By proposition 1.14 (a) it follows that

$$\begin{aligned} \frac{1}{p}\|z - x\|^p &= \frac{1}{p}\left\| \frac{P_C(x) - x}{2} + \frac{\tilde{P}_C(x) - x}{2} \right\|^p \\ &< \frac{\frac{1}{p}\|P_C(x) - x\|^p + \frac{1}{p}\|\tilde{P}_C(x) - x\|^p}{2} = \frac{1}{p} m^p \,, \end{aligned}$$

which leads to a contradiction. Let $x_0 \in C$ be the metric projection of x onto C. Then $\frac{1}{p}\|\big(x_0 + \lambda(y - x_0)\big) - x\|^p \geq \frac{1}{p}\|x_0 - x\|^p$ for any $y \in C$ and $\lambda \in (0, 1)$. By proposition 1.12 and the definition of subgradients (1.12) we get

$$\begin{aligned} 0 &\geq \frac{1}{p}\|x_0 - x\|^p - \frac{1}{p}\|\big(x_0 + \lambda(y - x_0)\big) - x\|^p \\ &\geq \Big\langle J^p\Big(\big(x_0 + \lambda(y - x_0)\big) - x\Big) \,\Big|\, -\lambda(y - x_0) \Big\rangle \end{aligned}$$

and therefore

$$\Big\langle J^p\Big(\big(x_0 + \lambda(y - x_0)\big) - x\Big) \,\Big|\, y - x_0 \Big\rangle \geq 0 \,.$$

According to proposition 1.14 (f) J^p is norm-to-weak-continuous and thus by letting $\lambda \to 0$ we arrive at

$$\langle J^p(x_0 - x) \,|\, y - x_0 \rangle \geq 0 \quad \text{for all} \quad y \in C \,.$$

Conversely if the above inequality is valid then we get

$$\|x_0 - x\|^p = \langle J^p(x_0 - x) \,|\, x_0 - x\rangle \le \langle J^p(x_0 - x) \,|\, y - x\rangle \le \|x_0 - x\|^{p-1}\|y - x\| \,.$$

Hence $\|x_0 - x\| \le \|y - x\|$ for all $y \in C$. □

In Hilbert spaces the metric projection operator P_C is known to be *non-expansive*, i.e.

$$\|P_C(x) - P_C(y)\| \le \|x - y\| \quad \text{for all} \quad x, y \in X \,.$$

This is a very useful property in applications since it preserves monotonicity of sequences in the form

$$\|P_C(x_n) - x\| \le \|x_n - x\| \quad \text{for all} \quad x \in C \,,$$

which ensures convergence of many optimization algorithms. In general Banach spaces the metric projection operator lacks this property. But we will see that Bregman projections behave better with respect to Bregman distances.

1.1.7 Characteristic Inequalities

The next two propositions provide us with inequalities which are of great relevance for proving the convergence of the iteration methods. These inequalities indeed completely characterize uniformly smooth resp. uniformly convex Banach spaces [48]. In the case of Hilbert spaces for the normalized duality mapping (i.e. the identity mapping) they reduce to the well-known polarisation identity

$$\|x - y\|^2 = \|x\|^2 - 2\langle x \,|\, y\rangle + \|y\|^2 \,.$$

Let again J^p be the duality mapping of X with gauge function $t \mapsto t^{p-1}$ and j^p denote a single-valued selection.

Proposition 1.16. *If X is uniformly convex then for all $x, y \in X$*

$$\|x - y\|^p \ge \|x\|^p - p\langle j^p(x) \,|\, y\rangle + \sigma_p(x, y) \tag{1.17}$$

with

$$\sigma_p(x, y) = pK_p \int_0^1 \frac{\big(\|x - ty\| \vee \|x\|\big)^p}{t} \delta_X \left(\frac{t\|y\|}{2\big(\|x - ty\| \vee \|x\|\big)} \right) \mathrm{d}t \,, \tag{1.18}$$

whereby

$$K_p = 4(2 + \sqrt{3}) \min \Big\{ \frac{1}{2} p(p-1) \wedge 1 \,, \left(\frac{1}{2} p \wedge 1\right)(p-1) \,, \\ (p-1)\left(1 - (\sqrt{3} - 1)^q\right) \,, 1 - \left(1 + (2 - \sqrt{3})q\right)^{1-p} \Big\} \,. \tag{1.19}$$

Proposition 1.17. *If X is uniformly smooth then for all $x, y \in X$*

$$\|x - y\|^p \leq \|x\|^p - p\langle J^p(x) \,|\, y\rangle + \tilde{\sigma}_p(x, y) \tag{1.20}$$

with

$$\tilde{\sigma}_p(x, y) = pG_p \int_0^1 \frac{\big(\|x - ty\| \vee \|x\|\big)^p}{t} \rho_X \left(\frac{t\|y\|}{\|x - ty\| \vee \|x\|} \right) \mathrm{d}t \,, \tag{1.21}$$

whereby $G_p = 8 \vee 64cK_p^{-1}$ with K_p defined according to (1.19) and

$$c = 4\frac{\tau_0}{\sqrt{1 + \tau_0^2} - 1} \prod_{j=1}^{\infty} \left(1 + \frac{15}{2^{j+2}}\tau_0\right) \quad \textit{with} \quad \tau_0 = \frac{\sqrt{339} - 18}{30} \,.$$

To see how such results can be obtained in special cases and since we need it in our applications, we prove the above proposition for L_p-spaces.

Proposition 1.18.

(a) In an L_p-space with $p \geq 2$ the following inequality is valid for all $x, y \in L_p$ and all $r \geq 2$:

$$\|x - y\|^r \leq \|x\|^r - r\langle J^r(x) \,|\, y\rangle + \frac{r}{2}\big((p \vee r) - 1\big)(\|x\| + \|y\|)^{r-2}\|y\|^2 \,. \tag{1.22}$$

Especially for the normalized duality mapping we have for all $x, y \in L_p$:

$$\|x - y\|^2 \leq \|x\|^2 - 2\langle J^2(x) \,|\, y\rangle + (p - 1)\|y\|^2 \,. \tag{1.23}$$

(b) In an L_p-space with $p \in (1, 2]$ the following inequality is valid for all $x, y \in L_p$:

$$\|x - y\|^p \leq \|x\|^p - p\langle J^p(x) \,|\, y\rangle + 2^{2-p}\|y\|^p \,, \tag{1.24}$$

whereby the "p" in "L_p" and "J^p" are the same.

Proof. (a) Let $x, y \in L_p$ ($p \geq 2$) be given such that $x - \lambda y \neq 0$ for all $\lambda \in [0, 1]$. As a consequence of proposition 1.14 (b) for all $\tilde{p} > 1$ the function

$$h_{\tilde{p}} : [0, 1] \longrightarrow \mathbb{R} \quad , \quad h_{\tilde{p}}(t) = \frac{1}{\tilde{p}}\|x - ty\|^{\tilde{p}}$$

is differentiable with

$$h_{\tilde{p}}(0) = \frac{1}{\tilde{p}}\|x\|^{\tilde{p}} \quad , \quad h_{\tilde{p}}(1) = \frac{1}{\tilde{p}}\|x - y\|^{\tilde{p}}$$

and

$$h'_{\tilde{p}}(t) = -\langle J^{\tilde{p}}(x - ty) \,|\, y\rangle \quad , \quad h'_{\tilde{p}}(0) = -\langle J^{\tilde{p}}(x) \,|\, y\rangle \,.$$

Therefore we can write

$$h_r(1) - h_r(0) - h_r'(0) = \int_0^1 \big(h_r'(t) - h_r'(0)\big)\,\mathrm{d}t\,.$$

We aim to find an upper estimate for the right hand side of the above equality. By proposition 1.10 (e) and example 1.11 (a) we get

$$\begin{aligned} h_r'(t) &= -\|x-ty\|^{r-p}\,\langle J^p(x-ty)\,|\,y\rangle \\ &= -\big(p\,h_p(t)\big)^{\frac{r-p}{p}}\,\big\langle |x-ty|^{p-1}\,\mathrm{sgn}(x-ty)\,\big|\,y\big\rangle\,. \end{aligned}$$

This function is again differentiable and by the sum rule we get

$$\begin{aligned} h_r''(t) &= -\frac{r-p}{p}\big(p\,h_p(t)\big)^{\frac{r-2p}{p}}\,p\,h_p'(t)\,\langle J^p(x-ty)\,|\,y\rangle \\ &\quad + \|x-ty\|^{r-p}(p-1)\,\big\langle |x-ty|^{p-2}y\,\big|\,y\big\rangle \\ &= (r-p)\|x-ty\|^{r-2p}\,|\langle J^p(x-ty)\,|\,y\rangle|^2 \\ &\quad + \|x-ty\|^{r-p}(p-1)\,\big\langle |x-ty|^{p-2}y\,\big|\,y\big\rangle\,, \end{aligned}$$

where by the Hölder inequality the element $|x-ty|^{p-2}y$ is in L_q with

$$\big\|\,|x-ty|^{p-2}y\big\| \le \|x-ty\|^{p-2}\|y\|\,.$$

If $p \ge r$ then the first summand is less than or equal to zero. Otherwise we can estimate it for all $t \in (0,1)$ by

$$(r-p)\|x-ty\|^{r-2p}\,|\langle J^p(x-ty)\,|\,y\rangle|^2 \le (r-p)(\|x\|+\|y\|)^{r-2}\|y\|^2\,.$$

For the second summand we get for all $r \ge 2$

$$\|x-ty\|^{r-p}(p-1)\,\big\langle |x-ty|^{p-2}y\,\big|\,y\big\rangle \le (p-1)(\|x\|+\|y\|)^{r-2}\|y\|^2\,.$$

All in all we get

$$h_r''(t) \le \begin{cases} (p-1)(\|x\|+\|y\|)^{r-2}\|y\|^2 & , \quad p \ge r \\ (r-1)(\|x\|+\|y\|)^{r-2}\|y\|^2 & , \quad p \le r \end{cases}.$$

For all $t \in (0,1)$ we can therefore find a $t_0 \in (0,t)$ such that

$$h_r'(t) - h_r'(0) = h_r''(t_0)\,t \le \big((p\vee r)-1\big)(\|x\|+\|y\|)^{r-2}\|y\|^2\,t\,.$$

Hence

$$\int_0^1 \big(h_r'(t) - h_r'(0)\big)\,\mathrm{d}t \le \frac{1}{2}\big((p\vee r)-1\big)(\|x\|+\|y\|)^{r-2}\|y\|^2\,,$$

from which the assertion follows. It remains to prove the inequality in case $x = \lambda y$ for a $\lambda \in [0,1]$; and by inserting such an x in (1.22) we see that it suffices to show that

$$(1-\lambda)^r \le \lambda^r - r\lambda^{r-1} + \frac{r}{2}(r-1)(1+\lambda)^{r-2}$$

for all $\lambda \in [0,1]$ and $r \ge 2$. To this end we show that the differentiable function $g_\lambda : [2,\infty) \longrightarrow \mathbb{R}$ defined by

$$g_\lambda(r) := \lambda^r - r\lambda^{r-1} + \frac{1}{2}(r^2-r)(1+\lambda)^{r-2} - (1-\lambda)^r$$

is greater than or equal to zero. We have

$$g_\lambda(2) = \lambda^2 - 2\lambda + 1 - (1-\lambda)^2 = 0$$

and g_λ is increasing since

$$\begin{aligned} g'_\lambda(r) &= \ln(\lambda)\lambda^r - \lambda^{r-1} - r\ln(\lambda)\lambda^{r-1} \\ &\quad + \frac{1}{2}(2r-1)(1+\lambda)^{r-2} + \frac{1}{2}(r^2-r)\ln(1+\lambda)(1+\lambda)^{r-2} \\ &\quad - \ln(1-\lambda)(1-\lambda)^r \\ &\ge -\frac{1}{r\,e} - 1 - 0 + \frac{1}{2}(4-1) + 0 - 0 \\ &= \frac{1}{2} - \frac{1}{r\,e} \ge 0\,, \end{aligned}$$

because $\ln(\lambda) \le 0$, $\ln(1+\lambda) \ge 0$, $\ln(1-\lambda) \le 0$ and as can be easily seen

$$0 \ge \ln(\lambda)\lambda^r \ge -\frac{1}{r\,e}$$

for all $\lambda \in [0,1]$.
To prove (b) we remark that the pointwise resp. componentwise inequality

$$\frac{\big||x|^{p-1}\operatorname{sgn}(x) - |y|^{p-1}\operatorname{sgn}(y)\big|}{|x-y|^{p-1}} \le 2^{2-p}$$

is valid for all $x \ne y \in \mathbb{R}$ and $p \in (1,2]$. Hence we get

$$\|J^p(x) - J^p(y)\| \le 2^{2-p}\|x-y\|^{p-1} \quad \text{for all} \quad x, y \in L_p\,.$$

Now we can estimate as in the beginning of the proof of (a):

$$\begin{aligned} \frac{1}{p}\|x-y\|^p - \frac{1}{p}\|x\|^p - \langle J^p(x)\,|\,y\rangle &= \int_0^1 \langle J^p(x-ty) - J^p(x)\,|\,y\rangle \,\mathrm{d}t \\ &\le \int_0^1 \|J^p(x-ty) - J^p(x)\|\;\|y\|\,\mathrm{d}t \\ &\le 2^{2-p}\|y\|^p \int_0^1 t^{p-1}\,\mathrm{d}t = \frac{2^{2-p}}{p}\|y\|^p\,. \end{aligned}$$

□

1.1.8 Positive Duality Mappings in Banach Lattices

Since we intend to include constraints of the form "$Ax \leq y$" in the split feasibility problem, we shortly introduce positive duality mappings on Banach lattices (see also [19, 22, 34]).

Definition 1.19. *A Banach space X with a partial order "$\leq$" is called a* Banach lattice *if for all $x, y, z \in X$ and $\lambda \geq 0$ the following holds:*

(i) $x \leq y$ *implies* $x + z \leq y + z$.
(ii) $\lambda x \geq 0$ *for* $x \geq 0$.
(iii) There exists a least upper bound, denoted by $x \vee y$, and a greatest lower bound, denoted by $x \wedge y$.
(iv) $\|x\| \leq \|y\|$ *whenever* $|x| \leq |y|$, *whereby the* absolute value $|x|$ *of x is defined by* $|x| := x \vee (-x)$.

Example 1.20. The L_p-spaces ($p \in [1, \infty]$) with "$\leq$" defined componentwise resp. pointwise almost everywhere are Banach lattices.

It is a convenient fact that in a Banach lattice every (in-)equality involving lattice operations and algebraic operations is valid if its analogue is valid in the real line, e.g.

$$|x - y| = |x \vee z - y \vee z| + |x \wedge z - y \wedge z| \,,$$

which together with property 1.19 (iv) implies that the lattice operations are continuous. Thus the set

$$\mathcal{P} := \{x \in X \,|\, x \geq 0\}$$

of all *positive* elements of X is a closed convex cone. It is called the *positive cone* of X. For $x \in X$ we set

$$x_+ := x \vee 0 \quad \text{and} \quad x_- := -(x \wedge 0) \,.$$

We obviously have $x_+, x_- \in \mathcal{P}$ and

$$x = x_+ - x_- \quad \text{and} \quad |x| = x_+ + x_- \,.$$

Two elements $x, y \in X$ for which $|x| \wedge |y| = 0$ are said to be *disjoint* and we write

$$\text{disj}(x) := \{y \in X \,|\, |x| \wedge |y| = 0\} \,. \tag{1.25}$$

It is not difficult to see that $x_- \in \text{disj}(x_+)$ and $x_+ \in \text{disj}(x_-)$. The dual X^* of a Banach lattice is also a Banach lattice provided its order is defined by

$$x^* \leq y^* \quad \text{iff} \quad \langle x^* \,|\, z\rangle \leq \langle y^* \,|\, z\rangle \quad \text{for all} \quad z \in \mathcal{P} \,.$$

The positive cone $\mathcal{P}^*$ of X^* is then given by

$$\mathcal{P}^* = \{x^* \in X^* \,|\, \langle x^* \,|\, x\rangle \geq 0 \text{ for every } x \in \mathcal{P}\} \,.$$

Shortly said, the positive duality mapping preserves positivity.

Definition 1.21. *Let J be the normalized duality mapping of a Banach lattice X. The mapping $J_+ : \mathcal{P} \longrightarrow 2^{X^*}$ defined by*

$$J_+(x) := \{x^* \in J(x) \mid x^* \geq 0 \text{ and } \langle x^* \mid y \rangle = 0 \text{ for all } y \in \operatorname{disj}(x)\}$$

is called positive duality mapping *of X.*

For every $x \in \mathcal{P}$ the set $J_+(x)$ is not empty. As a consequence of this and 1.14 (b) the normalized duality mapping of a smooth Banach lattice is a positive duality mapping. The single-valued selections of the normalized duality mapping in example 1.11 (b) and (c) also define selections of the positive duality mapping.

Proposition 1.22. *Let J_+ be the positive duality mapping of a Banach lattice X and let $f_+ : X \longrightarrow \mathbb{R}$ be defined by*

$$f_+(x) = \frac{1}{2}\|x_+\|^2 \quad , \quad x \in X\,.$$

Then $J_+(x_+) \subset \partial f_+$.

Proof. Obviously we have $J_+(x) \subset J(x)$ for all positive $x \in X$. Thus by 1.12 we get for all $x, y \in X$ and $x^* \in J_+(x_+)$

$$\begin{aligned} \frac{1}{2}\|y_+\|^2 - \frac{1}{2}\|x_+\|^2 &\geq \langle x^* \mid y_+ - x_+ \rangle \\ &= \langle x^* \mid y - x \rangle + \langle x^* \mid y_- \rangle - \langle x^* \mid x_- \rangle \\ &\geq \langle x^* \mid y - x \rangle\,, \end{aligned}$$

because x^* and y_- are positive and $x_- \in \operatorname{disj}(x_+)$. □

We also need the following characterization of positivity.

Proposition 1.23. *An element $x \in X$ is positive iff $\langle x^* \mid x \rangle \geq 0$ for all positive $x^* \in X^*$.*

Proof. Let $x \in X$ be such that $\langle x^* \mid x \rangle \geq 0$ for all positive $x^* \in X^*$. We write $x = x_+ - x_-$ and get

$$\langle x^* \mid x_+ \rangle \geq \langle x^* \mid x_- \rangle \quad \text{for all} \quad x^* \geq 0\,.$$

Especially for some $x^* \in J_+(x_-)$ this yields

$$0 = \langle x^* \mid x_+ \rangle \geq \langle x^* \mid x_- \rangle = \|x_-\|^2$$

and therefore $x = x_+ \geq 0$. The converse part is obvious due to the definition of positivity in the dual space. □

1.2 Bregman Distances and Bregman Projections

In this section we are concerned with Bregman distances that are induced by the functions $f(x) = \frac{1}{p}\|x\|^p$ and the related Bregman projections (which ALBER [3] also calls *generalized projections*). The idea to use such distances to design and analyse optimization algorithms goes back to LEV BREGMAN [11] and since then there has been an ever growing area of research in which his ideas are applied in various ways: for analysing feasibility in optimization, for projections onto convex sets, for approximating equilibria, for computing fixed points of nonlinear mappings, for analysing regularization methods,... [1, 2, 6, 8, 10, 12–15, 18, 23, 26, 31, 32, 38, 43, 44, 46].
We try to give a self-contained representation, because most results in the literature are presented in the more general context of *total convexity, essential strict convexity, essential strict smoothness, Legendre and Bregman functions* (and some of course don't hold in the general case). More information about this interesting topic can e.g. be found in [7, 9, 14, 16].

1.2.1 Bregman Distances

For a Gâteaux differentiable convex function $f : X \longrightarrow \mathbb{R}$ the function

$$\Delta_f(x,y) := f(y) - f(x) - \langle f'(x) \,|\, y - x\rangle \quad , \quad x, y \in X \tag{1.26}$$

is called the *Bregman distance* of x to y with respect to the function f.

Though it is not a metric in the usual sense – it is e.g. in general not symmetric – this function has some distance-like properties; it indicates how much $f(y)$ increases over $f(x)$ above linear growth with slope $f'(x)$. Because of proposition 1.14 (b) and (1.9), in smooth Banach spaces the Bregman-distance with respect to the function $f(x) = \frac{1}{p}\|x\|^p$ can be written as

$$\begin{aligned}\Delta_p(x,y) &= \frac{1}{q}\|x\|^p - \langle J^p(x) \,|\, y\rangle + \frac{1}{p}\|y\|^p \\ &= \frac{1}{p}\left(\|y\|^p - \|x\|^p\right) + \langle J^p(x) \,|\, x - y\rangle \\ &= \frac{1}{q}\left(\|x\|^p - \|y\|^p\right) + \langle J^p(y) - J^p(x) \,|\, y\rangle \quad , \quad x, y \in X\,. \end{aligned} \tag{1.27}$$

In a Hilbert space we just have $\Delta_2(x,y) = \frac{1}{2}\|x - y\|^2$.

In the next proposition we collect some important properties of Δ_p and point out its relationship to the norm in X (see also [6, 15, 46]).

Proposition 1.24. *Let X be reflexive, smooth and strictly convex. Then for all $x, y \in X$ and sequences $(x_n)_n$ in X the following holds:*

(a) $\Delta_p(x,y) \geq 0$ *and* $\Delta_p(x,y) = 0 \Leftrightarrow x = y$.
(b) $\Delta_p(x,y) + \Delta_p(y,x) = \langle J^p(x) - J^p(y) \,|\, x - y \rangle$.
(c) $\Delta_p(-x,-y) = \Delta_p(x,y)$ *and* Δ_p *is* positively homogenous *of degree* p*, i.e.*

$$\Delta_p(\lambda x, \lambda y) = \lambda^p \Delta_p(x,y) \quad \text{for all} \quad x, y \in X \,,\, \lambda \geq 0 \,.$$

(d) $\lim_{\|x_n\| \to \infty} \Delta_p(x_n, x) = \infty$ *and* $\lim_{\|x_n\| \to \infty} \Delta_p(x, x_n) = \infty$*, i.e. the sequence* $(x_n)_n$ *remains bounded if the sequence* $\big(\Delta_p(x_n, x)\big)_n$ *resp. the sequence* $\big(\Delta_p(x, x_n)\big)_n$ *is bounded.*
(e) Δ_p *is continuous in both arguments and it is strictly convex and Gâteaux differentiable with respect to the second variable with derivative*

$$\frac{\partial}{\partial y} \Delta_p(x,y) = J^p(y) - J^p(x) \,. \tag{1.28}$$

(f) Consider the following assertions:
(i) $\lim_{n\to\infty} \|x_n - x\| = 0$.
(ii) $\lim_{n\to\infty} \|x_n\| = \|x\|$ *and* $\lim_{n\to\infty} \langle J^p(x_n) \,|\, x \rangle = \langle J^p(x) \,|\, x \rangle$.
(iii) $\lim_{n\to\infty} \Delta_p(x_n, x) = 0$.
The implications (i) $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii) are valid. If X is uniformly convex then the assertions are equivalent.*
(g) If $(x_n)_n$ *is a Cauchy sequence then it is bounded and for all* $\epsilon > 0$ *there exists an* $n_0 \in \mathbb{N}$ *such that* $\Delta_p(x_k, x_l) < \epsilon$ *for all* $k, l \geq n_0$*. If X is uniformly convex then the converse is also true (and it suffices to show* $\Delta_p(x_k, x_l) < \epsilon$ *for all* $k \geq l \geq n_0$ *or* $l \geq k \geq n_0$*).*
(h) If X is uniformly convex and $M \subset X$ *is bounded, then for all* $\epsilon > 0$ *there exists a* $\delta > 0$ *such that* $\|x - y\| < \epsilon$ *for all* $x, y \in M$ *with* $\Delta_p(x,y) < \delta$.
(i) Let us write $\Delta_q^*(x^*, y^*) = \frac{1}{p}\|x^*\|_{X^*}^q - \langle J_{X^*}^q(x^*) \,|\, y^* \rangle + \frac{1}{q}\|y^*\|_{X^*}^q$ *for the Bregman distance on the dual space* X^* *with respect to the function* $f^*(x^*) = \frac{1}{q}\|x^*\|_{X^*}^q$[3]*. Then we have*

$$\Delta_p(x,y) = \Delta_q^*(y^*, x^*)$$

for $x^* = J_X^p(x) \Leftrightarrow J_{X^*}^q(x^*) = x$ *and* $y^* = J_X^p(y) \Leftrightarrow J_{X^*}^q(y^*) = y$.

Proof. At first we want to point out that by proposition 1.7 (a) X^* is also reflexive, smooth and strictly convex and thus all the assertions are valid for the dual distance Δ_q^* too. The relation $\Delta_p(x,y) = \Delta_q^*(y^*, x^*)$ in (i) is obvious. (1.27) and (1.9) yield

[3] This is in fact the conjugate function of f, whereby in general for a function $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ the *conjugate function* $f^* : X^* \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ is defined by $f^*(x^*) := \sup\limits_{x \in X} \big(\langle x^*, x \rangle - f(x)\big)$.

$$\begin{aligned}\Delta_p(x,y) &\geq \frac{1}{p}\|y\|^p - \|x\|^{p-1}\|y\| + \frac{1}{q}\|x\|^p \\ &= \|x\|^p\left(\frac{1}{p}\left(\frac{\|y\|}{\|x\|}\right)^p - \frac{\|y\|}{\|x\|} + \frac{1}{q}\right).\end{aligned}$$

Since the right hand side of the above inequality converges to infinity for $\|x\| \to \infty$ and fixed y or vice versa, we see that (d) holds. To prove (a) we consider $t = \frac{\|y\|}{\|x\|}$ and define $h : [0,\infty) \longrightarrow \mathbb{R}$ by

$$h(t) := \frac{1}{p}t^p - t + \frac{1}{q}.$$

From $h(0) = \frac{1}{q}$, $\lim_{t\to\infty} h(t) = \infty$, $h'(t) = t^{p-1} - 1 = 0 \Leftrightarrow t = 1$ and $h(1) = 0$ we conclude that $h(t) \geq 0$ and $h(t) = 0 \Leftrightarrow t = 1$. Therefore $\Delta_p(x,y) \geq 0$ and if $\Delta_p(x,y) = 0$ then $\|y\| = \|x\| \Leftrightarrow \|J^p(x)\| = \|x\|^{p-1} = \|y\|^{p-1}$ and thus also

$$\langle J^p(x) \,|\, y\rangle = \frac{1}{q}\|x\|^p + \frac{1}{p}\|y\|^p = \|y\|^p.$$

Hence $J^p(x) = J^p(y)$ and by proposition 1.14 (e) it follows that $x = y$. We directly calculate

$$\begin{aligned}&\Delta_p(x,y) + \Delta_p(y,x) \\ &= \frac{1}{q}\|x\|^p - \langle J^p(x)\,|\,y\rangle + \frac{1}{p}\|y\|^p + \frac{1}{q}\|y\|^p - \langle J^p(y)\,|\,x\rangle + \frac{1}{p}\|x\|^p \\ &= \|x\|^p + \|y\|^p - \langle J^p(x)\,|\,y\rangle - \langle J^p(y)\,|\,x\rangle \\ &= \langle J^p(x)\,|\,x\rangle - \langle J^p(x)\,|\,y\rangle + \langle J^p(y)\,|\,y\rangle - \langle J^p(y)\,|\,x\rangle \\ &= \langle J^p(x) - J^p(y)\,|\,x - y\rangle,\end{aligned}$$

which proves (b). (c) is a consequence of the homogeneity of J^p $\big($Prop. 1.10 (b)$\big)$. In (e) continuity of the function $y \mapsto \Delta_p(x,y)$ is obvious and its strict convexity follows by proposition 1.14 (a). The continuity of the function $x \mapsto \Delta_p(x,y)$ is a consequence of proposition 1.14 (f). The assertion about differentiability follows by proposition 1.14 (b) and straightforward calculation. The implication (i) $\Rightarrow$ (ii) in (f) is valid due to proposition 1.14 (f) and the implication (ii) $\Rightarrow$ (iii) follows directly from the first line in (1.27). Now let X be uniformly convex. Substituting $x - y$ for y in theorem 1.16 we arrive at

$$p\Delta_p(x,y) = \|y\|^p + (p-1)\|x\|^p - p\langle J^p(x)\,|\,y\rangle \geq \sigma_p(x, x-y).$$

With the explicit expression for σ_p (1.18) we have $\frac{1}{pK_p}\sigma_p(x,x-y) =$

$$\int_0^1 \frac{\big(\|x - t(x-y)\| \vee \|x\|\big)^p}{t}\,\delta_X\left(\frac{t\|x-y\|}{2\big(\|x - t(x-y)\| \vee \|x\|\big)}\right)\,\mathrm{d}t\,.$$

Since by proposition 1.5 (a) δ_X is nondecreasing and non-negative and

$$\|x - t(x-y)\| \vee \|x\| \le \|x\| + \|x-y\|$$

and

$$\|x - t(x-y)\| \vee \|x\| \ge \frac{t}{2}\|x-y\|$$

for all $t \in [0,1]$ $\big($in case $\|x\| \ge \frac{t}{2}\|x-y\|$ this is clear and otherwise $\|x - t(x-y)\| \ge t\|x-y\| - \|x\| \ge \frac{t}{2}\|x-y\|\big)$, we can estimate

$$\begin{aligned}\frac{2^p}{pK_p}\sigma_p(x, x-y) &\ge \|x-y\|^p \int_0^1 t^{p-1}\,\delta_X\left(\frac{t\|x-y\|}{2(\|x\|+\|x-y\|)}\right)\mathrm{d}t \\ &\ge \|x-y\|^p \int_{\frac{1}{2}}^1 t^{p-1}\,\delta_X\left(\frac{t\|x-y\|}{2(\|x\|+\|x-y\|)}\right)\mathrm{d}t \\ &\ge \|x-y\|^p\,\delta_X\left(\frac{\|x-y\|}{4(\|x\|+\|x-y\|)}\right)\frac{1}{p}\left(1-\frac{1}{2^p}\right).\end{aligned}$$

Putting all together we see that

$$\Delta_p(x,y) \ge C\|x-y\|^p\,\delta_X\left(\frac{\|x-y\|}{4(\|x\|+\|x-y\|)}\right)$$

with $C = \frac{K_p}{p2^p}\left(1-\frac{1}{2^p}\right)$. If $M \subset X$ is bounded, then there is a constant $R > 0$ such that $4(\|x\|+\|x-y\|) \le R$ for all $x, y \in M$. Suppose there is an $\epsilon > 0$ such that for all $\delta > 0$ we can find $x_\delta, y_\delta \in M$ with $\Delta_p(x_\delta, y_\delta) < \delta$ but $\|x_\delta - y_\delta\| \ge \epsilon$. Then by the monotonicity of δ_X and the uniform convexity of X (Def. 1.6) we get for all $\delta > 0$

$$\delta > \Delta_p(x_\delta, y_\delta) \ge C\epsilon^p\,\delta_X\left(\frac{\epsilon}{R}\right) > 0\,.$$

By letting $\delta \to 0$ this leads to a contradiction. Thus (h) is proven and the "converse"-part in (g) and the implication (iii) $\Rightarrow$ (i) in (f) for uniformly convex X are immediate consequences. Finally if $(x_n)_n$ is a Cauchy sequence then it is bounded and convergent and the rest of (g) follows by looking at the second line in (1.27). □

1.2.2 Bregman Projections

We can now define Bregman projections onto closed convex sets.

Proposition 1.25. *Let X be a reflexive, smooth and strictly convex Banach space and J^p be a duality mapping of X. Then for every $C \in \mathcal{C}(X)$ and $x \in X$ there exists a unique element $\Pi_C^p(x) \in C$ such that*

$$\Delta_p\big(x, \Pi_C^p(x)\big) = \min_{y \in C} \Delta_p(x,y)\,. \tag{1.29}$$

$\Pi_C^p(x)$ *is called the* Bregman projection *of x onto C (with respect to the function $f(x) = \frac{1}{p}\|x\|^p$). Moreover $x_0 \in C$ is the Bregman projection of x onto C iff*

$$\langle J^p(x_0) - J^p(x) \,|\, y - x_0 \rangle \geq 0 \tag{1.30}$$

or equivalently

$$\Delta_p(x_0, y) \leq \Delta_p(x, y) - \Delta_p(x, x_0) \tag{1.31}$$

for every $y \in C$.

Proof. There is a sequence $(x_n)_n \in C$ with

$$\lim_{n\to\infty} \Delta_p(x, x_n) = m := \inf_{y\in C} \Delta_p(x, y)\,.$$

In particular $(x_n)_n$ is bounded by propostion 1.24 (d) and by proposition 1.2 it therefore has a weakly convergent subsequence $(x_{n_k})_k$ with w.l.o.g. $\lim_{k\to\infty} \|x_{n_k}\| = R$ for an $R \geq 0$. Since C is convex and (weakly) closed, the weak limit point x_0 of $(x_{n_k})_k$ lies again in C. Hence $\Delta_p(x, x_0) \geq m$ and we have

$$\|x_0\|^p = \langle J^p(x_0) \,|\, x_0 \rangle = \lim_{k\to\infty} \langle J^p(x_0) \,|\, x_{n_k} \rangle \leq \|x_0\|^{p-1} R$$

and thus $\|x_0\| \leq R$. Suppose $\Delta_p(x, x_0) > m$. Then there exists a $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$

$$\begin{aligned}
&\Delta_p(x, x_{n_k}) \leq \Delta_p(x, x_0)\\
\Leftrightarrow\quad &\frac{1}{p}\|x_{n_k}\|^p - \langle J^p(x) \,|\, x_{n_k} \rangle \leq \frac{1}{p}\|x_0\|^p - \langle J^p(x) \,|\, x_0 \rangle\\
\Leftrightarrow\quad &\frac{1}{p}\left(\|x_{n_k}\|^p - \|x_0\|^p\right) \leq \langle J^p(x) \,|\, x_{n_k} - x_0 \rangle\,.
\end{aligned}$$

The right hand side converges to zero and the left hand side converges to $\frac{1}{p}\left(R^p - \|x_0\|^p\right)$ for $k \to \infty$ and we get $\|x_0\| \geq R$. Hence $\|x_0\| = R$ and thus

$$\begin{aligned}
\Delta_p(x, x_0) &= \frac{1}{q}\|x\|^p + \frac{1}{p}\|x_0\|^p - \langle J^p(x) \,|\, x_0 \rangle\\
&= \lim_{k\to\infty} \frac{1}{q}\|x\|^p + \frac{1}{p}\|x_{n_k}\|^p - \langle J^p(x) \,|\, x_{n_k} \rangle\\
&= \Delta_p(x, x_{n_k}) = m\,.
\end{aligned}$$

The uniqueness follows by the strict convexity of the function $y \mapsto \Delta_p(x, y)$ (proposition 1.24 (e)). Now let $\Pi_C^p(x) \in C$ be the Bregman projection of x onto C. Then for all $y \in C$ and $\lambda \in (0, 1)$ we have $\lambda \Pi_C^p(x) + (1-\lambda)y \in C$ and

$$\begin{aligned}
0 &\geq \Delta_p\big(x, \Pi_C^p(x)\big) - \Delta_p\big(x, \lambda\Pi_C^p(x) + (1-\lambda)y\big)\\
&\geq \left\langle \frac{\partial}{\partial y}\Delta_p\big(x, \lambda\Pi_C^p(x) + (1-\lambda)y\big) \,\middle|\, \Pi_C^p(x) - \big(\lambda\Pi_C^p(x) + (1-\lambda)y\big) \right\rangle\\
&= (1-\lambda)\left\langle J^p\big(\lambda\Pi_C^p(x) + (1-\lambda)y\big) - J^p(x) \,\middle|\, \Pi_C^p(x) - y \right\rangle
\end{aligned}$$

by the first part of proposition 1.24 (e). We divide by $1-\lambda$ and get

$$\big\langle J^p\big(\lambda \Pi_C^p(x) + (1-\lambda)y\big) - J^p(x) \,\big|\, y - \Pi_C^p(x)\big)\big\rangle \geq 0$$

whereby the left hand side converges to $\big\langle J^p\big(\Pi_C^p(x)\big) - J^p(x) \,\big|\, y - \Pi_C^p(x)\big)\big\rangle$ for $\lambda \to 1$ since J^p is norm-to-weak-continuous $\big($proposition 1.14 (f)$\big)$. Conversely let $x_0 \in C$ be such that $\langle J^p(x_0) - J^p(x) \,|\, y - x_0\rangle \geq 0$ for all $y \in C$. Then we get

$$\begin{aligned} \Delta_p(x,y) - \Delta_p(x,x_0) &\geq \left\langle \frac{\partial}{\partial y}\Delta_p(x,x_0) \,\middle|\, y - x_0 \right\rangle \\ &= \langle J^p(x_0) - J^p(x) \,|\, y - x_0\rangle \geq 0 \end{aligned}$$

and therefore $\Delta_p(x,x_0) \leq \Delta_p(x,y)$ for all $y \in C$. Finally we consider the following equivalent inequalities for every $x_0, y \in C$:

$$\begin{aligned} &\Delta_p(x_0,y) \leq \Delta_p(x,y) - \Delta_p(x,x_0) \\ \Leftrightarrow\quad &\frac{1}{q}\|x_0\|^p - \langle J^p(x_0) \,|y\rangle \leq -\langle J^p(x) \,|\, y\rangle + \langle J^p(x) \,|\, x_0\rangle - \frac{1}{p}\|x_0\|^p \\ \Leftrightarrow\quad &\langle J^p(x_0) \,|\, x_0\rangle - \langle J^p(x_0) \,|y\rangle \leq -\langle J^p(x) \,|\, y\rangle + \langle J^p(x) \,|\, x_0\rangle \\ \Leftrightarrow\quad &0 \leq \langle J^p(x_0) - J^p(x) \,|\, y - x_0\rangle\ . \end{aligned}$$

□

If $x \notin C$ then the variational characterization (1.30) is equivalent to saying that $x_0 \in C$ is the Bregman projection of x onto C iff

$$H_{\leq}(u^*, \alpha) \quad \text{with} \quad u^* = J^p(x) - J^p(x_0) \quad \text{and} \quad \alpha = \langle J^p(x) - J^p(x_0) \,|\, x_0\rangle$$

is a supporting halfspace of C at x_0. Likewise the variational characterization of the metric projection (1.16) is then equivalent to saying that $x_0 \in C$ is the metric projection of x onto C iff

$$H_{\leq}(u^*, \alpha) \quad \text{with} \quad u^* = J^p(x - x_0) \quad \text{and} \quad \alpha = \langle J^p(x - x_0) \,|\, x_0\rangle$$

is a supporting halfspace of C at x_0. In Hilbert spaces the Bregman projection with respect to the function $f(x) = \frac{1}{2}\|x\|^2$ coincides with the metric projection. As we already pointed out when we talked about metric projections, (1.31) is a property of Bregman projections which ensures monotonicity of the iteration methods.
We prove some further properties of Bregman projections with respect to the class of Bregman distances we use here. Although simple, we have not seen them explicitly stated elsewhere (except (a); but see also [3, 4]). Especially (b) answers the question asked in [1], whether their is some relationship between metric and Bregman projections (but we feel certain that we are missing some references).

Proposition 1.26. *Let X be a reflexive, smooth and strictly convex Banach space and $C \in \mathcal{C}(X)$ be given.*

(a) For $x \in X$ we have $\Pi_C^p(x) = x \Leftrightarrow x \in C$ and if $x \notin C$ then $\Pi_C^p(x)$ lies in the boundary of C.

(b) The Bregman projection and the metric projection are related via

$$P_C(x) - x = \Pi_{C-x}^p(0) \quad \text{for all} \quad x \in X\,.$$

Especially we have $P_C(0) = \Pi_C^p(0)$ and thus $\|\Pi_C^p(0)\| = \min_{y \in C} \|y\|$.

(c) Equivalent are
(i) $0 \in C$,
(ii) $\|\Pi_C^p(x)\| \le \|x\|$ for all $x \in X$.

(d) Π_C^p maps bounded sets onto bounded sets; more precisely we have

$$\|\Pi_C^p(x)\| \le \left(2^{q-1}\|x\|\right) \vee \left(3\,\|\Pi_C^p(0)\|\,\right) \quad \text{for all} \quad x \in X\,. \tag{1.32}$$

(e) The Bregman projection is parity- and scale-invariant in the sense that

$$\Pi_{\lambda C}^p(\lambda x) = \lambda \Pi_C^p(x) \quad \text{for every} \quad \lambda \in \mathbb{R}\,,\, x \in X\,.$$

Especially if C is a cone then $\lambda C = C$ for $\lambda > 0$ and thus Π_C^p is positively homogenous of degree 1; if C is also symmetric[4], then Π_C^p is homogenous of degree 1.

(f) The Bregman projections of points along the "dual ray" to the ray

$$z^*(\lambda) := J_X^p\left(\Pi_C^p(x)\right) + \lambda\Big(J_X^p(x) - J_X^p\left(\Pi_C^p(x)\right)\Big)$$

coincide, i.e.

$$\Pi_C^p\Big(J_{X^*}^q\left(z(\lambda)\right)\Big) = \Pi_C^p(x) \quad \text{for every} \quad \lambda \ge 0\,,\, x \in X\,.$$

(g) If we know $\Pi_C^p(x)$ then we obtain the Bregman projection of x onto the set $\lambda_x C$ with respect to the function $f(x) = \frac{1}{r}\|x\|^r$ $(r > 1)$ via

$$\Pi_{\lambda_x C}^r(x) = \lambda_x \Pi_C^p(x) \tag{1.33}$$

with

$$\lambda_x := \begin{cases} 1 & , \quad x = 0 \text{ or } \Pi_C^p(x) = 0 \\ \left(\frac{\|x\|}{\|\Pi_C^p(x)\|}\right)^{\frac{r-p}{r-1}} & , \quad \text{otherwise} \end{cases}\,.$$

Especially if C is a cone then

$$\Pi_C^r(x) = \lambda_x \Pi_C^p(x)\,. \tag{1.34}$$

Moreover if $x \in \lambda_x C$ then $x \in C$.

---
[4] i.e. $-C = C$

Proof. For $x \in C$ we have $\Delta_p(x,x) = 0 \le \Delta_p(x,y)$ for all $y \in C$ and therefore $x = \Pi_C^p(x)$. Conversely $x = \Pi_C^p(x)$ lies in C. In case $x \notin C$ then $H_{\le}(u^*,\alpha)$ with $u^* = J^p(x) - J^p(x_0)$ and $\alpha = \langle J^p(x) - J^p(x_0) \,|\, x_0 \rangle$ is a supporting halfspace of C at $x_0 := \Pi_C^p(x)$ and $x_0 \in H(u^*,\alpha)$. If x_0 were an interior point of C then we would also have

$$x_0 \in \mathrm{int}(C) \subset \mathrm{int}\big(H_{\le}(u^*,\alpha)\big) = H_{<}(u^*,\alpha)\,,$$

which contradicts $x_0 \in H(u^*,\alpha)$. To see (b) we compare the variational inequalities (1.16) and (1.30) for $x_0 \in C$ and $\tilde{x}_0 := x_0 - x \in \tilde{C} := C - x$:

$$\begin{aligned} &\langle J^p(x_0 - x) \,|\, y - x_0 \rangle \ge 0 \quad \text{for all} \quad y \in C \\ \Leftrightarrow\ &\langle J^p(x_0 - x) \,|\, (y - x) - (x_0 - x) \rangle \ge 0 \quad \text{for all} \quad y \in C \\ \Leftrightarrow\ &\langle J^p(\tilde{x}_0) \,|\, \tilde{y} - \tilde{x}_0 \rangle \ge 0 \quad \text{for all} \quad \tilde{y} \in \tilde{C}\,. \end{aligned}$$

If $0 \in C$ then by taking $y = 0$ in (1.30) we get

$$\begin{aligned} 0 &\ge \big\langle J^p\big(\Pi_C^p(x)\big) \,\big|\, \Pi_C^p(x) \big\rangle - \langle J^p(x) \,|\, \Pi_C^p(x) \rangle \\ &= \|\Pi_C^p(x)\|^p - \langle J^p(x) \,|\, \Pi_C^p(x) \rangle \end{aligned}$$

and therefore $\|\Pi_C^p(x)\|^p \le \langle J^p(x) \,|\, \Pi_C^p(x) \rangle \le \|x\|^{p-1}\, \|\Pi_C^p(x)\|$ which yields $\|\Pi_C^p(x)\| \le \|x\|$. Conversely if the above inequality is valid for all $x \in X$ then for $x = 0$ we get

$$0 = \|0\| \ge \|\Pi_C^p(0)\|\ .$$

Hence $0 = \Pi_C^p(0) \in C$. To prove (d) we transform (1.30) into

$$\begin{aligned} \|\Pi_C^p(x)\|^p &\le \big\langle J^p\big(\Pi_C^p(x)\big) \,\big|\, y \big\rangle - \langle J^p(x) \,|\, y \rangle + \langle J^p(x) \,|\, \Pi_C^p(x) \rangle \\ &\le \|\Pi_C^p(x)\|^{p-1}\, \|y\| + \|x\|^{p-1}\|y\| + \|x\|^{p-1}\, \|\Pi_C^p(x)\|\ . \end{aligned} \tag{1.35}$$

If $\|\Pi_C^p(x)\|^{p-1} < 2\|x\|^{p-1} \Leftrightarrow \|\Pi_C^p(x)\| < 2^{q-1}\|x\|$ then we are done. Otherwise $t_x := \left(\frac{\|\Pi_C^p(x)\|}{\|x\|}\right)^{p-1} \ge 2$ for $x \ne 0$ and by (1.35) we get

$$\begin{aligned} \|\Pi_C^p(x)\| \left(\|\Pi_C^p(x)\|^{p-1} - \|x\|^{p-1}\right) &\le \|y\| \left(\|\Pi_C^p(x)\|^{p-1} + \|x\|^{p-1}\right) \\ \Leftrightarrow \qquad \|\Pi_C^p(x)\| &\le \|y\| \frac{\|\Pi_C^p(x)\|^{p-1} + \|x\|^{p-1}}{\|\Pi_C^p(x)\|^{p-1} - \|x\|^{p-1}}\,. \end{aligned}$$

With $\|\Pi_C^p(0)\| = \min\limits_{y \in C} \|y\|$ and since the function $h(t) := \frac{t+1}{t-1}$ is decreasing for $t > 1$ we arrive at

$$\|\Pi_C^p(x)\| \le \|\Pi_C^p(0)\|\, h(t_x) \le \|\Pi_C^p(0)\|\, h(2) = 3\, \|\Pi_C^p(0)\|\ .$$

(e) is a consequence of the homogeneity of Δ_p $\big($Prop. 1.24 (c)$\big)$ because

$$\begin{aligned} &\Delta_p\big(\lambda x, \lambda \Pi_C^p(x)\big) \le \Delta_p(\lambda x, \lambda y) \quad \text{for all} \quad x \in X, \lambda y \in \lambda C \\ \Leftrightarrow \quad &\Delta_p\big(x, \Pi_C^p(x)\big) \le \Delta_p(x, y) \quad \text{for all} \quad x \in X, y \in C\,. \end{aligned}$$

To prove (f) we recall that $J_X^p\big(J_{X^*}^q(z^*)\big) = z^*$ by proposition 1.14 (e) and we check the validity of the variational inequality (1.30) with $\tilde{x} = J_{X^*}^q\big(z^*(\lambda)\big)$ and $\tilde{x_0} = \Pi_C^p(x)$ for all $y \in C$:

$$\begin{aligned} \langle J_X^p(\tilde{x_0}) - J_X^p(\tilde{x}) \,|\, y - \tilde{x_0}\rangle &= \big\langle J_X^p\big(\Pi_C^p(x)\big) - z^*(\lambda) \,\big|\, y - \Pi_C^p(x)\big\rangle \\ &= \lambda \big\langle J_X^p\big(\Pi_C^p(x)\big) - J_X^p(x) \,\big|\, y - \Pi_C^p(x)\big\rangle \ge 0\,. \end{aligned}$$

Due to proposition 1.10 (e) we see that for $x \neq 0$ and $\Pi_C^p(x) \neq 0$

$$\begin{aligned} &\big\langle J^r\big(\lambda_x \Pi_C^p(x)\big) - J^r(x) \,\big|\, \lambda_x y - \lambda_x \Pi_C^p(x)\big\rangle \\ &= \lambda_x \big\langle \lambda_x^{r-1} \|\Pi_C^p(x)\|^{r-p} J^p\big(\Pi_C^p(x)\big) - \|x\|^{r-p} J^p(x) \,\big|\, y - \Pi_C^p(x)\big\rangle \\ &= \lambda_x \|x\|^{r-p} \big\langle J^p\big(\Pi_C^p(x)\big) - J^p(x) \,\big|\, y - \Pi_C^p(x)\big\rangle \ge 0\,. \end{aligned}$$

Moreover $\Pi_C^r(0) = \min_{y \in C} \|y\| = \Pi_C^p(0)$ by (b) of this proposition and if $\Pi_C^p(x) = 0$ and $x \neq 0$ then for all $y \in C$

$$\langle J^r(0) - J^r(x) \,|\, y - 0\rangle = \|x\|^{r-p} \langle J^p(0) - J^p(x) \,|\, y - 0\rangle \ge 0\,,$$

which proves the first part of (g). Let x be in $\lambda_x C$. Then $x = \Pi_{\lambda_x C}^r(x) = \lambda_x \Pi_C^p(x)$. If $x = 0$ or $\Pi_C^p(x) = 0$ then $\lambda_x = 1$ and therefore $x = \Pi_C^p(x) \in C$. Otherwise we get

$$\|x\| = \lambda_x \|\Pi_C^p(x)\| = \left(\frac{\|x\|}{\|\Pi_C^p(x)\|}\right)^{\frac{r-p}{r-1}} \|\Pi_C^p(x)\|\,,$$

which gives $\|x\| = \|\Pi_C^p(x)\|$. Hence $\lambda_x = 1$ and $x = \Pi_C^p(x) \in C$. □

By the variational inequality (1.16) we can show that the metric projections of points along the ray

$$z(\lambda) := P_C(x) + \lambda\big(x - P_C(x)\big) \quad , \quad \lambda \ge 0\,,\, x \in X\,,$$

coincide (compare (f) of the above proposition). A useful translation-property of metric projections, which Bregman projections in general do not share, is contained in the following corollary.

Corollary 1.27. *For $z \in X$ we have*

$$P_{z+C}(x) = z + P_C(x - z)\,.$$

Proof. By (b) of the preceeding proposition we get

$$P_{z+C}(x) = x + \Pi_{z+C-x}^p(0) = z + (x - z) + \Pi_{C-(x-z)}^p(0) = z + P_C(x - z)\,.$$

□

In analogy to Hilbert spaces we characterize Bregman projections onto closed affine subspaces and prove a related decomposition theorem, in which the metric and the Bregman projection play a complementary role (compare with [4]). This enables us to use the same iterative scheme to compute metric as well as Bregman projections onto affine subspaces which are given via the nullspace or the range of a linear operator. To simplify notation we just write Π instead of Π^p for the Bregman projection in X with respect to the function $f(x) = \frac{1}{p}\|x\|^p$ and Π^* for the Bregman projection in the dual X^* with respect to the conjugate function $f^*(x^*) = \frac{1}{q}\|x^*\|^q$.

Lemma 1.28. *In a reflexive, smooth and strictly convex Banach space X the sets $J(C) \subset X^*$ are closed for every $C \in \mathcal{C}(X)$.*

Proof. If the sequence $\big(J(x_n)\big)_n$ with $x_n \in C$ converges to some $x^* \in X^*$ then the sequence $(x_n)_n = \Big(J^*\big(J(x_n)\big)\Big)_n$ converges weakly to $x := J^*(x^*)$ because J^* is norm-to-weak-continuous by 1.14 (f). Therefore x lies in C and $x^* = J(x)$ by 1.14 (e). □

Proposition 1.29. *Let X be a reflexive, smooth and strictly convex Banach space, $U \subset X$ a closed subspace and $x, y, z \in X$ be given.*

(a) If we write $x^* = J(x) \Leftrightarrow x = J^*(x^*)$ and analogously for y and z then the following assertions are equivalent:
- *(i) $x = \Pi_{z+U}(y)$,*
- *(ii) $x - z \in U$ and $J(x) - J(y) \in U^\perp$,*
- *(iii) $x^* = \Pi^*_{y^*+U^\perp}(z^*)$,*
- *(iv) $x^* - y^* \in U^\perp$ and $J^*(x^*) - J^*(z^*) \in U$.*

(b) X can be decomposed into the "orthogonal sum"

$$X = U \oplus J^*(U^\perp)\,,$$

i.e. every $x \in X$ can be uniquely written in the form

$$x = x_U + J^*(x^*_{U^\perp})$$

with $x_U \in U$ and $x^*_{U^\perp} \in U^\perp$. More precisely we have

$$x_U = P_U(x) \quad \text{and} \quad x^*_{U^\perp} = \Pi^*_{U^\perp}\big(J(x)\big)\,.$$

Proof. Since U is a subspace, x is the Bregman projection of y onto $z + U$ iff $x \in z + U$ (i.e. $x = z + u_x$ for some $u_x \in U$) and

$$\begin{aligned}
&\langle J(x) - J(y)\,|\,v - x\rangle \geq 0 \quad \text{for all} \quad v = z + u_v \in z + U \\
\Leftrightarrow\ &\langle J(x) - J(y)\,|\,(z + u_v) - (z + u_x)\rangle \geq 0 \quad \text{for all} \quad u_v \in U \\
\Leftrightarrow\ &\langle J(x) - J(y)\,|\,u\rangle \geq 0 \quad \text{for all} \quad u \in U \\
\Leftrightarrow\ &\langle J(x) - J(y)\,|\,u\rangle = 0 \quad \text{for all} \quad u \in U\,,
\end{aligned}$$

and by the definition of $U^\perp$ this is fulfilled iff $J(x) - J(y) \in U^\perp$. This proves (i) $\Leftrightarrow$ (ii). (iv) is just an equivalent reformulation of (ii) and since $(U^\perp)^\perp = U$ it follows that (iii) $\Leftrightarrow$ (iv). By proposition 1.26 (b) we have

$$P_U(x) - x = \Pi_{U-x}(0) = J^*\Big(J\big(\Pi_{-x+U}(0)\big)\Big).$$

By part (a) of this proposition we see that

$$J\big(\Pi_{-x+U}(0)\big) = \Pi^*_{U^\perp}\big(J(-x)\big)$$

and we arrive at $x = P_U(x) + J^*\Big(\Pi^*_{U^\perp}\big(J(x)\big)\Big)$. The uniqueness of the decomposition also follows by part (a) of this proposition. □

To justify a little more the notation "orhogonal sum", we point out that by lemma 1.28 $J^*(U^\perp)$ is a closed subset of X and that if $x \in U \cap J^*(U^\perp)$ then $\|x\|^p = \langle J(x) \,|\, x\rangle = 0$ and therefore $x = 0$. Of course we have $\langle x \,|\, J(y)\rangle = 0$ for all $x \in U$ and $y \in J^*(U^\perp)$ but in general it need *not* be that $\langle J(x) \,|\, y\rangle = 0$ as well.
In the next proposition we give a few examples; these shall also demonstrate that the metric projection and the Bregman projection Π^p sometimes might coincide for all choices of p, sometimes only for special choices of p, but that they differ in general (compare (b) with [15]).

Proposition 1.30. *Let X be reflexive, smooth and strictly convex.*

(a) The metric and the Bregman projection onto the ball around the origin with radius $c > 0$ coincide and are given by

$$\Pi^p_{B_c}(x) = P_{B_c}(x) = \lambda_x x \quad \textit{with} \quad \lambda_x = 1 \wedge \frac{c}{\|x\|}. \tag{1.36}$$

(b) Let $H(u^*, \alpha)$ be a hyperplane and for $x \in X$ let $h_x : \mathbb{R} \longrightarrow \mathbb{R}$ be the strictly convex, differentiable function

$$h_x(t) := \frac{1}{q}\,\|J^p_X(x) - tu^*\|^q + \alpha\, t \tag{1.37}$$

with continuous, strictly increasing derivative

$$h'_x(t) = -\left\langle u^* \,\middle|\, J^q_{X^*}\big(J^p_X(x) - tu^*\big)\right\rangle + \alpha. \tag{1.38}$$

The Bregman projection of x onto $H(u^*, \alpha)$ is then given by

$$\Pi^p_{H(u^*,\alpha)}(x) = J^q_{X^*}\big(J^p_X(x) - t_0 u^*\big), \tag{1.39}$$

whereby t_0 is the (necessarily existing) unique solution of the one-dimensional optimization problem

$$\min_{t \in \mathbb{R}} h_x(t). \tag{1.40}$$

Moreover if x is not contained in the halfspace $H_{\leq}(u^*, \alpha)$ then the Bregman projection of x onto $H_{\leq}(u^*, \alpha)$ is also given by

$$\Pi^p_{H_{\leq}(u^*,\alpha)}(x) = J^q_{X^*}\big(J^p_X(x) - t_0 u^*\big)\,, \tag{1.41}$$

whereby t_0 is then the necessarily positive solution of (1.40).

(c) The metric projection onto a hyperplane $H(u^*, \alpha)$ is given by

$$P_{H(u^*,\alpha)}(x) = x - \frac{\langle u^* \,|\, x\rangle - \alpha}{\|u^*\|^q} J^q_{X^*}(u^*)\,.$$

Moreover if x is not contained in the halfspace $H_{\leq}(u^*, \alpha)$ then $P_{H_{\leq}(u^*,\alpha)}(x)$ is also given by this formula.

(d) If X is an L_p-space ($1 < p < \infty$) and $[a, b] := \{x \in L_p \,|\, a \leq x \leq b\}$ is a closed "interval" with $0 \in [a, b]$ for extended real valued a, b[5] then the metric and the Bregman projection $\Pi^p_{[a,b]}$ onto $[a, b]$ coincide and are given by

$$\Pi^p_{[a,b]}(x) = P_{[a,b]}(x) = (a \vee x) \wedge b = a \vee (x \wedge b)\,, \tag{1.42}$$

whereby the "p" in "L_p" and "$\Pi^p_{[a,b]}$" are the same.

Proof. If $x \in B_c$ then $\Pi^p_{B_c}(x) = x$. If $\|x\| > c$ then $\lambda_x = \frac{c}{\|x\|} < 1$, $\|\lambda_x x\| = c$, i.e. $\lambda_x x \in B_c$ and for all $y \in B_c$ we have

$$\big\langle J^p\big(\lambda_x x\big) - J^p(x) \,\big|\, y - \lambda_x x\big\rangle = \big(\lambda_x^{p-1} - 1\big) \langle J^p(x) \,|\, y - \lambda_x x\rangle \geq 0\,,$$

because $\lambda_x^{p-1} - 1 < 0$ and

$$\begin{aligned}\langle J^p(x) \,|\, y - \lambda_x x\rangle &= \langle J^p(x) \,|\, y\rangle - \lambda_x \langle J^p(x) \,|\, x\rangle \\ &\leq \|x\|^{p-1}\|y\| - \lambda_x\|x\|^p \\ &= \|x\|^{p-1}(\|y\| - c) \leq 0\,,\end{aligned}$$

which proves (a) for the Bregman projection; the proof for the metric projection is analogous by using the variational inequality (1.16). In (b) differentiability and strict convexity of h_x and continuity of h'_x are consequences of 1.14 (a), (b) and (f). We show that h'_x is strictly increasing. For $t \in \mathbb{R}$ we set $x^*(t) := J(x) - tu^*$ and for $t_0, t \in \mathbb{R}$ with $t > t_0$ we get

$$\begin{aligned}h'(t) - h'(t_0) &= -\big\langle u^* \,\big|\, J^*\big(x^*(t)\big) - J^*\big(x^*(t_0)\big)\big\rangle \\ &= \frac{1}{t - t_0} \big\langle x^*(t) - x^*(t_0) \,\big|\, J^*\big(x^*(t)\big) - J^*\big(x^*(t_0)\big)\big\rangle > 0\end{aligned}$$

since J^* is strictly monotone by 1.14 (a). If we set $z := \frac{\alpha}{\|u^*\|^q} J^*(u^*)$ then we can write

$$H(u^*, \alpha) = z + H(u^*, 0)$$

[5] a and b need not be themselves elements of L_p.

with $H(u^*,0) = \big(\operatorname{span}\{u^*\}\big)^{\perp}$ being a closed subspace. By proposition 1.29 (a) we get

$$\begin{aligned} & x_0 := \Pi_{H(u^*,\alpha)}(x) = \Pi_{z+H(u^*,0)}(x) \\ \Leftrightarrow\ & x_0 \in H(u^*,\alpha) \quad \text{and} \quad J(x_0) \in J(x) + \operatorname{span}\{u^*\} \qquad (1.43) \\ \Leftrightarrow\ & J(x_0) = \Pi^*_{J(x)+\operatorname{span}\{u^*\}}\big(J(z)\big)\,. \qquad (1.44) \end{aligned}$$

(1.43) is equivalent to

$$x_0 = J^*\big(J(x) - t_0 u^*\big)$$

with $t_0 \in \mathbb{R}$ such that

$$0 = \alpha - \big\langle u^* \,\big|\, J^*\big(J(x) - t_0 u^*\big)\big\rangle = h'_x(t_0)$$

and (1.44) is then equivalent to t_0 being a solution of the optimization problem

$$\begin{aligned} & \min_{t\in\mathbb{R}} \left(\Delta^*\big(J(z), J(x) - tu^*\big) = \frac{1}{p}\|z\|^p - \langle z \,|\, J(x) - tu^*\rangle + \frac{1}{q}\|J(x) - tu^*\|^q \right) \\ \Leftrightarrow\ & \min_{t\in\mathbb{R}} \left(\alpha t + \frac{1}{q}\|J(x) - tu^*\|^q = h_x(t) \right)\,. \end{aligned}$$

Existence and uniqueness of t_0 are guaranteed by the existence and uniqueness of the Bregman projection. If x is not contained in the halfspace $H_{\le}(u^*,\alpha)$, i.e. $\langle u^* \,|\, x\rangle > \alpha$, then the Bregman projection $\tilde{x}_0$ of x onto $H_{\le}(u^*,\alpha)$ lies in the boundary $H(u^*,\alpha)$ of $H_{\le}(u^*,\alpha)$ by 1.26 (a) and thus $\tilde{x}_0$ coincides with x_0. Moreover the solution t_0 of (1.40) must be positive since $h'_x(0) = \alpha - \langle u^* \,|\, x\rangle < 0$ and h'_x is strictly increasing. For the metric projection we then get by 1.26 (b)

$$P_{H(u^*,\alpha)}(x) = x + \Pi_{H(u^*,\alpha)-x}(0) = x + \Pi_{H(u^*,\tilde{\alpha})}(0)$$

with $\tilde{\alpha} = \alpha - \langle u^* \,|\, x\rangle$. By what we have just shown for the Bregman projection we get

$$P_{H(u^*,\alpha)}(x) = x + J^*\big(J(0) - t_0 u^*\big) = x - t_0^q \operatorname{sgn}(t_0)\, J^*(u^*)$$

with

$$\begin{aligned} & 0 = \tilde{\alpha} - \big\langle u^* \,\big|\, J^*\big(J(0) - t_0 u^*\big)\big\rangle = \alpha - \langle u^* \,|\, x\rangle + t_0^q \operatorname{sgn}(t_0)\, \|u^*\|^q \\ \Leftrightarrow\ & t_0^q \operatorname{sgn}(t_0) = \frac{\langle u^* \,|\, x\rangle - \alpha}{\|u^*\|^q}\,. \end{aligned}$$

In (d) we at first point out that $[a,b]$ is indeed a closed convex subset of L_p, since convergence of a sequence in L_p implies the existence of a componentwise resp. pointwise almost everywhere convergent subsequence. We have $a \le 0 \le b$ since $0 \in [a,b]$. Therefore $(a \vee x) \wedge b = a \vee (x \wedge b)$ and

$$(a \vee x) \wedge b \ \le\ a \vee x \ \le\ 0 \vee x \ \le\ |x|$$

and

$$(a \vee x) \wedge b \;\geq\; (a \vee x) \wedge 0 \;\geq\; x \wedge 0 \;\geq\; -|x|\,,$$

from which we infer that $(a \vee x) \wedge b \in L_p$. By examples 1.1 (c) and 1.11 (a) it suffices to check the variational inequality (1.30) componentwise resp. pointwise almost everywhere:

$$\begin{aligned} &\Big(J^p\big(\Pi^p_{[a,b]}(x)\big) - J^p(x)\Big)\Big(y - \Pi^p_{[a,b]}(x)\Big) \\ &= \Big(|\Pi^p_{[a,b]}(x)|^{p-1}\operatorname{sgn}\big(\Pi^p_{[a,b]}(x)\big) - |x|^{p-1}\operatorname{sgn}(x)\Big)\Big(y - \Pi^p_{[a,b]}(x)\Big) \\ &= \begin{cases} \left(-|a|^{p-1} + |x|^{p-1}\right)(y-a) & , \quad x < a \\ 0 & , \quad a \leq x \leq b \\ \left(b^{p-1} - x^{p-1}\right)(y-b) & , \quad x > b \end{cases} \\ &\geq 0\,. \end{aligned}$$

Again the proof for the metric projection is analogous by using (1.16). □

1.3 Bounded Hausdorff Convergence and Continuity of the Projections

Another important topic is continuity of the projections with respect to the x-variable and how perturbations of the convex set C affect the projection. There are many notions of set convergence and we concentrate on *bounded Hausdorff convergence* [41] in $\mathcal{C}(X)$ because we find it convenient for our purposes. This notion of convergence generalizes convergence in the Hausdorff metric for unbounded sets by local versions of the Hausdorff metric and has also been used by Penot [39] in the context of metric projections. Alber [3] obtained results with respect to the Hausdorff metric and Resmerita [43] with respect to *Mosco convergence*. For more information about set convergence we refer the reader to [45] and the references cited therein.

1.3.1 Bounded Hausdorff Convergence

At first we recall the Hausdorff metric, which indicates how well two bounded closed convex sets "fit into each other". By $\mathcal{C}_b(X) \subset \mathcal{C}(X)$ we denote the set of all non-empty, bounded, closed and convex subsets of X and we write $B = B_1$ for the unit ball around the origin.

Proposition 1.31. *Let X be a reflexive Banach space. The mapping*

$$d : \mathcal{C}_b(X) \times \mathcal{C}_b(X) \longrightarrow [0, \infty)$$

defined by

$$d(C, D) := \min\left\{\lambda \geq 0 \,\middle|\, C \subset D + \lambda B \quad \text{and} \quad D \subset C + \lambda B\right\} \tag{1.45}$$

is called Hausdorff metric. *It is indeed a metric.*

Proof. Obviously we have $d(C,D) \geq 0$ and $d(C,D) = d(D,C)$ and the set on the right hand side of (1.45) is not empty for bounded sets C and D and thus $d(C,D)$ is finite. We show that the minimum in (1.45) indeed exists. If we set $\lambda := \inf \big\{ \lambda \geq 0 \,\big|\, C \subset D + \lambda B \text{ and } D \subset C + \lambda B \big\}$ for two sets $C, D \in \mathcal{C}_b(X)$ then there exists a sequence $(\lambda_n)_n$ of positive numbers greater than or equal to λ and converging to λ such that

$$C \subset D + \lambda_n B \quad \text{and} \quad D \subset C + \lambda_n B \quad \text{for all} \quad n \in \mathbb{N}\,.$$

So for an arbitrary element $x \in C$ and every $n \in \mathbb{N}$ we can find a $y_n \in D$ and a $b_n \in B$ with $x = y_n + \lambda_n b_n$. Since the sequences $(y_n)_n$ and $(b_n)_n$ lie in the bounded, closed and convex sets D and B and since X is assumed to be reflexive, by theorem 1.2 there exist subsequences $(y_{n_k})_k$ and $(b_{n_k})_k$ such that $(y_{n_k})_k$ converges weakly to an element $y \in D$ and $(b_{n_k})_k$ converges weakly to an element $b \in B$. Hence $x = y_{n_k} + \lambda_{n_k} b_{n_k}$ converges weakly to the element $x = y + \lambda b \in D + \lambda B$ and therefore $C \subset D + \lambda B$. Analoguously we can show that $D \subset C + \lambda B$ and conclude that λ is the minimum of the above set. From this it also follows that if $d(C,D) = 0$ then $C \subset D + 0B$ and $D \subset C + 0B$ and thus $C = D$. It remains to prove the triangle inequality. For $C, D, E \in \mathcal{C}_b(X)$ we have $C \subset E + d(C,E)B$ and $E \subset D + d(E,D)B$ which implies

$$C \subset D + \big(d(C,E) + d(E,D)\big)B\,.$$

In the same way we get

$$D \subset C + \big(d(D,E) + d(E,C)\big)B\,.$$

Since d is symmetric we conclude that $d(C,D) \leq d(C,E) + d(E,D)$. □

On the one hand having a metric is convenient, on the other hand some important classes of convex sets like e.g. cones are not bounded and thus we cannot directly measure their distance with the Hausdorff metric. Therefore we use local versions of the Hausdorff metric to extend the notion of convergence to unbounded sets. For $C, D \in \mathcal{C}(X)$ and $m \in \mathbb{N}$ we set

$$d_m(C,D) := \min \left\{ \lambda \geq 0 \,\middle|\, \begin{array}{rr} & C \cap B_m \subset D + \lambda B \\ \text{and} & D \cap B_m \subset C + \lambda B \end{array} \right\}. \tag{1.46}$$

and $\mathcal{C}_m(X) := \{C \in \mathcal{C}(X) \,|\, C \cap B_m \neq \emptyset\}$. Roughly speaking, we measure the Hausdorff distance on bounded parts and two sets will be close to each other if the distance is small on all these parts. Analogously to the case of the Hausdorff metric and by (c) of the following lemma the minimum in (1.46) indeed exists.

Lemma 1.32. *In a reflexive Banach space X the following is valid for all $C, D, E \in \mathcal{C}(X)$ and $m \in \mathbb{N}$:*

(a) $d_m(C,D) = d_m(D,C) \in [0,\infty)$ and $d_m(C,D) \leq d_{m+1}(C,D)$.

(b) $d_m(C,C) = 0$ *and if there exists some* $k_0 \in \mathbb{N}$ *such that* $d_k(C,D) = 0$ *for all* $k \geq k_0$ *then we have* $C = D$.

(c) The relation $\emptyset \neq C \cap B_m \subset D + \lambda B$ *for some* $\lambda \geq 0$ *implies* $D \cap B_{\tilde{m}} \neq \emptyset$ *and*

$$C \cap B_m \subset D \cap B_{\tilde{m}} + \lambda B \quad \textit{for every} \quad \tilde{m} \geq m + \lambda \tag{1.47}$$

(d) The "triangle inequality"

$$d_m(C,D) \leq \big(d_m(C,E) + d_{m_1}(E,D)\big) \vee \big(d_m(D,E) + d_{m_2}(E,C)\big) \tag{1.48}$$

is valid for every $m_1 \geq m + d_m(C,E)$ *and* $m_2 \geq m + d_m(D,E)$.

Proof. Since $C \cap B_m$ and $D \cap B_m$ are bounded, d_m is finite. The rest of (a) and $d_m(C,C) = 0$ is obvious. If $k_0 \in \mathbb{N}$ is such that $d_k(C,D) = 0$ for all $k \geq k_0$, then

$$C \cap B_k \subset D \quad \text{and} \quad D \cap B_k \subset C \quad \text{for all} \quad k \geq k_0\,,$$

from which we infer that $C = D$. In (c) $\emptyset \neq C \cap B_m \subset D + \lambda B$ implies that every $x \in C \cap B_m$ can be written as $x = y + \lambda b$ with some $y \in D$ and $b \in B$. We get $y = x - \lambda b \in B_m + \lambda B$ and therefore $y \in D \cap B_{\tilde{m}} \neq \emptyset$ for every $\tilde{m} \geq m + \lambda$. To prove the triangle inequality we observe that by what we have just shown we get

$$C \cap B_m \subset E \cap B_{m_1} + d_m(C,E)B \subset \big(D + d_{m_1}(E,D)B\big) + d_m(C,E)B$$

for every $m_1 \geq m + d_m(C,E)$ and we arrive at

$$C \cap B_m \subset D + \big(d_m(C,E) + d_{m_1}(E,D)\big)B\,.$$

(If the intersection $C \cap B_m$ is empty then this also holds trivially.) Analogously we get

$$D \cap B_m \subset C + \big(d_m(D,E) + d_{m_2}(E,C)\big)B$$

for every $m_2 \geq m + d_m(D,E)$. □

Definition 1.33. *In a reflexive Banach space* X *a sequence* $(C_n)_n$ *in* $\mathcal{C}(X)$ *is said to be* boundedly convergent*, if there exists a* $C \in \mathcal{C}(X)$ *such that*

$$\lim_{n\to\infty} d_m(C, C_n) = 0 \quad \textit{for all} \quad m \in \mathbb{N}\,. \tag{1.49}$$

We examine some properties of boundedly convergent sequences and the relationship to the Hausdorff metric in case of bounded sets.

Proposition 1.34. *Let* X *be a reflexive Banach space.*

(a) The limit C *of a boundedly convergent sequence* $(C_n)_n$ *in* $\mathcal{C}(X)$ *is unique. Moreover if* C *is bounded then almost all* C_n *are uniformly bounded, i.e. there exist* $n_0, m_0 \in \mathbb{N}$ *such that* $C_n \subset B_{m_0}$ *for all* $n \geq n_0$.

(b) A sequence in $\mathcal{C}_b(X)$ converges boundedly to a bounded set $C \in \mathcal{C}_b(X)$ iff it converges with respect to the Hausdorff metric, i.e. for bounded sets these notions coincide.

(c) If $(x_n)_n$ is a bounded sequence in X with $x_n \in C_n$ for a boundedly convergent sequence $(C_n)_n \in \mathcal{C}(X)$, then all weak cluster points of the sequence $(x_n)_n$ lie in the limit C of the sequence $(C_n)_n$. Especially if $(x_n)_n$ is convergent, then its limit x lies in C.

(d) The sequence $(C_n + D_n)_n$ converges boundedly to $C + D$ whenever $(C_n)_n$ converges boundedly to C and $(D_n)_n$ converges boundedly to a bounded set D. Especially if $(x_n)_n$ converges to x in X then $(C_n + x_n)_n$ converges boundedly to $C + x$.

(e) The sequence $(\lambda_n C_n)_n$ converges boundedly to λC whenever $(C_n)_n$ converges boundedly to C and $(\lambda_n)_n$ converges in $\mathbb{R}$ to some $\lambda \neq 0$. Moreover $(\lambda_n C_n)_n$ converges boundedly to $\{0\}$ in case $\lambda = 0$ and C is bounded.

(f) If $(C_n)_n$ converges boundedly to C then there are $n_0, k \in \mathbb{N}$ such that $C, C_n \in \mathcal{C}_k(X)$ for all $n \geq n_0$.

Proof. Let $(C_n)_n$ be a boundedly convergent sequence in $\mathcal{C}(X)$. Suppose there are two limit sets C and $\tilde{C}$. For all $m \in \mathbb{N}$ there is then an $n_m \in \mathbb{N}$ such that for all $n \geq n_m$ we have $d_m(C, C_n) \leq 1$ and by 1.32 (a) and the triangle inequality (1.48) we get

$$d_m(C, \tilde{C}) \leq d_{m+1}(C, C_n) + d_{m+1}(C_n, \tilde{C}) \longrightarrow 0 \quad \text{for} \quad n \to \infty \,.$$

By lemma 1.32 (b) we conclude that $C = \tilde{C}$. If in addition C is bounded then w.l.o.g.

$$C \subset C + 2B \subset B_{m_0} \,. \tag{1.50}$$

Let $n_0 \in \mathbb{N}$ be such that $d_{m_0}(C, C_n) \leq \frac{1}{4}$ for all $n \geq n_0$. We show that all C_n with $n \geq n_0$ are contained in B_{m_0}. Our choice of n_0 and m_0 gives

$$C_n \cap B_{m_0} \subset C + \frac{1}{4}B \tag{1.51}$$

and

$$C = C \cap B_{m_0} \subset C_n + \frac{1}{4}B \,. \tag{1.52}$$

Suppose there is an $x \in C_n$ which is not an element of B_{m_0}. It can be easily seen that the set $C + \frac{1}{4}B$ is convex and closed and by proposition 1.3 we can therefore find a $y \in C + \frac{1}{4}B$ with

$$\|x - y\| = d := \min_{z \in C + \frac{1}{4}B} \|x - z\| \,. \tag{1.53}$$

Moreover $d \geq 1$ by (1.50) and since $x \notin B_{m_0}$. Hence $\lambda := \frac{1}{d + \frac{1}{2}} \in (0, 1)$. From (1.52) it follows that

$$C + \frac{1}{4}B \subset C_n + \frac{1}{2}B$$

and thus we find $\tilde{y} \in C_n$ and $b \in B$ such that $y = \tilde{y} + \frac{1}{2}b$. From this we also get $\tilde{y} = y - \frac{1}{2}b \subset C + \frac{3}{4}B$ and $\|y - \tilde{y}\| \leq \frac{1}{2}$. The element $z := \lambda x + (1-\lambda)\tilde{y}$ then lies in C_n since $x, \tilde{y} \in C_n$. And if we write $z = \tilde{y} + \lambda(x - \tilde{y})$, we see that $z \in C + 2B \subset B_{m_0}$ since $\tilde{y} \in C + \frac{3}{4}B$ and

$$\|\lambda(x - \tilde{y})\| \leq \frac{1}{d + \frac{1}{2}}\big(\|x - y\| + \|y - \tilde{y}\|\big) \leq \frac{1}{d + \frac{1}{2}}\left(d + \frac{1}{2}\right) = 1\,.$$

Together we get $z \in C_n \cap B_{m_0} \subset C + \frac{1}{4}B$ by (1.51). On the other hand we have

$$\|x - z\| = (1-\lambda)\|x - \tilde{y}\| \leq \frac{d - \frac{1}{2}}{d + \frac{1}{2}}\left(d + \frac{1}{2}\right) < d$$

and thus z cannot lie in $C + \frac{1}{4}B$ by (1.53), which leads to a contradiction. Another direct consequence of the preceeding considerations is, that a boundedly convergent sequence in $\mathcal{C}_b(X)$ with a bounded limit C is uniformly bounded and thus we can find an $m_0 \in \mathbb{N}$ with $C \subset B_{m_0}$ and $C_n \subset B_{m_0}$ for all $n \in \mathbb{N}$. Hence

$$d(C, C_n) = d_{m_0}(C, C_n) \longrightarrow 0 \quad \text{for} \quad n \to \infty\,.$$

Conversely if the sequence converges in the Hausdorff metric, then there is an $n_0 \in \mathbb{N}$ with $d(C, C_n) \leq 1$ for all $n \geq n_0$. This implies

$$C_n \subset C + d(C, C_n)B \subset C + B \subset B_{m_0}$$

for an m_0 with $C \subset B_{m_0 - 1}$. We get for all $m \geq m_0$

$$d_m(C, C_n) = d(C, C_n) \longrightarrow 0 \quad \text{for} \quad n \to \infty\,,$$

which together with 1.32 (a) proves (b). Let $(x_n)_n$ be contained in B_{m_0} and w.l.o.g. be weakly convergent to some $x \in X$ and $(C_n)_n$ converge boundedly to $C \in \mathcal{C}(X)$ with $x_n \in C_n$ for all $n \in \mathbb{N}$. To an arbitrary $k \in \mathbb{N}$ we thus find an $n_k \geq k$ such that

$$x_{n_k} \in C_{n_k} \cap B_{m_0} \subset C \cap B_{m_0+1} + \frac{1}{k}B$$

and therefore $x_{n_k} = y_{n_k} + \frac{1}{k}b_{n_k}$ with $y_{n_k} \in C \cap B_{m_0+1}$ and $b_{n_k} \in B$. These sequences are also bounded and thus have weakly convergent subsequences with weak limit points $y \in C \cap B_{m_0+1}$ resp. $b \in B$. Since $\frac{1}{k}b_{n_k}$ converges weakly to zero, we conclude that $x = y \in C \cap B_{m_0+1} \subset C$, from which assertion (c) follows. Now let $(C_n)_n$ converge boundedly to C. If $(D_n)_n$ converges boundedly to a bounded set D then by (a) and (b) of this proposition we find $m_0, n_0 \in \mathbb{N}$ such that D and all D_n are contained in B_{m_0} for all $n \geq n_0$ and $(D_n)_{n \geq n_0}$ converges in the Hausdorff metric to D. For all $m \in \mathbb{N}$ and $n \geq n_0$ we get

$$(C_n + D_n) \cap B_m \subset C_n \cap B_{m+m_0} + D_n \subset C + D + \big(d_{m+m_0}(C, C_n) + d(D, D_n)\big)B$$

and

$$(C+D)\cap B_m \subset C\cap B_{m+m_0}+D \subset C_n+D_n+\big(d_{m+m_0}(C,C_n)+d(D,D_n)\big)B\,.$$

Hence $d_m(C+D,C_n+D_n) \le d_{m+m_0}(C,C_n)+d(D,D_n) \longrightarrow 0$ for $n\to\infty$. If $\lim_{n\to\infty}\lambda_n=\lambda\neq 0$ then there are $m_0,n_0\in\mathbb{N}$ such that for all $n\ge n_0$ we have

$$|\lambda|,|\lambda_n|,\frac{1}{|\lambda|},\frac{1}{|\lambda_n|}\le m_0\,. \tag{1.54}$$

Let $m\in\mathbb{N}$ and $\epsilon>0$ be given. We choose $n_0\in\mathbb{N}$ such that for all $n\ge n_0$

$$|\lambda-\lambda_n|<\frac{\epsilon}{2m_0m}\qquad d_{m_0m}(C,C_n)<\frac{\epsilon}{2m_0}\,. \tag{1.55}$$

Let $y=\lambda_nx_n$ with $x_n\in C_n$ be an arbitrary element of $(\lambda_nC_n)\cap B_m$. By (1.54) and (1.55) we get $\|x_n\|=\frac{1}{|\lambda_n|}\|y\|\le m_0m$ and therefore

$$x_n\in C_n\cap B_{m_0m}\subset C+\frac{\epsilon}{2m_0}B$$

and

$$\|(\lambda_n-\lambda)x_n\|\le\frac{\epsilon}{2m_0m}m_0m=\frac{\epsilon}{2}\,.$$

This yields

$$y=\lambda_nx_n=\lambda x_n+(\lambda_n-\lambda)x_n\in\lambda\left(C+\frac{\epsilon}{2m_0}B\right)+\frac{\epsilon}{2}B$$

and since by (1.54)

$$\left|\lambda\frac{\epsilon}{2m_0}\right|\le\frac{\epsilon}{2}$$

we arrive at $y\in\lambda C+\epsilon B$. Hence

$$(\lambda_nC_n)\cap B_m\subset\lambda C+\epsilon B\,.$$

Analogously we get

$$(\lambda C)\cap B_m\subset\lambda_nC_n+\epsilon B\,.$$

In case C is bounded and $\lambda=0$ we may again assume $C_n\subset B_{m_0}$ for all $n\ge n_0$ and $(C_n)_{n\ge n_0}$ converges in the Hausdorff metric to C. Thus we get

$$\lambda_nC_n\subset\lambda_nB_{m_0}=\{0\}+\lambda_nm_0B$$

and

$$\{0\}=\{\lambda_nx_n-\lambda_nx_n\}\subset\lambda_nC_n+\lambda_nC_n\subset\lambda_nC_n+\lambda_nm_0B$$

for some $x_n\in C_n$. It follows that $d(\{0\},\lambda_nC_n)\le\lambda_nm_0\longrightarrow 0$ for $n\to\infty$. In (f) we find $k\in\mathbb{N}$ with $C\cap B_{k-1}\neq\emptyset$ and $n_0\in\mathbb{N}$ with $d_{k-1}(C,C_n)\le 1$ for all $n\ge n_0$. Lemma 1.32 (c) then ensures that $C_n\cap B_k\neq\emptyset$ for all $n\ge n_0$. $\square$

1.3.2 Continuity of the Projections

To formulate the desired continuity result we write $\Pi^p(x,C) := \Pi^p_C(x)$ and $\Pi^p_x(C) := \Pi^p(x,C)$ for $x \in X$ and $C \in \mathcal{C}(X)$.

Proposition 1.35. *In a uniformly smooth and uniformly convex Banach space X the mapping $\Pi^p : X \times \mathcal{C}(X) \longrightarrow X$ is uniformly continuous on bounded sets in the following sense: For all $n,k \in \mathbb{N}$ we can find an $m_0 \in \mathbb{N}$ such that for every $\epsilon > 0$ there exist $\delta > 0$ and $\tilde{\delta} > 0$ such that $\|\Pi^p(x,C) - \Pi^p(y,D)\| < \epsilon$ for all $x,y \in B_n$ with $\|x-y\| < \delta$ and for every $m \geq m_0$ and all $C,D \in \mathcal{C}_k(X)$ with $d_m(C,D) < \tilde{\delta}$.*

Proof. At first we recall that $C \in \mathcal{C}_k(X)$ means that $C \cap B_k \neq \emptyset$ and thus $\min_{y\in C} \|y\| \leq k$. By proposition 1.26 (b) and (d) for all $C \in \mathcal{C}_k(X)$ and $x \in B_n$ we get

$$\|\Pi^p(x,C)\| \leq \big(2^{q-1}\|x\|\big) \vee \left(3 \min_{y\in C} \|y\|\right) \leq \big(2^{q-1}n\big) \vee (3k)$$

and therefore $\Pi^p(x,C) \in B_{m_0}$ for an $m_0 \in \mathbb{N}$ with $m_0 \geq \big(2^{q-1}n\big) \vee (3k)$. Furthermore for $R := n^{p-1} \vee m_0 \vee m_0^{p-1}$ we have

$$\|J^p(x)\|\,,\ \|\Pi^p(x,C)\|\ ,\ \big\|J^p\big(\Pi^p(x,C)\big)\big\| \leq R \quad \text{for all} \quad x \in B_n\,,\ C \in \mathcal{C}_k(X)\,.$$

Let $\epsilon > 0$ be given. Since X is uniformly convex, by proposition 1.24 (h) we find a $\tilde{\delta} > 0$ such that

$$\|\tilde{x} - \tilde{y}\| < \epsilon \quad \text{for all} \quad \tilde{x},\tilde{y} \in B_{m_0} \quad \text{with} \quad \Delta_p(\tilde{x},\tilde{y}) < 6R\tilde{\delta}\,. \tag{1.56}$$

Since further J^p is uniformly continuous on bounded sets in a uniformly smooth X (Prop. 1.14 (d)), we find a $\delta > 0$ such that

$$\|J^p(x) - J^p(y)\| < \tilde{\delta} \quad \text{for all} \quad x,y \in B_n \quad \text{with} \quad \|x-y\| < \delta\,. \tag{1.57}$$

By proposition 1.24 (a) and (b) for such x and y we can estimate

$$\begin{aligned}
&\Delta_p\big(\Pi^p(x,C), \Pi^p(y,D)\big)\\
&\leq \big\langle J^p\big(\Pi^p(x,C)\big) - J^p\big(\Pi^p(y,D)\big) \,\big|\, \Pi^p(x,C) - \Pi^p(y,D)\big\rangle\\
&= -\big\langle J^p\big(\Pi^p(x,C)\big) - J^p(x) \,\big|\, \Pi^p(y,D) - \Pi^p(x,C)\big\rangle\\
&\quad + \langle J^p(x) - J^p(y) \,|\, \Pi^p(x,C) - \Pi^p(y,D)\rangle\\
&\quad - \big\langle J^p\big(\Pi^p(y,D)\big) - J^p(y) \,\big|\, \Pi^p(x,C) - \Pi^p(y,D)\big\rangle\,.
\end{aligned} \tag{1.58}$$

We estimate the first summand: For every $m \geq m_0$ and $C,D \in \mathcal{C}_k(X)$ with $d_m(C,D) < \tilde{\delta}$ we have

$$C \cap B_m \subset D + \tilde{\delta}B \quad \text{and} \quad D \cap B_m \subset C + \tilde{\delta}B\,.$$

Since $\Pi^p(y,D)$ lies in $D \cap B_{m_0} \subset D \cap B_m$ for $y \in B_n$, we therefore find $\tilde{y} \in C$ and $b \in B$ with $\Pi^p(y,D) = \tilde{y} + \tilde{\delta} b$. We keep in mind the validity of the variational inequality (1.30) for $\Pi^p(x,C)$ and get

$$\begin{aligned} &- \big\langle J^p\big(\Pi^p(x,C)\big) - J^p(x) \,\big|\, \Pi^p(y,D) - \Pi^p(x,C)\big\rangle \\ =& - \big\langle J^p\big(\Pi^p(x,C)\big) - J^p(x) \,\big|\, \tilde{y} - \Pi^p(x,C)\big\rangle \\ &- \Big\langle J^p\big(\Pi^p(x,C)\big) - J^p(x) \,\Big|\, \tilde{\delta} b\Big\rangle \\ \leq& \, \big\| J^p\big(\Pi^p(x,C)\big) - J^p(x)\big\| \, \tilde{\delta} \leq 2R\tilde{\delta}\,. \end{aligned}$$

Analogously the third summand in (1.58) can also be estimated from above by $2R\tilde{\delta}$. For the second summand we get by (1.57)

$$\begin{aligned} &+ \langle J^p(x) - J^p(y) \,|\, \Pi^p(x,C) - \Pi^p(y,D)\rangle \\ \leq& \, \|J^p(x) - J^p(y)\| \, \|\Pi^p(x,C) - \Pi^p(y,D)\| \\ \leq& \, \tilde{\delta} 2R\,. \end{aligned}$$

Altogether we arrive at $\Delta_p\big(\Pi^p(x,C), \Pi^p(y,D)\big) < 6R\tilde{\delta}$ and by (1.56) we conclude that $\|\Pi^p(x,C) - \Pi^p(y,D)\| < \epsilon$. □

This implies the following continuity results for the case that one of the variables is fixed (for the last part in (b) see also 1.34 (f)).

Corollary 1.36. *Let X be a uniformly smooth and uniformly convex Banach space.*

(a) For every $C \in \mathcal{C}(X)$ the mapping Π^p_C is uniformly continuous on bounded sets.

(b) For every $x \in X$ the mapping Π^p_x is uniformly continuous on $\mathcal{C}_k(X)$ for all $k \in \mathbb{N}$ in the sense that we can find an $m_0 \in \mathbb{N}$ such that for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\Pi^p_x(C) - \Pi^p_x(D)\| < \epsilon$ for every $m \geq m_0$ and all $C, D \in \mathcal{C}_k(X)$ with $d_m(C,D) < \delta$. Especially if a sequence $(C_n)_n$ in $\mathcal{C}(X)$ converges boundedly to $C \in \mathcal{C}(X)$ then $\big(\Pi^p_x(C_n)\big)_n$ converges to $\Pi^p_x(C)$.

By the relation $P_C(x) = x + \Pi^p_{C-x}(0)$ (1.26 (b)) and 1.34 (d) we obtain the same continuity results for the metric projection.

Corollary 1.37. *The assertions of 1.35 and 1.36 remain valid for the metric projection.*

2

SFP and Projections onto Affine Subspaces

In this chapter we develop and discuss the iteration methods for the solution of the *split feasibility problem* (SFP) and the computation of metric and Bregman projections onto affine subspaces. At first we examine what operators may be used in the iterative process to handle different kinds of constraints appearing in the SFP. The ones related to constraints in the range of a linear operator depend on a positive parameter which in general has to be chosen a posteriori. In section 2.2 we show how these parameters can be chosen in case of exact as well as approximate data to ensure convergence of the methods. In case of approximate or noisy data this choice is linked to a discrepancy priniple. The iteration methods for the SFP are analyzed in section 2.3. They produce sequences which in general have weak accumulation points that are solutions of the SFP. In the following section we show that the same iterative scheme can be used to compute metric and Bregman projections onto affine subspaces that are given via the nullspace or the range of a linear operator. For this case we can even prove strong convergence. In the last two sections we are concerned with possibilities to efficiently implement the methods. We show that the choice of parameters can be replaced by line searches and propose generalized conjugate gradient and sequential subspace methods for the computation of projections onto affine subspaces in case of exact data.

2.1 Convex Constraints and Related Operators

We intend to examine a little more closely the operators we will deal with. First we recall some facts about linear operators [25, 47]. By $\mathcal{L}(X,Y)$ we denote the Banach space of all continuous linear operators $A : X \longrightarrow Y$ endowed with the *operator norm*

$$\|A\| := \sup_{\|x\|\leq 1} \|Ax\| \,. \tag{2.1}$$

The *dual operator* $A^* \in \mathcal{L}(Y^*, X^*)$ is defined by

$$\langle A^* y^* \,|\, x \rangle := \langle y^* \,|\, Ax \rangle \quad \text{for all} \quad x \in X,\ y^* \in Y^* \tag{2.2}$$

and the equalities $\|A^*\| = \|A\|$ and $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$ are valid. In case X is reflexive we also have $\mathcal{N}(A) = \mathcal{R}(A^*)^\perp$ and $\mathcal{N}(A)^\perp = \overline{\mathcal{R}(A^*)}$ and in case Y is reflexive we also have $\mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$. An operator $A \in \mathcal{L}(X, Y)$ is called *compact*, if the image $A(B_X)$ of the unit ball of X is a relatively compact subset of Y. It is a fact that A is compact iff A^* is compact and that a compact operator A is weak-to-norm-continuous, i.e. if $(x_n)_n$ is a sequence in X which converges weakly to some $x \in X$, then $(Ax_n)_n$ converges strongly to Ax.

From now on we assume that **X** is a **smooth** and **uniformly convex** Banach space with a (bijective) duality mapping J_X with gauge function $t \mapsto t^{p-1}$. If J_Y is a set-valued duality mapping of another Banach space Y and we write "$J_Y(y)$" for some $y \in Y$, then we mean that $J_Y(y)$ is allowed to be any element in the set $J_Y(y)$. The **additional assumptions** in the following definition will be used for the different kinds of contraints in case of exact and approximate data.

Definition 2.1. *We call assumption*

(C) *X is uniformly smooth and a set $C \in \mathcal{C}(X)$ is given.*

(A, Q) *Given are: a uniformly smooth and uniformly convex Banach space Y with duality mapping J_Y (with gauge function $t \mapsto t^{r-1}$), a compact operator $0 \neq A \in \mathcal{L}(X, Y)$, a set $Q \in \mathcal{C}(Y)$ and a constant $\gamma \in (0, 1)$. The set*

$$M_{Ax \in Q} := \{x \in X \,|\, Ax \in Q\}$$

is not empty.

(A, y) *Given are: an arbitrary Banach space Y with duality mapping J_Y (with gauge function $t \mapsto t^{r-1}$), an operator $0 \neq A \in \mathcal{L}(X, Y)$, an element $y \in Y$ and a constant $\gamma \in (0, 1)$. The set*

$$M_{Ax = y} := \{x \in X \,|\, Ax = y\}$$

is not empty.

$(A, y, +)$ *Given are: an arbitrary Banach lattice Y with positive duality mapping J_+, an operator $0 \neq A \in \mathcal{L}(X, Y)$, an element $y \in Y$ and a constant $\gamma \in (0, 1)$. The set*

$$M_{Ax \le y} := \{x \in X \,|\, Ax \le y\}$$

is not empty.

(C_i) *In addition to assumption (C) a constant $\beta \in (0, 1)$ and convex sets $C_i \in \mathcal{C}(X)$ are given with*

$$d_m(C, C_i) \le \epsilon_i^m$$

and

$$\lim_{i \to \infty} \epsilon_i^m = 0 \quad \textit{for all} \quad m \in \mathbb{N}\,.$$

(A_j, Q_k) *In addition to assumption* (A, Q) *a constant* $\beta \in (0,1)$*, compact operators* $0 \neq A_j \in \mathcal{L}(X,Y)$ *and sets* $Q_k \in \mathcal{C}(Y)$ *are given with*

$$\|A - A_j\| \leq \eta_j \leq \eta_{j-1} \quad , \quad d_m(Q, Q_k) \leq \delta_k^m \leq \delta_{k-1}^m$$

and

$$\lim_{j\to\infty} \eta_j = 0 \quad , \quad \lim_{k\to\infty} \delta_k^m = 0 \quad \textit{for all} \quad m \in \mathbb{N}\,.$$

(A_j, y_k) *In addition to assumption* (A, y) *a constant* $\beta \in (0,1)$*, operators* $0 \neq A_j \in \mathcal{L}(X,Y)$ *and elements* $y_k \in Y$ *are given with*

$$\|A - A_j\| \leq \eta_j \leq \eta_{j-1} \quad , \quad \|y - y_k\| \leq \delta_k \leq \delta_{k-1}$$

and

$$\lim_{j\to\infty} \eta_j = 0 \quad , \quad \lim_{k\to\infty} \delta_k = 0\,.$$

$(A_j, y_k, +)$ *In addition to assumption* $(A, y, +)$ *the same holds as under assumption* (A_j, y_k).

Under assumption (C) *we define the operator*

$$T_C : X \longrightarrow X$$

by

$$T_C(x) := \Pi_C^p(x)\,. \tag{2.3}$$

Under assumption (A, Q) *we define for* $\mu > 0$ *the operators*

$$T_{A,Q,\Pi}^\mu\,,\, T_{A,Q,P}^\mu : X \longrightarrow X$$

by

$$T_{A,Q,\Pi}^\mu(x) := J_X^*\Big(J_X(x) - \mu A^*\Big(J_Y(Ax) - J_Y\big(\Pi_Q^r(Ax)\big)\Big)\Big)\,, \tag{2.4}$$

and

$$T_{A,Q,P}^\mu(x) := J_X^*\Big(J_X(x) - \mu A^* J_Y\big(Ax - P_Q(Ax)\big)\Big)\,, \tag{2.5}$$

whereby Π^r *is the Bregman projection and* P *is the metric projection in* Y*. For* $Q = \{y\}$ *with some* $y \in Y$ *under assumption* (A, y) *we get the (possibly set-valued) operator*

$$T_{A,y}^\mu : X \longrightarrow 2^X$$

with $\big(T_{A,\{y\},P}^\mu(x) = \big)$

$$T_{A,y}^\mu(x) := J_X^*\big(J_X(x) - \mu A^* J_Y(Ax - y)\big)\,. \tag{2.6}$$

Under assumption $(A, y, +)$ *we define for* $\mu > 0$ *the (possibly set-valued) operator*

$$T_{A,y,+}^\mu : X \longrightarrow 2^X$$

by

$$T_{A,y,+}^\mu(x) := J_X^*\Big(J_X(x) - \mu A^* J_+\big((Ax - y)_+\big)\Big)\,. \tag{2.7}$$

In Hilbert spaces $T^{\mu}_{A,y}$ and $T^{\mu}_{A,Q,P}$ are just the familiar operators

$$T^{\mu}_{A,y}(x) = x - \mu A^*(Ax - y) \quad \text{and} \quad T^{\mu}_{A,Q}(x) = x - \mu A^*\big(Ax - P_Q(Ax)\big)\,,$$

which appear in the ordinary Landweber methods and the CQ algorithm for the SFP. Operator $T^{\mu}_{A,Q,\Pi}$ may also be useful in the context of more general Bregman projections.

For an operator $T : X \longrightarrow 2^X$ we denote by

$$\mathrm{Fix}(T) := \{x \in X \,|\, x \in T(x)\}$$

the set of all *fixed points* of T and by

$$\text{S-Fix}(T) := \{x \in X \,|\, x = T(x)\}$$

the set of all *strong fixed points* of T. Obviously $\text{S-Fix}(T) \subset \mathrm{Fix}(T)$ and if T is single-valued then these sets coincide.

Proposition 2.2.

(a) Under assumption (C) *we have*

$$\mathrm{Fix}(T_C) = C\,.$$

(b) Under assumption (A, Q) *and for all* $\mu > 0$ *we have*

$$\mathrm{Fix}(T^{\mu}_{A,Q,\Pi}) = \mathrm{Fix}(T^{\mu}_{A,Q,P}) = M_{Ax\in Q}\,.$$

(c) Under assumption (A, y) *and for all* $\mu > 0$ *we have*

$$\mathrm{Fix}(T^{\mu}_{A,y}) = \text{S-Fix}(T^{\mu}_{A,y}) = M_{Ax=y}\,.$$

(d) Under assumption $(A, y, +)$ *and for all* $\mu > 0$ *we have*

$$\mathrm{Fix}(T^{\mu}_{A,y,+}) = \text{S-Fix}(T^{\mu}_{A,y,+}) = M_{Ax\le y}\,.$$

Proof. (a) is just 1.26 (a). For $x \in M_{Ax\in Q}$ we have $\Pi^r_Q(Ax) = Ax = P_Q(Ax)$ and thus $T^{\mu}_{A,Q,\Pi}(x) = x = T^{\mu}_{A,Q,P}(x)$. Hence $M_{Ax\in Q} \subset \mathrm{Fix}(T^{\mu}_{A,Q,\Pi})$ and $M_{Ax\in Q} \subset \mathrm{Fix}(T^{\mu}_{A,Q,P})$. Conversely for $x \in \mathrm{Fix}(T^{\mu}_{A,Q,\Pi})$ we get

$$\begin{aligned} & x = T^{\mu}_{A,Q,\Pi}(x) \\ \Leftrightarrow\ & J_X(x) = J_X(x) - \mu A^*\Big(J_Y(Ax) - J_Y\big(\Pi^r_Q(Ax)\big)\Big) \\ \Leftrightarrow\ & A^*\Big(J_Y(Ax) - J_Y\big(\Pi^r_Q(Ax)\big)\Big) = 0\,. \end{aligned}$$

Since $M_{Ax\in Q}$ is supposed to be non-empty, we take some $z \in X$ with $Az \in Q$ and get

$$\begin{aligned}
0 &= \Big\langle A^*\Big(J_Y(Ax) - J_Y\big(\varPi^r_Q(Ax)\big)\Big) \Big| \, x - z\Big\rangle \\
&= \big\langle J_Y(Ax) - J_Y\big(\varPi^r_Q(Ax)\big) \,\big|\, Ax - Az\big\rangle \\
&= \big\langle J_Y(Ax) - J_Y\big(\varPi^r_Q(Ax)\big) \,\big|\, Ax - \varPi^r_Q(Ax)\big\rangle \\
&\quad + \big\langle J_Y(Ax) - J_Y\big(\varPi^r_Q(Ax)\big) \,\big|\, \varPi^r_Q(Ax) - Az\big\rangle \\
&\geq \big\langle J_Y(Ax) - J_Y\big(\varPi^r_Q(Ax)\big) \,\big|\, Ax - \varPi^r_Q(Ax)\big\rangle \,,
\end{aligned}$$

because of the validity of the variational inequality (1.30) for $\varPi^r_Q(Ax)$ and $Az \in Q$. Since Y is strictly convex by 1.14 (a) the above inequality gives $Ax = \varPi^r_Q(Ax) \in Q$. The inclusion $\mathrm{Fix}(T^\mu_{A,Q,P}) \subset M_{Ax\in Q}$ can be shown similarly. In (b) it suffices to show

$$\mathrm{Fix}(T^\mu_{A,y}) \subset M_{Ax=y} \subset \text{S-Fix}(T^\mu_{A,y})\,,$$

because $\text{S-Fix}(T^\mu_{A,y}) \subset \mathrm{Fix}(T^\mu_{A,y})$. If $x \in M_{Ax=y}$ then we have $J_y(Ax-y) = 0$ and it follows that $M_{Ax=y} \subset \text{S-Fix}(T^\mu_{A,y})$. Conversely for $x \in \mathrm{Fix}(T^\mu_{A,y})$ we find some $u^* \in J_Y(Ax-y)$ such that

$$x = T^\mu_{A,y}(x) \;\Leftrightarrow\; J_X(x) = J_X(x) - \mu A^* J_Y(Ax-y) \;\Leftrightarrow\; A^* J_Y(Ax-y) = 0\,.$$

Since $M_{Ax=y}$ is supposed to be non-empty, we take some $z \in X$ with $Az = y$ and get

$$0 = \langle A^* J_Y(Ax-y) \,|\, x - z\rangle = \langle J_Y(Ax-y) \,|\, Ax - y\rangle = \|Ax - y\|^r\,,$$

which gives $Ax = y$ and thus $\mathrm{Fix}(T^\mu_{A,y}) \subset M_{Ax=y}$. In (c) it again suffices to show

$$\mathrm{Fix}(T^\mu_{A,y,+}) \subset M_{Ax\leq y} \subset \text{S-Fix}(T^\mu_{A,y,+})\,.$$

If $Ax \leq y$ then we get $(Ax-y)_+ = 0$ and thus $J_+\big((Ax-y)_+\big) = 0$, which yields $T^\mu_{A,y,+}(x) = x$. Hence $M_{Ax\leq y} \subset \text{S-Fix}(T^\mu_{A,y,+})$. Conversely for $x \in \mathrm{Fix}(T^\mu_{A,y,+})$ we find some $u^* \in J_+\big((Ax-y)_+\big)$ such that

$$x = J^*_X\big(J_X(x) - \mu A^* u^*\big) \quad\Leftrightarrow\quad J_X(x) = J_X(x) - \mu A^* u^* \quad\Leftrightarrow\quad A^* u^* = 0\,.$$

Since $M_{Ax\leq y}$ is supposed to be non-empty, we find some $z \in X$ with $y - Az \geq 0$ and by the properties of the positive duality mapping 1.21 we get

$$\begin{aligned}
0 &= \langle A^* u^* \,|\, x - z\rangle \\
&= \langle u^* \,|\, Ax - y\rangle + \langle u^* \,|\, y - Az\rangle \\
&= \langle u^* \,|\, (Ax-y)_+\rangle + \langle u^* \,|\, y - Az\rangle \\
&= \|(Ax-y)_+\|^2 + \langle u^* \,|\, y - Az\rangle \\
&\geq \|(Ax-y)_+\|^2\,,
\end{aligned}$$

from which we infer that $Ax - y \leq 0$ and thus $\mathrm{Fix}(T^\mu_{A,y,+}) \subset M_{Ax\leq y}$. □

The operators are also linked to subdifferentials of certain functionals.

Proposition 2.3. *We assume* (A, Q), (A, y) *or* $(A, y, +)$ *and accordingly define the functions* $f_{A,Q,P}$, $f_{A,y}$, $f_{A,y,+} : X \longrightarrow \mathbb{R}$ *by*

$$
\begin{aligned}
f_{A,Q,P}(x) &:= \frac{1}{r} \|Ax - P_Q(Ax)\|^r , \\
f_{A,y}(x) &:= \frac{1}{r} \|Ax - y\|^r , \\
f_{A,y,+}(x) &:= \frac{1}{2} \|(Ax - y)_+\|^2 .
\end{aligned}
$$

Then we have for all $x \in X$

$$
\begin{aligned}
A^* J_Y\big(Ax - P_Q(Ax)\big) \quad &\subset \quad \partial f_{A,Q,P}(x) , \\
A^* J_Y(Ax - y) \quad &\subset \quad \partial f_{A,y}(x) , \\
A^* J_+\big((Ax - y)_+\big) \quad &\subset \quad \partial f_{A,y,+}(x) .
\end{aligned}
$$

Proof. The assertions for $f_{A,y}$ and $f_{A,y,+}$ follow immediately by 1.12 and 1.22. We prove the assertion for $f_{A,Q,P}$. For all $x, y \in X$ we get by 1.12 and the variational inequality for the metric projection (1.16)

$$
\begin{aligned}
& f_{A,Q,P}(y) - f_{A,Q,P}(x) \\
= & \frac{1}{r} \|Ay - P_Q(Ay)\|^r - \frac{1}{r} \|Ax - P_Q(Ax)\|^r \\
\geq & \big\langle J_Y\big(Ax - P_Q(Ax)\big) \,\big|\, \big(Ay - P_Q(Ay)\big) - \big(Ax - P_Q(Ax)\big) \big\rangle \\
= & \big\langle A^* J_Y\big(Ax - P_Q(Ax)\big) \,\big|\, y - x \big\rangle \\
& + \big\langle J_Y\big(Ax - P_Q(Ax)\big) \,\big|\, P_Q(Ax) - P_Q(Ay) \big\rangle \\
\geq & \big\langle A^* J_Y\big(Ax - P_Q(Ax)\big) \,\big|\, y - x \big\rangle .
\end{aligned}
$$

□

We do not know whether operator $T^{\mu}_{A,Q,\Pi}$ is also linked to a subdifferential of a functional $f_{A,Q,\Pi}$. The canonical candidates $f_{A,Q,\Pi}(x) = \Delta_r\big(Ax, \Pi^r_Q(Ax)\big)$ or $f_{A,Q,\Pi}(x) = \Delta_r\big(\Pi^r_Q(Ax), Ax\big)$ do not seem to work (do they?). However by 1.24 (e) for fixed $z \in X$ we have

$$
A^* \Big(J_Y(Az) - J_Y\big(\Pi^r_Q(Az)\big) \Big) \quad \subset \quad \partial f_{A,Q,\Pi^r_Q(Az)}(z)
$$

with

$$
f_{A,Q,\Pi^r_Q(Az)}(x) = \Delta_r\big(\Pi^r_Q(Az), Ax\big) \quad , \quad x \in X .
$$

2.2 Choice of Parameters

The methods we will discuss are all based on an iterative scheme

$$x_{n+1} = T_n(x_n)\,,$$

whereby each T_n is one of the operators introduced in 2.1. The key for proving the convergence of these methods is a monotonicity estimate of the form

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - S_n(x_n)$$

with a remainder term $S_n(x_n) \ge 0$ and for all z in a subset of the fixed points of T_n. This relation results in $(x_n)_n$ having (weak) cluster points and $\big(S_n(x_n)\big)_n$ converging to zero, whereby $S_n(x_n)$ is of such a form that this eventually forces the cluster points to be the sought after solutions. In the following "cluster" of lemmas we derive these estimates.

Lemma 2.4. *We assume* (C). *Let* $x_n \in X$ *for some* $n \in \mathbb{N}$ *be given. We set* $x_{n+1} := T_C(x_n)$ *and*

$$R_C(x_n) := \Delta_p(x_n, x_{n+1})\,. \tag{2.8}$$

Then we have $x_n \in C \Leftrightarrow R_C(x_n) = 0$ *and the following estimate is valid for all* $z \in C$*:*

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - R_C(x_n)\,. \tag{2.9}$$

Proof. This is just a reformulation and direct consequence of (1.31), 1.26 (a) and 1.24 (a). □

In case of approximate data $(C_i)_i$ we must also adjust the choice of the sets C_i.

Lemma 2.5. *We assume* (C_i) *and that* $C \cap B_{m_0}$ *is not empty for some* $m_0 \in \mathbb{N}$. *Let* $x_n \in X$ *and* $i_{n-1} \in \mathbb{N}$ *for some* $n \in \mathbb{N}$ *be given. We set*

$$R_{C_i}(x_n) := \Delta_p\left(x_n, \Pi^p_{C_i}(x_n)\right)\,. \tag{2.10}$$

If for all $i > i_{n-1}$

$$R_{C_i}(x_n) \le \frac{1}{\beta}\epsilon_i^{m_0}\left\|J^p\big(\Pi^p_{C_i}(x_n)\big) - J^p(x_n)\right\| \tag{2.11}$$

then x_n *lies in* C. *In this case we choose* $i_n > i_{n-1}$ *and set* $x_{n+1} := x_n$. *Otherwise we find* $i_n > i_{n-1}$ *with*

$$0 \le \epsilon_{i_n}^{m_0}\left\|J^p\big(\Pi^p_{C_{i_n}}(x_n)\big) - J^p(x_n)\right\| < \beta R_{C_{i_n}}(x_n)\,. \tag{2.12}$$

In this case we set $x_{n+1} := T_{C_{i_n}}(x_n)$ *and the following estimate is valid for all* $z \in C \cap B_{m_0}$*:*

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - (1-\beta)R_{C_{i_n}}(x_n)\,. \tag{2.13}$$

Proof. Since $(C_i)_i$ converges boundedly to C and X is uniformly smooth and uniformly convex, corollary 1.36 (b) ensures that $\left(\Pi^p_{C_i}(x_n)\right)_i$ converges to $\Pi^p_C(x_n)$ for $i \to \infty$. Therefore $\left\|J^p\left(\Pi^p_{C_i}(x_n)\right) - J^p(x_n)\right\|$ remains bounded and the right hand side in (2.11) converges to zero for $i \to \infty$ and so does $R_{C_i}(x_n) = \Delta_p\left(x_n, \Pi^p_{C_i}(x_n)\right)$. By 1.24 (h) in a uniformly convex X we conclude that the sequence $\left(\Pi^p_{C_i}(x_n)\right)_i$ converges to x_n. Since it also converges to $\Pi^p_C(x_n)$, we get $x_n = \Pi^p_C(x_n) \in C$.
In case of (2.12) we set $x_{n+1} := T_{C_{i_n}}(x_n)$. Since $d_{m_0}(C, C_{i_n}) \le \epsilon^{m_0}_{i_n}$, for every $z \in C \cap B_{m_0}$ we can find some $z_{i_n} \in C_{i_n}$ with $\|z - z_{i_n}\| \le \epsilon^{m_0}_{i_n}$. By (1.31) and (2.12) we get

$$\begin{aligned}
\Delta_p(x_{n+1}, z) &= \Delta_p(x_{n+1}, z_{i_n}) + \frac{1}{p}\left(\|z\|^p - \|z_{i_n}\|^p\right) + \langle J^p(x_{n+1}) \,|\, z_{i_n} - z\rangle \\
&\le \Delta_p(x_n, z_{i_n}) - \Delta_p(x_n, x_{n+1}) \\
&\quad + \frac{1}{p}\left(\|z\|^p - \|z_{i_n}\|^p\right) + \langle J^p(x_{n+1}) \,|\, z_{i_n} - z\rangle \\
&= \Delta_p(x_n, z) - \Delta_p(x_n, x_{n+1}) + \langle J^p(x_{n+1}) - J^p(x_n) \,|\, z_{i_n} - z\rangle \\
&\le \Delta_p(x_n, z) - \Delta_p(x_n, x_{n+1}) + \epsilon^{m_0}_{i_n}\|J^p(x_{n+1}) - J^p(x_n)\| \\
&\le \Delta_p(x_n, z) - (1-\beta) R_{C_{i_n}}(x_n)\,.
\end{aligned}$$

□

The other operators all depend on a parameter $\mu > 0$. The monotonicity estimate does not hold for all μ, but we show that there always exists a $\mu_n > 0$ for which the relation is valid. This parameter choice is linked to the modulus of smoothness of the dual X^* because we use the characteristic inequality (prop. 1.17) to estimate from above terms of the form

$$\frac{1}{q}\left\|J_X(x_n) - \mu_n A^*_{j_n} w^*_n\right\|^q \,.$$

Therefore these lemmas are quite technical. Later on we will give an example of how these parameters may look like if we have at hand a concrete version of the characteristic inequality as in 1.18. Even better, we will see that it suffices to know that these parameters exist and that we can replace their choice by line searches. The following lemma covers a situation occurring in the consecutive proofs.

Lemma 2.6. *Let X be a uniformly convex Banach space with duality mapping J (with gauge function $t \mapsto t^{p-1}$) and $\tilde{\sigma}_q$ be the function (1.21) appearing in proposition 1.17 for the characteristic inequality of the uniformly smooth dual X^*. If $0 \ne x \in X$, $0 \ne A \in \mathcal{L}(X,Y)$ and $0 \ne y^* \in Y^*$ with an arbitrary Banach space Y are given and $\mu > 0$ is defined by*

$$\mu := \frac{\tau}{\|A\|}\frac{\|x\|^{p-1}}{\|y^*\|} \quad \textit{for some} \quad \tau \in (0,1] \tag{2.14}$$

then the following estimate is valid:

$$\frac{1}{q}\tilde{\sigma}_q\big(J(x),\mu A^* y^*\big) \le 2^q G_q \|x\|^p \rho_{X^*}(\tau)\,, \tag{2.15}$$

whereby G_q is the constant appearing in (1.21) and ρ_{X^*} is the modulus of smoothness of X^*.

Proof. According to (1.21) we have $\frac{1}{q}\tilde{\sigma}_q\big(J(x),\mu A^* y^*\big) =$

$$G_q \int_0^1 \frac{\big(\|J(x) - t\mu A^* y^*\| \vee \|J(x)\|\big)^q}{t} \rho_{X^*}\left(\frac{t\mu\|A^* y^*\|}{\|J(x) - t\mu A^* y^*\| \vee \|J(x)\|}\right) \mathrm{d}t\,.$$

By the choice of μ (2.14) and $\tau \in (0,1]$ we estimate for every $t \in [0,1]$

$$\|J(x) - t\mu A^* y^*\| \le \|x\|^{p-1} + \mu\|A\|\,\|y^*\| \le 2\|x\|^{p-1}$$

and get

$$\|J(x) - t\mu A^* y^*\| \vee \|J(x)\| \begin{cases} \le 2\|x\|^{p-1} \\ \ge \|x\|^{p-1} \end{cases}.$$

Since ρ_{X^*} is nondecreasing $\big($prop. 1.5 (b)$\big)$ we see that

$$\rho_{X^*}\left(\frac{t\mu\|A^* y^*\|}{\|J(x) - t\mu A^* y^*\| \vee \|J(x)\|}\right) \le \rho_{X^*}\left(\frac{t\mu\|A\|\,\|y^*\|}{\|x\|^{p-1}}\right) \le \rho_{X^*}(t\tau)$$

and we arrive at

$$\begin{aligned} \frac{1}{q}\tilde{\sigma}_q\big(J(x),\mu A^* y^*\big) &\le 2^q G_q \|x\|^p \int_0^1 \frac{\rho_{X^*}(t\tau)}{t}\,\mathrm{d}t\,. \\ &= 2^q G_q \|x\|^p \int_0^\tau \frac{\rho_{X^*}(t)}{t}\,\mathrm{d}t\,. \\ &\le 2^q G_q \|x\|^p \rho_{X^*}(\tau)\,, \end{aligned}$$

because also the function $\tau \mapsto \frac{\rho_X(\tau)}{\tau}$ is nondecreasing $\big($prop. 1.5 (c)$\big)$. □

Both operator $T^\mu_{A,Q,\Pi}$ and operator $T^\mu_{A,Q,P}$ are suitable for situation (A,Q). The use of $T^\mu_{A,Q,P}$ seems to be more simple, but as we have already mentioned, $T^\mu_{A,Q,\Pi}$ may be used in the context of more general Bregman projections. We at first treat $T^\mu_{A,Q,\Pi}$.

Lemma 2.7. *We assume (A,Q). Let $x_n \in X$ for some $n \in \mathbb{N}$ be given. If $\big\|J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big)\big\| = 0$ $\Big(\Leftrightarrow x_n \in M_{Ax\in Q}\Big)$ then we set $x_{n+1} := x_n$ and $R_{A,Q,\Pi}(x_n) := 0$.*
Otherwise we set

$$R_{A,Q,\Pi}(x_n) := \frac{\big\langle J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big) \,\big|\, Ax_n - \Pi^r_Q(Ax_n)\big\rangle}{\Big\|J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big)\Big\|} \tag{2.16}$$

and

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A\|} \dfrac{\|x_n\|^{p-1}}{\left\| J_Y(Ax_n) - J_Y\big(\Pi_Q^r(Ax_n)\big) \right\|} & , x_n \neq 0 \\ \dfrac{1}{\|A\|^p} \dfrac{R_{A,Q,\Pi}(x_n)^{p-1}}{\left\| J_Y(Ax_n) - J_Y\big(\Pi_Q^r(Ax_n)\big) \right\|} & , x_n = 0 \end{cases} , \tag{2.17}$$

whereby $\tau_n \in (0,1]$ *is chosen such that*

$$\frac{\rho_{X^*}\big(\tau_n\big)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q \|A\|} \frac{R_{A,Q,\Pi}(x_n)}{\|x_n\|} \right) . \tag{2.18}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A,Q,\Pi}(x_n)$ *and the following estimate is valid for all* $z \in M_{Ax\in Q}$*:*

$$\Delta_p\big(x_{n+1}, z\big) \leq \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)}{\|A\|} \tau_n \|x_n\|^{p-1} R_{A,Q,\Pi}(x_n) & , x_n \neq 0 \\ \frac{1}{p} \left(\frac{1}{\|A\|} R_{A,Q,\Pi}(x_n) \right)^p & , x_n = 0 \end{cases} . \tag{2.19}$$

Proof. This follows from the next lemma. □

Again approximate data must be adjusted.

Lemma 2.8. *We assume* (A_j, Q_k) *and that* $M_{Ax\in Q} \cap B_{m_0}$ *is not empty for some* $m_0 \in \mathbb{N}$*. Let* $x_n \in X$ *and* $j_{n-1}, k_{n-1} \in \mathbb{N}$ *for some* $n \in \mathbb{N}$ *be given. We choose* $m \geq \|A\| m_0$ *and set*

$$R_{A_j,Q_k,\Pi}(x_n) := \frac{\Big\langle J_Y(A_j x_n) - J_Y\big(\Pi^r_{Q_k}(A_j x_n)\big) \,\Big|\, A_j x_n - \Pi^r_{Q_k}(A_j x_n) \Big\rangle}{\left\| J_Y(A_j x_n) - J_Y\big(\Pi^r_{Q_k}(A_j x_n)\big) \right\|} \tag{2.20}$$

if the denominator is not equal to zero and $R_{A_j,Q_k,\Pi}(x_n) := 0$ *otherwise.*
If for all $j > j_{n-1}$ *and* $k > k_{n-1}$

$$R_{A_j,Q_k,\Pi}(x_n) \leq \frac{1}{\beta}(\eta_j m_0 + \delta_k^m) \tag{2.21}$$

then x_n *lies in the set* $M_{Ax\in Q}$*. In this case we choose* $j_n > j_{n-1}$*,* $k_n > k_{n-1}$ *and set* $x_{n+1} := x_n$*.*
Otherwise we find $j_n > j_{n-1}$ *and* $k_n > k_{n-1}$ *with*

$$0 \leq \eta_{j_n} m_0 + \delta^m_{k_n} < \beta R_{A_{j_n},Q_{k_n},\Pi}(x_n) \tag{2.22}$$

and for those indices we set

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A_{j_n}\|} \dfrac{\|x_n\|^{p-1}}{\left\| J_Y(A_j x_n) - J_Y\big(\Pi^r_{Q_k}(A_j x_n)\big) \right\|} & , x_n \neq 0 \\ \dfrac{(1-\beta)^{p-1}}{\|A_{j_n}\|^p} \dfrac{R_{A_{j_n},Q_{k_n},\Pi}(x_n)^{p-1}}{\left\| J_Y(A_j x_n) - J_Y\big(\Pi^r_{Q_k}(A_j x_n)\big) \right\|} & , x_n = 0 \end{cases} , \tag{2.23}$$

whereby $\tau_n \in (0,1]$ *is chosen such that*

$$\frac{\rho_{X^*}\big(\tau_n\big)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma(1-\beta)}{2^q G_q \|A_{j_n}\|} \frac{R_{A_{j_n},Q_{k_n},\Pi}(x_n)}{\|x_n\|}\right) . \tag{2.24}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n)$ *and the following estimate is valid for all* $z \in M_{Ax\in Q} \cap B_{m_0}$*:*

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|}\tau_n \|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},\Pi}(x_n) & , x_n \neq 0 \\ \frac{1}{p}\left(\frac{1-\beta}{\|A_{j_n}\|} R_{A_{j_n},Q_{k_n},\Pi}(x_n)\right)^p & , x_n = 0 \end{cases} . \tag{2.25}$$

Proof. We at first point out that $R_{A_j,Q_k,\Pi}(x_n) \ge 0$ in (2.20) by the monotonicity of J_Y $\big($prop. 1.10 (c)$\big)$. Suppose (2.21) is fulfilled, i.e.

$$R_{A_j,Q_k,\Pi}(x_n) \le \frac{1}{\beta}(\eta_j m + \delta_k^m)$$

for all $j > j_{n-1}$ and $k > k_{n-1}$. Since the right hand side converges to zero for $j,k \to \infty$, so does $R_{A_j,Q_k,\Pi}(x_n)$. Moreover the denominator in $R_{A_j,Q_k,\Pi}(x_n)$ converges to $\big\|J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big)\big\|$ by 1.35, because $(A_j x_n)_j$ converges to Ax_n and $(Q_k)_k$ converges boundedly to Q. If we have $\big\|J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big)\big\| = 0$ then this is already equivalent to $Ax_n \in Q$. Otherwise the numerator in (2.20) converges to zero. Since the numerator also converges to $\big\langle J_Y(Ax_n) - J_Y\big(\Pi^r_Q(Ax_n)\big) \,\big|\, Ax_n - \Pi^r_Q(Ax_n)\big\rangle$, this expression equals zero. By the strict convexity of Y we get $Ax_n = \Pi^r_Q(Ax_n) \in Q$.
In the case of (2.22) we set μ_n according to (2.23) with $\tau_n \in (0,1]$ according to (2.24). This choice of τ_n is possible since X^* is uniformly smooth $\big($def. 1.6 (d)$\big)$ and because of 1.5 (b) and (c). We at first consider $x_n \neq 0$. We set

$$w_n^* := J_Y(A_{j_n} x_n) - J_Y\big(\Pi^r_{Q_{k_n}}(A_{j_n} x_n)\big) \tag{2.26}$$

and get

$$\begin{aligned} &\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \\ &= \frac{1}{q}\left\|J_X(x_n) - \mu_n A^*_{j_n} w_n^*\right\|^q + \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle + \mu_n \langle w_n^* \,|\, A_{j_n} z\rangle . \end{aligned}$$

We estimate the last summand (whereby we use (2.20) in the last line):

$$\begin{aligned} \langle w_n^* \,|\, A_{j_n} z\rangle &= \langle w_n^* \,|\, (A_{j_n} - A)z\rangle + \left\langle w_n^* \,\middle|\, Az - \Pi^r_{Q_{k_n}}(A_{j_n} x_n)\right\rangle \\ &\quad + \left\langle w_n^* \,\middle|\, \Pi^r_{Q_{k_n}}(A_{j_n} x_n) - A_{j_n} x_n\right\rangle + \langle w_n^* \,|\, A_{j_n} x_n\rangle \\ &= \langle w_n^* \,|\, (A_{j_n} - A)z\rangle + \left\langle w_n^* \,\middle|\, Az - \Pi^r_{Q_{k_n}}(A_{j_n} x_n)\right\rangle \\ &\quad - \|w_n^*\| R_{A_{j_n},Q_{k_n},\Pi}(x_n) + \langle w_n^* \,|\, A_{j_n} x_n\rangle . \end{aligned}$$

Since z lies in $M_{Ax\in Q}\cap B_{m_0}$, we can write $Az = q$ with some $q \in Q$. Moreover by the coice of m we have

$$\|q\| = \|Az\| \leq \|A\| m_0 \leq m\,.$$

Thus we find $\tilde{q} \in Q_{k_n}$ with $\|\tilde{q} - q\| \leq \delta^m_{k_n}$ and get

$$\begin{aligned}
&\langle w_n^* \,|\, A_{j_n} z\rangle \\
&\leq \|w_n^*\|\eta_{j_n} m_0 + \left\langle w_n^* \,\middle|\, \tilde{q} - \Pi^r_{Q_{k_n}}(A_{j_n}x_n)\right\rangle + \langle w_n^* \,|\, q - \tilde{q}\rangle \\
&\quad -\|w_n^*\| R_{A_{j_n},Q_{k_n},\Pi}(x_n) + \langle w_n^* \,|\, A_{j_n} x_n\rangle \\
&\leq -\|w_n^*\| R_{A_{j_n},Q_{k_n},\Pi}(x_n) + \langle w_n^* \,|\, A_{j_n} x_n\rangle + \|w_n^*\|(\eta_{j_n} m_0 + \delta^m_{k_n}) \\
&\leq \langle w_n^* \,|\, A_{j_n} x_n\rangle - \|w_n^*\|(1-\beta) R_{A_{j_n},Q_{k_n},\Pi}(x_n)\,,
\end{aligned}$$

since $\tilde{q} \in Q_{k_n}$ and because of the variational inequality for the Bregman projection (1.30). Inserting this above yields ($\mu_n \geq 0$)

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \\
&\leq \frac{1}{q}\left\|J_X(x_n) - \mu_n A^*_{j_n} w_n^*\right\|^q + \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle \\
&\quad +\mu_n \left(\langle w_n^* \,|\, A_{j_n} x_n\rangle - \|w_n^*\|(1-\beta) R_{A_{j_n},Q_{k_n},\Pi}(x_n)\right)\,. \qquad (2.27)
\end{aligned}$$

We estimate the first summand by the characteristic inequality for the dual X^* (prop. 1.17) and get

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \\
&\leq \frac{1}{q}\|J_X(x_n)\|^q - \mu_n \langle w_n^* \,|\, A_{j_n} x_n\rangle + \frac{1}{q}\tilde{\sigma}_q\big(J_X(x_n), \mu_n A^*_{j_n} w_n^*\big) \\
&\quad + \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle + \mu_n \left(\langle w_n^* \,|\, A_{j_n} x_n\rangle - \|w_n^*\|(1-\beta) R_{A_{j_n},Q_{k_n},\Pi}(x_n)\right) \\
&= \Delta_p(x_n, z) + \frac{1}{q}\tilde{\sigma}_q\big(J_X(x_n), \mu_n A^*_{j_n} w_n^*\big) - \mu_n(1-\beta)\|w_n^*\| R_{A_{j_n},Q_{k_n},\Pi}(x_n)\,.
\end{aligned}$$

We estimate the second summand via lemma 2.6 and by the definitions of μ_n (2.23) and τ_n (2.24) we finally arrive at

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \\
&\leq \Delta_p(x_n, z) + \tau_n 2^q G_q \|x_n\|^p \frac{\rho_{X^*}(\tau_n)}{\tau_n} \\
&\quad -(1-\beta)\frac{\tau_n}{\|A_{j_n}\|}\|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},\Pi}(x_n) \\
&\leq \Delta_p(x_n, z) + \tau_n \frac{\gamma(1-\beta)}{\|A_{j_n}\|}\|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},\Pi}(x_n) \\
&\quad -(1-\beta)\frac{\tau_n}{\|A_{j_n}\|}\|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},\Pi}(x_n) \\
&= \Delta_p(x_n, z) - \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|}\tau_n \|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},\Pi}(x_n)\,.
\end{aligned}$$

For $x_n = 0$ we have $\Delta_p(x_n, z) = \frac{1}{p}\|z\|^p$ and thus

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \\
&= \frac{1}{q}\left\|\mu_n A^*_{j_n} w^*_n\right\|^q + \frac{1}{p}\|z\|^p + \mu_n \langle w^*_n \mid A_{j_n} z\rangle \\
&= \Delta_p(x_n, z) + \frac{1}{q}\left\|\mu_n A^*_{j_n} w^*_n\right\|^q + \mu_n \langle w^*_n \mid A_{j_n} z\rangle \\
&\le \Delta_p(x_n, z) + \frac{1}{q}\mu_n^q \|A_{j_n}\|^q \|w^*_n\|^q - \mu_n(1-\beta)\|w^*_n\| R_{A_{j_n},Q_{k_n},\Pi}(x_n)\,,
\end{aligned}$$

whereby we estimated the last summand analogously to the case $x_n \neq 0$. By the definition of μ_n (2.23) and observing that

$$\mu_n^q = \frac{(1-\beta)^p}{\|A_{j_n}\|^{p+q}} \frac{R_{A_{j_n},Q_{k_n},\Pi}(x_n)^p}{\|w^*_n\|^q}$$

we get

$$\Delta_p\big(T^{\mu_n}_{A_{j_n},Q_{k_n},\Pi}(x_n), z\big) \le \Delta_p(x_n, z) - \frac{1}{p}\left(\frac{1-\beta}{\|A_{j_n}\|} R_{A_{j_n},Q_{k_n},\Pi}(x_n)\right)^p .$$

□

Now we treat $T^{\mu}_{A,Q,P}$.

Lemma 2.9. *We assume (A,Q). Let $x_n \in X$ for some $n \in \mathbb{N}$ be given. We set*

$$R_{A,Q,P}(x_n) := \|Ax_n - P_Q(Ax_n)\|\,. \tag{2.28}$$

If $R_{A,Q,P}(x_n) = 0$ $\Big(\Leftrightarrow x_n \in M_{Ax\in Q}\Big)$ then we set $x_{n+1} := x_n$. Otherwise we define

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A\|} \dfrac{\|x_n\|^{p-1}}{R_{A,Q,P}(x_n)^{r-1}} \;, x_n \neq 0 \\ \dfrac{1}{\|A\|^p} R_{A,Q,P}(x_n)^{p-r} \;, x_n = 0 \end{cases} , \tag{2.29}$$

whereby $\tau_n \in (0,1]$ is chosen such that

$$\frac{\rho_{X^*}(\tau_n)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q \|A\|} \frac{R_{A,Q,P}(x_n)}{\|x_n\|}\right) . \tag{2.30}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A,Q,P}(x_n)$ and the following estimate is valid for all $z \in M_{Ax\in Q}$:

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)}{\|A\|}\tau_n \|x_n\|^{p-1} R_{A,Q,P}(x_n) \;, x_n \neq 0 \\ \frac{1}{p}\left(\frac{1}{\|A\|} R_{A,Q,P}(x_n)\right)^p \;, x_n = 0 \end{cases} . \tag{2.31}$$

Proof. This follows from the next lemma. □

Lemma 2.10. *We assume (A_j, Q_k) and that $M_{Ax\in Q} \cap B_{m_0}$ is not empty for some $m_0 \in \mathbb{N}$. Let $x_n \in X$ and $j_{n-1}, k_{n-1} \in \mathbb{N}$ for some $n \in \mathbb{N}$ be given. We choose $m \geq \|A\| m_0$ and set*

$$R_{A_j,Q_k,P}(x_n) := \|A_j x_n - P_{Q_k}(A_j x_n)\| . \tag{2.32}$$

If for all $j > j_{n-1}$ and $k > k_{n-1}$

$$R_{A_j,Q_k,P}(x_n) \leq \frac{1}{\beta}(\eta_j m_0 + \delta_k^m) \tag{2.33}$$

then x_n lies in the set $M_{Ax\in Q}$. In this case we choose $j_n > j_{n-1}$, $k_n > k_{n-1}$ and set $x_{n+1} := x_n$.
Otherwise we find $j_n > j_{n-1}$ and $k_n > k_{n-1}$ with

$$0 \leq \eta_{j_n} m_0 + \delta_{k_n}^m < \beta R_{A_{j_n},Q_{k_n},P}(x_n) \tag{2.34}$$

and for those indices we set

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A_{j_n}\|} \dfrac{\|x_n\|^{p-1}}{R_{A_{j_n},Q_{k_n},P}(x_n)^{r-1}} \ , \ x_n \neq 0 \\ \dfrac{(1-\beta)^{p-1}}{\|A_{j_n}\|^p} R_{A_{j_n},Q_{k_n},P}(x_n)^{p-r} \ , \ x_n = 0 \end{cases} , \tag{2.35}$$

whereby $\tau_n \in (0,1]$ is chosen such that

$$\frac{\rho_{X^*}(\tau_n)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma(1-\beta)}{2^q G_q \|A_{j_n}\|} \frac{R_{A_{j_n},Q_{k_n},P}(x_n)}{\|x_n\|} \right) . \tag{2.36}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A_{j_n},Q_{k_n},P}(x_n)$ and the following estimate is valid for all $z \in M_{Ax\in Q} \cap B_{m_0}$:

$$\Delta_p\big(x_{n+1}, z\big) \leq \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|} \tau_n \|x_n\|^{p-1} R_{A_{j_n},Q_{k_n},P}(x_n) \ , \ x_n \neq 0 \\ \frac{1}{p} \left(\frac{1-\beta}{\|A_{j_n}\|} R_{A_{j_n},Q_{k_n},P}(x_n) \right)^p \ , \ x_n = 0 \end{cases} . \tag{2.37}$$

Proof. We can prove this as in the case of the Bregman projection by setting

$$w_n^* := J_Y\big(A_{j_n} x_n - P_{Q_{k_n}}(A_{j_n} x_n)\big) \tag{2.38}$$

with $\|w_n^*\| = R_{A_{j_n},Q_{k_n},P}(x_n)^{r-1}$. □

We turn to linear equality constraints.

Lemma 2.11. *We assume (A, y). Let $x_n \in X$ for some $n \in \mathbb{N}$ be given. We set*

$$R_{A,y}(x_n) := \|Ax_n - y\| . \tag{2.39}$$

If $R_{A,y}(x_n) = 0$ $\big(\Leftrightarrow x_n \in M_{Ax=y}\big)$ then we set $x_{n+1} := x_n$.
Otherwise we define

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A\|} \dfrac{\|x_n\|^{p-1}}{R_{A,y}(x_n)^{r-1}} \ , x_n \neq 0 \\ \dfrac{1}{\|A\|^p} R_{A,y}(x_n)^{p-r} \ , x_n = 0 \end{cases} , \tag{2.40}$$

whereby $\tau_n \in (0, 1]$ is chosen such that

$$\frac{\rho_{X^*}\big(\tau_n\big)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q \|A\|} \frac{R_{A,y}(x_n)}{\|x_n\|} \right) . \tag{2.41}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A,y}(x_n)$ and the following estimate is valid for all $z \in M_{Ax=y}$:

$$\varDelta_p\big(x_{n+1}, z\big) \leq \varDelta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)}{\|A\|} \tau_n \|x_n\|^{p-1} R_{A,y}(x_n) \ , x_n \neq 0 \\ \frac{1}{p} \left(\frac{1-\beta}{\|A\|} R_{A,y}(x_n) \right)^p \ , x_n = 0 \end{cases} . \tag{2.42}$$

Proof. This follows from the next lemma. □

Lemma 2.12. *We assume (A_j, y_k) and that $M_{Ax=y} \cap B_{m_0}$ is not empty for some $m_0 \in \mathbb{N}$. Let $x_n \in X$ and $j_{n-1}, k_{n-1} \in \mathbb{N}$ for some $n \in \mathbb{N}$ be given. We set*

$$R_{A_j,y_k}(x_n) := \|A_j x_n - y_k\| . \tag{2.43}$$

If for all $j > j_{n-1}$ and $k > k_{n-1}$

$$R_{A_j,y_k}(x_n) \leq \frac{1}{\beta}(\eta_j m_0 + \delta_k) \tag{2.44}$$

then x_n lies in the set $M_{Ax=y}$. In this case we choose $j_n > j_{n-1}$ and $k_n > k_{n-1}$ and set $x_{n+1} := x_n$.
Otherwise we find $j_n > j_{n-1}$ and $k_n > k_{n-1}$ with

$$0 \leq \eta_{j_n} m_0 + \delta_{k_n} < \beta R_{A_{j_n},y_{k_n}}(x_n) \tag{2.45}$$

and for those indices we set

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A_{j_n}\|} \dfrac{\|x_n\|^{p-1}}{R_{A_{j_n},y_{k_n}}(x_n)^{r-1}} \ , x_n \neq 0 \\ \dfrac{(1-\beta)^{p-1}}{\|A_{j_n}\|^p} R_{A_{j_n},y_{k_n}}(x_n)^{p-r} \ , x_n = 0 \end{cases} , \tag{2.46}$$

whereby $\tau_n \in (0,1]$ *is chosen such that*

$$\frac{\rho_{X^*}\big(\tau_n\big)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma(1-\beta)}{2^q G_q \|A_{j_n}\|} \frac{R_{A_{j_n},y_{k_n}}(x_n)}{\|x_n\|} \right) . \tag{2.47}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A_{j_n},y_{k_n}}(x_n)$ *and the following estimate is valid for all* $z \in M_{Ax=y} \cap B_{m_0}$*:*

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|} \tau_n \|x_n\|^{p-1} R_{A_{j_n},y_{k_n}}(x_n) & , x_n \neq 0 \\ \frac{1}{p}\left(\frac{1-\beta}{\|A_{j_n}\|} R_{A_{j_n},y_{k_n}}(x_n) \right)^p & , x_n = 0 \end{cases} . \tag{2.48}$$

Proof. The proof is quite similar to the proofs of 2.10 and 2.14 by setting

$$w_n^* := J_Y(A_{j_n} x_n - y_{k_n}) \tag{2.49}$$

with $\|w_n^*\| = R_{A_{j_n},y_{k_n}}(x_n)^{r-1}$ and keeping in mind that $Az = y$ for all $z \in M_{Ax=y} \cap B_{m_0}$. □

Finally we consider linear inequality constraints.

Lemma 2.13. *We assume* $(A, y, +)$*. Let* $x_n \in X$ *for some* $n \in \mathbb{N}$ *be given. We set*

$$R_{A,y,+}(x_n) := \|(Ax_n - y)_+\| . \tag{2.50}$$

If $R_{A,y}(x_n) = 0$ $\big(\Leftrightarrow x_n \in M_{Ax\le y}\big)$ *then we set* $x_{n+1} := x_n$*.*
Otherwise we define

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A\|} \dfrac{\|x_n\|^{p-1}}{R_{A,y,+}(x_n)} & , x_n \neq 0 \\ \dfrac{1}{\|A\|^p} R_{A,y,+}(x_n)^{p-2} & , x_n = 0 \end{cases} , \tag{2.51}$$

whereby $\tau_n \in (0,1]$ *is chosen such that*

$$\frac{\rho_{X^*}\big(\tau_n\big)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q \|A\|} \frac{R_{A,y,+}(x_n)}{\|x_n\|} \right) . \tag{2.52}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A,y,+}(x_n)$ *and the following estimate is valid for all* $z \in M_{Ax\le y}$*:*

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)}{\|A\|} \tau_n \|x_n\|^{p-1} R_{A,y,+}(x_n) & , x_n \neq 0 \\ \frac{1}{p}\left(\frac{1-\beta}{\|A\|} R_{A,y,+}(x_n) \right)^p & , x_n = 0 \end{cases} . \tag{2.53}$$

Proof. This follows from the next lemma. □

Lemma 2.14. *We assume* $(A_j, y_k, +)$ *and that* $M_{Ax\le y} \cap B_{m_0}$ *is not empty for some* $m_0 \in \mathbb{N}$. *Let* $x_n \in X$ *and* $j_{n-1}, k_{n-1} \in \mathbb{N}$ *for some* $n \in \mathbb{N}$ *be given. We set*

$$R_{A_j,y_k,+}(x_n) := \|(A_j x_n - y_k)_+\| \,. \tag{2.54}$$

If for all $j > j_{n-1}$ *and* $k > k_{n-1}$

$$R_{A_j,y_k,+}(x_n) \le \frac{1}{\beta}(\eta_j m_0 + \delta_k) \tag{2.55}$$

then x_n *lies in the set* $M_{Ax\le y}$. *In this case we choose* $j_n > j_{n-1}$ *and* $k_n > k_{n-1}$ *and set* $x_{n+1} := x_n$.
Otherwise we find $j_n > j_{n-1}$ *and* $k_n > k_{n-1}$ *with*

$$0 \le \eta_{j_n} m_0 + \delta_{k_n} < \beta R_{A_{j_n},y_{k_n},+}(x_n) \tag{2.56}$$

and for those indices we set

$$\mu_n := \begin{cases} \dfrac{\tau_n}{\|A_{j_n}\|} \dfrac{\|x_n\|^{p-1}}{R_{A_{j_n},y_{k_n},+}(x_n)} & , x_n \ne 0 \\ \dfrac{(1-\beta)^{p-1}}{\|A_{j_n}\|^p} R_{A_{j_n},y_{k_n},+}(x_n)^{p-2} & , x_n = 0 \end{cases} , \tag{2.57}$$

whereby $\tau_n \in (0,1]$ *is chosen such that*

$$\frac{\rho_{X^*}(\tau_n)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma(1-\beta)}{2^q G_q \|A_{j_n}\|} \frac{R_{A_{j_n},y_{k_n},+}(x_n)}{\|x_n\|} \right) . \tag{2.58}$$

In this case we set $x_{n+1} := T^{\mu_n}_{A_{j_n},y_{k_n},+}(x_n)$ *and the following estimate is valid for all* $z \in M_{Ax\le y} \cap B_{m_0}$:

$$\Delta_p(x_{n+1}, z) \le \Delta_p(x_n, z) - \begin{cases} \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|} \tau_n \|x_n\|^{p-1} R_{A_{j_n},y_{k_n},+}(x_n) & , x_n \ne 0 \\ \frac{1}{p}\left(\frac{1-\beta}{\|A_{j_n}\|} R_{A_{j_n},y_{k_n},+}(x_n) \right)^p & , x_n = 0 \end{cases} . \tag{2.59}$$

Proof. Suppose (2.55) is valid, i.e.

$$R_{A_j,y_k,+}(x_n) \le \frac{1}{\beta}(\eta_j m_0 + \delta_k) \quad \text{for all} \quad j > j_{n-1} \,,\ k > k_{n-1} \,.$$

The right hand side converges to zero for $j,k \to \infty$ and so does $R_{A_j,y_k,+}(x_n)$. Since the mapping $x \mapsto (x)_+$ is continuous in a Banach lattice, $R_{A_j,y_k,+}(x_n)$ also converges to $\|(Ax_n - y)_+\|$. Hence $Ax_n \le y$.
In the case of (2.56) we choose μ_n according to (2.57) with $\tau_n \in (0,1]$ according to (2.58). We at first consider $x_n \ne 0$. We set

$$w_n^* := J_+\big((A_{j_n} x_n - y_{k_n})_+\big) \ge 0 \,. \tag{2.60}$$

Then we have $\|w_n^*\| = R_{A_{j_n},y_{k_n},+}(x_n)$ and get for all $z \in M_{Ax\le y} \cap B_{m_0}$

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},y_{k_n},+}(x_n), z\big)\\
&= \frac{1}{q}\left\|J_X(x_n) - \mu_n A^*_{j_n} w_n^*\right\|^q + \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle + \mu_n \langle w_n^* \,|\, A_{j_n} z\rangle \ .
\end{aligned}$$

Since w_n^* is positive and $Az - y \le 0$ we can estimate the last summand by

$$\begin{aligned}
&\langle w_n^* \,|\, A_{j_n} z\rangle\\
&= \langle w_n^* \,|\, (A_{j_n} - A)z\rangle + \langle w_n^* \,|\, Az - y\rangle + \langle w_n^* \,|\, y - y_{k_n}\rangle\\
&\quad + \langle w_n^* \,|\, y_{k_n} - A_{j_n} x_n\rangle + \langle w_n^* \,|\, A_{j_n} x_n\rangle\\
&\le \|w_n^*\|(\eta_{j_n} m_0 + \delta_{k_n}) + \langle w_n^* \,|\, y_{k_n} - A_{j_n} x_n\rangle + \langle w_n^* \,|\, A_{j_n} x_n\rangle\\
&\le \beta R_{A_{j_n},y_{k_n},+}(x_n)^2 + \langle w_n^* \,|\, y_{k_n} - A_{j_n} x_n\rangle + \langle w_n^* \,|\, A_{j_n} x_n\rangle \ .
\end{aligned}$$

Moreover we can write

$$\langle w_n^* \,|\, y_{k_n} - A_{j_n} x_n\rangle = -\,\langle w_n^* \,|\, (A_{j_n} x_n - y_{k_n})_+\rangle = -R_{A_j,y_k,+}(x_n)^2$$

because $(A_{j_n} x_n - y_{k_n})_- \in \operatorname{disj}\big((A_{j_n} x_n - y_{k_n})_+\big)$. Since $\mu_n \ge 0$ we get

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},y_{k_n},+}(x_n), z\big)\\
&\le \frac{1}{q}\left\|J_X(x_n) - \mu_n A^*_{j_n} w_n^*\right\|^q + \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle\\
&\quad + \mu_n\big(\langle w_n^* \,|\, A_{j_n} x_n\rangle - (1-\beta) R_{A_{j_n},y_{k_n},+}(x_n)^2\big) \ .
\end{aligned} \tag{2.61}$$

As in the proof of lemma 2.8 we estimate the first summand via the characteristic inequality for the dual X^* and lemma 2.6 and get

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},y_{k_n},+}(x_n), z\big)\\
&\le \Delta_p(x_n, z) - \mu_n(1-\beta) R_{A_{j_n},y_{k_n},+}(x_n)^2 + 2^q G_q \|x_n\|^p \rho_{X^*}(\tau_n) \ .
\end{aligned}$$

By the definitions of μ_n (2.57) and τ_n (2.58) we arrive at

$$\begin{aligned}
&\Delta_p\big(T^{\mu_n}_{A_{j_n},y_{k_n},+}(x_n), z\big)\\
&\le \Delta_p(x_n, z) - (1-\beta)\frac{\tau_n}{\|A_{j_n}\|}\|x_n\|^{p-1} R_{A_{j_n},y_{k_n},+}(x_n) + \tau_n 2^q G_q \|x_n\|^p \frac{\rho_{X^*}(\tau_n)}{\tau_n}\\
&\le \Delta_p(x_n, z) - \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|}\tau_n \|x_n\|^{p-1} R_{A_{j_n},y_{k_n},+}(x_n) \ .
\end{aligned}$$

The case $x_n = 0$ can be treated as in the proof of 2.8. □

We exemplify the parameter choice in L_p-spaces. At first we consider $p \le 2$ (dual space L_q with $q \ge 2$) with the normalized duality mapping[1]. We enter

[1] We remind that this "p" in "L_p" has nothing to do with the "p" we have used the whole time. The latter corresponds to the weight of the duality mapping; since in this example we use the normalized duality mapping, this weight is "$p = q = 2$".

the above proofs where we used the characteristic inequality, e.g. (2.27) and (2.61). With the respective T_n, w_n^* and R_n we have

$$\begin{aligned}\Delta_2\big(T_n^{\mu_n}(x_n), z\big) \le\ & \frac{1}{2}\left\|J_X(x_n) - \mu_n A_{j_n}^* w_n^*\right\|^2 + \frac{1}{2}\|z\|^2 - \langle J_X(x_n) \,|\, z\rangle \\ & + \mu_n \left(\langle w_n^* \,|\, A_{j_n} x_n\rangle - \|w_n^*\|(1-\beta)R_n\right) .\end{aligned}$$

By 1.18 (a) we get

$$\begin{aligned}\Delta_2\big(T_n^{\mu_n}(x_n), z\big) \le\ & \frac{1}{2}\|x_n\|^2 - \mu_n\ \langle w_n^* \,|\, A_{j_n} x_n\rangle + \frac{q-1}{2}\|A_{j_n}^* w_n^*\|^2\, \mu_n^2 \\ & + \mu_n\big(\langle w_n^* \,|\, A_{j_n} x_n\rangle - (1-\beta)\|w_n^*\| R_n\big) \\ & + \frac{1}{2}\|z\|^2 - \langle J_X(x_n) \,|\, z\rangle \\ =\ & \Delta_2(x_n, z) - (1-\beta)\|w_n^*\| R_n\, \mu_n + \frac{q-1}{2}\|A_{j_n}^* w_n^*\|^2\, \mu_n^2 .\end{aligned}$$

The right hand side is a quadratic function in μ_n and is minimal for

$$\mu_n := \frac{1-\beta}{q-1}\,\frac{\|w_n^*\| R_n}{\|A_{j_n}^* w_n^*\|^2}\,, \tag{2.62}$$

which yields

$$\Delta_2\big(T_n^{\mu_n}(x_n), z\big) \le \Delta_2(x_n, z) - \frac{(1-\beta)^2}{2(q-1)}\,\frac{\|w_n^*\|^2 R_n^2}{\|A_{j_n}^* w_n^*\|^2}\,. \tag{2.63}$$

Since $\|A_{j_n}^* w_n^*\| \le \|A_{j_n}\|\,\|w_n^*\|$ we can further estimate

$$\Delta_2\big(T_n^{\mu_n}(x_n), z\big) \le \Delta_2(x_n, z) - \frac{(1-\beta)^2}{2(q-1)\|A_{j_n}\|^2} R_n^2$$

to see more easily that this ensures convergence (but of course (2.63) is better). Except for $T^{\mu}_{A,Q,\Pi}$ we have $\|w_n^*\| = R_n^{r-1}$ and thus we can also write

$$\mu_n = \frac{1-\beta}{q-1}\,\frac{R_n^r}{\|A_{j_n}^* w_n^*\|^2}\,. \tag{2.64}$$

With the above estimate for $\|A_{j_n}^* w_n^*\|$ we could also take

$$\mu_n = \frac{1-\beta}{(q-1)\|A_{j_n}\|^2 R_n^{r-2}}\,. \tag{2.65}$$

In Hilbert spaces with normalized duality mappings (for Y as well, i.e. $r = 2$) and $T^{\mu}_{A,y}$ in case of exact data (A, y) $(\beta = 0)$ we would get $w_n^* = Ax_n - y$ and $R_n = \|Ax_n - y\|$ and thus (2.64) and (2.65) would result in

$$\mu_n = \frac{R_n^2}{\|A^*(Ax_n - y)\|^2} \quad \text{resp.} \quad \mu_n = \frac{1}{\|A\|^2}$$

($\rightsquigarrow$ *steepest descent method* resp. ordinary Landweber method for solving operator equations).
For $p \geq 2$ (dual space L_q with $q \leq 2$) with the same weight p we get by inequality 1.18 (b)

$$\Delta_p\big(T_n^{\mu_n}(x_n), z\big) \leq \Delta_p(x_n, z) - (1-\beta)\|w_n^*\| R_n\, \mu_n + \frac{2^{2-q}}{q}\|A_{j_n}^* w_n^*\|^q\, \mu_n^q\,.$$

The right hand side is minimal for

$$\mu_n^{q-1} = \frac{1-\beta}{2^{2-q}}\frac{\|w_n^*\| R_n}{\|A_{j_n}^* w_n^*\|^q} \quad \Leftrightarrow \quad \mu_n := \frac{(1-\beta)^{p-1}}{2^{p-2}}\frac{\|w_n^*\|^{p-1} R_n^{p-1}}{\|A_{j_n}^* w_n^*\|^p}\,, \qquad (2.66)$$

which yields

$$\Delta_p\big(T_n^{\mu_n}(x_n), z\big) \leq \Delta_p(x_n, z) - \frac{(1-\beta)^p}{p 2^{p-2}}\frac{\|w_n^*\|^p R_n^p}{\|A_{j_n}^* w_n^*\|^p}\,. \qquad (2.67)$$

The next lemma shows that "$R_n \to 0$" implies "$\lim_{n\to\infty} x_n \in M$" for the respective set M. We prove this here for the constraints in the range of a linear operator. The case of constraints C in X will be treated while proving the convergence of the iteration methods.

Lemma 2.15. *Let* $\big(R_n(x_n)\big)_n$ *be a sequence with* $x_n \in X$ *and* $R_n(x_n)$ *having one of the forms* $R_{A,Q,\Pi}(x_n)$, $R_{A,Q,P}(x_n)$, $R_{A,y}(x_n)$, $R_{A,y,+}(x_n)$, $R_{A_{j_n},Q_{k_n},\Pi}(x_n)$, $R_{A_{j_n},Q_{k_n},P}(x_n)$, $R_{A_{j_n},y_{k_n}}(x_n)$ *or* $R_{A_{j_n},y_{k_n},+}(x_n)$ *(but all of the same form) as in the previous lemmas. If the sequence* $(x_n)_n$ *is bounded and converges weakly to some* $x \in X$ *and* $\big(R_n(x_n)\big)_n$ *converges to zero then* x *lies in the corresponding set* $M_{Ax\in Q}$, $M_{Ax=y}$ *or* $M_{Ax\leq y}$.

Proof. The assertion for $R_n(x_n) = R_{A_{j_n},Q_{k_n},\Pi}(x_n)$ follows analogously to the beginning of the proof of 2.8, when we show that $(A_{j_n}x_n)_n$ converges to Ax. Since A is supposed to be compact, $\lim_{n\to\infty}\|A - A_{j_n}\| = 0$ and $(x_n)_n$ converges weakly to x and is bounded, say $\|x_n\| \leq c$, we get

$$\begin{aligned}\|Ax - A_{j_n}x_n\| &\leq \|(A - A_{j_n})x_n\| + \|A(x_n - x)\| \\ &\leq \|A - A_{j_n}\|c + \|A(x_n - x)\| \quad \longrightarrow 0 \quad \text{for} \quad n\to\infty\,.\end{aligned}$$

Hence $(A_{j_n})_n$ converges to Ax. Therefore also the cases $R_n(x_n) = R_{A,Q,\Pi}(x_n)$, $R_n(x_n) = R_{A,Q,P}(x_n)$ and $R_n(x_n) = R_{A_{j_n},Q_{k_n},P}(x_n)$ follow similarly. In case $R_n(x_n) = R_{A_{j_n},y_{k_n}}(x_n) = \|A_{j_n}x_n - y_{k_n}\|$ we take some $z \in M_{Ax=y}$ and get

$$\begin{aligned}\|Ax - y\|^p &= \langle J_X(Ax-y)\,|\,Ax-y\rangle \\ &= \langle A^* J_X(Ax-y)\,|\,x-z\rangle \\ &= \lim_{n\to\infty}\langle A^* J_X(Ax-y)\,|\,x_n-z\rangle \\ &= \lim_{n\to\infty}\langle J_X(Ax-y)\,|\,Ax_n-y\rangle = 0\,,\end{aligned}$$

because

$$\begin{aligned}\|Ax_n - y\| &\le |(A - A_{j_n})x_n\| + \|A_{j_n}x_n - y_{k_n}\| + \|y_{k_n} - y\| \\ &\le \|A - A_{j_n}\|c + R_{A_{j_n},y_{k_n}}(x_n) + \|y_{k_n} - y\| \longrightarrow 0 \quad \text{for} \quad n \to \infty\,.\end{aligned}$$

The case $R_n(x_n) = R_{A,y}(x_n)$ is an immediate consequence. In case $R_n(x_n) = R_{A_{j_n},y_{k_n},+}(x_n) = \|(A_{j_n}x_n - y_{k_n})_+\|$ by 1.23 it suffices to show that

$$\langle z^* \mid Ax - y\rangle \le 0 \quad \text{for all positive} \quad z^* \in Y^*\,.$$

At first we observe that

$$\begin{aligned}\langle z^* \mid A_{j_n}x_n - y_{k_n}\rangle &= \langle z^* \mid (A_{j_n}x_n - y_{k_n})_+\rangle - \langle z^* \mid (A_{j_n}x_n - y_{k_n})_-\rangle \\ &\le \langle z^* \mid (A_{j_n}x_n - y_{k_n})_+\rangle \\ &\le \|z^*\| R_{A_{j_n},y_{k_n},+}(x_n)\,.\end{aligned}$$

Therefore we get

$$\begin{aligned}\langle z^* \mid Ax - y\rangle &= \langle z^* \mid y_{k_n} - y\rangle + \langle z^* \mid A_{j_n}x_n - y_{k_n}\rangle \\ &\quad + \langle z^* \mid (A - A_{j_n})x_n\rangle + \langle A^*z^* \mid x - x_n\rangle \\ &\le \langle z^* \mid y_{k_n} - y\rangle + \|z^*\| R_{A_{j_n},y_{k_n},+}(x_n) \\ &\quad + \|z^*\|\, \|A - A_{j_n}\|c + \langle A^*z^* \mid x - x_n\rangle\ ,\end{aligned}$$

whereby the right hand side converges to zero for $n \to \infty$. The case $R_n(x_n) = R_{A,y,+}(x_n)$ is an immediate consequence. □

2.3 Split Feasibility Problem

The *convex feasibility problem* (CFP) consists of finding a common point in the intersection of finitely many closed convex sets. Such sets typically arise as constraints in a convex optimization problem. A classical procedure for its solution in Hilbert spaces is the method of cyclic orthogonal projections [30], where a convergent sequence is generated by projecting cyclically onto the individual sets. Alber and Butnariu [1] used *Bregman projections* to solve the CFP in Banach spaces. In applications such projection algorithms are efficient if the projections onto the individual sets are relatively simple to calculate. If the sets arise by imposing constraints in the range of a linear operator, like equality constraints $M_{Ax=y}$, inequality constraints $M_{Ax\le y}$ or sets of the form $M_{Ax\in Q}$, then it is in general too difficult or too costly to project onto these sets in each iterative step. This special case of the CFP, where some of the convex sets are related to constraints in the range of a linear operator, was also called the *split feasibility problem* (SFP) by Censor and Elfving [20]. We are concerned with its solution in Banach spaces via a generalization of the *CQ* algorithm suggested by Byrne [17], which has the iterative form

$$x_{n+1} = P_C\Big(x_n - \mu A^*\big(Ax_n - P_Q(Ax_n)\big)\Big)\,.$$

Let us formulate the SFP.

Problem 2.16 (SFP). Let finitely many closed convex sets $C_\iota \in \mathcal{C}(X)$ be given $\big(\iota \in I := \{0,\ldots,N-1\}\big)$ and assume that

$$\mathbf{C} := \bigcap_{\iota\in I} C_\iota \neq \emptyset\,.$$

Find some $x \in \mathbf{C}$.

Of course it looks like the CFP. But in the solution methods the sets C_ι will be treated differently depending on their structure. If the Bregman projection onto C_ι is relatively simple to calculate or if we even have a closed form expression like in 1.30, then we use operator $T_C = \Pi_C^p$. If C_ι is of the form $M_{Ax\in Q}$, $M_{Ax=y}$ or $M_{Ax\leq y}$ then we use the other respective operators T (thereby several operators $A \in \mathcal{L}(X,Y)$ and spaces Y are allowed). More precisely to each set C_ι we associate an operator T_ι due to the structure of C_ι, i.e. T_ι has the form $T^\mu_{A,Q,\Pi}$, $T^\mu_{A,Q,P}$, $T^\mu_{A,y}$, $T^\mu_{A,y,+}$ or T_C, depending on whether C_ι is of the form $M_{Ax\in Q}$, $M_{Ax=y}$, $M_{Ax\leq y}$ or just of general form, whereby we assume (A,Q), (A,y), $(A,y,+)$ or (C) respectively.

Let $\iota : \mathbb{N} \longrightarrow I$ be the *cyclic control mapping*

$$\iota(n) := n \mod N\,.$$

Method 2.17. *We set* $x_0 := 0$ *and iteratively define*

$$x_{n+1} := T_{\iota(n)}(x_n) \quad , \quad n \in \mathbb{N}$$

according to lemma 2.4, 2.7, 2.9, 2.11 or 2.13 respectively.

For instance suppose that we are given two sets $C_0 = C$ and $C_1 = M_{Ax\in Q}$. We use $T_0 = T_C = \Pi_C^p$ and $T_1 = T^\mu_{A,Q,P}$ and since $\iota(2n) = 0$ and $\iota(2n+1) = 1$ we would get

$$x_{(2n)+1} = \Pi_C^p(x_{2n})$$

and

$$x_{(2n+1)+1} = J_X^*\Big(J_X(x_{2n}) - \mu_{2n}A^*J_Y\big(Ax_{2n} - P_Q(Ax_{2n})\big)\Big)\,,$$

whereby the parameter μ_{2n} has to be chosen according to lemma 2.9. This can also be written more conveniently in a closed form

$$x_{n+1} = \Pi_C^p\bigg(J_X^*\Big(J_X(x_n) - \mu_n A^*J_Y\big(Ax_n - P_Q(Ax_n)\big)\Big)\bigg)\,,$$

which in case of Hilbert spaces reduces to the CQ algorithm.

Proposition 2.18. *The sequence $(x_n)_n$ generated by method 2.17 has the following properties.*

(a) It is bounded and therefore has weak cluster points. Moreover for all $z \in \boldsymbol{C}$ the sequence $\big(\Delta_p(x_n, z)\big)_n$ is decreasing.
(b) Every (weak) cluster point x is a solution of problem 2.16 and fulfills

$$\|x\| \le q \, \|\Pi^p_{\boldsymbol{C}}(0)\| = q \min_{z \in \boldsymbol{C}} \|z\| .$$

(c) If it has a strongly convergent subsequence then the whole sequence converges strongly. Especially this is the case if X is finite dimensional or one of the sets C_ι is boundedly compact[2].
(d) If the duality mapping of X is weak-to-weak-continuous[3] then the whole sequence converges weakly.

Proof. According to (2.9), (2.19), (2.31), (2.42) or (2.53) respectively we have for all $z \in C_{\iota(n)}$

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - S_n \tag{2.68}$$

with some $S_n > 0$ in case $x_n \notin C_{\iota(n)}$ and $S_n = 0$ in case $x_n \in C_{\iota(n)}$ (which by the fixed point relations 2.2 is equivalent to $x_{n+1} = T_{\iota(n)}(x_n) = x_n$). Therefore (2.68) especially holds for all $z \in \mathbf{C}$ and $n \in \mathbb{N}$, i.e. the sequence $\big(\Delta_p(x_n, z)\big)_n$ is decreasing and thus convergent to some $\Delta_z \ge 0$. Especially it is bounded and 1.24 (d) then ensures that the sequence $(x_n)_n$ is bounded, say $\|x_n\| \le d$, and consequently it has weak cluster points. If $0 \in \mathbf{C} \subset C_\iota$ for all $\iota \in I$ then $x_0 = 0$ is our solution and we are done. In the interesting case $0 \notin \mathbf{C}$ we show that there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that $\|x_n\| \ge c$ for all $n \ge n_0$. We w.l.o.g. assume $x_0 = 0 \notin C_{\iota(0)}$ and thus $S_0 > 0$. By (2.68) we get

$$\Delta_z \le \Delta_p(x_1, z) < \Delta_p(x_0, z) = \Delta_p(0, z) .$$

So if $(x_n)_n$ had a subsequence $(x_{n_k})_k$ converging to zero, this would lead to the contradiction

$$\Delta_p(0, z) = \lim_{k \to \infty} \Delta_p(x_{n_k}, z) = \Delta_z < \Delta_p(0, z) .$$

Now suppose that $z_0 \in X$ is a weak cluster point, say $x_{n_k} \to z_0$ weakly for $k \to \infty$. We show that z_0 lies in $\mathbf{C}$. Since ι is the cyclic control mapping we may assume $\iota(n_k + j) = j$ for $j \in \{0, \dots, N-1\}$ and all $k \in \mathbb{N}$. Hence we have $C_{\iota(n_k+j)} = C_j$. By passing to subsequences if necessary we may further assume that the sequence $(x_{n_k+j})_k$ converges weakly to some $z_j \in X$. We at first show that z_0 lies in C_0 and $z_0 = z_1$. We take some $z \in \mathbf{C}$ and by (2.68) we get

[2] i.e. every bounded closed subset is compact

[3] i.e. $\big(J_X(x_n)\big)_n$ converges weakly to $J_X(x)$ if $(x_n)_n$ converges weakly to x. The duality mappings of the L_p-sequence spaces have this property, whereas this is not true for the L_p-function spaces [22].

$$S_{n_k} \leq \Delta_p(x_{n_k}, z) - \Delta_p\big(x_{n_k+1}, z\big) \longrightarrow \Delta_z - \Delta_z = 0 \quad \text{for} \quad k \to \infty\,. \tag{2.69}$$

(If $x_{n_k+1} = x_{n_k}$ infinitely often then the assertions follow trivially.) In case C_0 is of general form and therefore T_0 of the form $T_{C_0} = \Pi^p_{C_0}$, all iterates $x_{n_k+1} = \Pi^p_{C_0}(x_n)$ lie in C_0 and $S_{n_k} = \Delta_p(x_{n_k}, x_{n_k+1})$ (see 2.4). Hence $z_1 \in C_0$ and by $\lim_{k\to\infty} \Delta_p(x_{n_k}, x_{n_k+1}) = \lim_{k\to\infty} S_{n_k} = 0$ and 1.24 (h) we conclude that $z_0 = z_1 \in C_0$. In case C_0 is of one of the special forms, S_{n_k} has the form

$$S_{n_k} = \frac{(1-\gamma)}{\|A\|} \tau_{n_k} \|x_{n_k}\|^{p-1} R(x_{n_k})$$

for all k big enough such that $\|x_{n_k}\| \geq c > 0$ (see 2.7, 2.9, 2.11 and 2.13). Suppose $\liminf_{k\to\infty} R(x_{n_k}) \neq 0$. Then especially $R(x_{n_k})$ remains bounded away from zero and by the definition of τ_{n_k} and 1.5 (c), so does τ_{n_k}. But this implies that also S_{n_k} remains bounded away from zero which contradicts (2.69). So w.l.o.g. $\big(R(x_{n_k})\big)_k$ converges to zero and by 2.15 we conclude that $z_0 \in C_0$. Moreover by the definition of T_0 (see 2.1) we have

$$J_X(x_{n_k+1}) = J_X(x_{n_k}) - \mu_{n_k} A^* w^*_{n_k}$$

with $w^*_{n_k} \in Y^*$, which by the choice of μ_{n_k} (2.17), (2.29), (2.40) or (2.51) yields

$$\|J_X(x_{n_k+1}) - J_X(x_{n_k})\| \leq \mu_{n_k} \|A\|\, \|w^*_{n_k}\| \leq \tau_{n_k} \|x_{n_k}\|^{p-1} \leq \tau_{n_k} d^{p-1}\,.$$

Since $\big(R(x_{n_k})\big)_k$ converges to zero, $\|x_{n_k}\| \geq c$ for k big enough and X^* is uniformly smooth $\big(1.6\text{ (d)}\big)$, also τ_{n_k} converges to zero and therefore, so does $\|J_X(x_{n_k+1}) - J_X(x_{n_k})\|$. By proposition 1.24 (b) we have

$$\begin{aligned} \Delta_p(x_{n_k}, x_{n_k+1}) &\leq \langle J_X(x_{n_k+1}) - J_X(x_{n_k}) \,|\, x_{n_k+1} - x_{n_k} \rangle \\ &\leq \|J_X(x_{n_k+1}) - J_X(x_{n_k})\|\, 2d \quad \longrightarrow 0 \quad \text{for} \quad k \to \infty \end{aligned}$$

and since X is uniformly convex by 1.24 (h) we conclude that $z_0 = z_1$. Thus we have shown in every case that $z_0 \in C_0$ and $z_0 = z_1$. In the same way we can show $z_1 \in C_1$ and $z_1 = z_2$ and inductively we get $z_0 = z_j \in C_j$ for all $j \in I$. Hence $z_0 \in \mathbf{C}$. Moreover we have for all $z \in \mathbf{C}$

$$\begin{aligned} &\Delta_p(x_{n_k}, z) \leq \Delta_p(0, z) \\ \Rightarrow\ &\frac{1}{q} \|x_{n_k}\|^p \leq \langle J_X(x_{n_k}) \,|\, z \rangle \leq \|x_{n_k}\|^{p-1} \|z\| \\ \Rightarrow\ &\|x_{n_k}\| \leq q\|z\| \end{aligned}$$

and therefore $\|z_0\| \leq \liminf_{k\to\infty} \|x_{n_k}\| \leq q\|z\|$. Hence $\|z_0\| \leq q\ \min_{z\in\mathbf{C}} \|z\|$. We show that the whole sequence converges strongly if it has a strongly convergent subsequence. Suppose $x_{n_k} \to z_0 \in X$ strongly for $k \to \infty$. By what we have shown above, we know that $z_0 \in \mathbf{C}$ and thus the sequence

$\big(\Delta_p(x_n,z_0)\big)_n$ converges (to Δ_{z_0}). On the other hand by 1.24 (f) the subsequence $\big(\Delta_p(x_{n_k},z_0)\big)_k$ converges to zero and therefore the whole sequence must converge to zero. Again by 1.24 (f) and since X is uniformly convex, we conclude that the whole sequence $(x_n)_n$ converges strongly to $z_0 \in \mathbf{C}$. Finally suppose that J_X is weak-to-weak-continuous. Let $z_1, z_2 \in \mathbf{C}$ be two weak cluster points. We show that they coincide. We have

$$\Delta_p(x_n,z_1) - \Delta_p(x_n,z_2) - \frac{1}{p}\|z_1\|^p + \frac{1}{p}\|z_2\|^p = \langle J_X(x_n) \,|\, z_2 - z_1\rangle \ .$$

Since the left hand side converges to $\Delta := \Delta_{z_1} - \Delta_{z_2} - \frac{1}{p}\|z_1\|^p + \frac{1}{p}\|z_2\|^p$, so does the right hand side. Let $(x_{n_k})_k$ converge weakly to z_1 and $(x_{m_l})_l$ converge weakly to z_2. Then $\big(J_X(x_{n_k})\big)_k$ converges weakly to $J_X(z_1)$ and $\big(J_X(x_{m_l})\big)_l$ converges weakly to $J_X(z_2)$ and we get

$$\begin{aligned} &\langle J_X(z_2) - J_X(z_1) \,|\, z_2 - z_1\rangle \\ &= \langle J_X(z_2) \,|\, z_2 - z_1\rangle - \langle J_X(z_1) \,|\, z_2 - z_1\rangle \\ &= \lim_{l\to\infty} \langle J_X(x_{m_l}) \,|\, z_2 - z_1\rangle - \lim_{k\to\infty} \langle J_X(x_{n_k}) \,|\, z_2 - z_1\rangle \\ &= \Delta - \Delta = 0\,. \end{aligned}$$

By the strict convexity of X we conclude that $z_1 = z_2$. □

Now we examine stability and regularizing properties of method 2.17 and how we may include approximate data. If X and all spaces Y_ι connected to the sets C_ι are uniformly smooth and uniformly convex then every iterate

$$x_n = T_{\iota(n-1)}(x_{n-1}) = \ldots = T_{\iota(n-1)} \ldots T_{\iota(0)}(x_0)$$

depends continuously on all data C, A, Q, y because of the continuity properties of all mappings involved, i.e. the iterates behave stable with respect to small perturbations. But also in the general case we can show that method 2.17 has some regularizing properties.

Suppose the sets C_ι are only approximately given under the assumptions (C_i), (A_j, Q_k), (A_j, y_k) and $(A_j, y_k, +)$, so that we actually know estimates $(\epsilon_i,\ \eta_k,\ \ldots)$ for the deviations of the approximate data from the exact ones. Moreover we assume that we know an estimate for the norm of some members of $\mathbf{C}$, i.e. we know an $m_0 \in \mathbb{N}$ such that

$$\mathbf{C} \cap B_{m_0} \neq \emptyset\,. \tag{2.70}$$

Again we associate to each set C_ι the appropriate sequence of operators $\big(T_{C_i}\big)_i$, $\big(T^\mu_{A_j,Q_k,\Pi}\big)_{j,k}$, $\big(T^\mu_{A_j,Q_k,P}\big)_{j,k}$, $\big(T^\mu_{A_j,y_k}\big)_{j,k}$ or $\big(T^\mu_{A_j,y_k,+}\big)_{j,k}$ and consider the following method.

Method 2.19. *We set* $x_0 := 0$*, all starting indices* $i_{-1}, j_{-1}, k_{-1} := 0$ *and iteratively define*

$$x_{n+1} := T_{\iota(n)}(x_n) \quad , \quad n \in \mathbb{N}$$

and i_n, j_n, k_n *according to lemma 2.5, 2.8, 2.10, 2.12 and 2.14, whereby the respective indices "* i_{n-1} *", "* j_{n-1} *", "* k_{n-1} *" are the ones defined in iteration* $n-N$*, when the operator* $T_{\iota(n-N)}$ *linked to the same set* $C_{\iota(n)}$ *was applied.*

We recall that according to the listed lemmas we set $x_{n+1} := x_n$ if relation (2.11), (2.21), (2.33), (2.44) or (2.55) holds.

Proposition 2.20. *For the sequence* $(x_n)_n$ *generated by method 2.19 all assertions of proposition 2.18 remain valid.*

Proof. We can prove this in the same way as 2.18 when we take (2.70) and the following into account: The sequences $\big(\|A_{j_{n_k}}\|\big)_k$ converge to $\|A\| \neq 0$ and therefore the case "C_0 is of special form" can be treated analoguosly. In case "C_0 is of general form", and thus $x_{n_k+1} = \Pi^p_{C_{i_{n_k}}}(x_{n_k})$, we also have

$$\lim_{k\to\infty} \Delta_p(x_{n_k}, x_{n_k+1}) = \lim_{k\to\infty} S_{n_k} = 0$$

and therefore $z_1 = z_0$. (Here we slightly abuse notation: The sets $C_{i_{n_k}}$ are meant to be the ones converging to the (fixed) set $C_0 = C_{\iota(n_k)}$ and should not be confused with the other sets C_ι.) Moreover since $(x_{n_k+1})_k$ is bounded and converges weakly to z_1 with $x_{n_k+1} \in C_{i_{n_k}}$ for all $k \in \mathbb{N}$ and $\big(C_{i_{n_k}}\big)_k$ converges boundedly to C_0, by 1.34 (c) we conclude that $z_1 \in C_0$. Hence also in this case we have $z_0 = z_1 \in C_0$. $\square$

In the proofs of 2.17 and 2.19 we have also seen that the whole sequence of the remainders converges to zero, i.e.

$$\lim_{n\to\infty} R(x_n) = 0$$

(Because every subsequence $\big(R(x_{n_k})\big)_k$ has in turn a subsequence converging to zero). Indeed even the stronger relation

$$\sum_{n=0}^{\infty} S_n < \infty$$

holds, since by (2.68) we get for all $k \in \mathbb{N}$

$$\begin{aligned}\sum_{n=0}^{k} S_n &\le \sum_{n=0}^{k} \Big(\Delta_p(x_n, z) - \Delta_p\big(x_{n+1}, z\big)\Big)\\ &= \Delta_p(x_0, z) - \Delta_p\big(x_{k+1}, z\big)\\ &\le \Delta_p(x_0, z)\,.\end{aligned}$$

Now we consider the case of noisy data, i.e. rather than approximations C_i, A_j, Q_k, y_k we are given $C_\epsilon, A_\eta, Q_\delta, y_\delta$ with known noise levels

$$d_{m_0}(C, C_\epsilon) \leq \epsilon \quad , \quad \|A - A_\eta\| \leq \eta \quad , \quad \|y - y_\delta\| \leq \delta \, ,$$

$$d_m(Q, Q_\delta) \leq \delta \quad \text{with some} \quad m \geq \|A\| m_0 \, .$$

We apply method 2.19 with $\epsilon_i = \epsilon, \eta_k = \eta, \ldots$ and use the *discrepancy principle* [28, 35] as a stopping rule: We terminate the iteration when for the first time all remainders $R(x_n)$ in a cycle of N successive iterations fulfill the relations (2.11), (2.21), (2.33), (2.44) or (2.55) respectively. Because whenever such a relations is not fulfilled, (2.13), (2.25), (2.37), (2.48) and (2.59) guarantee that

$$\Delta_p\big(x_{n+1}, z\big) < \Delta_p(x_n, z) \quad \text{for all} \quad z \in \mathbf{C} \cap B_{m_0} \, ,$$

which means in this sense x_{n+1} is a better approximation to the set of exact solutions than is x_n. Moreover we can interpret proposition 2.20 in the following way.

Proposition 2.21. *Together with the discrepancy principle method 2.19 is a regularization method for problem 2.16 in the following sense: Let $n(\epsilon, \eta, \ldots)$ be the stopping index according to the discrepancy principle for the noise levels $\epsilon, \eta, \ldots$. Then all assertions of 2.18 are valid for the sequence $\big(x_{n(\epsilon,\eta,\ldots)}\big)_{\epsilon,\eta,\ldots}$ if the noise levels $\epsilon, \eta, \ldots$ tend to zero.*

2.4 Projections onto Affine Subspaces

We want to use a special case of method 2.17 to compute Bregman and metric projections onto affine subspaces which are given via a linear operator, i.e. onto sets of the form $x + \mathcal{N}(A)$ $\big(\stackrel{\triangle}{=} M_{Ax=y}\big)$ or $y + \overline{\mathcal{R}(A)}$. In [46] we already used this method with starting point $x_0 = 0$ to compute minimum norm solutions of linear operator equations and proved its strong convergence. Here we improve and complement our work by placing it in the context of projections. This enables us to use the same method (with arbitrary starting points) for a broader class of problems.

Method 2.22. *We assume (A, y), choose an arbitrary starting point $x_0 \in X$ and iteratively define*

$$x_{n+1} := T^{\mu_n}_{A,y}(x_n) = J^*_X\big(J_X(x_n) - \mu_n A^* J_Y(Ax_n - y)\big)$$

according to lemma 2.11.

Proposition 2.23. *The sequence* $(x_n)_n$ *generated by method 2.22 converges strongly to the Bregman projection of* x_0 *onto the set* $M_{Ax=y}$. *Except in the following case: Suppose that* $y \neq 0$ *and* x_0 *is such that* $\Delta_p(x_0, z) > \Delta_p(0, z)$ *for all* $z \in M_{Ax=y}$. *Then it might happen that* $\liminf_{n\to\infty} R_{A,y}(x_n) \neq 0$. *But then we also have* $\lim_{n\to\infty} x_n = 0$ *and the Bregman projections of* 0 *and* x_0 *onto the set* $M_{Ax=y}$ *must already coincide.*

The exceptional case in fact poses no problem. When while running the iteration it becomes obvious that $\liminf_{n\to\infty} R_{A,y}(x_n) \neq 0$ and (then necessarily also) $\lim_{n\to\infty} x_n = 0$, then we know that $\Pi^p_{M_{Ax=y}}(0) = \Pi^p_{M_{Ax=y}}(x_0)$; so we just have to restart with starting point $x_0 = 0$ (in which case $\Delta_p(x_0, z) = \Delta_p(0, z)$ for all $z \in M_{Ax=y}$ and the exception cannot occur).

Proof (of 2.23). We at first point out that by the definition of the iterates we inductively get

$$\begin{aligned} J_X(x_{n+1}) - J_X(x_0) &= J_X(x_n) - J_X(x_0) - \mu_n A^* J_Y(Ax_n - y) \\ &= \dots \\ &= -\sum_{k=0}^{n} \mu_k A^* J_Y(Ax_k - y) \end{aligned}$$

and thus

$$J_X(x_n) - J_X(x_0) \in \overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^{\perp} \quad \text{for all} \quad n \in \mathbb{N}\,. \tag{2.71}$$

Hence if $R_{A,y}(x_n) = 0 \Leftrightarrow x_n \in M_{Ax=y}$ for some $n \in \mathbb{N}$ then by 1.29 (a) we already have $x_n = \Pi^p_{M_{Ax=y}}(x_0)$ and we can stop the iteration. So we assume $R_{A,y}(x_n) > 0$ for all $n \in \mathbb{N}$. As in the proof of 2.18 and by (2.42) we get for all $z \in M_{Ax=y}$ the recursive inequality

$$\Delta_p\big(x_{n+1}, z\big) \le \Delta_p(x_n, z) - S_n \tag{2.72}$$

with

$$0 < S_n = \begin{cases} \frac{(1-\gamma)}{\|A\|} \tau_n \|x_n\|^{p-1} R_{A,y}(x_n) & , x_n \neq 0 \\ \frac{1}{p} \left(\frac{1-\beta}{\|A\|} R_{A,y}(x_n) \right)^p & , x_n = 0 \end{cases},$$

which implies that the limit $\Delta_z = \lim_{n\to\infty} \Delta_p(x_n, z)$ exists, that the sequence $(x_n)_n$ is bounded and has weak cluster points, that $\lim_{n\to\infty} S_n = 0$ and $\sum\limits_{n=0}^{\infty} S_n < \infty$. Moreover it ensures that all iterates are different from each other. Therefore we assume $x_n \neq 0$ for all $n \in \mathbb{N}$ and thus

$$S_n = \frac{(1-\gamma)}{\|A\|} \tau_n \|x_n\|^{p-1} R_{A,y}(x_n) \longrightarrow 0 \quad \text{for} \quad n \to \infty\,. \tag{2.73}$$

and

$$\sum_{n=0}^{\infty} \tau_n \|x_n\|^{p-1} R_{A,y}(x_n) < \infty \,. \tag{2.74}$$

So what happens when $\liminf_{n\to\infty} R_{A,y}(x_n) \neq 0$? Then especially $R_{A,y}(x_n)$ remains bounded away from zero and by the boundedness of the sequence $(x_n)_n$, the definition of τ_n (2.41) and 1.5 (c), so does τ_n. But then (2.73) forces $(x_n)_n$ to converge to zero. Thus we have for all $z \in M_{Ax=y}$

$$\Delta_p(0,z) = \lim_{n\to\infty} \Delta_p(x_n,z) \leq \Delta_p(x_1,z) < \Delta_p(x_0,z)\,,$$

and this can only happen if $\Delta_p(x_0,z) > \Delta_p(0,z)$ for all $z \in M_{Ax=y}$ and

$$\|y\| = \lim_{n\to\infty} \|Ax_n - y\| = \lim_{n\to\infty} R_{A,y}(x_n) \neq 0\,.$$

We show that in this exceptional case the Bregman projections of 0 and x_0 onto the set $M_{Ax=y}$ already coincide. Let z be an arbitrary element in $M_{Ax=y}$. Then $z - \Pi^p_{M_{Ax=y}}(0) \in \mathcal{N}(A) = \mathcal{R}(A^*)^\perp$ and by (2.71) we get for all $n \in \mathbb{N}$

$$\begin{aligned} &\left\langle J_X\big(\Pi^p_{M_{Ax=y}}(0)\big) - J_X(x_0) \,\middle|\, z - \Pi^p_{M_{Ax=y}}(0)\right\rangle \\ =\ &\left\langle J_X\big(\Pi^p_{M_{Ax=y}}(0)\big) - J_X(x_n) \,\middle|\, z - \Pi^p_{M_{Ax=y}}(0)\right\rangle \\ \longrightarrow &\left\langle J_X\big(\Pi^p_{M_{Ax=y}}(0)\big) - J_X(0) \,\middle|\, z - \Pi^p_{M_{Ax=y}}(0)\right\rangle \geq 0 \quad \text{for} \quad n \to \infty\,. \end{aligned}$$

Hence $\Pi^p_{M_{Ax=y}}(0) = \Pi^p_{M_{Ax=y}}(x_0)$ by (1.30).

Now let $\liminf_{n\to\infty} R_{A,y}(x_n) = 0$. Then we can choose a subsequence $\big(R_{A,y}(x_{n_k})\big)_k$ with the property that

$$\begin{aligned} & R_{A,y}(x_{n_k}) \longrightarrow 0 \quad \text{for} \quad k \to \infty \\ \text{and} \qquad & R_{A,y}(x_{n_k}) < R_{A,y}(x_n) \quad \text{for all} \quad n < n_k\,. \end{aligned} \tag{2.75}$$

The same property also holds for every subsequence of $\big(R_{A,y}(x_{n_k})\big)_k$. We show that $(x_n)_n$ has a Cauchy subsequence. By the boundedness of $(x_n)_n$ and $\big(J_X(x_n)\big)_n$ we can find a subsequence $(x_{n_k})_k$ with

(S.1) the sequence of the norms $\big(\|x_{n_k}\|\big)_k$ is convergent,
(S.2) the sequence $\big(J_X(x_{n_k})\big)_k$ is weakly convergent and
(S.3) the sequence $\big(R_{A,y}(x_{n_k})\big)_k$ fulfills (2.75).

We show that $(x_{n_k})_k$ is a Cauchy sequence. With (1.27) we have for all $l, k \in \mathbb{N}$ with $k > l$

$$\Delta_p(x_{n_l}, x_{n_k}) = \frac{1}{q}\left(\|x_{n_l}\|^p - \|x_{n_k}\|^p\right) + \langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k}\rangle\ .$$

Because of (S.1) the first summand converges to zero for $l \to \infty$. We fix some $z \in M_{Ax=y}$ and write the second summand as

$$\begin{aligned}&\langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k}\rangle \\ &= \langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, z\rangle + \langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k} - z\rangle \;. \end{aligned} \tag{2.76}$$

Again the first summand converges to zero for $l \to \infty$ by (S.2). We estimate the second summand:

$$|\langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k} - z\rangle| = \left| \sum_{n=n_l}^{n_k-1} \langle J_X(x_{n+1}) - J_X(x_n) \,|\, x_{n_k} - z\rangle \right| .$$

The recursive definition of method 2.22 and the definition of μ_n (2.40) yield

$$\begin{aligned}|\langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k} - z\rangle| &= \left| \sum_{n=n_l}^{n_k-1} \mu_n \langle J_Y(Ax_n - y) \,|\, Ax_{n_k} - y\rangle \right| \\ &\le \sum_{n=n_l}^{n_k-1} \mu_n \|J_Y(Ax_n - y)\| \, \|Ax_{n_k} - y\| \\ &= \sum_{n=n_l}^{n_k-1} \mu_n R_{A,y}(x_n)^{r-1} R_{A,y}(x_{n_k}) \\ &= \frac{1}{\|A\|} \sum_{n=n_l}^{n_k-1} \tau_n \|x_n\|^{p-1} R_{A,y}(x_{n_k}) \,. \end{aligned}$$

(S.3) ensures that

$$|\langle J_X(x_{n_k}) - J_X(x_{n_l}) \,|\, x_{n_k} - x\rangle| \le \frac{1}{\|A\|} \sum_{n=n_l}^{n_k-1} \tau_n \|x_n\|^{p-1} R_{A,y}(x_n) \,.$$

By (2.74) the right side converges to zero for $l \to \infty$ and therefore so does $\Delta_p(x_{n_l}, x_{n_k})$. By proposition 1.24 (g) we conclude that $(x_{n_k})_k$ is a Cauchy sequence and thus convergent to an $x \in X$. Moreover we have $x \in M_{Ax=y}$ by 2.15, and as in the proof of 2.18 we see that the whole sequence $(x_n)_n$ converges strongly to x. It remains to prove that $x = \Pi^p_{M_{Ax=y}}(x_0)$. Since J_X is norm-to-weak continuous and $\overline{\mathcal{R}(A^*)}$ weakly closed, (2.71) guarantees that $J_X(x) - J_X(x_0) \in \overline{\mathcal{R}(A^*)} = \mathcal{N}(A)^\perp$. By proposition 1.29 (a) we conclude that indeed $x = \Pi^p_{M_{Ax=y}}(x_0)$. $\square$

Of course we are also allowed to use approximate data.

Method 2.24. *We assume (A_j, y_k) and that $M_{Ax=y} \cap B_{m_0}$ is not empty and that the operators $(A_j)_j$ fulfill the condition $\overline{\mathcal{R}(A_j^*)} \subset \overline{\mathcal{R}(A^*)}$. We set $j_{-1} := 0$, $k_{-1} := 0$, choose an arbitrary starting point $x_0 \in X$ and iteratively define*

$$x_{n+1} := T^{\mu_n}_{A_{j_n}, y_{k_n}}(x_n) = J_X^*\big(J_X(x_n) - \mu_n A_{j_n}^* J_Y(A_{j_n} x_n - y_{k_n})\big)$$

and j_n, k_n according to lemma 2.12.

The condition $\overline{\mathcal{R}(A_j^*)} \subset \overline{\mathcal{R}(A^*)}$ assures that the iterates $J_X(x_n)$ remain in $\overline{\mathcal{R}(A^*)}$.

Proposition 2.25. *For the sequence $(x_n)_n$ generated by method 2.24 the assertion of 2.23 remains valid.*

Proof. We can prove this analogously to the case of exact data if we keep in mind that $\big(A_{j_n}\big)_n$ converges to $A \neq 0$ and $\big(y_{k_n}\big)_n$ converges to y and consider the following modifications. According to lemma 2.12, if 2.44 is fulfilled for some $n \in \mathbb{N}$, then x_n lies in $M_{Ax=y}$, from which we infer, as in the case of exact data, that x_n is our solution and we can stop iterating. Otherwise (2.45) is valid, i.e.

$$0 \leq \eta_{j_n} m_0 + \delta_{k_n} < \beta R_{A_{j_n}, y_{k_n}}(x_n)\,. \tag{2.77}$$

So we may also assume $R_{A_{j_n}, y_{k_n}}(x_n) > 0$ for all $n \in \mathbb{N}$. For $z \in M_{Ax=y} \cap B_{m_0}$ $\big(\Leftrightarrow Az = y \text{ and } \|z\| \leq m_0\big)$ the second summand in (2.76) can be estimated as follows.

$$\begin{aligned}
&|\langle J_X(x_{n_i}) - J_X(x_{n_l}) \,|\, x_{n_i} - z\rangle| \\
&= \left|\sum_{n=n_l}^{n_i-1} \mu_n \,\langle J_Y(A_{j_n}x_n - y_{k_n}) \,|\, A_{j_n}x_{n_i} - A_{j_n}z\rangle\right| \\
&\leq \sum_{n=n_l}^{n_i-1} \mu_n \Big(\big|\big\langle J_Y(A_{j_n}x_n - y_{k_n}) \,\big|\, A_{j_{n_i}}x_{n_i} - y_{k_{n_i}}\big\rangle\big| \\
&\quad + \big|\big\langle J_Y(A_{j_n}x_n - y_{k_n}) \,\big|\, y_{k_{n_i}} - y\big\rangle\big| \\
&\quad + |\langle J_Y(A_{j_n}x_n - y_{k_n}) \,|\, Az - A_{j_n}z\rangle| \\
&\quad + \big|\big\langle J_Y(A_{j_n}x_n - y_{k_n}) \,\big|\, \big(A_{j_n} - A\big)x_{n_i}\big\rangle\big| \\
&\quad + \big|\big\langle J_Y(A_{j_n}x_n - y_{k_n}) \,\big|\, \big(A_{j_{n_i}} - A\big)x_{n_i}\big\rangle\big| \Big)\,.
\end{aligned}$$

Since the sequences $(\eta_n)_n$ and $(\delta_n)_n$ are supposed to be nonincreasing we get

$$\begin{aligned}
&|\langle J_X(x_{n_i}) - J_X(x_{n_l}) \,|\, x_{n_i} - z\rangle| \\
&\leq \sum_{n=n_l}^{n_i-1} \mu_n R_{A_{j_n}, y_{k_n}}(x_n)^{r-1} \big(R_{A_{j_{n_i}}, y_{k_{n_i}}}(x_{n_i}) + \delta_{k_n} + \eta_{j_n} m_0 + 2\eta_{j_n}\|x_{n_i}\|\big)\,.
\end{aligned}$$

For $m \in \mathbb{N}$ with $\|x_n\| \leq m$ for all $n \in \mathbb{N}$ relation (2.77) yields

$$2\eta_{j_n}\|x_{n_i}\| \leq 2\beta \frac{m}{m_0} R_{A_{j_n}, y_{k_n}}(x_n)\,.$$

Therefore by the definition of μ_n (2.46) and property (S.3) of the sequence $\big(R_{A_{j_{n_i}}, y_{k_{n_i}}}(x_{n_i})\big)_i$ we finally get

$$|\langle J_X(x_{n_i}) - J_X(x_{n_l}) \,|\, x_{n_i} - x\rangle| \leq \frac{1 + \beta + 2\beta\frac{m}{m_0}}{\|A_{j_n}\|} \sum_{n=n_l}^{n_i-1} \tau_n \|x_n\|^{p-1} R_{A_{j_n}, y_{k_n}}(x_n)$$

and the assertion follows as in the proof of 2.23 (we remind that the condition $\overline{\mathcal{R}(A_j^*)} \subset \overline{\mathcal{R}(A^*)}$ assures that the iterates $J_X(x_n)$ remain in $\overline{\mathcal{R}(A^*)}$). $\square$

In the case of noisy data A_η, y_δ we again use the discrepancy principle, i.e. we terminate the iteration when for the first time

$$R_{A_\eta, y_\delta}(x_n) \leq \frac{1}{\beta}(\eta m_0 + \delta)\,,$$

and call the respective stopping index $n(\eta, \delta)$.

Proposition 2.26. *Together with the discrepancy principle method 2.24 is a regularization method for finding the Bregman projection of an arbitrary point* $x_0 \in X$ *onto the set* $M_{Ax=y}$*, i.e.*

$$\lim_{(\eta,\delta)\to 0} \|x_{n(\eta,\delta)} - \Pi^p_{M_{Ax=y}}(x_0)\| = 0\,.$$

The same method enables us to compute metric projections onto the affine subspace $M_{Ax=y}$. To this end we recall that by (1.26) (b) we have

$$P_{M_{Ax=y}}(\tilde{x}_0) = \tilde{x}_0 + \Pi^p_{M_{Ax=y}-\tilde{x}_0}(0) = \tilde{x}_0 + \Pi^p_{M_{Ax=\tilde{y}}}(0) \tag{2.78}$$

with $\tilde{y} = y - A\tilde{x}_0$ for some $\tilde{x}_0 \in X$. So we only have to start method 2.22 with $x_0 = 0$ and $\tilde{y} = y - A\tilde{x}_0$ instead of y (or method 2.24 with $x_0 = 0$ and $\tilde{y}_k = y_k - A_j\tilde{x}_0$ instead of y_k). And since the sequence $(x_n)_n$ converges strongly to $\Pi^p_{M_{Ax=\tilde{y}}}(0)$, the sequence $(\tilde{x}_0 + x_n)_n$ converges strongly to $P_{M_{Ax=y}}(\tilde{x}_0)$.

Often the element $\tilde{x}_0$ is also called *initial discrepancy*, and $P_{M_{Ax=y}}(\tilde{x}_0)$ is referred to as the $\tilde{x}_0$*-minimum norm solution* of the operator equation $Ax = y$, i.e. the (unique) element $x \in X$ with

$$Ax = y \quad \text{and} \quad \|x - \tilde{x}_0\| = \min\{\|z\| \mid Az = y\}\,.$$

In this context we can formulate the following corollary.

Corollary 2.27. *Together with the discrepancy principle method 2.24 with starting point* $x_0 = 0$ *and* $\tilde{y}_{\tilde{\delta}} = y_\delta - A_\eta\tilde{x}_0$ *(with* $\tilde{\delta} \leq \delta + \eta\|\tilde{x}_0\|$ *) is a regularization method for finding the* $\tilde{x}_0$*-minimum norm solution of the operator equation* $Ax = y$*.*

To tackle the problem of projecting onto an affine subspace of the form $y + \overline{\mathcal{R}(A)} \subset Y$ we revisit proposition 1.29 and see that

$$\Pi_{y+\overline{\mathcal{R}(A)}}(y_0) = J^*\Big(\Pi^*_{J(y_0)+\mathcal{N}(A^*)}\big(J(y)\big)\Big) = J^*\Big(\Pi^*_{M_{A^*y^*=x^*}}\big(J(y)\big)\Big) \tag{2.79}$$

with $x^* = A^*J(y_0)$, whereby J is the duality mapping and Π the Bregman projection in Y and J^* is the duality mapping and Π^* the Bregman projection in Y^*. Thus we can solve this problem by method 2.22 (or 2.24), too, when instead of $A \in \mathcal{L}(X,Y)$ we use the operator $A^* \in \mathcal{L}(Y^*,X^*)$ and iterate in Y^*. Nevertheless we must look more closely at the assumptions assuring convergence. Assumption (A,y) $\big($resp. $(A_j,y_k)\big)$ translates into (A^*,x^*) $\big($resp. $(A_j^*,x_k^*)\big)$, i.e. now we have to require the space Y^* to be smooth and uniformly convex and X is allowed to be an arbitrary Banach space. Then by 2.23 (resp. 2.25) the sequence $(y_n^*)_n$ generated in Y^* converges strongly to $\Pi^*_{M_{A^*y^*=x^*}}\big(J(y)\big)$. Since by 1.14 (f) J^* is norm-to-weak-continuous, this at least implies that the sequence $\big(J^*(y_n^*)\big)_n$ converges weakly to $J^*\Big(\Pi^*_{M_{A^*y^*=x^*}}\big(J(y)\big)\Big) = \Pi_{y+\overline{\mathcal{R}(A)}}(y_0)$. Hence for strong convergence we should additionally require that Y^* is uniformly smooth $\big($prop. 1.14 (d)$\big)$.

The same considerations hold for the metric projection which we can compute as before via

$$P_{y+\overline{\mathcal{R}(A)}}(\tilde{y}_0) = \tilde{y}_0 + \Pi_{(y-\tilde{y}_0)+\overline{\mathcal{R}(A)}}(0) = \tilde{y}_0 + J^*\Big(\Pi^*_{M_{A^*y^*=0}}\big(J(y-\tilde{y}_0)\big)\Big)$$

and especially

$$P_{\overline{\mathcal{R}(A)}}(\tilde{y}_0) = \tilde{y}_0 - J^*\Big(\Pi^*_{M_{A^*y^*=0}}\big(J(\tilde{y}_0)\big)\Big)\,. \tag{2.80}$$

Now that we know that method 2.22 enables us to compute Bregman (metric) projections onto $y+\overline{\mathcal{R}(A)}$ by producing sequences $\big(y_n = J^*(y_n^*)\big)_n$ in Y which converge to $\tilde{y} := \Pi_{y+\overline{\mathcal{R}(A)}}(y_0)$, the question arises whether we can also produce sequences $(x_n)_n$ in X such that $(y+Ax_n)_n$ converges to $\tilde{y}$ or that even the sequence $(x_n)_n$ itself converges to some $\tilde{x} \in X$ with $\tilde{y} = y + A\tilde{x}$. We can partially answer this question by considering the following iteration in X:

$$x_{n+1} := x_n - \mu_n J_X^*\big(A^*J_Y(y+Ax_n) - A^*J_Y(y_0)\big) \quad , \quad x_0 := 0\,. \tag{2.81}$$

We apply A and add y on both sides and get

$$y + Ax_{n+1} = y + Ax_n - \mu_n A J_X^*\big(A^*J_Y(y+Ax_n) - A^*J_Y(y_0)\big)\,.$$

By setting $y_n^* := J_Y(y+Ax_n) \Leftrightarrow J_Y^*(y_n^*) = y + Ax_n$ and $x^* := A^*J_Y(y_0)$ we can write this as

$$y_{n+1}^* = J_Y\big(J_Y^*(y_n^*) - \mu_n A J_X^*(A^*y_n^* - x^*)\big) \quad , \quad y_0^* = J_Y(y)\,.$$

This is just method 2.22 and thus the sequence $\big(y + Ax_n = J_Y^*(y_n^*)\big)$ converges weakly to $\tilde{y}$ (strongly for uniformly smooth Y^*). We even get a stronger result when A is injective and X is finite dimensional (but still allowed to be endowed with an arbitrary norm). In that case we have $\overline{\mathcal{R}(A)} = \mathcal{R}(A)$ and there exists a unique $\tilde{x} \in X$ with $\tilde{y} = y + A\tilde{x}$. Therefore the sequence $\big(A(x_n - \tilde{x}) = y + Ax_n - \tilde{y}\big)_n$ converges in $\mathcal{R}(A)$ to zero, from which we infer that $(x_n)_n$ converges to $\tilde{x}$.

2.5 Line Searches

In [46] we claimed that we are thinking about ways to easily and efficiently implement the methods. In this and the next section we offer some possibilities. We examine more closely what is needed to derive efficient algorithms. Each iteration $x_{n+1} = T_n(x_n)$ only consists of applying an operator T_n which has a rather simple structure. Of course we must know the duality mappings of the spaces we work in (e.g. 1.11) since they appear in the very definition of the operators $T^{\mu}_{A,Q,\Pi}$, $T^{\mu}_{A,Q,P}$, $T^{\mu}_{A,y}$ and $T^{\mu}_{A,y,+}$. Moreover we should be able to efficiently compute the projections involved (e.g. 1.30). The main work left is then the application of the linear operators A and A^* ($\rightsquigarrow$ matrix-vector-products). Only the choice of the parameters μ_n (resp. τ_n) seems to be a little difficult in general. But as we have already mentioned before we can overcome this problem by a line search method. When we look again at what is necessary to prove the (weak) convergence of the methods, we see that we may take as μ_n any $t \in \mathbb{R}$ such that for

$$x_{n+1} = x_{n+1}(t) = J^*_X\big(J_X(x_n) - tA^*_{j_n} w^*_n\big)$$

(with the appropriate $w^*_n \in Y^*$) and all $z \in M\big(\cap B_{m_0}\big)$ a relation of the form

$$\Delta_p\big(x_{n+1}(t), z\big) \leq \Delta_p(x_n, z) - \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|} \tau_n \|x_n\|^{p-1} R_n(x_n) \tag{2.82}$$

holds (which is fulfilled for $t = \mu_n$). We arrived at this relation in 2.7-2.12 by at first estimating $\Delta_p\big(x_{n+1}(t), z\big)$ from above by

$$\begin{aligned} &\Delta_p\big(x_{n+1}(t), z\big) \\ \leq\; &\frac{1}{q}\, \big\|J_X(x_n) - tA^*_{j_n} w^*_n\big\|^q + t\big(\, \langle w^*_n \,|\, A_{j_n} x_n\rangle - (1-\beta)\|w^*_n\| R_n(x_n)\big) \\ &+ \frac{1}{p}\|z\|^p - \langle J_X(x_n) \,|\, z\rangle\,, \end{aligned}$$

whereby $t \geq 0$ was necessary to obtain the estimate

$$t\,\langle w^*_n \,|\, A_{j_n} z\rangle \leq t\big(\,\langle w^*_n \,|\, A_{j_n} x_n\rangle - (1-\beta)\|w^*_n\| R_n(x_n)\big)\,. \tag{2.83}$$

We continued by estimating from above the terms in which the parameter t occurs, i.e. the function

$$h_n(t) := \frac{1}{q}\,\|J_X(x_n) - t u^*_n\|^q + t\,\alpha_n \tag{2.84}$$

with

$$u^*_n := A^*_{j_n} w^*_n \quad \text{and} \quad \alpha_n := \langle w^*_n \,|\, A_{j_n} x_n\rangle - (1-\beta)\|w^*_n\| R_n(x_n)\,. \tag{2.85}$$

We did this via the characteristic inequality for the dual X^* and showed that the choice $t = \mu_n$ leads to the desired relation (2.82). Instead we may as well take any $t \geq 0$ such that

$$h_n(t) \leq h_n(\mu_n)$$

and we can find such a t by solving the one-dimensional optimization problem

$$\min_{t \geq 0} h_n(t)\,. \tag{2.86}$$

The important thing about h_n is, that it is independent of the (unknown) elements $z \in M\big(\cap B_{m_0}\big)$. We can interpret (2.86) in a familiar way by proposition 1.30 (b): Solving (2.86) is equivalent to computing the Bregman projection of x_n onto the halfspace $H_{\leq}(u_n^*, \alpha_n)$ if x_n does not yet lie in $H_{\leq}(u_n^*, \alpha_n)$. This is the case if $x_n \notin M\big(\cap B_{m_0}\big)$, because in the lemmas 2.7-2.14 we have shown that then we have $\|w_n^*\| R_n(x_n) \neq 0$ and thus

$$\langle u_n^* \,|\, x_n \rangle = \langle w_n^* \,|\, A_{j_n} x_n \rangle > \alpha_n\,.$$

Moreover every $z \in M\big(\cap B_{m_0}\big)$ lies in $H_{\leq}(u_n^*, \alpha_n)$ by (2.83), i.e.

$$\mathbf{C} \subset \bigcap_{n \in \mathbb{N}} H_{\leq}(u_n^*, \alpha_n)\,.$$

If $x_n \in M\big(\cap B_{m_0}\big)$ then x_n is already the desired point and we set $x_{n+1} := x_n$. Therefore we may as well solve

$$\min_{t \in \mathbb{R}} h_n(t)\,, \tag{2.87}$$

i.e.

$$x_{n+1} = \Pi^p_{H(u_n^*, \alpha_n)}(x_n)\,,$$

because we know that the solution t will be positive. With regard to 2.3 we may view the methods as subgradient methods and thus (2.87) as a *line search* along a subgradient of an appropriate functional as *search direction* u_n^*. All in all this enables us to easily implement the methods by solving in each iterative step a one-dimensional minimization problem with a well behaving function h_n $\big($convex, continuously differentiable $\big($see 1.30 (b)$\big)$, with low costs for function and gradient evaluations ($J_X(x_n)$, u_n^* and α_n are fixed)$\big)$. Of course in practice we are in general not able to perform exact line searches. But by (2.82) the condition

$$h_n(t) \leq h_n(0) - \frac{(1-\gamma)(1-\beta)}{\|A_{j_n}\|} \tau_n \|x_n\|^{p-1} R_n(x_n)$$

with some $\tau_n \in (\tau, 1)$, $\tau > 0$, ensures convergence also for inexact line searches. However we should mention that in the infinite-dimensional case we could not prove the strong convergence of method 2.22 and 2.24 with line searches, since for the part in which we showed that the sequence $(x_n)_n$ has a Cauchy subsequence we somehow needed $t = \mu_n$; nevertheless it remains true that every subsequence has in turn a subsequence converging weakly to the same point $\Pi^p_{M_{Ax=y}}(x_0)$ and thus the whole sequence converges weakly.

What do we gain when we know more about the space we are iterating in? When we want to perform a line search, it is advantageous to have a better starting point than the canonical $t_0 = 0$ for minimizing h_n. A good starting point would be $t_0 = \mu_n$ since it already guarantees convergence. So whenever we can easily compute μ_n, e.g. as we have done for L_p-spaces in (2.62) and (2.66), we should use it. In this case we can even use method 2.22 to perform the line search. This procedure is contained in the following two propositions.

Proposition 2.28. *Let $H(u^*, \alpha)$ be a hyperplane in an L_p-space X ($p \leq 2$) with the normalized duality mapping and $x_0 \in X$ be given. Then the sequence $(x_n)_n$ iteratively defined by*

$$x_{n+1} := J^*\big(J(x_n) - s_n u^*\big) \tag{2.88}$$

with

$$s_n := \frac{\langle u^* \,|\, x_n \rangle - \alpha}{(q-1)\|u^*\|^2} \tag{2.89}$$

converges strongly to the Bregman projection of x_0 onto $H(u^*, \alpha)$. If we set

$$h(t) := \frac{1}{2}\,\|J(x_0) - tu^*\|^2 + t\,\alpha \quad , \quad t \in \mathbb{R}$$

and

$$t_n := \sum_{k=0}^{n} s_n \,, \tag{2.90}$$

then we have

$$x_{n+1} = J^*\big(J(x_0) - t_n u^*\big)\,, \tag{2.91}$$

the sequence $(t_n)_n$ converges to the unique solution $\tilde{t}$ of

$$\min_{t \in \mathbb{R}} h(t)$$

and the sequence $\big(h(t_n)\big)_n$ converges decreasingly to $h(\tilde{t})$ with

$$\begin{aligned} h(t_n) &\leq h(t_{n-1}) - \frac{|\,\langle u^* \,|\, x_n \rangle - \alpha|^2}{2(q-1)\|u^*\|^2} \\ &\leq h(0) - \frac{1}{2(q-1)\|u^*\|^2} \sum_{k=0}^{n} |\,\langle u^* \,|\, x_k \rangle - \alpha|^2 \,. \end{aligned} \tag{2.92}$$

Proof. If we set $Ax := \langle u_n^* \,|\, x \rangle$ for $x \in X$, $Y := (\mathbb{R}, |.|)$ and $y := \alpha_n$ then we have $A \in \mathcal{L}(X,Y)$ and

$$\Pi^p_{H(u_n^*,\alpha_n)}(x_n) = \Pi^p_{M_{Ax=y}}(x_n)\,.$$

Therefore we can apply method 2.22 and get the iteration

$$x_{n+1} = J^*\big(J(x_n) - \mu_n A^* J_Y(Ax_n - y)\big) = J^*\big(J(x_n) - \mu_n u^*(\langle u^* \,|\, x_n \rangle - \alpha)\big)\,,$$

whereby we can use the μ_n of (2.64) under assumption (A,y) with $r = 2$:

$$\mu_n = \frac{1}{q-1}\frac{R_n^2}{\|A^* J_Y(Ax_n - y)\|^2} = \frac{1}{q-1}\frac{|\,\langle u^* \,|\, x_n \rangle - \alpha|^2}{\|u^*(\langle u^* \,|\, x_n \rangle - \alpha)\big)\|^2} = \frac{1}{(q-1)\|u^*\|^2}\,.$$

Hence the iteration has the form (2.88) with s_n as in (2.89) and the assertion about convergence is valid (from (2.92) it follows that $R_n = |\,\langle u^* \,|\, x_n \rangle - \alpha|$ converges to zero and the exceptional case in 2.23 cannot happen). The form (2.91) is obvious by (2.88) and (2.90). By (2.91) we get

$$t_n u^* = J(x_0) - J(x_{n+1})\,.$$

We apply $J^*(u^*)$ on both sides and arrive at

$$t_n = \frac{\langle J(x_0) - J(x_{n+1}) \,|\, J^*(u^*) \rangle}{\|u^*\|^q}\,,$$

whereby the right hand side converges. Since $\tilde{t}$ is unique, the limit of $(t_n)_n$ must be $\tilde{t}$. Finally by 1.18 (a) we get

$$\begin{aligned}
h(t_n) &= \frac{1}{2}\,\|J(x_0) - t_n u^*\|^2 + t_n\,\alpha \\
&= \frac{1}{2}\,\big\|\big(J(x_0) - t_{n-1}u^*\big) - s_n u^*\big\|^2 + t_{n-1}\,\alpha + s_n\,\alpha \\
&\le \frac{1}{2}\,\|J(x_0) - t_{n-1}u^*\|^2 - s_n\,\langle u^* \,|\, x_n \rangle + \frac{q-1}{2}\,\|u^*\|^2\; s_n^2 \\
&\quad + t_{n-1}\,\alpha + s_n\,\alpha \\
&= h(t_{n-1}) - s_n(\langle u^* \,|\, x_n \rangle - \alpha) + \frac{q-1}{2}\,\|u^*\|^2\; s_n^2 \\
&= h(t_{n-1}) - \frac{|\,\langle u^* \,|\, x_n \rangle - \alpha|^2}{2(q-1)\|u^*\|^2}\,.
\end{aligned}$$

□

The proof of the next proposition for the case $p \ge 2$ is quite similar by using 1.18 (b) and (2.66).

Proposition 2.29. *Let $H(u^*, \alpha)$ be a hyperplane in an L_p-space X $(p \geq 2)$ with duality mapping with the same weight p and $x_0 \in X$ be given. Then the sequence $(x_n)_n$ iteratively defined by*

$$x_{n+1} := J^q\big(J^p(x_n) - s_n u^*\big) \tag{2.93}$$

with

$$s_n := \frac{|\langle u^* \,|\, x_n\rangle - \alpha|^{p-1}}{2^{p-2}\|u^*\|^p} \operatorname{sgn}\big(\langle u^* \,|\, x_n\rangle - \alpha\big) \tag{2.94}$$

converges strongly to the Bregman projection of x_0 onto $H(u^*, \alpha)$. If we set

$$h(t) := \frac{1}{q}\,\|J^p(x_0) - tu^*\|^q + t\,\alpha \quad , \quad t \in \mathbb{R}$$

and

$$t_n := \sum_{k=0}^{n} s_n\,, \tag{2.95}$$

then we have

$$x_{n+1} = J^q\big(J^p(x_0) - t_n u^*\big)\,, \tag{2.96}$$

the sequence $(t_n)_n$ converges to the unique solution $\tilde{t}$ of

$$\min_{t \in \mathbb{R}} h(t)$$

and the sequence $\big(h(t_n)\big)_n$ converges decreasingly to $h(\tilde{t})$ with

$$\begin{aligned} h(t_n) &\leq h(t_{n-1}) - \frac{|\,\langle u^* \,|\, x_n\rangle - \alpha|^p}{p2^{p-2}\|u^*\|^p} \\ &\leq h(0) - \frac{1}{p2^{p-2}\|u^*\|^p}\sum_{k=0}^{n}|\,\langle u^* \,|\, x_k\rangle - \alpha|^p\,. \end{aligned} \tag{2.97}$$

2.6 Generalized CG and Sequential Subspace Methods

Even with line searches the above discussed methods are more or less steepest descent methods and it is well known that these can be arbitrarily slow (nevertheless with appropriate stopping rules like the discrepancy principle they often prove superior to faster methods with respect to regularization, where only some iterations are needed to obtain good results).

We concentrate again on the special case of computing projections onto affine subspaces in case of exact data. At first we show that in the finite-dimensional case we can formulate this as an in general non-linear optimization problem which can be solved by standard optimization routines like non-linear *CG* methods and then propose alternative generalized *CG* and sequential subspace methods which also work in the infinite-dimensional case (but we do not claim that they are superior in applications). Let us first consider the problem of projecting onto the intersection of finitely many hyperplanes.

Proposition 2.30. *Let X be reflexive, smooth and strictly convex, and vectors $u_1^*, \dots, u_N^* \in X^*$ and $\alpha_1, \dots, \alpha_N \in \mathbb{R}$ be given. For $x \in X$ let $h_x : \mathbb{R}^N \longrightarrow \mathbb{R}$ be the convex, differentiable function*

$$h_x(t) := \frac{1}{q} \left\| J^p(x) - \sum_{k=1}^N t_k\, u_k^* \right\|^q + \sum_{k=1}^N t_k\, \alpha_k \quad , \quad t = (t_1, \dots, t_N) \in \mathbb{R}^N$$

with continuous partial derivatives

$$\partial_j h_x(t) = -\left\langle u_j^* \,\middle|\, J^q \left(J^p(x) - \sum_{k=1}^N t_k\, u_k^* \right) \right\rangle + \alpha_j \quad , \quad j = 1, \dots, N\,.$$

If the intersection

$$H := \bigcap_{k=1}^N H(u_k^*, \alpha_k)$$

of the hyperplanes $H(u_k^*, \alpha_k)$ is not empty then the Bregman projection of x onto H is given by

$$\Pi_H^p(x) = J^q \left(J^p(x) - \sum_{k=1}^N t_k^0\, u_k^* \right) ,$$

whereby $t_0 = (t_1^0, \dots, t_N^0)$ is a (necessarily existing) solution of the N-dimensional optimization problem

$$\min_{t \in \mathbb{R}^N} h_x(t)\,. \tag{2.98}$$

Moreover for all $z \in H$ the Bregman projection in X^* of $J^p(z)$ onto the affine subspace $J^p(x) + \operatorname{span}\{u_1^*, \dots, u_N^*\}$ is given by

$$\Pi^q_{J^p(x)+\operatorname{span}\{u_1^*,\dots,u_N^*\}}\big(J^p(z)\big) = J^p(x) - \sum_{k=1}^N t_k^0\, u_k^*\,.$$

If the vectors $u_1^*, \dots, u_N^*$ are linearly independent then t_0 is unique.

Proof. We can prove this in the way of 1.30 (b) by using 1.29 (a). For all $z \in H$ we can write

$$H = z + \big(\operatorname{span}\{u_1^*, \dots, u_N^*\}\big)^\perp\,.$$

Thus an element $x_0 \in X$ is the Bregman projection of x onto H iff $x_0 \in H$ and $J^p(x_0) \in J^p(x) + \operatorname{span}\{u_1^*, \dots, u_N^*\}$ iff $J^p(x_0) = J^p(x) - \sum_{k=1}^N t_k^0\, u_k^*$ with some $t_1^0, \dots, t_N^0 \in \mathbb{R}$ such that $\langle u_k^* \,|\, x_0 \rangle = \alpha_k$ for all $k = 1, \dots, N$ (uniquely determined in case $u_1^*, \dots, u_N^*$ are linearly independent) iff $t_0 = (t_1^0, \dots, t_N^0) \in \mathbb{R}^N$ is a (unique) solution of the optimization problem

$$\min_{t\in\mathbb{R}^N} \Delta_q^*\left(J^p(z), J^p(x) - \sum_{k=1}^N t_k\, u_k^*\right)$$
$$= \min_{t\in\mathbb{R}^N} \frac{1}{p}\|z\|^p - \langle z \,|\, J^p(x)\rangle + \sum_{k=1}^N t_k\, \langle z \,|\, u_k^*\rangle + \frac{1}{q}\left\| J^p(x) - \sum_{k=1}^N t_k\, u_k^* \right\|^q ,$$

which is equivalent to (2.98) and $J^p(x) - \sum_{k=1}^N t_k^0\, u_k^*$ being the Bregman projection in X^* of $J^p(z)$ onto $J^p(x) + \mathrm{span}\{u_1^*, \ldots, u_N^*\}$. □

We can use this to compute Bregman projections onto affine subspaces of the form $M_{Ax=y}$.

Corollary 2.31. *We assume (A, y). Let Y be of finite dimension N and $w_1^*, \ldots, w_M^* \in Y^*$ $(M \le N)$ be given such that*

$$\mathcal{R}(A^*) = \mathrm{span}\{A^* w_1^*, \ldots, A^* w_M^*\}\,.$$

The Bregman projection of x onto $M_{Ax=y}$ is then given by

$$\Pi^p_{M_{Ax=y}}(x) = J^q\left(J^p(x) - \sum_{k=1}^M t_k^0\, A^* w_k^*\right) ,$$

whereby $t_0 = (t_1^0, \ldots, t_M^0)$ is a (necessarily existing) solution of the M-dimensional optimization problem

$$\min_{t\in\mathbb{R}^M} h_x(t) \tag{2.99}$$

with h_x defined as in 2.30, whereby we set $u_k^* = A^* w_k^*$ and $\alpha_k := \langle w_k^* \,|\, y\rangle$ for $k = 1, \ldots, M$. Moreover for all $z \in M_{Ax=y}$ the Bregman projection in X^* of $J^p(z)$ onto the affine subspace $J^p(x) + \mathrm{span}\{A^* w_1^*, \ldots, A^* w_M^*\}$ is given by

$$\Pi^q_{J^p(x)+\mathrm{span}\{A^*w_1^*,\ldots,A^*w_M^*\}}\big(J^p(z)\big) = J^p(x) - \sum_{k=1}^N t_k^0\, A^* w_k^*\,.$$

If the vectors $u_1^*, \ldots, u_M^*$ are linearly independent then t_0 is unique.

Proof. We only remark that z lies in $M_{Ax=y}$ iff $\langle A^* w_k^* \,|\, z\rangle = \langle w_k^* \,|\, y\rangle$ for all $k = 1, \ldots, M$ and therefore $\bigcap_{k=1}^M H(u_k^*, \alpha_k) = M_{Ax=y} \neq \emptyset$. □

We recall that Bregman projections onto $z + \overline{\mathcal{R}(A)}$ (with then finite-dimensional X) and the respective metric projections can be computed via (2.79), (2.78) and (2.80). Problem (2.99) is an in general nonlinear, convex, smooth unconstrained optimization problem which can be solved by standard optimization routines (see e.g. [37] and [36], where we have taken our ideas from).

We now propose and develop methods in the spirit of conjugate gradient (CG) and sequential subspace methods. They are based on the iterative process of the previously discussed methods. Therefore we once again look at the iterates in case of exact data under assumption (A, y):

$$x_{n+1} = x_{n+1}(t) = J_X^*\big(J_X(x_n) - tA^*w_n^*\big)$$

with

$$w_n^* = J_Y(Ax_n - y) \quad \text{and} \quad R_n := R_{A,y}(x_n) = \|Ax_n - y\| .$$

For all $z \in M_{Ax=y}$ we have

$$\langle A^*w_n^* \,|\, z\rangle = \langle w_n^* \,|\, y\rangle = \langle A^*w_n^* \,|\, x_n\rangle + \langle w_n^* \,|\, y - Ax_n\rangle = \langle A^*w_n^* \,|\, x_n\rangle - R_n^r$$

$\big($with equality! in contrast to "$\leq$" in case of (A, Q) or $(A, y, +)\big)$ and thus

$$\Delta_p\big(x_{n+1}(t), z\big) = h_n(t) - \langle J_X(x_n) \,|\, z\rangle + \frac{1}{p}\|z\|^p \tag{2.100}$$

with

$$h_n(t) = \frac{1}{q}\,\|J_X(x_n) - tA^*w_n^*\|^q + t\,\alpha_n$$

and

$$\alpha_n = \langle A^*w_n^* \,|\, x_n\rangle - R_n^r = \langle w_n^* \,|\, y\rangle\ .$$

Because of equality in (2.100) we can directly measure the decrease of $\Delta_p\big(x_{n+1}(t), z\big)$ via $h_n(t)$ and thus the solution $t_0 \in \mathbb{R}$ of

$$\min_{t\in\mathbb{R}} h_n(t)$$

fulfills for all $z \in M_{Ax=y}$

$$\Delta_p\big(x_{n+1}(t_0), z\big) \leq \Delta_p\big(x_{n+1}(\mu_n), z\big)$$

with the μ_n of (2.40). Hence (weak) convergence is ensured. Although we know that t_0 is in fact positive, we do not have to impose this restriction here $\big($as we have to in case of (A, Q) or $(A, y, +)\big)$. Proposition 2.30 and 2.31 suggest that the more search directions A^*w_k we use to find x_{n+1} , the better x_{n+1} will approximate the exact solutions $M_{Ax=y}$.

Method 2.32. *We assume (A, y), choose an arbitrary starting point $x_0 \in X$ and for $n = 0, 1, 2, \ldots$ repeat the following steps:*
We set

$$R_n := \|Ax_n - y\| .$$

If $R_n = 0$ then STOP, else we set

$$w_n^* := J_Y(Ax_n - y) \tag{2.101}$$

and

$$\alpha_k^n := \begin{cases} \alpha_k^{n-1} & , \quad k = 0, \ldots, n-1 \\ \langle A^* w_n^* \,|\, x_n \rangle - R_n^r & , \quad k = n \end{cases} , \tag{2.102}$$

and define

$$x_{n+1} := x_{n+1}(t_n) = J_X^* \left(J_X(x_n) - \sum_{k=0}^{n} t_k^n \, A^* w_k^* \right) , \tag{2.103}$$

whereby $t_n = (t_0^n, \ldots, t_n^n)$ *is the (necessarily existing) unique solution of the* $(n+1)$*-dimensional optimization problem*

$$\min_{t \in \mathbb{R}^{n+1}} h_n(t) \tag{2.104}$$

with

$$h_n(t) := \frac{1}{q} \left\| J_X(x_n) - \sum_{k=0}^{n} t_k \, A^* w_k^* \right\|^q + \sum_{k=0}^{n} t_k \, \alpha_k^n \quad , \quad t = (t_0, \ldots, t_n) \in \mathbb{R}^{n+1} .$$

Obviously it suffices to store the vectors $u_k^* := A^* w_k^*$ instead of both u_k^* and w_k^*. Existence and uniqueness of t_n are consequences of 2.30 and the following proposition.

Proposition 2.33. *Method 2.32 either stops after a finite number* $n \in \mathbb{N}$ *of iterations (in case* $R_n = 0$*) with* x_n *being the Bregman projection* $\tilde{x}$ *of* x_0 *onto the set* $M_{Ax=y}$ *or the sequence of the iterates* $(x_n)_n$ *converges weakly to* $\tilde{x}$ *(with the exceptional case as in 2.23). Moreover as long as* $R_n \neq 0$ *the following holds:*

(a) The vectors $\{w_0^*, \ldots, w_n^*\} \subset Y^*$ *and* $\{A^* w_0^*, \ldots, A^* w_n^*\} \subset X^*$ *are linearly independent.*

(b) For

$$H_n := \bigcap_{k=0}^{n} H(A^* w_k^*, \alpha_k^n) \quad \text{and} \quad U_n^* := \operatorname{span}\{A^* w_0^*, \ldots, A^* w_n^*\}$$

we have

$$U_n^* = \operatorname{span}\{A^* J_Y(Ax_0 - y), \ldots, A^* J_Y(Ax_n - y)\}$$

$$M_{Ax=y} \subset H_n \quad , \quad x_{n+1} = \Pi_{H_n}^p(x_n)$$

and

$$J_X(x_{n+1}) = \Pi_{J_X(x_0)+U_n^*}^q \big(J_X(z) \big) \quad \text{for all} \quad z \in M_{Ax=y} . \tag{2.105}$$

(c) We have $\langle w_k^* \,|\, Ax_{n+1} - y \rangle = 0$ *for all* $k = 0, \ldots, n$.

Proof. The assertions about convergence follow by 2.23 and from what we have just discussed above, because for all $z \in M_{Ax=y}$ we have

$$\Delta_p\big(x_{n+1}(t_n), z\big) \le \Delta_p\Big(x_{n+1}\big(t = (0, \ldots, 0, \mu_n)\big), z\Big)$$

with the μ_n of (2.40) (and thus every subsequence of $(x_n)_n$ has in turn a subsequence converging weakly to the same point $\tilde{x} = \Pi^p_{M_{Ax=y}}(x_0)$, from which we infer that the whole sequence converges weakly). (b) is clear by the definition of the vectors w_k^* (2.101) and by 2.30, since

$$\langle A^* w_k^* \,|\, z \rangle = \langle w_k^* \,|\, y \rangle = \langle A^* w_k^* \,|\, x_k \rangle - R_k^r = \alpha_k^n$$

for all $z \in M_{Ax=y}$ and inductively by the definition of the iterates (2.103)

$$J_X(x_n) + U_n^* = J_X(x_0) + U_n^* \,.$$

In (a) it suffices to show that the vectors $\{A^* w_0^*, \ldots, A^* w_n^*\}$ are linearly independent. We show this inductively in combination with (c). If $R_0 \neq 0$ then $A^* w_0^* \neq 0$, because otherwise we take some $z \in M_{Ax=y}$ and get

$$0 = \langle A^* w_0^* \,|\, x_0 - z \rangle = \langle J_X(Ax_0 - y) \,|\, Ax_0 - Az \rangle = \|Ax_0 - y\|^r = R_0^r \,.$$

Hence $w_0^* \neq 0$ as well and t_0 fulfills

$$\begin{aligned} 0 = h_0'(t_0) &= -\left\langle A^* w_0^* \,|\, J_X^* \left(J_X(x_0) - t_0\, A^* w_0^*\right)\right\rangle + \alpha_0^0 \\ &= -\left\langle w_0^* \,|\, Ax_1 \right\rangle + \left\langle w_0^* \,|\, y \right\rangle \\ &= -\left\langle w_0^* \,|\, Ax_1 - y \right\rangle \,. \end{aligned}$$

Now let $R_n \neq 0$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be given such that

$$\sum_{k=0}^{n} \lambda_k\, A^* w_k^* = 0 \,.$$

We take some $z \in M_{Ax=y}$ and get

$$\begin{aligned} 0 &= \sum_{k=0}^{n} \lambda_k\, \langle A^* w_k^* \,|\, x_n - z \rangle = \sum_{k=0}^{n} \lambda_k\, \langle w_k^* \,|\, Ax_n - y \rangle \\ &= \lambda_n\, \langle w_n^* \,|\, Ax_n - y \rangle = \lambda_n\, R_n^r \,. \end{aligned}$$

Thus $\lambda_n = 0$ and we continue with

$$0 = \sum_{k=0}^{n-1} \lambda_k\, \langle A^* w_k^* \,|\, x_{n-1} - z \rangle = \lambda_{n-1}\, R_{n-1}^r \,.$$

Inductively we get $\lambda_k = 0$ for all $k = 0, \ldots, n$. Hence the vectors $\{A^* w_0^*, \ldots, A^* w_n^*\}$ are linearly independent. Finally we get for all $j = 0, \ldots, n$

$$
\begin{aligned}
0 = \partial_j h_n(t_n) &= - \left\langle A^* w_j^* \,\middle|\, J_X^* \left(J_X(x_n) - \sum_{k=0}^{n} t_k^n \, A^* w_k^* \right) \right\rangle + \alpha_j^n \\
&= - \left\langle w_j^* \,\middle|\, A x_{n+1} \right\rangle + \left\langle w_j^* \,\middle|\, y \right\rangle \\
&= - \left\langle w_j^* \,\middle|\, A x_{n+1} - y \right\rangle .
\end{aligned}
$$

□

If Y is of finite dimension N then by (a) and (c) of the above proposition the method stops after at most N iterations (indeed after at most $M = \dim\big(\mathcal{R}(A^*)\big)$ iterations). But of course the hope is that we are done much earlier ($n \ll M$); which by (2.105) should be the case if H_n is already a good approximation to the solution set $M_{Ax=y}$ or equivalently $J_X(x_0) + U_n^*$ is a good approximation to $J_X(M_{Ax=y})$. Therefore we could try to build up U_n^* in such a way that it somehow better fits the solution set. We try to motivate our approach. Suppose we already have vectors $\{w_0^*, \dots, w_{n-1}^*\} \subset Y^*$ at hand such that (a), (b) and (c) of 2.33 are fulfilled. We intend to construct a vector $w_n^* \in Y^*$ which incorporates some information about the solution set of its predecessors so that we will find a better approximation x_{n+1} more quickly (and such that (a), (b) and (c) still hold). Since we have seen that the choice $w_n^* = J_Y(Ax_n - y)$ is already a good one (ensuring convergence) we make the ansatz

$$
w_n^* = J_Y(Ax_n - y) - \sum_{k=0}^{n-1} s_k^n \, w_k^* . \tag{2.106}
$$

We know that x_n is the optimal approximation with respect to U_{n-1}^*, more exactly with respect to the search directions $A^* w_0^*, \dots, A^* w_{n-1}^*$. Therefore the new direction $A^* w_n^*$ should be “as different as possible” from the latter; so that when we search x_{n+1} via t_n the main change will be along $A^* w_n^*$, i.e. in the component t_n^n, and only minor corrections will have to be done along the old directions, i.e. the components $t_0^n, \dots, t_{n-1}^n$ will differ only little from zero. We propose

$$
A^* w_n^* = A^* J_Y(Ax_n - y) - P_{U_{n-1}^*}\big(A^* J_Y(Ax_n - y)\big) ,
$$

i.e. $s^n = (s_0^n, \dots, s_{n-1}^n) \in \mathbb{R}^n$ is the solution of

$$
\min_{s \in \mathbb{R}^n} g_n(s)
$$

with

$$
g_n(s) = \frac{1}{q} \| A^* w_n^*(s) \|^q = \frac{1}{q} \left\| A^* J_Y(Ax_n - y) - \sum_{k=0}^{n-1} s_k \, A^* w_k^* \right\|^q .
$$

By proposition 1.26 (b) and 1.29 (a) we also have

$$
J_X^*(A^* w_n^*) = \Pi^p_{(U_{n-1}^*)^\perp} \Big(J_X^*\big(A^* J_Y(Ax_n - y)\big)\Big)
$$

and thus

$$\langle A^* w_k^* \,|\, J_X^*(A^* w_n^*) \rangle = 0 \quad \text{for all} \quad k = 0, \ldots, n-1 \tag{2.107}$$

(which is equivalent to $\partial_k \, g_n(s^n) = 0$ for all $k = 0, \ldots, n-1$). We look more closely at how this choice of the search directions may qualitatively affect the minimization of the function h_n. By the characteristic inequality for the dual X^*, property (c) of 2.33 and (2.106) we get

$$\begin{aligned}
h_n(t) &= \frac{1}{q} \left\| J_X(x_n) - \sum_{k=0}^{n} t_k \, A^* w_k^* \right\|^q + \sum_{k=0}^{n} t_k \, \langle w_k^* \,|\, y \rangle \\
&\leq \frac{1}{q} \|x_n\|^p - \sum_{k=0}^{n} t_k \, \langle w_k^* \,|\, A x_n - y \rangle + \frac{1}{q} \tilde{\sigma}_q \left(J_X(x_n), \sum_{k=0}^{n} t_k \, A^* w_k^* \right) \\
&= \frac{1}{q} \|x_n\|^p - t_n \, \langle w_n^* \,|\, A x_n - y \rangle + \frac{1}{q} \tilde{\sigma}_q \left(J_X(x_n), \sum_{k=0}^{n} t_k \, A^* w_k^* \right) \\
&= \frac{1}{q} \|x_n\|^p - t_n R_n^r + \frac{1}{q} \tilde{\sigma}_q \left(J_X(x_n), \sum_{k=0}^{n} t_k \, A^* w_k^* \right) ,
\end{aligned}$$

whereby the size of the last summand essentially depends on the norms $\|x_n\|$ and $\|\sum_{k=0}^{n} t_k \, A^* w_k^*\|$ (see also 1.18). Therefore with (2.107) we likewise estimate

$$\begin{aligned}
\frac{1}{q} \left\| \sum_{k=0}^{n} t_k \, A^* w_k^* \right\|^q &= \frac{1}{q} \left\| t_n \, A^* w_n^* + \sum_{k=0}^{n-1} t_k \, A^* w_k^* \right\|^q \\
&\leq \frac{|t_n|^q}{q} \, \|A^* w_n^*\|^q + t_n^{q-1} \, \operatorname{sgn}(t_n) \sum_{k=0}^{n-1} t_k \, \langle A^* w_k^* \,|\, J_X^*(A^* w_n^*) \rangle \\
&\quad + \frac{1}{q} \tilde{\sigma}_q \left(t_n \, A^* w_n^*, \sum_{k=0}^{n-1} t_k \, A^* w_k^* \right) \\
&= \frac{|t_n|^q}{q} \, \|A^* w_n^*\|^q + \frac{1}{q} \tilde{\sigma}_q \left(t_n \, A^* w_n^*, \sum_{k=0}^{n-1} t_k \, A^* w_k^* \right) .
\end{aligned}$$

Inductively we get

$$h_n(t) \leq \frac{1}{q} \|x_n\|^p - t_n R_n^r + \phi\big(\|x_n\|; |t_n| \cdot \|A^* w_n^*\|, \ldots, |t_0| \cdot \|A^* w_0^*\| \big)$$

with a function $\phi \geq 0$ which decreases with $\big(|t_n| \cdot \|A^* w_n^*\|, \ldots, |t_0| \cdot \|A^* w_0^*\| \big)$. Hence h_n is likely to get minimal for small values of $|t_0|, \ldots, |t_{n-1}|$ and $\|A^* w_n^*\|$. For example by 1.18 in an L_p-space ($p \leq 2$, dual space L_q with $q \geq 2$) with the normalized duality mapping we would get $\big($see also around (2.63)$\big)$

$$h_n(t) \le \tilde{h}_n(t) := \frac{1}{2}\|x_n\|^2 - t_n R_n^r + \frac{1}{2}\sum_{k=0}^{n}(q-1)^{n+1-k}\|A^* w_k^*\|^2\, t_k^2\,,$$

and $\tilde{h}_n(t)$ is minimal for

$$\tilde{t}_n = \left(0,\dots,0,\frac{R_n^r}{(q-1)\|A^* w_n^*\|^2}\right)$$

with

$$\tilde{h}_n(\tilde{t}_n) = \frac{1}{2}\left(\|x_n\|^2 - \frac{R_n^{2r}}{(q-1)\|A^* w_n^*\|^2}\right)\,.$$

This also shows that $\tilde{t}_n$ would be a good starting value for minimizing h_n and $h_n(t) \le c\cdot\tilde{h}_n(\tilde{t}_n)$ with $c \in (0,1)$ a good stopping criterion for (the realistic case of) inexact minimization of h_n. We summarize this in the following method.

Method 2.34. *The same as method 2.32 but the choice of w_n^* (2.101) is replaced by*

$$w_n^* := J_Y(Ax_n - y) - \sum_{k=0}^{n-1} s_k^n\, w_k^*\,. \tag{2.108}$$

whereby $s_n = (s_0^n,\dots,s_{n-1}^n)$ is the (necessarily existing) unique solution of the n-dimensional optimization problem

$$\min_{s\in\mathbb{R}^n} g_n(s) \tag{2.109}$$

with

$$g_n(s) := \frac{1}{q}\left\|A^* J_Y(Ax_n - y) - \sum_{k=0}^{n-1} s_k\, A^* w_k^*\right\|^q \quad,\quad s = (s_0,\dots,s_{n-1}) \in \mathbb{R}^n\,.$$

Existence and uniqueness of s_n are consequences of 2.30 and the following proposition.

Proposition 2.35. *The assertions of 2.33 remain valid for method 2.34. Additionally the following holds as long as $R_n \ne 0$:*

(d) $\langle A^* w_k^* \,|\, J_X^*(A^* w_n^*)\rangle = 0$ *for all* $k = 0,\dots,n-1$.

(e)

$$A^* w_n^* = A^* J_Y(Ax_n - y) - P_{U_{n-1}^*}\big(A^* J_Y(Ax_n - y)\big)$$

and

$$J_X^*(A^* w_n^*) = \Pi^p_{(U_{n-1}^*)^\perp}\Big(J_X^*\big(A^* J_Y(Ax_n - y)\big)\Big)\,.$$

Proof. Inductively we get $U_n^* = \operatorname{span}\{A^* J_Y(Ax_0 - y),\dots,A^* J_Y(Ax_n - y)\}$ by the definition of the vectors w_k^* (2.108). Therefore the assertions about convergence remain valid, because for all $z \in M_{Ax=y}$ we have

$$\Delta_p\big(x_{n+1}(t_n), z\big) \le \Delta_p\big(x_{n+1}(\tilde{t}), z\big)$$

with $\tilde{t}$ such that $\mu_n A^* J_Y(Ax_n - y) = \sum_{k=0}^{n} \tilde{t}_k \, A^* w_k^*$. Moreover we have $R_n^r = \langle w_n^* \,|\, Ax_n - y\rangle$ by 2.33 (c) and thus the relation

$$\langle A^* w_k^* \,|\, z\rangle = \langle w_k^* \,|\, y\rangle = \langle A^* w_k^* \,|\, x_k\rangle - R_k^r = \alpha_k^n$$

for all $z \in M_{Ax=y}$ remains unchanged. Hence (b) holds. In (a) it again suffices to show that the vectors $\{A^* w_0^*, \ldots, A^* w_n^*\}$ are linearly independent. We show this inductively in combination with (c) and (d). Let $R_n \neq 0$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be given such that

$$\sum_{k=0}^{n} \lambda_k \, A^* w_k^* = 0 \,.$$

We apply $J_X^*(A^* w_n^*)$ on both sides and by (d) we get

$$0 = \sum_{k=0}^{n} \lambda_k \, \langle A^* w_k^* \,|\, J_X^*(A^* w_n^*)\rangle = \lambda_n \, \langle A^* w_n^* \,|\, J_X^*(A^* w_n^*)\rangle = \lambda_n \, \|A^* w_n^*\|^q \,.$$

Suppose $A^* w_n^* = 0$. Then we take some $z \in M_{Ax=y}$ and get

$$0 = \langle A^* w_n^* \,|\, x_n - z\rangle = \langle w_n^* \,|\, Ax_n - y\rangle = R_n^r \,,$$

which contradicts $R_n \neq 0$. Hence $\lambda_n = 0$. Inductively we get $\lambda_k = 0$ for all $k = 0, \ldots, n$ and thus the vectors $\{A^* w_0^*, \ldots, A^* w_n^*\}$ are linearly independent. (c) follows as in 2.33 and likewise (d) by

$$\begin{aligned} 0 = \partial_j g_n(s_n) &= -\left\langle A^* w_j^* \,\middle|\, J_X^*\left(A^* J_Y(Ax_n - y) - \sum_{k=0}^{n-1} s_k^n \, A^* w_k^*\right)\right\rangle \\ &= -\big\langle A^* w_j^* \,\big|\, J_X^*(A^* w_n^*)\big\rangle \end{aligned}$$

for all $j = 0, \ldots, n-1$. (e) is just the definition of w_n^*. □

The nice behaviour of the above methods is of course only of theoretical nature. Even in case of exact data we will only be able to perform the minimizations of h_n and g_n inexactly. Rounding errors are also unavoidable. As a result the orthogonality relations in (c) and (d) will not hold exactly and errors made in previous iterations will be carried on. Moreover the number of vectors to be stored and the dimension of the otimization sub-problems increases with n. Therefore in practice we should keep only a few of the old directions, say the last M ones $A^* w_{n-1}^*, \ldots, A^* w_{n-M}^*$.

Method 2.36. *The same as the methods 2.32 or 2.34 but with the respective modifications (for some $M \in \mathbb{N}$)*

$$w_n^* := J_Y(Ax_n - y) - \sum_{k=0\vee(n-M)}^{n-1} s_k^n\, w_k^*\,,$$

$$g_n(s) := \frac{1}{q}\left\| A^* J_Y(Ax_n - y) - \sum_{k=0\vee(n-M)}^{n-1} s_k\, A^* w_k^* \right\|^q\,,$$

$$x_{n+1} := J_X^*\left(J_X(x_n) - \sum_{k=0\vee(n-M)}^{n} t_k^n\, A^* w_k^* \right)\,,$$

$$h_n(t) := \frac{1}{q}\left\| J_X(x_n) - \sum_{k=0\vee(n-M)}^{n} t_k\, A^* w_k^* \right\|^q + \sum_{k=0\vee(n-M)}^{n} t_k\, \alpha_k^n\,.$$

Corollary 2.37. *The assertions of 2.33 and 2.35 remain valid for method 2.36 with the modifications*

(a) The vectors $\{w_{0\vee(n-M)}^*, \ldots, w_n^*\}$ *and* $\{A^* w_{0\vee(n-M)}^*, \ldots, A^* w_n^*\}$ *are linearly independent.*

(b) For

$$H_n := \bigcap_{k=0\vee(n-M)}^{n} H(A^* w_k^*, \alpha_k^n)$$

and

$$U_n^* := \operatorname{span}\{A^* w_{0\vee(n-M)}^*, \ldots, A^* w_n^*\}$$

we have

$$\operatorname{span}\{A^* J_Y(Ax_{0\vee(n-M)} - y), \ldots, A^* J_Y(Ax_n - y)\} \subset U_n^*$$

$$M_{Ax=y} \subset H_n \quad , \quad x_{n+1} = \Pi_{H_n}^p(x_n)$$

and

$$J_X(x_{n+1}) = \Pi_{J_X(x_n)+U_n^*}^q\big(J_X(z)\big) \quad \textit{for all} \quad z \in M_{Ax=y}\,.$$

(c) $\langle w_k^* \,|\, Ax_{n+1} - y\rangle = 0$ *for all* $k = 0 \vee (n-M), \ldots, n$.

(d) $\langle A^* w_k^* \,|\, J_X^*(A^* w_n^*)\rangle = 0$ *for all* $k = 0 \vee (n-M), \ldots, n-1$.

(e) For

$$\tilde{U}_{n-1}^* := \operatorname{span}\{A^* w_{0\vee(n-M)}^*, \ldots, A^* w_{n-1}^*\}$$

we have

$$A^* w_n^* = A^* J_Y(Ax_n - y) - P_{\tilde{U}_{n-1}^*}\big(A^* J_Y(Ax_n - y)\big)$$

and

$$J_X^*(A^* w_n^*) = \Pi_{(\tilde{U}_{n-1}^*)^\perp}^p\Big(J_X^*\big(A^* J_Y(Ax_n - y)\big)\Big)\,.$$

As in CG methods we examine what happens when we only use the previous search direction $u_{n-1}^* = A^* w_{n-1}^*$ to determine the new one $u_n^* = A^* w_n^*$ and minimize along that one only (instead of $\operatorname{span}\{u_n^*, u_{n-1}^*\}$ as in method 2.36).

Method 2.38. *We assume (A, y), choose an arbitrary starting point $x_0 \in X$ and for $n = 0, 1, 2, \dots$ repeat the following steps:*
We set

$$R_n := \|Ax_n - y\| .$$

If $R_n = 0$ then STOP, else we set

$$u_n^* := A^* J_Y(Ax_n - y) - s_n\, u_{n-1}^* . \tag{2.110}$$

whereby s_n is the (necessarily existing) unique solution of the one-dimensional optimization problem

$$\min_{s \in \mathbb{R}} g_n(s) \tag{2.111}$$

with

$$g_n(s) := \frac{1}{q} \left\| A^* J_Y(Ax_n - y) - s\, u_{n-1}^* \right\|^q \quad , \quad s \in \mathbb{R} .$$

We set

$$\alpha_n := \langle u_n^* \,|\, x_n \rangle - R_n^r \tag{2.112}$$

and define

$$x_{n+1} := J_X^* \left(J_X(x_n) - t_n\, u_n^* \right) , \tag{2.113}$$

whereby t_n is the (necessarily existing) unique positive solution of the one-dimensional optimization problem

$$\min_{t \in \mathbb{R}} h_n(t) \tag{2.114}$$

with

$$h_n(t) := \frac{1}{q} \left\| J_X(x_n) - t\, u_n^* \right\|^q + t\, \alpha_n \quad , \quad t \in \mathbb{R} .$$

If both X and Y are Hilbert spaces (with the identity as normalized duality mapping) then this is just the well-known CG method for finding the minimum-norm solution of operator equations.

Proposition 2.39. *Method 2.38 either stops after a finite number $n \in \mathbb{N}$ of iterations (in case $R_n = 0$) with x_n being the Bregman projection $\tilde{x}$ of x_0 onto the set $M_{Ax=y}$ or the sequence of the iterates $(x_n)_n$ converges weakly to $\tilde{x}$ (with the exceptional case as in 2.23).*

Proof. We point out that this is not an immediate corollary of 2.37, because we do not minimize h_n over $\operatorname{span}\{u_n^*, u_{n-1}^*\}$ in which the convergence ensuring descent direction $A^* J_Y(Ax_n - y)$ is contained. We show that u_n^* retains this property. As in the previous methods for $w_n^* := J_Y(Ax_n - y) - s_n\, w_{n-1}^*$ we have $u_n^* = A^* w_n^*$, $\left\langle w_{n-1}^* \,\middle|\, Ax_n - y \right\rangle = 0$, $\left\langle u_{n-1}^* \,\middle|\, J_X^*(u_n^*) \right\rangle = 0$ and therefore $\alpha_n = \langle w_n^* \,|\, y \rangle$, $R_n^r = \langle w_n^* \,|\, Ax_n - y \rangle$ and $u_n^* \neq 0$ if $R_n \neq 0$. We once again do the by now familiar estimations for $h_n(t)$ (compare 2.8, 2.11, 2.18, 2.23). We w.l.o.g. assume $x_n \neq 0$, $u_n^* \neq 0$ and set

$$\tilde{t}_n := \frac{\tau_n \|x_n\|^{p-1}}{\|u_n^*\|}$$

with $\tau_n \in (0, 1]$ chosen such

$$\frac{\rho_{X^*}\left(\tau_n\right)}{\tau_n} = \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q} \frac{R_n^r}{\|x_n\| \, \|u_n^*\|} \right) .$$

By (2.114) we get

$$\begin{aligned} h_n(t_n) &\leq h_n(\tilde{t}_n) \\ &\leq \frac{1}{q}\|x_n\|^p - \tilde{t}_n \, \langle u_n^* \,|\, x_n \rangle + \frac{1}{q}\tilde{\sigma}_q\left(J_X(x_n), \tilde{t}_n \, u_n^*\right) + \langle w_n^* \,|\, y \rangle \\ &= \frac{1}{q}\|x_n\|^p - \tilde{t}_n R_n^r + \frac{1}{q}\tilde{\sigma}_q\left(J_X(x_n), \tilde{t}_n \, u_n^*\right) \\ &= \frac{1}{q}\|x_n\|^p - \tau_n \|x_n\|^{p-1} \frac{R_n^r}{\|u_n^*\|} + \frac{1}{q}\tilde{\sigma}_q\left(J_X(x_n), \tilde{t}_n \, u_n^*\right) . \end{aligned}$$

If we look at the proof of 2.6 we see that we can estimate the last summand by

$$\begin{aligned} \frac{1}{q}\tilde{\sigma}_q\left(J_X(x_n), \tilde{t}_n \, u_n^*\right) &\leq \tau_n 2^q G_q \|x_n\|^p \frac{\rho_{X^*}(\tau_n)}{\tau_n} \\ &\leq \tau_n 2^q G_q \|x_n\|^p \frac{\gamma}{2^q G_q} \frac{R_n^r}{\|x_n\| \, \|u_n^*\|} \\ &= \gamma \tau_n \|x_n\|^{p-1} \frac{R_n^r}{\|u_n^*\|} . \end{aligned}$$

Hence

$$h_n(t_n) \leq \frac{1}{q}\|x_n\|^p - (1-\gamma)\tau_n \|x_n\|^{p-1} \frac{R_n^r}{\|u_n^*\|} .$$

By (2.114) we have

$$\|u_n^*\| \leq \|A^* J_Y(Ax_n - y)\| \leq \|A\| \, R_n^{r-1}$$

and thus

$$h_n(t_n) \leq \frac{1}{q}\|x_n\|^p - \frac{1-\gamma}{\|A\|}\tau_n \|x_n\|^{p-1} R_n .$$

Since this τ_n also fulfills

$$\frac{\rho_{X^*}\left(\tau_n\right)}{\tau_n} \geq \rho_{X^*}(1) \wedge \left(\frac{\gamma}{2^q G_q \|A\|} \frac{R_n}{\|x_n\|} \right) ,$$

i.e. τ_n remains bounded away from zero if R_n does, the assertions about convergence follow as in the proofs of 2.18, 2.23 and 2.33. □

If X is an L_p-space with $p \leq 2$ and the normalized duality mapping then we may use the line search 2.28 for (2.111) and (2.114):

$$u^*_{n,0} := A^* J_Y(Ax_n - y)\,,$$

$$u^*_{n,k+1} := u^*_{n,k} - \tilde{s}_k\, u^*_{n-1} \quad \text{with} \quad \tilde{s}_k := \frac{\Big\langle u^*_{n-1} \,\Big|\, J^*_X(u^*_{n,k}) \Big\rangle}{(q-1)\|u^*_{n-1}\|^2}\,,$$

$$u^*_n := \lim_{k\to\infty} u^*_{n,k}$$

and

$$x_{n+1,0} := x_n\,,$$

$$x_{n+1,k+1} := J^*_X\big(J_X(x_{n+1,k}) - \tilde{t}_k\, u^*_n\big) \quad \text{with} \quad \tilde{t}_k := \frac{\langle u^*_n \,|\, u_{n,k}\rangle - \alpha_n}{(q-1)\|u^*_n\|^2}\,,$$

$$x_{n+1} := \lim_{k\to\infty} x_{n+1,k}\,.$$

In an L_p-space with $p \geq 2$ and duality mapping with the same weight p we take

$$\tilde{s}_k := \frac{\Big|\Big\langle u^*_{n-1} \,\Big|\, J^*_X(u^*_{n,k}) \Big\rangle\Big|^{p-1}}{2^{p-2}\|u^*_{n-1}\|^p}\, \operatorname{sgn}\big(\big\langle u^*_{n-1} \,\big|\, J^*_X(u^*_{n,k}) \big\rangle\big)$$

and

$$\tilde{t}_k := \frac{|\langle u^*_n \,|\, u_{n,k}\rangle - \alpha_n|^{p-1}}{2^{p-2}\|u^*_n\|^p}\, \operatorname{sgn}\big(\langle u^*_n \,|\, u_{n,k}\rangle - \alpha_n\big)\,.$$

3

Conclusions and Outlook

The goals of our study was to develop iterative methods for the solution of the split feasibility problem (SFP) in Banach spaces and to analyze their regularizing properties and stability with respect to noisy and approximate data. To this end we have generalized the CQ algorithm via duality mappings, metric and Bregman projections. We have shown that the resulting methods solve the SFP and that in combination with a discrepancy principle they have good regularizing properties and we may also use approximate data. Moreover we have seen how the same iterative scheme can be used to compute metric and Bregman projections onto affine subspaces that are given via a linear operator.
Although we have mainly concentrated on theoretical aspects and proven convergence in infinite-dimensional spaces, the methods are finally intended to solve real world problems and we think that they will prove efficient with the line search, conjugate gradient and sequential subspace techniques presented in the last two sections. Our preliminary results in [46] were promising and thorough numerical tests are the topic of current research. In this context it is also important to have criteria for the quality of the solutions. A suitable criterion for the SFP is the size of the remainder terms. For the projections onto affine subspaces it would be preferable to have something like convergence rates, possibly in dependence of source conditions (see also [35, 44]).
It is also challenging to examine whether similar methods can be used to solve problems in non-smooth Banach spaces X, which e.g. arise in image processing [38]. As our results in [46] indicate, already the use of non-smooth spaces Y can have a non-smoothing effect on the solution, which may be used to handle discontinuities.
Of course to apply the methods it is necessary to know the duality mappings of the spaces involved and useful to have concrete versions of the characteristic inequalities. Therefore a list of Banach spaces with their duality mappings and characteristic inequalities would be desirable.
Finally we want to mention that we think that many of our ideas and re-

sults can be carried over to the case of more general Bregman distances and projections.

References

[1] Y. Alber and D. Butnariu. Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces. *Journal of Optimization Theory and Applications*, 92(1):33–61, 1997.

[2] Y. I. Alber. Iterative regularization in Banach spaces. *Soviet Math. (Iz. VUZ)*, 30:1–8, 1986.

[3] Y. I. Alber. Metric and generalized projection operators in Banach spaces: Properties and applications. In A. Kartsatos, editor, *Theory and Applications of Nonlinear Operators of Monotone and Accretive Type*, pages 15–50. Marcel Dekker, New York, 1996.

[4] Y. I. Alber. Generalized projections, decompositions, and the pythagorean-type theorem in Banach spaces. *Appl. Math. Lett.*, 11(6): 115–121, 1998.

[5] Y. I. Alber and A. I. Notik. On some estimates for projection operators in Banach spaces. *Comm. Appl. Nonlinear Anal.*, 2(1):47–55, 1995.

[6] Y. I. Alber, A. G. Kartsatos, and E. Litsyn. Iterative solution of unstable variational inequalities on approximately given sets. *Abstract and Applied Analysis*, 1(1):45–64, 1996.

[7] H. H. Bauschke and J. M. Borwein. Legendre functions and the method of random Bregman projections. *Journal of Convex Analysis*, 4(1):27–67, 1997.

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

[9] 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(4):615–647, 2001.

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

[11] L. M. Bregman. The relaxation method for finding common points of convex sets and its application to the solution of problems in convex program-

ming. *USSR Computational Mathematics and Mathematical Physics*, 7: 200–217, 1967.

[12] R. S. Burachik and S. Scheimberg. A proximal point method for the variational inequality problem in Banach spaces. *SIAM J. Control Optim.*, 39(5):1633–1649, 2001.

[13] D. Butnariu and A. N. Iusem. *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*. Kluwer Academic Publishers, Dordrecht, 2000.

[14] D. Butnariu and E. Resmerita. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis*, 2006:Article ID 84919, 39 pages, 2006.

[15] D. Butnariu, A. N. Iusem, and E. Resmerita. Total convexity for powers of the norm in uniformly convex Banach spaces. *Journal of Convex Analysis*, 7(2):319–334, 2000.

[16] D. Butnariu, A. N. Iusem, and C. Zălinescu. On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces. *Journal of Convex Analysis*, 10 (1):35–61, 2003.

[17] C. Byrne. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems*, 18:441–453, 2002.

[18] C. Byrne and Y. Censor. Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization. *Annals of Operations Research*, 105:77–98, 2001.

[19] B. D. Calvert. Nonlinear equations of evolution. *Pacific J. Math.*, 39(2): 293–350, 1971.

[20] Y. Censor and T. Elfving. A multiprojection algorithm using Bregman projections in a product space. *Numer. Algorithms*, 8:221–239, 1994.

[21] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld. The multiple-sets split feasibility problem and its applications for inverse problems. *Inverse Problems*, 21:2071–2084, 2005.

[22] I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Kluwer Academic Publishers, Dordrecht, 1990.

[23] M. Collins, R. E. Schapire, and Y. Singer. Logistic regression, Adaboost and Bregman distances. *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, 2000.

[24] J. Diestel. *Geometry of Banach space-selected topics*, volume 485 of *Lecture Notes in Mathematics*. Springer, New York, 1975.

[25] N. Dunford and J. T. Schwartz. *Linear Operators, I*. Wiley–Interscience, New York, 1958.

[26] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.*, 18 (1):202–226, 1993.

[27] B. Eicke. Iteration methods for convexly constrained ill-posed problems in Hilbert space. *Numer. Funct. Anal. and Optimiz.*, 13:413–429, 1992.

[28] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[29] T. Figiel. On the moduli of convexity and smoothness. *Stud. Math.*, 56: 121–155, 1976.

[30] L. G. Gubin, B. T. Polyak, and E. V. Raik. The method of projections for finding the common point of convex sets. *USSR Computational Mathematics and Mathematical Physics*, 7:1–24, 1967.

[31] K. C. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.*, 35(4):1142–1168, 1997.

[32] J. D. Lafferty, S. D. Pietra, and V. D. Pietra. Statistical learning algorithms based on Bregman distances. *Proceedings of the Canadian Workshop on Information Theory*, 1997.

[33] J. Lindenstrauß and L. Tzafriri. *Classical Banach Spaces, I.* Springer, Berlin, 1977.

[34] J. Lindenstrauß and L. Tzafriri. *Classical Banach Spaces, II.* Springer, Berlin, 1979.

[35] A. K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[36] G. Narkiss and M. Zibulevsky. Sequential subspace optimization method for large-scale unconstrained optimization. Technical report, Technion - The Israel Institute of Technology, Department of Electrical Engineering, 2005.

[37] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[38] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *Multiscale Modelling and Simulation*, 4:460–489, 2005.

[39] J. P. Penot. Continuity properties of projection operators. *Journal of Inequalities and Applications*, 5:509–521, 2003.

[40] J. P. Penot and R. Ratsimahalo. Characterizations of metric projections in Banach spaces and applications. *Abstr. Appl. Anal.*, 3(1-2):85–103, 1998.

[41] J. P. Penot and C. Zălinescu. Bounded (Hausdorff) convergence: basic facts and applications. In F. Giannessi and A. Maugeri, editors, *Variational Analysis and Applications*, pages 827–853. Proc. Erice Conference 2003, Kluwer, Dordrecht, 2005.

[42] M. M. Rao and Z. D. Ren. *Theory of Orlicz Spaces*. Pure and Applied Mathematics. Marcel Dekker, Inc., New York, 1991.

[43] E. Resmerita. On total convexity, Bregman projections and stability in Banach spaces. *Journal of Convex Analysis*, 11(1):1–16, 2004.

[44] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21:1303–1314, 2005.

[45] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.

[46] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22: 311–329, 2006.

[47] H. Schröder. *Funktionalanalysis*. Harri Deutsch, Frankfurt am Main, 2000.

[48] Z.-B. Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications*, 157:189–210, 1991.

[49] J. Zhao and Q. Yang. Several solution methods for the split feasibility problem. *Inverse Problems*, 21:1791–1799, 2005.