J. Inv. Ill-Posed Problems **15** (2007), 1–29

DOI 10.1515 / JIP.2007.

Metric and Bregman Projections onto Affine Subspaces and their Computation via Sequential Subspace Optimization Methods

F. Schöpfer, T. Schuster, and A.K. Louis

Communicated by

Abstract. In this article we investigate and prove relationships between metric and Bregman projections induced by powers of the norm of a Banach space. We consider Bregman projections onto affine subspaces of Banach spaces and deduce some interesting analogies to results which are well known for Hilbert spaces. Using these concepts as well as ideas from sequential subspace optimization techniques we construct efficient iterative methods to compute Bregman projections onto affine subspaces that are connected to linear, bounded operators between Banach spaces. Especially these methods can be used to compute minimum-norm solutions of linear operator equations or best approximations in the range of a linear operator. Numerical experiments illuminate the performance of our iterative algorithms and demonstrate a significant acceleration compared to the Landweber method.

Key words. metric projection, Bregman projection, duality mapping, affine subspaces, sequential subspace optimization methods.

AMS classification. 46B20,65K10,65R30.

1. Introduction

Projections onto affine subspaces are an important ingredient to solve constrained optimization methods or to develop efficient iterative solvers for linear operator equations. The classical method of conjugate gradients, where one computes projections onto an affine Krylov subspace, can be seen as a typical example for such an iterative solution scheme. Consider a continuous linear operator $A : X \to Y$ between Banach spaces X and Y. We are interested in computing projections onto affine subspaces of the form

$$z + \mathcal{N}(A) \quad \text{and} \quad z + \overline{\mathcal{R}(A)}\,,$$

where $\mathcal{N}(A)$ is the nullspace and $\overline{\mathcal{R}(A)}$ is the closure of the range of A. These projections are essential if one wants to iteratively approximate the minimum-norm solution $x^\dagger$ of

$$Ax = y\,.$$

In [20] we suggested to approximate $x^\dagger$ by means of a nonlinear Landweber method

$$x_{n+1} = J_q^*\big(J_p(x_n) - \mu_n A^* J_r(Ax_n - y)\big) \quad n = 0, 1, \dots\,, \qquad x_0 = 0\,.$$

Here J_p, J_q^*, J_r are duality mappings of the corresponding Banach spaces. The step size μ_n must be properly chosen in order to achieve convergence. We again refer to [20]

for an exhaustive convergence and stability analysis of the method. As could have been expected from results in Hilbert spaces the method showed good regularization properties but the convergence is tremendously slow. Hence the idea came up to use more search directions w_n^* rather than only using the single direction $w_n^* = A^* J_r(Ax_n - y)$ in order to get a faster iterative scheme. Following ideas of the well-known CG-method we compute a projection onto an affine subspace U_n^* of $\mathcal{R}(A^*)$ in each iteration step. An appropriate choice of U_n^* accelerates the convergence significantly.
As projections we use metric and Bregman projections which are tightly connected by the important relation

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

Here C is a closed, convex subset of X, P_C means the metric projection onto C and Π^p_{C-x} the Bregman projection onto $C - x$. Thus Bregman projections can be used to compute metric projections.
The aim of this article is twofold. On the one hand we illuminate the connections between Bregman and metric projections. We show known results and deduce new relations. Moreover we demonstrate how metric and Bregman projections can be computed numerically. On the other hand we expand ideas from sequential subspace optimization techniques to accelerate the convergence of our iterative computation of $x^\dagger$.
Sequential subspace optimization methods (SESOP) were considered by NARKISS, ZIBULEVSKY [16] and ELAD ET. AL. [11] to solve large-scale unconstrained optimization problems in $\mathbb{R}^n$. Details about affine subspaces applied in conjugate gradient methods are contained in the article [22] of STOER, YUAN. A concise overview about numerical optimization is given in the book [18] of NOCEDAL, WRIGHT. Compared to the optimization techniques outlined in these references our subspace methods extend to arbitrary, but smooth Banach spaces which do not need to have finite dimension. There is a prospective need for efficient solvers of operator equations in Banach spaces, since a Banach space setting sometimes allows a more realistic modelling of problems arising in applications from industry and natural sciences. Hence, the contents of the paper are interesting not only from a theoretical point of view but also to tackle real world problems. Moreover we point out that it is important to consider problems even in infinite dimensional spaces, since a discretization always veils the nature of an inverse problem. By doing so we follow arguments which also DEUFLHARD [9] used to construct solvers for nonlinear problems.
We give a brief summary of the paper's subjects. In sections 2 and 3 we give a short survey of duality mappings, metric projections and Bregman projections induced by powers of the norm. Such Bregman projections are also called *generalized projections*, see ALBER [2]. We characterize Bregman projections onto closed affine subspaces and prove new relations between Bregman and metric projections (propositions 3.6, 3.7, 3.8, and 3.10). Our investigations lead to an extension of the recently established decomposition theorems in ALBER [3], SONG, CAO [21] to the affine case. More precisely, we show

$$X = U \oplus J_q^*(z^* + U^\perp)\,,$$

where $U \subset X$ is a closed subspace of a reflexive, smooth and strictly convex Banach space X. The second part of the article starts with section 4 and consists of the development of sequential subspace optimization methods in Banach spaces followed by some convergence results (propositions 4.1 and 4.2). The performance and advantages of the method are finally demonstrated in section 5.

2. Duality Mappings

We recall the definition of duality mappings and some of their properties, all of which can be found in the book of CIORANESCU [8]. Throughout the paper X is a real Banach space with norm $\|.\|$ and dual X^* and we write $\langle x^*|x\rangle = x^*(x)$ for the application of $x^* \in X^*$ on $x \in X$. Moreover we always assume $p, q \in (1, \infty)$ to be conjugate such that $\frac{1}{p} + \frac{1}{q} = 1$.

Definition 2.1. The mapping $J_p : X \longrightarrow 2^{X^*}$ defined by

$$J_p(x) = \left\{ x^* \in X^* \,\middle|\, \langle x^*|x\rangle = \|x\|^p \,,\ \|x^*\| = \|x\|^{p-1} \right\} \tag{2.1}$$

is the *duality mapping* of X with gauge function $t \mapsto t^{p-1}$.

J_2 is also called the *normalized* duality mapping. By J_p^* we denote a duality mapping of the dual X^*. In general duality mappings are set-valued and by the Hahn-Banach theorem the sets $J_p(x)$ are not empty for all $x \in X$. By checking (2.1) we see that the following mappings are indeed duality mappings.

Example 2.2.

(a) In a Hilbert space the normalized duality mapping is just the identity mapping.

(b) For $p \in (1, \infty)$ we have

$$J_p(x)(t) = |x(t)|^{p-1} \operatorname{sign}\big(x(t)\big)$$

in L_p-function spaces and

$$\big(J_p(x)\big)_n = |x_n|^{p-1} \operatorname{sign}(x_n)$$

in l_p-sequence spaces, where $\operatorname{sign}(x) := \frac{x}{|x|}$ for $0 \neq x \in \mathbb{R}$ and $\operatorname{sign}(0) := 0$.

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

$$\big(0, \dots, 0, x_k, 0, \dots, 0\big) \in J_2(x)\,,$$

where k is any index such that $|x_k| = \|x\|_\infty$.

(d) In $(\mathbb{R}^n, \|.\|_1)$ we may take

$$\big(\|x\|_1 \operatorname{sign}(x_1), \dots, \|x\|_1 \operatorname{sign}(x_N)\big) \in J_2(x)\,.$$

The duality mapping J_p is *homogenous* of degree $p-1$, i.e.

$$J_p(\lambda x) = |\lambda|^{p-1}\,\mathrm{sign}(\lambda) J_p(x) \quad \text{for all} \quad x \in X\,,\ \lambda \in \mathbb{R}\,, \tag{2.2}$$

and duality mappings with different gauges $p, r \in (1, \infty)$ differ only by a (non-constant) factor

$$J_r(x) = \|x\|^{r-p} J_p(x)\,. \tag{2.3}$$

In fact duality mappings are subdifferentials of convex functions. 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

$$f(y) - f(x) \geq \langle x^* \,|\, y - x\rangle \quad \text{for all} \quad y \in X\,. \tag{2.4}$$

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. Now let $f_p : X \longrightarrow \mathbb{R}$ be the function

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

Then by the theorem of Asplund, see e.g. [8], we have

$$J_p = \partial f_p\,.$$

As a consequence every duality mapping J_p is *monotone*, i.e.

$$\langle x^* - y^* \,|\, x - y\rangle \geq 0 \quad \text{for all} \quad x, y \in X\,,\ x^* \in J_p(x)\,,\ y^* \in J_p(y)\,.$$

In the following proposition smoothness and convexity of a Banach space X are characterized by properties of the functions f_p and $J_p = \partial f_p$.

Proposition 2.3.

(a) *X is strictly convex iff f_p is strictly convex iff J_p is* strictly monotone, *i.e.*

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

(b) *X is smooth iff f_p is Gâteaux differentiable iff J_p is single-valued. In this case we have $\partial f_p(x) = f_p'(x) = J_p(x)$.*

(c) *X is uniformly convex iff f_p is uniformly convex.*

(d) *X is uniformly smooth iff f_p is uniformly Fréchet differentiable on the unit sphere iff J_p is single-valued and uniformly continuous on bounded sets.*

(e) *X is reflexive, strictly convex and smooth iff J_p is bijective. And in this case we have $(J_p)^{-1} = J_q^*$.*

(f) *If X is reflexive and smooth then J_p is norm-to-weak-continuous, i.e. sequences converging in norm are mapped to weakly convergent sequences.*

Uniform smoothness implies reflexivity and smoothness, uniform convexity implies reflexivity and strict convexity, and in finite dimensions the converse holds as well. Moreover smoothness and convexity are dual concepts, i.e. a Banach space X is uniformly smooth (uniformly convex) iff its dual X^* is uniformly convex (uniformly smooth) and in case X is reflexive we also have X is smooth (strictly convex) iff X^* is strictly convex (smooth).
It is known that L_p-, l_p-spaces with $p \in (1, \infty)$ are uniformly smooth and uniformly convex whereas L_1, l_1 and L_∞, l_∞ are neither smooth nor strictly convex.
We will show the convergence of the sequential subspace methods in spaces with a q-smooth dual. X is said to be *q-smooth* if there exists a constant $C > 0$ such that

$$\rho_X(\tau) \leq C\,\tau^q \quad \text{for all} \quad \tau \in [0, \infty)\,,$$

where the function $\rho_X : [0, \infty) \longrightarrow [0, \infty)$ is the *modulus of smoothness* of X, defined by

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

A Banach space X is said to be uniformly smooth iff

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

Hence q-smooth spaces are especially uniformly smooth (recall that $q \in (1, \infty)$). It is well known that L_p-, l_p-spaces with $1 < p \leq 2$ are p-smooth with

$$\rho_X(\tau) \leq \frac{1}{p}\,\tau^p$$

and L_p-, l_p-spaces with $p \geq 2$ are 2-smooth with

$$\rho_X(\tau) \leq \frac{p-1}{2}\,\tau^2\,.$$

For more information about geometry of Banach spaces we refer to CIORANESCU [8], DIESTEL [10] , FIGIEL [12], LINDENSTRAUSS and TZAFRIRI [14]. The following inequality can be found in XU and ROACH [23]. It plays a central role in our convergence proofs.

Proposition 2.4 ([23]). *Let X be q-smooth. Then there exists a constant $C > 0$ such that for all $x, y \in X$*

$$\frac{1}{q}\|x - y\|^q \leq \frac{1}{q}\|x\|^q - \langle J_q(x) \,|\, y \rangle + \frac{C}{q}\|y\|^q\,. \tag{2.5}$$

3. Metric and Bregman Projections

We are concerned with two different kind of projections: Metric projections and Bregman projections. The latter ones arise by minimizing a Bregman distance induced by powers of the norm, which are also called *generalized projections* by ALBER [2]. We recall some known facts and extend the existing theory by some further contributions. Throughout this section X is supposed to be reflexive, smooth and strictly convex and $C \neq \emptyset$ be a closed convex subset of X. Recall that if X is reflexive, smooth and strictly convex, then this is valid for the dual X^*, too.

Definition 3.1. The *metric projection* of $x \in X$ onto C is the unique element $P_C(x) \in C$ such that

$$\|x - P_C(x)\| = \min_{y \in C} \|x - y\| \,. \tag{3.1}$$

Obviously we have $P_C(x) = x \Leftrightarrow x \in C$ and thus $P_C^2 = P_C$ and $\mathcal{R}(P_C) = C$, where by $\mathcal{R}$ we denote the range of a mapping. The metric projection can also be characterized by a variational inequality.

Proposition 3.2. *Let J_p be any duality mapping of X. Then an element $\tilde{x} \in C$ is the metric projection of x onto C iff*

$$\langle J_p(\tilde{x} - x) \,|\, y - \tilde{x} \rangle \geq 0 \quad \textit{for all} \quad y \in C \,. \tag{3.2}$$

The proof is done as in the case of the normalized duality mapping, which can be found in LIONS [15], see also PENOT, RATSIMAHALO [19] for a more general treatment of metric projections.
Bregman projections are defined as minimizers of Bregman distances which go back to BREGMAN [6].
For a Gâteaux differentiable convex function $f : X \longrightarrow \mathbb{R}$ the function

$$\Delta_f(x, y) := f(y) - f(x) - \langle f'(x) \,|\, y - x \rangle \quad , \quad x, y \in X \tag{3.3}$$

is called the *Bregman distance* of x to y with respect to the function f. Here we consider Bregman distances with respect to the functions $f_p(x) = \frac{1}{p}\|x\|^p$ with $f_p' = J_p$. In this case (3.3) can be written as

$$\Delta_p(x, y) = \frac{1}{q}\|x\|^p - \langle J_p(x) \,|\, y \rangle + \frac{1}{p}\|y\|^p \,. \tag{3.4}$$

In Hilbert spaces we get

$$\Delta_2(x, y) = \frac{1}{2}\|x - y\|^2 \,.$$

Let us write Δ_q^* for the Bregman distance in the dual X^* with respect to the function $f_q^*(x^*) = \frac{1}{q}\|x^*\|^q$. Then it is easy to see that

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

In the next proposition we collect some properties of Δ_p. We only prove (e), because it is new and we will need it in our convergence proof.

Proposition 3.3. *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) $\Delta_p(-x,-y) = \Delta_p(x,y)$ *and* $\Delta_p(\lambda x, \lambda y) = \lambda^p \Delta_p(x,y)$ *for all* $\lambda \geq 0$*.*
(c) $\lim_{\|x_n\|\to\infty} \Delta_p(x_n, x) = \infty$*, i.e. the sequence* $(x_n)_n$ *remains bounded if the sequence* $\big(\Delta_p(x_n,x)\big)_n$ *is bounded.*
(d) Δ_p *is continuous in both arguments. It is strictly convex, weakly lower semi-continuous and Gâteaux differentiable with respect to the second variable with* $\frac{\partial}{\partial y}\Delta_p(x,y) = J_p(y) - J_p(x)$*.*
(e) Let X *be uniformly convex. If* $(x_n)_n$ *converges weakly to* x *and the sequence* $\big(\Delta_p(y,x_n)\big)_n$ *converges to* $\Delta_p(y,x)$ *then* $(x_n)_n$ *converges strongly to* x*.*

Proof. [of (e)] The left hand side of

$$\Delta_p(y,x_n) - \Delta_p(y,x) + \langle J_p(y) \,|\, x_n - x\rangle = \frac{1}{p}\|x_n\|^p - \frac{1}{p}\|x\|^p$$

converges to zero and therefore so does the right hand side. Hence the sequence of the norms $\big(\|x_n\|\big)_n$ converges to $\|x\|$. Together with the weak convergence of $(x_n)_n$ to x this implies the strong convergence of $(x_n)_n$ to x in a uniformly convex space, see CIORANESCU [8].
Concerning the proofs of (a), (c) and (d) we refer to ALBER [2] and SCHÖPFER et. al. [20]. Part (b) is obvious.

Definition 3.4. The *Bregman projection* of $x \in X$ onto C with respect to the function $f_p(x) = \frac{1}{p}\|x\|^p$ is the unique element $\Pi_C^p(x) \in C$ such that

$$\Delta_p\big(x, \Pi_C^p(x)\big) = \min_{y\in C} \Delta_p(x,y)\,. \tag{3.5}$$

We also write Π^q for the Bregman projection in the dual X^* with respect to f_q^*. Obviously we have $\Pi_C^p(x) = x \Leftrightarrow x \in C$ and thus $(\Pi_C^p)^2 = \Pi_C^p$ and $\mathcal{R}(\Pi_C^p) = C$. Similar to the metric projection Bregman projections can be characterized by a variational inequality, too. Moreover they have an important descent property with respect to the Bregman distance. The proof of existence and uniqueness of the Bregman projection as well as the proof of the next proposition are contained in ALBER, BUTNARIU [1]. There and e.g. in BAUSCHKE ET. AL. [4] as well as in BUTNARIU, RESMERITA [7] the reader can gain more insight into Bregman distances and projections with respect to more general functions than powers of the norm of a Banach space.

Proposition 3.5. *An element* $\tilde{x} \in C$ *is the Bregman projection of* x *onto* C *with respect to the function* f_p *iff*

$$\langle J_p(\tilde{x}) - J_p(x) \,|\, y - \tilde{x}\rangle \geq 0 \quad \textit{for all} \quad y \in C\,. \tag{3.6}$$

Moreover this variational inequality is equivalent to

$$\Delta_p(\tilde{x},y) \leq \Delta_p(x,y) - \Delta_p(x,\tilde{x}) \quad \textit{for all} \quad y \in C\,. \tag{3.7}$$

In Hilbert spaces the Bregman projection with respect to the function f_2 coincides with the metric projection. But in general they differ from each other, as an example given in [1] demonstrates. In the same paper the authors asked whether in general Banach spaces there is a relationship between metric and Bregman projections onto closed convex sets. The following proposition gives answers.

Proposition 3.6.

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

$$P_C(x) - x = \Pi^p_{C-x}(0) \quad \textit{for all} \quad x \in X\,. \tag{3.8}$$

Especially we have $P_C(0) = \Pi^p_C(0)$.

(b) The metric projection has the translation property

$$P_{y+C}(x) = y + P_C(x-y) \quad \textit{for all} \quad x, y \in X\,. \tag{3.9}$$

This property indeed distinguishes the metric from the Bregman projection since we have

$$\Pi^p_{y+C}(x) = y + \Pi^p_C(x-y) \quad \textit{for all} \quad x, y \in X$$

implying

$$\Pi^p_{y+C}(x) = P_{y+C}(x) \quad \textit{for all} \quad x, y \in X\,.$$

(c) We have $\|\Pi^p_C(x)\| \le \|x\|$ *for all* $x \in X$ *iff* $0 \in C$.

(d) The projections satisfy

$$\Pi^p_{\lambda\,C}(\lambda\,x) = \lambda\,\Pi^p_C(x) \quad \textit{and} \quad P_{\lambda\,C}(\lambda\,x) = \lambda\,P_C(x) \quad \textit{for every} \quad \lambda \in \mathbb{R}\,.$$

Especially if C *is a cone, then* $\lambda C = C$ *for* $\lambda > 0$ *and thus the projections onto a cone are positively homogenous of degree* 1. *Projections onto a symmetric cone, i.e.* $-C = C$, *are homogenous of degree* 1.

(e) Suppose we know $\Pi^p_C(x)$ *and set*

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

Then we obtain the Bregman projection of x *onto the set* $\lambda_x C$ *with respect to the function* f_r $(r > 1)$ *via*

$$\Pi^r_{\lambda_x C}(x) = \lambda_x \Pi^p_C(x)\,. \tag{3.10}$$

Moreover if $x \in \lambda_x C$ *then* $x \in C$. *If* C *is a cone, then*

$$\Pi^r_C(x) = \lambda_x \Pi^p_C(x)\,. \tag{3.11}$$

Proof. To see (a) we compare the variational inequalities (3.2) and (3.6) for $\tilde{x} \in C$ and $\tilde{z} := \tilde{x} - x \in \tilde{C} := C - x$ and obtain the equivalences:

$$\begin{aligned} & \langle J_p(\tilde{x} - x) \,|\, y - \tilde{x} \rangle \geq 0 \quad \text{for all} \quad y \in C \\ \Leftrightarrow \quad & \langle J_p(\tilde{x} - x) \,|\, (y - x) - (\tilde{x} - x) \rangle \geq 0 \quad \text{for all} \quad y \in C \\ \Leftrightarrow \quad & \langle J_p(\tilde{z}) \,|\, \tilde{y} - \tilde{z} \rangle \geq 0 \quad \text{for all} \quad \tilde{y} \in \tilde{C}\,. \end{aligned}$$

We show (b) by using (3.8):

$$P_{y+C}(x) = x + \Pi^p_{(y+C)-x}(0) = y + (x - y) + \Pi^p_{C-(x-y)}(0) = y + P_C(x - y)\,.$$

Now let the Bregman projection fullfill $\Pi^p_{y+C}(x) = y + \Pi^p_C(x - y)$ for all $x, y \in X$. Then we get

$$P_{y+C}(x) = x + \Pi^p_{-x+(y+C)}(0) = x + \Big(- x + \Pi^p_{y+C}\big(0 - (-x)\big)\Big) = \Pi^p_{y+C}(x)\,.$$

(c) If $0 \in C$ then by taking $y = 0$ in (3.6) we get

$$0 \geq \big\langle J_p\big(\Pi^p_C(x)\big) \,\big|\, \Pi^p_C(x) \big\rangle - \langle J_p(x) \,|\, \Pi^p_C(x) \rangle = \|\Pi^p_C(x)\|^p - \langle J_p(x) \,|\, \Pi^p_C(x) \rangle$$

and therefore

$$\|\Pi^p_C(x)\|^p \leq \langle J_p(x) \,|\, \Pi^p_C(x) \rangle \leq \|x\|^{p-1} \, \|\Pi^p_C(x)\| \ ,$$

which yields $\|\Pi^p_C(x)\| \leq \|x\|$. Conversely if this inequality is valid for all $x \in X$ then for $x = 0$ we have $\|\Pi^p_C(0)\| \leq \|0\| = 0$ and thus $0 = \Pi^p_C(0) \in C$.
The homogeneity (d) is a consequence of proposition 3.3 (b) and (3.8), because

$$\begin{aligned} & \Delta_p\big(\lambda x, \lambda\, \Pi^p_C(x)\big) && \leq && \Delta_p(\lambda\, x, \lambda\, y) \quad \text{for all} \quad x \in X, \lambda\, y \in \lambda\, C \\ \Leftrightarrow \quad & \Delta_p\big(x, \Pi^p_C(x)\big) && \leq && \Delta_p(x, y) \quad \text{for all} \quad x \in X, y \in C \end{aligned}$$

and thus also

$$P_{\lambda\, C}(\lambda\, x) = \lambda\, x + \Pi^p_{\lambda\, C - \lambda\, x}(0) = \lambda\, x + \lambda\, \Pi^p_{C-x}(0) = \lambda\, P_C(x)\,.$$

(e) Due to the homogeneity of the duality mapping (2.2) and relation (2.3) we see that for $x \neq 0$ and $\Pi^p_C(x) \neq 0$ and all $y \in C$ we have

$$\begin{aligned} & \big\langle J_r\big(\lambda_x \Pi^p_C(x)\big) - J_r(x) \,\big|\, \lambda_x y - \lambda_x \Pi^p_C(x) \big\rangle \\ = \quad & \lambda_x \left\langle \lambda_x^{r-1} \|\Pi^p_C(x)\|^{r-p} J_p\big(\Pi^p_C(x)\big) - \|x\|^{r-p} J_p(x) \,\big|\, y - \Pi^p_C(x) \right\rangle \\ = \quad & \lambda_x \|x\|^{r-p} \left\langle J_p\big(\Pi^p_C(x)\big) - J_p(x) \,\big|\, y - \Pi^p_C(x) \right\rangle \geq 0\,. \end{aligned}$$

Moreover $\Pi^r_C(0) = P_C(0) = \Pi^p_C(0)$ by (a) of this proposition and if $\Pi^p_C(x) = 0$ and $x \neq 0$ then for all $y \in C$

$$\langle J_r(0) - J_r(x) \,|\, y - 0 \rangle = \|x\|^{r-p} \, \langle J_p(0) - J_p(x) \,|\, y - 0 \rangle \geq 0\,.$$

This proves the first part of (g).
Now let x be in $\lambda_x C$. Then $x = \Pi^r_{\lambda_x C}(x) = \lambda_x \Pi^p_C(x)$. If $x = 0$ or $\Pi^p_C(x) = 0$ then $\lambda_x = 1$ and therefore $x = \Pi^p_C(x) \in C$. Otherwise we get

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

which gives $\|x\| = \|\Pi^p_C(x)\|$. Hence $\lambda_x = 1$ and $x = \Pi^p_C(x) \in C$.

We characterize Bregman projections onto closed affine cones and closed affine subspaces. See also ALBER [3] and SONG and CAO [21] for the non-affine case. In light of prop. 3.6 (b) considering the affine case is meaningful. And in combination with (3.8) this especially 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.
For a subspace $U \subset X$ the set $U^\perp \subset X^*$ is the *annihilator* of U

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

For a (convex) cone $K \subset X$ the set $K^+ \subset X^*$ is the *dual cone* of K

$$K^+ := \{x^* \in X^* \,|\, \langle x^* \,|\, k\rangle \geq 0 \quad \text{for every} \quad k \in K\}$$

and $K^\circ = -K^+$ is the *polar cone* of K.

Proposition 3.7. *Let $U \subset X$ be a closed subspace, $K \subset X$ be a closed cone and $x, y, z \in X$ be given.*

(a) The following assertions are equivalent to each other:

(i) $x = \Pi^p_{z+K}(y)$,
(ii) $x - z \in K$ *and* $J_p(x) - J_p(y) \in K^+$ *and* $\langle J_p(x) - J_p(y) \,|\, x - z\rangle = 0$,
(iii) $J_p(x) = \Pi^q_{J_p(y)+K^+} J_p(z)$.

(b) In case of a subspace U the equivalencies read as

(i) $x = \Pi^p_{z+U}(y)$,
(ii) $x - z \in U$ *and* $J_p(x) - J_p(y) \in U^\perp$,
(iii) $J_p(x) = \Pi^q_{J_p(y)+U^\perp} J_p(z)$.

Proof. Assertion (b) is a consequence of (a), because a subspace is especially a cone with $U^+ = U^\perp$. Let us prove (a) by the variational inequality (3.6). An element x is the Bregman projection of y onto $z + K$ iff $x - z \in K$ and

$$\begin{aligned} &\langle J_p(x) - J_p(y) \,|\, (z + k) - x\rangle \geq 0 \quad \text{for all} \quad k \in K \\ \Leftrightarrow \quad &\langle J_p(x) - J_p(y) \,|\, z - x\rangle + \langle J_p(x) - J_p(y) \,|\, k\rangle \geq 0 \quad \text{for all} \quad k \in K \end{aligned} \tag{3.12}$$

Suppose that $\langle J_p(x) - J_p(y) \,|\, k_0 \rangle < 0$ for some $k_0 \in K$. Since K is a cone we get $\lambda\, k_0 \in K$ for all $\lambda > 0$ and thus by (3.12)

$$\langle J_p(x) - J_p(y) \,|\, z - x \rangle + \lambda \,\langle J_p(x) - J_p(y) \,|\, k_0 \rangle \geq 0 \quad \text{for all} \quad \lambda > 0\,.$$

But then the left hand side converges to $-\infty$ for $\lambda \to \infty$, which leads to a contradiction. Therefore inequality (3.12) can be fulfilled for all $k \in K$ only if

$$\langle J_p(x) - J_p(y) \,|\, k \rangle \geq 0 \quad \text{for all} \quad k \in K\,,$$

i.e. if $J_p(x) - J_p(y) \in K^+$. Since $x - z \in K$ this implies $\langle J_p(x) - J_p(y) \,|\, x - z \rangle \geq 0$. But by choosing $k = 0$ in (3.12) we also get $\langle J_p(x) - J_p(y) \,|\, z - x \rangle \geq 0$ and thus $\langle J_p(x) - J_p(y) \,|\, x - z \rangle = 0$. Hence (i) $\Rightarrow$ (ii). And since (ii) implies the validity of (3.12) we also have (ii) $\Rightarrow$ (i). Finally by the equality $(K^+)^+ = K$ it follows that (ii) $\Leftrightarrow$ (iii) is just assertion (ii) $\Leftrightarrow$ (i) in the dual space.

By means of these characterizations we are able to deduce the affine version of the decomposition theorems established in ALBER [3] and SONG, CAO [21]. See also

Proposition 3.8. *Let $U \subset X$ be a closed subspace, $K \subset X$ be a closed cone and $z^* \in X^*$ be given..*

(a) X can be decomposed into $\quad X = K \oplus J_q^*(z^* + K^\circ) \quad$, i.e. every $x \in X$ can be uniquely written in the form

$$x = x_K + J_q^*(x^*_{z^*+K^\circ})$$

with $x_K \in K$, $x^*_{z^*+K^\circ} \in z^* + K^\circ$ and $\langle x^*_{z^*+K^\circ} - z^* \,|\, x_K \rangle = 0$. More precisely we have

$$x_K = x - \Pi^p_{x-K}\, J_q^*(z^*) \quad \textit{and} \quad x^*_{z^*+K^\circ} = \Pi^q_{z^*+K^\circ}\, J_p(x)\,.$$

If $z^* = 0$ then $x_K = P_K(x)$.

(b) X can be decomposed into $\quad X = U \oplus J_q^*(z^* + U^\perp) \quad$, i.e. every $x \in X$ can be uniquely written in the form

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

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

$$x_U = x - \Pi^p_{x+U}\, J_q^*(z^*) \quad \textit{and} \quad x^*_{z^*+U^\perp} = \Pi^q_{z^*+U^\perp}\, J_p(x)\,.$$

If $z^* = 0$ then $x_U = P_U(x)$.

Proof. We only prove (a) since again (b) follows from (a). We set

$$x^*_{z^*+K^\circ} := \Pi^q_{z^*+K^\circ}\, J_p(x) \in z^* + K^\circ \quad \text{and} \quad x_K := x - \Pi^p_{x-K}\, J_q^*(z^*) \in K\,.$$

By proposition 3.7 (a) and $(-K)^+ = K^\circ$ we have

$$J_q^*(x^*_{z^*+K^\circ}) = J_q^*\, \Pi^q_{z^*+K^\circ}\, J_p(x) = \Pi^p_{x-K}\, J_q^*(z^*)$$

with $\big\langle x^*_{z^*+K^\circ} - z^* \,\big|\, J^*_q(x^*_{z^*+K^\circ}) - x \big\rangle = 0$ and thus

$$x = \big(x - J^*_q(x^*_{z^*+K^\circ})\big) + J^*_q(x^*_{z^*+K^\circ}) = x_K + J^*_q(x^*_{z^*+K^\circ})$$

with $\big\langle x^*_{z^*+K^\circ} - z^* \,\big|\, x_K \big\rangle = 0$.
The decomposition is unique, because if $x = x_K + J^*_q(x^*_{z^*+K^\circ})$ with some $x_K \in K$, $x^*_{z^*+K^\circ} \in z^* + K^\circ$ and $\langle x^*_{z^*+K^\circ} - z^* \,|\, x_K\rangle = 0$ then we have

$$J^*_q(x^*_{z^*+K^\circ}) - J^*_q\big(J_p(x)\big) = -x_K \in -K = (K^\circ)^+$$

and

$$\big\langle x^*_{z^*+K^\circ} - z^* \,\big|\, J^*_q(x^*_{z^*+K^\circ}) - J^*_q\big(J_p(x)\big)\big\rangle = -\langle x^*_{z^*+K^\circ} - z^*, \,|\, x_K\rangle = 0\,.$$

Applying proposition 3.7 (a) we conclude that indeed $x^*_{z^*+K^\circ} = \Pi^q_{z^*+K^\circ}\, J_p(x)$ and

$$x_k = x - J^*_q(x^*_{z^*+K^\circ}) = x - J^*_q\, \Pi^q_{z^*+K^\circ}\, J_p(x) = x - \Pi^p_{x-K}\, J^*_q(z^*)\,.$$

We continue by showing how far results concerning orthogonal projections in Hilbert spaces can be carried over to Bregman projections onto affine subspaces. So far we do not know whether similar results can be formulated for the metric projection, too, unless proposition 3.10 (b), which has been proven for the metric projection for $z = 0$ in ALBER [3].

Lemma 3.9. *Let $U \subset X$ be a closed subspace and $z \in X$ be given. Then*

$$\mathcal{R}\big(J_p(.) - J_p\, \Pi^p_{z+U}(.)\big) = U^\perp\,.$$

Proof. By prop. 3.7 (b) we have for all $x \in X$

$$J_p(x) - J_p\, \Pi^p_{z+U}(x) = J_p(x) - \Pi^q_{J_p(x)+U^\perp}\, J_p(z) \in U^\perp\,.$$

Conversely for $y^* \in U^\perp$ we set $J_p(x) := y^* + \Pi^q_{U^\perp}\, J_p(z) \in U^\perp$ and get

$$J_p(x) - \Pi^q_{J_p(x)+U^\perp}\, J_p(z) = J_p(x) - \Pi^q_{U^\perp}\, J_p(z) = y^*\,.$$

Proposition 3.10. *Let $U, V \subset X$ be closed subspaces and $z \in X$ be given. The following assertions hold true.*

(a) *We have $\Pi^p_{z+U}(x) = \Pi^p_{z+U}(y)$ iff $J_p(x) - J_p(y) \in U^\perp$.*
(b) *The Bregman projection Π^p_{z+U} is $J_p(.) - J_p(z)$-positive, i.e.*

$$\langle \Pi^p_{z+U}(x) - z \,|\, J_p(x) - J_p(z)\rangle \geq 0 \quad \textit{for all} \quad x \in X\,.$$

(c) $\Pi^p_{z+U}\, \Pi^p_{z+V} = \Pi^p_{z+U} \quad \Leftrightarrow \quad U \subset V.$

(d) The composition $\Pi^p_{z+U}\,\Pi^p_{z+V}$ *is again a Bregman projection* Π^p_M *onto some closed affine subspace* $M \subset X$ *iff*

$$\mathcal{R}(\Pi^p_{z+U}\,\Pi^p_{z+V}) \subset z + U \cap V\,.$$

In this case we have $\Pi^p_{z+U}\,\Pi^p_{z+V} = \Pi^p_{z+U\cap V}$.

(e) Let $\Pi : X \longrightarrow X$ *be any mapping with the properties*

(i) $\Pi(x) = \Pi(y) \Leftrightarrow J_p(x) - J_p(y) \in U^\perp$,
(ii) $\Pi^2 = \Pi$,
(iii) $\Pi(.) - z$ *is* $J_p(.) - J_p(z)$-positive.

Then $\Pi = \Pi^p_{z+U}$. *Hence (i)-(iii) characterize Bregman projections onto closed affine subspaces.*

(f) The sum $\Pi^p_{z+U}(.) + \Pi^p_{z+V}(.) - z$ *is again a Bregman projection* Π^p_M *onto some closed affine subspace* $M \subset X$ *iff*

$$\Delta_p(z, z+u+v) = \Delta_p(z, z+u) + \Delta_p(z, z+v) \quad \text{for all} \quad u \in U\,,\ v \in V\,. \tag{3.13}$$

And in this case we have $\Pi^p_{z+U}(.) + \Pi^p_{z+V}(.) - z = \Pi^p_{z+U+V}(.)$.

Proof. (a) We set $\tilde{x} := \Pi^p_{z+U}(x)$. If $\Pi^p_{z+U}(y) = \tilde{x}$ then by applying part (b) of proposition 3.7 we get

$$J_p(x) - J_p(y) = \big(J_p(x) - J_p(\tilde{x})\big) + \big(J_p(\tilde{x}) - J_p(y)\big) \in U^\perp + U^\perp = U^\perp\,.$$

Conversely if $J_p(x) - J_p(y) \in U^\perp$ then

$$J_p(\tilde{x}) - J_p(y) = \big(J_p(\tilde{x}) - J_p(x)\big) + \big(J_p(x) - J_p(y)\big) \in U^\perp + U^\perp = U^\perp\,,$$

and since $\tilde{x} - z \in U$ we again by proposition 3.7 (b) conclude that $\tilde{x} = \Pi^p_{z+U}(y)$.
(b) The Bregman projection Π^p_{z+U} is indeed $J_p(.) - J_p(z)$-positive, because applying the decomposition in prop. 3.8 (b) to $J_p(x) \in X^* = U^\perp \oplus J_p(z+U)$ together with the monotonicity of the duality mapping yields

$$\langle \Pi^p_{z+U}(x) - z \,|\, J_p(x) - J_p(z)\rangle = \langle \Pi^p_{z+U}(x) - z \,|\, J_p\,\Pi^p_{z+U}(x) - J_p(z)\rangle \geq 0\,.$$

To prove (c) we consider the following

$$\begin{aligned} & \Pi^p_{z+U}\,\Pi^p_{z+V}(x) = \Pi^p_{z+U}(x) \quad \text{for every} \quad x \in X \\ \Leftrightarrow\quad & J_p(x) - J_p\,\Pi^p_{z+V}(x) \in U^\perp \quad \text{for every} \quad x \in X \\ \Leftrightarrow\quad & V^\perp \subset U^\perp\,. \end{aligned}$$

The first equivalence is satisfied due to (a) and the second one is valid because of lemma 3.9.
(d) For all $x \in X$ we have

$$J_p(x) - J_p\,\Pi^p_{z+U}\,\Pi^p_{z+V}(x) = J_p(x) - J_p\,\Pi^p_{z+V}(x) + J_p\,\Pi^p_{z+V}(x) - J_p\,\Pi^p_{z+U}\,\Pi^p_{z+V}(x)\,.$$

By lemma 3.9 the right hand side lies in $V^\perp + U^\perp \subset (U\cap V)^\perp$ and by (a) we conclude that

$$\Pi^p_{z+U\cap V}(x) = \Pi^p_{z+U\cap V}\,\Pi^p_{z+U}\,\Pi^p_{z+V}(x)\,.$$

In case $\mathcal{R}(\Pi^p_{z+U}\,\Pi^p_{z+V}) \subset z + U\cap V$ we have

$$\Pi^p_{z+U\cap V}\,\Pi^p_{z+U}\,\Pi^p_{z+V}(x) = \Pi^p_{z+U}\,\Pi^p_{z+V}(x)$$

yielding

$$\Pi^p_{z+U\cap V}(x) = \Pi^p_{z+U}\,\Pi^p_{z+V}(x)\,.$$

Now suppose that $\Pi^p_{z+U}\,\Pi^p_{z+V} = \Pi^p_M$ for some closed affine subspace $M \subset X$. Then we obviously have

$$z + U\cap V \subset M \subset z + U\,.$$

Especially we get $z \in M$ and therefore $M = z + W$ for some closed subspace $W \subset X$. Hence

$$\Pi^p_{z+W}\,\Pi^p_{z+V} = \Pi^p_M\,\Pi^p_{z+V} = \Pi^p_{z+U}\,\Pi^p_{z+V}\,\Pi^p_{z+V} = \Pi^p_{z+U}\,\Pi^p_{z+V} = \Pi^p_M = \Pi^p_{z+W}$$

and by (c) we conclude that $W \subset V$ and thus $M = z + W \subset z + V$. We finally arrive at

$$z + U\cap V \subset M \subset (z+U)\cap(z+V) = z + U\cap V\,,$$

from which we infer that $M = z + U\cap V$.
(e) We observe that (i) and (ii) imply that $J_p(x) - J_p\,\Pi(x) \in U^\perp$ for all $x \in X$. Writing

$$J_p(x) = \big(J_p(x) - J_p\,\Pi(x)\big) + J_p\,\Pi(x)$$

and keeping in mind the uniqueness of the decomposition in proposition 3.8 (b) we see that in order to show $\Pi(x) = \Pi^p_{z+U}(x)$ it suffices to show $\Pi(x) - z \in U$. Let $y^* \in U^\perp$ be arbitrary. Then by (iii), (i) and (ii) we deduce

$$\begin{aligned}
0 &\le \big\langle \Pi\, J_q^*\big(y^* + J_p\,\Pi(x)\big) - z \,\big|\, \big(y^* + J_p\,\Pi(x)\big) - J_p(z)\big\rangle\\
&= \big\langle \Pi\big(\Pi(x)\big) - z \,\big|\, y^* + J_p\,\Pi(x) - J_p(z)\big\rangle\\
&= \langle \Pi(x) - z \,|\, y^* + J_p\,\Pi(x) - J_p(z)\rangle\\
&= \langle \Pi(x) - z \,|\, y^*\rangle + \langle \Pi(x) - z \,|\, J_p\,\Pi(x) - J_p(z)\rangle\ .
\end{aligned}$$

Since $y^* \in U^\perp$ is arbitrary this implies $\langle \Pi(x) - z \,|\, y^*\rangle = 0$ for all $y^* \in U^\perp$. Hence $\Pi(x) - z \in (U^\perp)^\perp = U$.
(f) Suppose that $\Pi^p_{z+U}(.) + \Pi^p_{z+V}(.) - z = \Pi^p_M(.)$. We at first show that in this case we must have $\Pi^p_M = \Pi^p_{z+U+V}$ and to this end we check (i)-(iii) in (e). (ii) is obvious and (iii) holds because

$$\Pi^p_M(.) - z = \Pi^p_{z+U}(.) - z + \Pi^p_{z+V}(.) - z$$

is $J_p(.) - J_p(z)$-positive. To see (i) we at first show $U\cap V = \{0\}$. For any $x \in U\cap V$ we get $\Pi^p_M(z+x) = (z+x) + (z+x) - z = z + 2x$ and thus

$$z + 2x = \Pi^p_M(z+x) = \Pi^p_M\,\Pi^p_M(z+x) = \Pi^p_M(z+2x) = z + 4x\,.$$

Hence $x = 0$ and therefore $U \cap V = \{0\}$. With this and (a) we get

$$\begin{array}{lrcl} & \Pi^p_M(x) & = & \Pi^p_M(y) \\ \Leftrightarrow & U \ni \Pi^p_{z+U}(x) - \Pi^p_{z+U}(y) & = & \Pi^p_{z+V}(y) - \Pi^p_{z+V}(x) \in V \\ \Leftrightarrow & \Pi^p_{z+U}(x) - \Pi^p_{z+U}(y) = 0 & \text{and} & \Pi^p_{z+V}(y) - \Pi^p_{z+V}(x) = 0 \\ \Leftrightarrow & J_p(x) - J_p(y) \in U^\perp \cap V^\perp & = & (U+V)^\perp . \end{array}$$

By (e) we conclude that $\Pi^p_M = \Pi^p_{z+\overline{U+V}}$ and since $\mathcal{R}(\Pi^p_M) \subset z + U + V$ we finally get $\Pi^p_M = \Pi^p_{z+U+V}$.
It remains to show (3.13). Let $u \in U$, $v \in V$ be arbitrary. According to proposition 3.3 (d) the function $f(t) := \Delta_p(z, z + t\,u + v) - \Delta_p(z, z + t\,u)$, $t \in \mathbb{R}$ is differentiable with

$$f'(t) = \langle J_p(z + t\,u + v) - J_p(z + t\,u) \,|\, u \rangle \ .$$

Since we have for all $u \in U$, $v \in V$

$$\begin{array}{rcl} \Pi^p_{z+U}(z+u+v) + \Pi^p_{z+V}(z+u+v) - z & = & z+u+v \\ \Rightarrow \qquad U \ni z + u - \Pi^p_{z+U}(z+u+v) & = & \Pi^p_{z+V}(z+u+v) - z - v \in V \\ \Rightarrow \qquad z + u - \Pi^p_{z+U}(z+u+v) & = & 0 \\ \Rightarrow \qquad J_p(z+u+v) - J_p(z+u) \in U^\perp , & & \end{array}$$

we conclude that $f'(t) = 0$. Hence f is constant and we get

$$\Delta_p(z, z+u+v) - \Delta_p(z, z+u) = f(1) = f(0) = \Delta_p(z, z+v) \, .$$

Conversely suppose that (3.13) holds. Then for all $u, w \in U$, $v \in V$ and with

$$g(t) := \Delta_p(z, z + u + t\,w + v) = \Delta_p(z, z + u + t\,w) + \Delta_p(z, z+v)$$

we get

$$\langle J_p(z+u+v) - J_p(z) \,|\, w \rangle = g'(0) = \langle J_p(z+u) - J_p(z) \,|\, w \rangle$$

and consequently $J_p(z+u+v) - J_p(z+u) \in U^\perp$. That also proves $J_p(z+u+v) - J_p(z+v) \in V^\perp$ and thus $J_p(z+u) - J_p(z) \in V^\perp$. For $x \in U \cap V$ we then get

$$0 = \langle J_p(z+x) - J_p(z) \,|\, x \rangle = \langle J_p(z+x) - J_p(z) \,|\, (z+x) - z \rangle$$

and by the strict monotonicity of J_p we conclude that $z + x = z$ whence $x = 0$ and thus $U \cap V = \{0\}$.
In a similar way we check (i)-(iii) in (e) in order to show that $\Pi = \Pi^p_{z+U+V}$ for $\Pi(x) := \Pi^p_{z+U}(x) + \Pi^p_{z+V}(x) - z$.

Relation (3.13) can also be written in the form

$$\|z+u+v\|^p + \|z\|^p = \|z+u\|^p + \|z+v\|^p \quad \text{for all} \quad u \in U \, , \, v \in V \, . \qquad (3.14)$$

In Hilbert spaces this is equivalent to $U \perp V$ if $p = 2$. But in general (3.14) seems to be stronger than requiring $J_p(U) \subset V^{\perp}$ and $J_p(V) \subset U^{\perp}$. Because of the pointwise (componentwise) definition of the duality mapping In L_p- (l_p-) spaces (3.14) is satisfied if the power p is the same number as the p defining L_p (l_p) and if z and all $u \in U$, $v \in V$ are functions (vectors) with pairwise disjoint support. Combining this with proposition 3.7 (b) we obtain the following examples.

Example 3.11.

(a) For any $u \in X$ we have

$$\Pi^p_{\text{span}\{u\}}(x) = \frac{|\langle J_p(x) \mid u\rangle|^{q-1}}{\|u\|^q} \,\text{sign}(\langle J_p(x) \mid u\rangle)\, u\,.$$

(b) In $(\mathbb{R}^3, \|.\|_p)$ for $U := \text{span}\{(0,0,1)\}$ and $V := \text{span}\{(1,1,0)\}$ we get

$$\Pi^p_{U+V}\big((x_1,x_2,x_3)\big) = (0,0,x_3) + \frac{|\phi(x_1,x_2)|^{q-1}}{2^{q-1}} \,\text{sign}(\phi(x_1,x_2))\,(1,1,0)$$

with

$$\phi(x_1,x_2) := |x_1|^{p-1}\,\text{sign}(x_1) + |x_2|^{p-1}\,\text{sign}(x_2)\,.$$

In the next proposition we show how Bregman projections onto a special kind of affine subspaces, namely finite intersections of hyperplanes, can be computed via solving a finite dimensional optimization problem. The subproblems we have to solve in the sequential subspace optimization methods will be of such a form.
For $0 \neq u^* \in X^*$ and $\alpha \in \mathbb{R}$ we denote by $H(u^*, \alpha)$ the *hyperplane*

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

Proposition 3.12. *Let $H(u_1^*, \alpha_1), \ldots, H(u_N^*, \alpha_N)$ be hyperplanes in a reflexive, smooth and strictly convex Banach space X such that the intersection*

$$H := \bigcap_{k=1}^{N} H(u_k^*, \alpha_k)$$

is not empty. For $x_0 \in X$ let $h : \mathbb{R}^N \longrightarrow \mathbb{R}$ be the convex function

$$h(t) := \frac{1}{q}\left\| J_p(x_0) - \sum_{k=1}^{N} t_k\, u_k^* \right\|^q + \sum_{k=1}^{N} t_k\, \alpha_k \quad , \quad t = (t_1, \ldots, t_N) \in \mathbb{R}^N$$

with continuous partial derivatives

$$\partial_j h(t) = -\left\langle u_j^* \,\middle|\, J_q^*\left(J_p(x_0) - \sum_{k=1}^{N} t_k\, u_k^* \right) \right\rangle + \alpha_j \quad , \quad j = 1, \ldots, N\,.$$

Then the Bregman projection of x_0 onto H is given by

$$\Pi_H^p(x_0) = J_q^* \left(J_p(x_0) - \sum_{k=1}^{N} \tilde{t}_k \, u_k^* \right) ,$$

where $\tilde{t} = (\tilde{t}_1, \ldots, \tilde{t}_N)$ is a solution of the N-dimensional optimization problem

$$\min_{t \in \mathbb{R}^N} h(t) . \tag{3.15}$$

Moreover if the vectors $u_1^*, \ldots, u_N^*$ are linearly independent then h is strictly convex and $\tilde{t}$ is unique.

Proof. Convexity of h is obvious, differentiability and continuity of the partial derivatives are consequences of parts (b), (e) and (f) of proposition 2.3. For any $z \in H$ we can write

$$H = z + \left(\text{span}\{u_1^*, \ldots, u_N^*\} \right)^{\perp} .$$

Thus in view of proposition 3.7 (b) an element $\tilde{x} \in X$ is the Bregman projection of x_0 onto H iff $\tilde{x} \in H$ and $J_p(\tilde{x}) \in J_p(x_0) + \text{span}\{u_1^*, \ldots, u_N^*\}$, i.e.

$$J_p(\tilde{x}) = J_p(x_0) - \sum_{k=1}^{N} \tilde{t}_k \, u_k^*$$

with some $\tilde{t}_1, \ldots, \tilde{t}_N \in \mathbb{R}$ such that $\langle u_k^* \,|\, \tilde{x} \rangle = \alpha_k$ for all $k = 1, \ldots, N$. The coefficients $\tilde{t}_k$ are uniquely determined in case the vectors $u_1^*, \ldots, u_N^*$ are linearly independent. This is equivalent to

$$J_p(\tilde{x}) = \Pi^q_{J_p(x_0) + \text{span}\{u_1^*, \ldots, u_N^*\}} \, J_p(z) ,$$

i.e. $\tilde{t} = (\tilde{t}_1, \ldots, \tilde{t}_N) \in \mathbb{R}^N$ is a solution of the optimization problem

$$\begin{aligned} & \min_{t \in \mathbb{R}^N} \Delta_q^* \left(J_p(z), J_p(x_0) - \sum_{k=1}^{N} t_k \, u_k^* \right) \\ = \quad & \min_{t \in \mathbb{R}^N} \frac{1}{p} \|z\|^p - \langle z \,|\, J_p(x_0) \rangle + \sum_{k=1}^{N} t_k \, \langle z \,|\, u_k^* \rangle + \frac{1}{q} \left\| J_p(x_0) - \sum_{k=1}^{N} t_k \, u_k^* \right\|^q , \end{aligned}$$

which in turn is equivalent to (3.15) since $\langle z \,|\, u_k^* \rangle = \alpha_k$ for $z \in H$.

Due to the nice properties of the function h, which is strictly convex and continuously differentiable with known partial derivatives, the optimization problem (3.15) can be efficiently solved by standard optimization routines like Newton's method, nonlinear conjugate gradient or variable metric methods, see e.g. JARRE, STOER [13] or NOCEDAL [17] for an overview.

4. Sequential Subspace Optimization Methods

Let $A : X \to Y$ be a continuous linear operator between Banach spaces X, Y and $A^* : Y^* \to X^*$ be its adjoint. We are interested in computing projections onto affine subspaces of the form

$$z + \mathcal{N}(A) \quad \text{and} \quad z + \overline{\mathcal{R}(A)}\,,$$

where $\mathcal{N}(A)$ is the nullspace and $\overline{\mathcal{R}(A)}$ is the closure of the range of A. At first we observe that it suffices to know a procedure to compute Bregman projections onto sets of the form $z + \mathcal{N}(A)$, because in light of (3.8) and proposition 3.7 (b) we have

$$P_{z+\mathcal{N}(A)}(x) = x + \Pi^p_{z-x+\mathcal{N}(A)}(0)\,,$$

$$\Pi^p_{z+\overline{\mathcal{R}(A)}}(y) = J_q^* \, \Pi^q_{J(y)+\mathcal{N}(A^*)} \, J(z)\,,$$

$$P_{z+\overline{\mathcal{R}(A)}}(y) = y + \Pi^p_{z-y+\overline{\mathcal{R}(A)}}(0) = y + J_q^* \, \Pi^q_{\mathcal{N}(A^*)} \, J(z-y)\,.$$

Furthermore if $z \in X$ is any solution of the operator equation $Ax = y$ for some given $y \in \mathcal{R}(A)$ then we can also write

$$z + \mathcal{N}(A) = M_{Ax=y} := \{x \in X \mid Ax = y\}\,.$$

Hence solving the constraint optimization problem

$$\min f(x) \quad \text{s.t.} \quad Ax = y$$

with $f(x) = \|x_0 - x\|$ ($f(x) = \Delta_p(x_0, x)$) is equivalent to computing the metric projection (Bregman projection) of $x_0 \in X$ onto the set $M_{Ax=y}$. The element $P_{M_{Ax=y}}(0)$ is also called the *minimum norm solution* of the operator equation $Ax = y$.

In SCHÖPFER et. al. [20] we have already analyzed a generalization of the well-known Landweber method for the computation of minimum-norm solutions of linear operator equations in Banach spaces. The iteration method reads as

$$x_{n+1} = J_q^*\big(J_p(x_n) - t_n A^* J_r(Ax_n - y)\big) \quad n = 0, 1, \ldots \quad x_0 = 0 \tag{4.1}$$

with appropriately chosen parameters t_n. We have shown the strong convergence of the method for smooth and uniformly convex X and arbitrary Banach spaces Y as well as its regularizing properties in case of noisy data y_δ and disturbed A_η by applying a discrepancy principle. The method turned out to have good regularizing properties but the convergence is rather slow. Interpreting $A^* J_r(Ax_n - y)$ as a search direction and adopting ideas from sequential subspace optimization methods, see NARKISS, ZIBULEVSKY [16], NOCEDAL, WRIGHT [18], STOER, YUAN [22]), we propose a modification of this method to accelerate convergence. We shortly motivate our approach.

One step towards proving the convergence of the above method was to show that the parameters t_n can always be chosen in such a way that the sequence of the Bregman distances of the iterates to potential solution points $z \in M_{Ax=y}$ is decreasing sufficiently, i.e.

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

with some $S_n > 0$. After a short calculation this can be seen to be equivalent to

$$\frac{1}{q}\,\|J_p(x_n) - t_n\,A^*J_r(Ax_n - y)\|^q + t_n\,\langle J_r(Ax_n - y)\,|\,y\rangle \le \frac{1}{q}\,\|x_n\|^p - S_n\,.$$

The important thing is that this relation is independent of the in general unknown points z. Hence as t_n we might as well take the minimizer of the functional

$$h_n(t) := \frac{1}{q}\,\|J_p(x_n) - t\,A^*w_n^*\|^q + t\,\langle w_n^*\,|\,y\rangle \quad , \quad t \in \mathbb{R}$$

with $w_n^* := J_r(Ax_n - y)$. In view of proposition 3.12 this just means that we obtain the next iterate x_{n+1} by computing the Bregman projection of x_n onto the hyperplane $H_n := H\big(A^*w_n^*, \langle w_n^*\,|\,y\rangle\big)$, which contains the set of potential solutions $M_{Ax=y}$ because $\langle A^*w_n^*\,|\,z\rangle = \langle w_n^*\,|\,y\rangle$ for $z \in M_{Ax=y}$. Rather than using a single search direction $A^*w_n^*$ in each iteration a finite-dimensional search space

$$U_n^* = \mathrm{span}\{A^*w_{n,1}^*, \ldots, A^*w_{n,N_n}^*\} \subset \mathcal{R}(A^*)$$

is used so that we minimize

$$h_n(t_1, \ldots, t_{N_n}) := \frac{1}{q}\left\|J_p(x_n) - \sum_{k=1}^{N_n} t_k\,A^*w_k^*\right\|^q + \sum_{k=1}^{N_n} t_k\,\langle w_k^*\,|\,y\rangle\,.$$

to get a vector of step sizes $(\mu_{n,1}, \ldots, \mu_{n,N_n})$. That means that we project x_n onto

$$H_n := \bigcap_{k=1}^{N_n} H\big(A^*w_k^*, \langle w_k^*\,|\,y\rangle\big) \quad \supset \quad M_{Ax=y}$$

by iterating

$$x_{n+1} = J_q^*\Big(J_p(x_n) - \sum_{k=1}^{N_n} \mu_{n,k}\,A^*w_{n,k}^*\Big)\,, \quad n = 0, 1, \ldots$$

Since we already know that the direction $A^*J_r(Ax_n - y)$ assures convergence, U_n^* should contain that direction. Furthermore, in order to guarantee that the new iterate remains optimal with respect to the old search space U_{n-1}^* and the optimization achieved so far is not spoiled by searching in new directions, we choose $U_{n-1}^* \subset U_n^* \subset \mathcal{R}(A^*)$ implying $H_{n-1} \supset H_n \supset M_{Ax=y}$. Doing so we hope that already after a relatively small number of iterations n the set H_n is a good approximation to $M_{Ax=y}$.

In the following X is assumed to be smooth with q-smooth dual, which implies that X is reflexive and uniformly convex, whereas Y is allowed to be an arbitrary Banach space. If the duality mapping J_r of Y is set-valued then we also write $J_r(y)$ for an arbitrary but fixed element in the set $J_r(y)$. To compute the Bregman projection $\Pi^p_{M_{Ax=y}}(x_0)$ of $x_0 \in X$ onto the set $M_{Ax=y}$ for some given $y \in \mathcal{R}(A)$ we consider the following sequential subspace optimization method.

Method 1.

(S0) Take x_0 as initial value, set $n := 0$, $U^*_{-1} := \{0\}$ and repeat the following steps:
(S1) If $R_n := \|Ax_n - y\| = 0$ then STOP else goto (S2).
(S2) Choose the search space $U^*_n \subset \mathcal{R}(A^*)$ and N_n search directions $A^* w^*_{n,k} \in U^*_n$, $k = 1, \ldots, N_n$, such that

$$A^* J_r(Ax_n - y) \in U^*_n \quad \textit{and} \quad U^*_{n-1} \subset U^*_n = \mathrm{span}\{A^* w^*_{n,1}, \ldots, A^* w^*_{n,N_n}\}\,.$$

(S3) Compute the new iterate:

$$x_{n+1} := J^*_q\left(J_p(x_n) - \sum_{k=1}^{N_n} \mu_{n,k}\, A^* w^*_{n,k}\right)\,, \tag{4.2}$$

where $\mu_n = (\mu_{n,1}, \ldots, \mu_{n,N_n})$ is a solution of the N_n-dimensional optimization problem

$$\min_{t\,\in\mathbb{R}^{N_n}}\ h_n(t)$$

with

$$h_n(t) := \frac{1}{q}\left\|J_p(x_n) - \sum_{k=1}^{N_n} t_k\, A^* w^*_{n,k}\right\|^q + \sum_{k=1}^{N_n} t_k\, \langle w^*_{n,k} \,|\, y\rangle\,.$$

(S4) Set $n \leftarrow n+1$ and goto (S1).

A natural choice for U^*_n fulfilling the requirements in (S2) is

$$U^*_n = \mathrm{span}\{A^* J_r(Ax_0 - y), \ldots, A^* J_r(Ax_n - y)\}\,.$$

Proposition 4.1. *Method 1 either stops after a finite number n^* of iterations in case $R_{n^*} = 0$ where $x_{n^*} = \Pi^p_{M_{Ax=y}}(x_0)$ or the sequence $(x_n)_n$ converges strongly to $\Pi^p_{M_{Ax=y}}(x_0)$. Moreover the following holds:*

(a) For $H_n := \bigcap_{k=1}^{N_n} H\big(A^* w^*_{n,k}, \langle w^*_{n,k} \,|\, y\rangle\big)$ and all $z \in M_{Ax=y}$ we have

$$M_{Ax=y} \subset H_n \subset H_{n-1}\ ,\ \ x_{n+1} = \Pi^p_{H_n}(x_0)\ ,\ \ J_p(x_{n+1}) = \Pi^q_{J_p(x_0)+U^*_n} J_p(z)\,.$$

(b) We have $\left\langle w^*_{n,k} \,\middle|\, Ax_{n+1} - y\right\rangle = 0$ for all $k = 1, \ldots, N_n$.
(c) For all $z \in M_{Ax=y}$ the estimate

$$\Delta_p(x_{n+1}, z) \le \Delta_p(x_n, z) - \frac{R^p_n}{p\, C^{p-1} \|A\|^p}\,,$$

is valid, which can also be expressed in terms of the function h_n by

$$h_n(t_n) \le h_n(0) - \frac{R^p_n}{p\, C^{p-1} \|A\|^p}\,,$$

where $C > 0$ is the constant appearing in (2.5) for the q-smooth dual X^*.

Proof. To see that $M_{Ax=y} \subset H_n$ we pick any $z \in M_{Ax=y}$ and get for all $k = 1, \dots, N_n$

$$\langle A^* w^*_{n,k} \,|\, z \rangle = \langle w^*_{n,k} \,|\, Az \rangle = \langle w^*_{n,k} \,|\, y \rangle \,.$$

Recalling that $U^*_{n-1} \subset U^*_n$ this implies

$$H_n = z + (U^*_n)^\perp \subset z + (U^*_{n-1})^\perp = H_{n-1} \,.$$

Computing the new iterate via solving $\min\limits_{t \in \mathbb{R}^{N_n}} h_n(t)$ we get for all $j = 1, \dots, N_n$

$$\begin{aligned} 0 = \partial_j h_n(t_n) &= -\left\langle A^* w^*_{n,j} \,\middle|\, J^*_q \left(J_p(x_n) - \sum_{k=1}^{N_n} t_{n,k}\, A^* w^*_{n,k} \right) \right\rangle + \left\langle w^*_{n,j} \,\middle|\, y \right\rangle \\ &= -\left\langle w^*_{n,j} \,\middle|\, A x_{n+1} - y \right\rangle \,. \end{aligned} \tag{4.3}$$

Proposition 3.12 then yields $x_{n+1} = \Pi^p_{H_n}(x_n)$. Applying proposition 3.7 (b) and $H_n = z + (U^*_n)^\perp$ for any $z \in M_{Ax=y} \subset H_n$ we furher get

$$J_p(x_{n+1}) = \Pi^q_{J_p(x_n) + U^*_n}\, J_p(z) \,.$$

Inductively this leads to

$$J_p(x_n) + U^*_n = J_p(x_0) + U^*_n$$

by (4.2) and finally we get

$$J_p(x_{n+1}) = \Pi^q_{J_p(x_0) + U^*_n}\, J_p(z) \,.$$

Hence we indeed have $x_{n+1} = \Pi^p_{H_n}(x_0)$ and (a) is proven.
Part (b) follows from (4.3).
It remains to prove (c) and the convergence statement. As shown above we have

$$J_p(x_n) - J_p(x_0) \in U^*_n \subset \overline{\mathcal{R}(A)} = \mathcal{N}(A)^\perp \,.$$

In case $R_{n^*} = 0$ for some n^* we have $x_{n^*} \in M_{Ax=y}$ and by proposition 3.7 (b) we conclude that $x_{n^*} = \Pi^p_{M_{Ax=y}}(x_0)$, i.e. we are done.
Thus, let us assume $R_n \neq 0$ for all $n \in \mathbb{N}$ and let $z \in M_{Ax=y}$ be arbitrary. We will at first establish the monotonicity estimate

$$\Delta_p(x_{n+1}, z) \le \Delta_p(x_n, z) - \frac{R^p_n}{p\, C^{p-1} \|A\|^p} < \Delta_p(x_n, z) \,.$$

Since the search space U^*_n is chosen such that it contains $A^* J_r(Ax_n - y)$ we have for all $\bar{\mu} \ge 0$

$$J_p(x_n) - \bar{\mu}\, A^* J_r(Ax_n - y) \in J_p(x_n) + U^*_n \,.$$

Together with $J_p(x_{n+1}) = \Pi^q_{J_p(x_n)+U_n^*} J_p(z)$ we get

$$\begin{aligned}\Delta_p(x_{n+1},z) &= \Delta_q^*\big(J_p(z), J_p(x_{n+1})\big)\\ &\leq \Delta_q^*\big(J_p(z), J_p(x_n) - \overline{\mu}\, A^* J_r(Ax_n - y)\big)\\ &= \frac{1}{p}\|z\|^p - \langle J_p(x_n)\,|\,z\rangle + \overline{\mu}\,\langle J_r(Ax_n - y)\,|\,y\rangle\\ &\quad + \frac{1}{q}\|J_p(x_n) - \overline{\mu}\, A^* J_r(Ax_n - y)\|^q\,.\end{aligned}$$

We estimate the last summand by (2.5) and with $\|A^*\| = \|A\|$ we obtain

$$\begin{aligned}\Delta_p(x_{n+1},z) &\leq \frac{1}{p}\|z\|^p - \langle J_p(x_n)\,|\,z\rangle + \overline{\mu}\,\langle J_r(Ax_n - y)\,|\,y\rangle\\ &\quad + \frac{1}{q}\|x_n\|^p - \overline{\mu}\,\langle J_r(Ax_n - y)\,|\,Ax_n\rangle\\ &\quad + \frac{C}{q}\overline{\mu}^q\|A\|^q\|Ax_n - y\|^{(r-1)q}\\ &= \Delta_p(x_n,z) - \overline{\mu}R_n^r + \frac{C}{q}\overline{\mu}^q\|A\|^q R_n^{(r-1)q}\,.\end{aligned}$$

The right hand side of the above inequality is easy to minimize as a function of $\overline{\mu}$ by setting its derivative equal to zero, which yields

$$\overline{\mu}^{q-1} = \frac{R_n^r}{C\|A\|^q R_n^{(r-1)q}} \quad \Leftrightarrow \quad \overline{\mu} = \frac{R_n^{p-r}}{C^{p-1}\|A\|^p}\,. \tag{4.4}$$

Inserting this $\overline{\mu}$ we arrive at the announced monotonicity estimate

$$\Delta_p(x_{n+1},z) \leq \Delta_p(x_n,z) - \frac{R_n^p}{p\,C^{p-1}\|A\|^p} < \Delta_p(x_n,z)\,.$$

Hence $(R_n)_n$ converges to zero and the sequence $\big(\Delta_p(x_n,z)\big)_n$ is strictly decreasing. Proposition 3.3 (c) assures that the sequence $(x_n)_n$ remains bounded. To prove the convergence of $(x_n)_n$ to $\Pi^p_{M_{Ax=y}}(x_0)$ it suffices to show that every subsequence has in turn a subsequence converging strongly to $\Pi^p_{M_{Ax=y}}(x_0)$. Let $(x_{n_k})_k$ be an arbitrary subsequence. Since it is bounded we may assume that it converges weakly to some $\tilde{x} \in X$, and since the mapping $x \mapsto \|Ax - y\|$ is weakly lower semicontinuous we have

$$\|A\tilde{x} - y\| \leq \liminf_{k\to\infty} \|Ax_{n_k} - y\| = \liminf_{k\to\infty} R_{n_k} = 0\,,$$

whence $\tilde{x} \in M_{Ax=y}$. Now we use the weak lower semicontinuity of Δ_p in the second argument, as stated in proposition 3.3 (d) and the already proven relations of part (a)

$$x_{n_k} = \Pi^p_{H_{n_k-1}}(x_0) \quad \text{and} \quad \Pi^p_{M_{Ax=y}}(x_0) \in M_{Ax=y} \subset H_{n_k-1}\,,$$

to arrive at

$$\begin{aligned}\Delta_p\big(x_0, \Pi^p_{M_{Ax=y}}(x_0)\big) &\leq \Delta_p(x_0, \tilde{x}) \\ &\leq \liminf_{k\to\infty} \Delta_p(x_0, x_{n_k}) \leq \limsup_{k\to\infty} \Delta_p(x_0, x_{n_k}) \\ &\leq \Delta_p\big(x_0, \Pi^p_{M_{Ax=y}}(x_0)\big)\,.\end{aligned}$$

From this we infer that the sequence $\big(\Delta_p(x_0, x_{n_k})\big)_k$ converges and that its limit is equal to

$$\Delta_p\big(x_0, \Pi^p_{M_{Ax=y}}(x_0)\big) = \Delta_p(x_0, \tilde{x})\,.$$

Since the Bregman projection is unique we must have $\tilde{x} = \Pi^p_{M_{Ax=y}}(x_0)$. Finally by means of proposition 3.3 (e) we conclude that $(x_{n_k})_k$ indeed converges strongly to $\Pi^p_{M_{Ax=y}}(x_0)$.

Although we have restricted ourselves here to the case of a q-smooth dual, convergence can also be proven for the general case of a uniformly smooth dual with similar techniques as in BONESKY ET. AL. [5], where we analyzed iteration methods for the minimization of Tikhonov functionals in Banach spaces.
We emphasize that the costly application of the operators A and A^* has to be done only in step (S2) but not in (S3). Therefore the numerical effort to solve the minimization problems in (S3) will be significantly minor as long as the dimension N_n is low. It may therefore be more reasonable to drop the requirement $U_{n-1} \subset U_n^*$, which together with $A^*J_r(Ax_n - y) \in U_n^*$ implies that N_n increases, and bound the dimension N_n. However our proof of the strong convergence essentially relied on the requirement $U_{n-1} \subset U_n^*$. Nevertheless we still get the following, weaker result.

Proposition 4.2. *Consider method 1 when for the choice of the search space U_n^* in* (S2) *we only demand that it contains $A^*J_r(Ax_n - y)$ and that the dimension N_n does not exceed some fixed upper bound $N \in \mathbb{N}$. Moreover let the duality mapping J_p be sequentially weak-to-weak-continuous, i.e. it maps weakly convergent sequences in X to weakly convergent sequences in X^*. Then the method either stops after a finite number n^* of iterations in case $R_{n^*} = 0$ where $x_{n^*} = \Pi^p_{M_{Ax=y}}(x_0)$ or the sequence $(x_n)_n$ converges weakly to $\Pi^p_{M_{Ax=y}}(x_0)$. Moreover we have*

$$M_{Ax=y} \subset H_n\,, \quad x_{n+1} = \Pi^p_{H_n}(x_n)\,, \quad J_p(x_{n+1}) = \Pi^q_{J_p(x_n)+U_n^*} J_p(z)\,,\, z \in M_{Ax=y}\,.$$

Assertions (b) *and* (c) *of proposition* 4.1 *remain valid.*

Proof. The proof is done similar to that of proposition 4.1 taking according modifications into account. Note that the relations $x_{n+1} = \Pi^p_{H_n}(x_0)$ and $J_p(x_{n+1}) = \Pi^q_{J_p(x_0)+U_n^*} J_p(z)$, which were necessary to prove strong convergence, do not longer hold. Nevertheless it remains true that every subsequence of $(x_n)_n$ has in turn a subsequence $(x_{n_k})_k$ converging weakly to some $\tilde{x} \in M_{Ax=y}$ and that $J_p(x_{n_k}) - J_p(x_0) \in \overline{\mathcal{R}(A^*)}$. Since $\overline{\mathcal{R}(A^*)}$ is weakly closed this implies that $J_p(\tilde{x}) - J_p(x_0) \in \overline{\mathcal{R}(A^*)}$

in case the duality mapping J_p is sequentially weak-to-weak-continuous. Hence $\tilde{x} = \Pi^p_{M_{Ax=y}}(x_0)$ and we conclude that $(x_n)_n$ converges weakly to $\Pi^p_{M_{Ax=y}}(x_0)$.

Remark 4.3. We remind that in finite dimensions weak and strong convergence coincide. Furthermore the duality mappings of the l_p-sequence spaces are sequentially weak-to-weak-continuous. This is not valid for the L_p-function spaces, see CIORANESCU [8].

5. Numerical Experiments

To illustrate the advantage of using sequential subspace methods when dealing with large scale problems we computed the minimum p-norm solutions $x_p^\dagger$ of matrix equations $Ax = y$ for different values p and dimensions N of the search spaces. The matrix A was a randomly generated 1000×5000-matrix with entries in $[-1, 1]$. To obtain systems with known solutions of minimum p-norm equal to 1 we generated a random vector y^* in $[-1, 1]^{1000}$, set

$$x_p^\dagger := \frac{J_q^*(A^* y^*)}{\|J_q^*(A^* y^*)\|_p} \quad \text{and} \quad y := A x_p^\dagger \,.$$

We implemented method 1 in MATLAB where the minimization subproblems in (S2) were solved with the function FMINUNC, which is a BFGS Quasi-Newton method. As search spaces we used

$$U_n^* = \text{span}\{A^* J_r(A x_{\max\{0, n-N+1\}} - y), \ldots, A^* J_r(A x_n - y)\} \,,$$

whence $\dim(U_n^*) \leq N$ with $N = 2, 4, 6$. Depending on the smoothness of the dual $(\mathbb{R}^{5000}, \|.\|_q)$ of $(\mathbb{R}^{5000}, \|.\|_p)$ we used J_2 and Δ_2 in case $p < 2$ and J_p and Δ_p in case $p > 2$. The algorithm was terminated when

$$\frac{\|A x_n - y\|_2}{\|y\|_2} \leq 10^{-4} \,.$$

Table 1 lists the number of outer iterations n for different values of p and N, showing that already using search spaces with low dimensions helps to reduce the number of the costly applications of A and A^* tremendously. The convergence of the Landweber method (4.1) where only one single search direction was used, i.e. $N = 1$, was significantly slower as tests in [20] demonstrated where we needed thousands of iterations to get reasonable approximations to $x_p^\dagger$. Figures 1-2 confirm that the Bregman distance of the iterates x_n to the solution $x_p^\dagger$ is indeed decreasing in each iteration as stated in proposition 4.1 (c), but this need not be true for the norm distance or the residuals. Moreover they demonstrate that in general $\Delta_p(x, y)^{\frac{1}{p}}$ is not proportional to $\|x - y\|$, whereas this is valid in Hilbert spaces for $p = 2$.

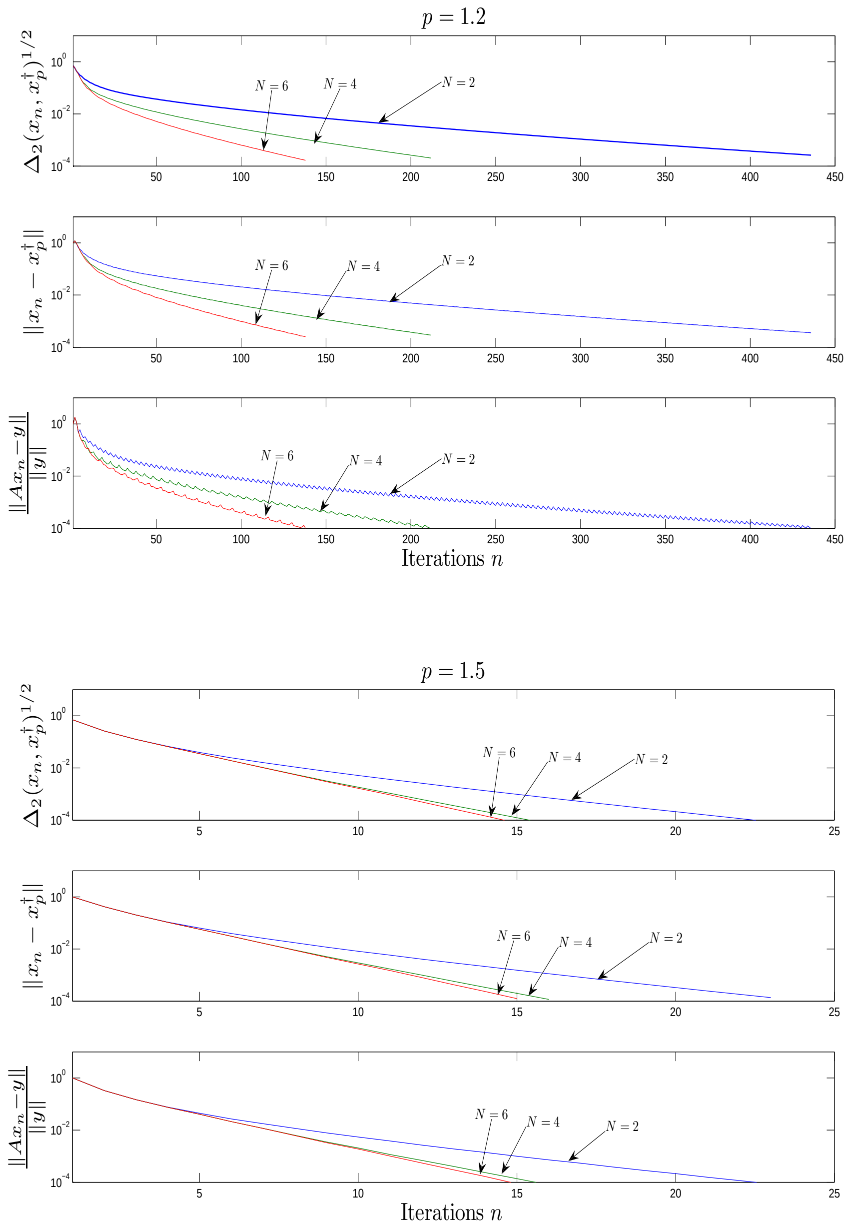

Figure 1: Relative error of the residual, norm and Bregman distance of the iterates x_n to the minimum p-norm solution $x_p^\dagger$ (log-scale) vs. number of iterations n for $p = 1.2$ (top), $p = 1.5$ (bottom) and dimensions $N = 2, 4, 6$ of the search spaces

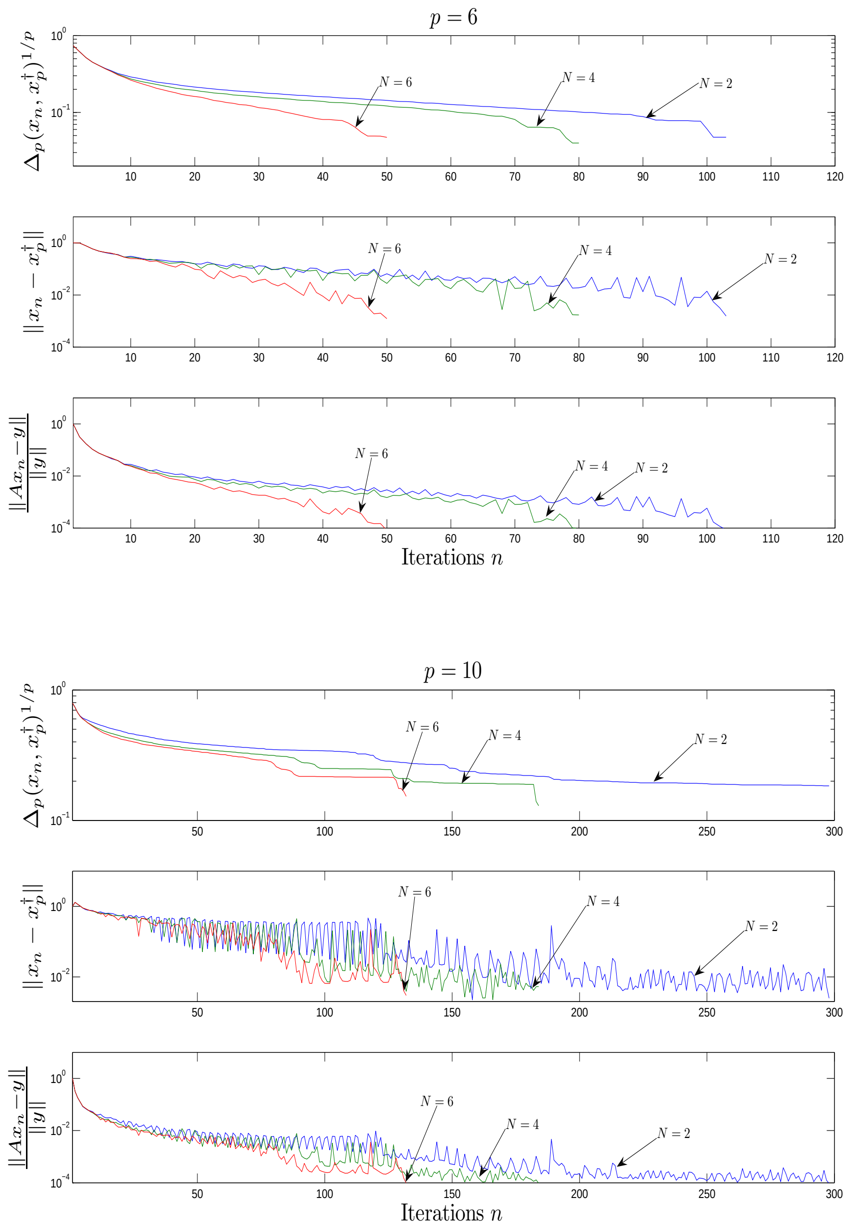

Figure 2: Relative error of the residual, norm and Bregman distance of the iterates x_n to the minimum p-norm solution $x_p^\dagger$ (log-scale) vs. number of iterations n for $p = 6$ (top), $p = 10$ (bottom) and dimensions $N = 2, 4, 6$ of the search spaces

	$p = 1.2$	$p = 1.5$	$p = 6$	$p = 10$
$N = 2$	435	22	102	297
$N = 4$	211	15	79	183
$N = 6$	137	14	49	131

Table 1: Number of iterations for different p-norms and dimensions N of the search spaces

6. Conclusions

We have placed the problem of computing minimum-norm solutions of linear operator equations in the context of computing metric and Bregman projections onto affine subspaces. Using the simple relationship $P_C(x) - x = \Pi^p_{C-x}(0)$ enabled us to use the same iterative method for the computation of metric and Bregman projections onto affine subspaces which are given by the nullspace or the range of a linear operator.
Furthermore we modified an earlier proposed method of Landweber type using ideas from sequential subspace optimization methods to accelerate convergence and to obtain a powerful iteration scheme. The construction of that scheme followed ideas from the CG-algorithm where in each step several search directions are taken into account. The convergence could be distinctively accelerated by using this advanced strategy. We at last mention that the developed method must not be seen as being just an improvement of the Landweber method but as general optimization approach leading to a highly efficient solver for linear inverse problems in Banach spaces.
Further research should include the investigation of the regularizing property of the subspace method, convergence rates in connection with appropriate source conditions, as well as the application to real-world problems.

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** (1997), 33–61.

2. Y. I. Alber, *Metric and Generalized Projection Operators in Banach Spaces: Properties and Applications*. Theory and Applications of Nonlinear Operators of Monotone and Accretive Type (A. Kartsatos, ed.), 15–50. Marcel Dekker, New York, 1996.

3. ______, James orthogonality and orthogonal decompositions of Banach spaces. *J. Math. Anal. Appl.* **312** (2005), 330–342.

4. 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** (2001), 615–647.

5. T. Bonesky, K. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, *Minimization of Tikhonov Functionals in Banach Spaces*, to appear, 2007.

6. L. M. Bregman, The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* **7** (1967), 200–217.

7. 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** (2006), Article ID 84919, 39 pages.

8. I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Kluwer Academic Publishers, Dordrecht, 1990.

9. P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*. Springer, New York, 2004.

10. J. Diestel, 485, *Geometry of Banach space-selected topics*. Springer, New York, 1975.

11. M. Elad, B. Matalon, and M. Zibulevsky, Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Appl. Comput. Harmon. Anal.* **23** (2007), 346–367.

12. T. Figiel, On the moduli of convexity and smoothness. *Stud. Math.* **56** (1976), 121–155.

13. F. Jarre and J. Stoer, *Optimierung*. Springer, Berlin, 2004.

14. J. Lindenstrauß and L. Tzafriri, *Classical Banach Spaces*, *II*. Springer, Berlin, 1979.

15. J. L. Lions, *Quelques methodes de résolution des problemes aux limites non linéaires*. Dunod Gauthier-Villars, Paris, 1969.

16. G. Narkiss and M. Zibulevsky, *Sequential subspace optimization method for large-scale unconstrained optimization*, Technion - The Israel Institute of Technology, Report, Department of Electrical Engineering, 2005.

17. J. Nocedal, Theory of Algorithms for Unconstrained Optimization. *Acta Numerica* (1991), 199–242.

18. J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.

19. J. P. Penot and R. Ratsimahalo, Characterizations of metric projections in Banach spaces and applications. *Abstr. Appl. Anal.* **3** (1998), 85–103.

20. F. Schöpfer, A. K. Louis, and T. Schuster, Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems* **22** (2006), 311–329.

21. W. Song and Z. Cao, The generalized decomposition theorem in Banach spaces and its applications. *Journal of Approximation Theory* **129** (2004), 167–181.

22. J. Stoer and Y. Yuan, A Subspace Study on Conjugate Gradient Algorithms. *Z. angew. Math. Mech.* **75** (1995), 69–77.

23. Z.-B. Xu and G. F. Roach, Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications* **157** (1991), 189–210.

Received

Author information

F. Schöpfer, Fakultät für Maschinenbau, Helmut-Schmidt-Universität, Holstenhofweg 85, 22043 Hamburg, Germany.
Email: schoepfer@hsu-hh.de

T. Schuster, Fakultät für Maschinenbau, Helmut-Schmidt-Universität, Holstenhofweg 85, 22043 Hamburg, Germany.
Email: schuster@hsu-hh.de

A.K. Louis, Fakultät für Mathematik und Informatik, Universität des Saarlandes, Postfach 15 11 50, 66041 Saarbrücken, Germany.
Email: louis@num.uni-sb.de