

Bregman Distances in Inverse Problems and Partial Differential Equations

Martin Burger

Abstract The aim of this paper is to provide an overview of recent development related to Bregman distances outside its native areas of optimization and statistics. We discuss approaches in inverse problems and image processing based on Bregman distances, which have evolved to a standard tool in these fields in the last decade. Moreover, we discuss related issues in the analysis and numerical analysis of non-linear partial differential equations with a variational structure. For such problems Bregman distances appear to be of similar importance, but are currently used only in a quite hidden fashion. We try to work out explicitly the aspects related to Bregman distances, which also lead to novel mathematical questions and may also stimulate further research in these areas.

1 Introduction

Bregman distances for (differentiable) convex functionals, originally introduced in the study of proximal algorithms in [11] and named in [25], are a well-established concept in continuous and discrete optimization in finite dimension. A classical example is the celebrated Bregman projection algorithm for finding points in the intersection of affine subspaces (cf., e.g., [26]). We refer to [26, 53] for introductory and exhaustive views on Bregman distances in optimization.

Although convex functionals play a role in many other branches of mathematics, e.g., in many variational problems and partial differential equations, the suitability of Bregman distances in such fields was hardly investigated for several decades. In mathematical imaging and inverse problems the situation changed with the rediscovery and further development of Bregman iterations as an iterative image restoration technique in the case of frequently used regularization techniques such as total variation (cf. [50]), which led to significantly improved results compared to standard variational models and could eliminate systematic errors to a certain extent

M. Burger (✉)

Institut für Numerische und Angewandte Mathematik, Westfälische Wilhelms-Universität (WWU) Münster. Einsteinstr. 62, D 48149 Münster, Germany
e-mail: martin.burger@wwu.de

© Springer International Publishing Switzerland 2016

J.-B. Hiriart-Urruty et al. (eds.), *Advances in Mathematical Modeling, Optimization and Optimal Control*, Springer Optimization and Its Applications 109,
DOI 10.1007/978-3-319-30785-5_2

(cf. [9, 16]). Another key observation increasing the interest in Bregman distances in these fields was that they can be employed for error estimation, in particular for not strictly convex and nonsmooth functionals (cf. [14]), which prevent norm estimates.

Although there are many obvious links to the main route of research in Bregman distances and related optimization algorithms, there are several peculiar aspects that deserve particular discussion. Besides missing smoothness of the considered functionals and the fact that problems in imaging, inverse problems and partial differential equations are naturally formulated in infinite-dimensional Banach spaces such as the space of functions of bounded variation or Sobolev spaces, which have only been considered in few instances before, a key point is that the motivation for using Bregman distances in these fields often differs significantly from those in optimization and statistics. In the following we want to provide an overview of such questions and consequent developments, keeping an eye on potential directions and questions for future research. We start with a section including definitions, examples, and some general properties of Bregman distances, before we survey aspects of Bregman distances in inverse problems and imaging developed in the last decade. Then we proceed to a discussion of Bregman distances in partial differential equations, which is less explicit and hence the main goal is to highlight hidden use of Bregman distances and make the idea more directly accessible for future research. Finally we conclude with a section on related recent developments.

2 Bregman Distances and Their Basic Properties

We start with a definition of a Bregman distance. In the remainder of this paper, let X be a Banach space and $J : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex functionals. We first recall the definition of subdifferential, respectively, subgradients.

Definition 1. The subdifferential of a convex functional J is defined by

$$\partial J(u) = \{p \in X^* \mid J(u) + \langle p, v - u \rangle \leq J(v) \text{ for all } v \in X\}. \quad (1)$$

An element $p \in \partial J(u)$ is called subgradient.

Having defined a subdifferential we can proceed to the definition of Bregman distances, respectively, generalized Bregman distances according to [44].

Definition 2. The (generalized) Bregman distance related to a convex functional J with subgradient p is defined by

$$D_J^p(v, u) = J(v) - J(u) - \langle p, v - u \rangle, \quad (2)$$

where $p \in \partial J(u)$. The symmetric Bregman distance is defined by

$$D_J^{p,q}(u, v) = D_J^p(v, u) + D_J^q(u, v) = \langle p - q, u - v \rangle, \quad (3)$$

where $p \in \partial J(u)$, $q \in \partial J(v)$.

Note that in the differentiable case, i.e., $\partial J(u)$ being a singleton, we can omit the special subgradient and write $D_J(v, u)$ or $D_J^{J'(u)}(v, u)$.

By the definition of subgradients the nonnegativity is apparent:

Proposition 1. *Let J be convex and $p \in \partial J(u)$. Then*

$$D_J^p(v, u) \geq 0 \quad \forall v \in X$$

and

$$D_J^p(u, u) = 0.$$

If J is strictly convex, then $D_J^p(v, u) > 0$ for $v \neq u$.

We can further characterize vanishing Bregman distances as sharing a subgradient:

Proposition 2. *Let J be convex and $p \in \partial J(u)$. Then $D_J^p(v, u) = 0$ if and only if $p \in \partial J(v)$.*

Since Bregman distances are convex with respect to the first argument, we can also compute a subdifferential with respect to that variable, which is simply a shift of the subdifferential of J :

Proposition 3. *Let J be convex, $p \in \partial J(u)$. Then*

$$\partial_v D_J^p(v, u) = \partial J(v) - p.$$

Concerning existence proofs for variational problems involving Bregman distance it is often useful to investigate lower semicontinuity properties. Since Bregman distances can be considered as affinely linear perturbations of the functional J it is natural that these properties carry over:

Proposition 4. *Let J be convex and $q \in \partial J(v)$. Then the functional H defined by*

$$H(u) = D_J^q(u, v)$$

is convex. Hence, if X is reflexive, then H is weakly lower semicontinuous. If X is the dual of some Banach space Z and J is the convex conjugate of a functional on Z , then $q \in Z$ implies that H is lower semicontinuous in the weak star topology.

2.1 Examples of Bregman Distances

In the following we provide several examples of Bregman distances as frequently found in literature as well as some that received recent attention. This shall provide further insights into the relation to other distance measures and the basic properties of Bregman distances:

Example 1. Let X be a Hilbert space and $J(u) = \frac{1}{2} \|u\|_X^2$. Then $\partial J(u) = \{u\}$ and hence

$$D_J^u(v, u) = \frac{1}{2} \|u - v\|_X^2. \quad (4)$$

Example 2. Let I be a countable index set and $X = \ell^1(I)$ with

$$J(u) = \|u\|_{\ell^1} = \sum_{i \in I} |u_i|.$$

Then the Bregman distance is given by

$$D_J^p(v, u) = \sum_{i \in I} (q_i - p_i) v_i = \sum_{i, v_i > 0} (1 - p_i) |v_i| + \sum_{i, v_i < 0} (1 + p_i) |v_i|. \quad (5)$$

Note that the above sums have nonzero entries only if the sign of u_i does not match the sign of v_i , since $p_i = 1$ if $u_i > 0$ and $p_i = -1$ if $u_i < 0$.

Example 3. Let $X = \ell_+^1(\{1, \dots, N\})$ with

$$J(u) = \sum_{i=1}^N u_i \log u_i + 1 - u_i,$$

which is called the logarithmic entropy (or Boltzmann entropy). Then the Bregman distance is given by

$$D_J^p(v, u) = \sum_{i=1}^N v_i \log \frac{v_i}{u_i} + u_i - v_i, \quad (6)$$

which is known as Kullback–Leibler divergence. An analogous treatment applies to $X = L_+^1(\Omega)$, for a bounded domain Ω , and the continuous version

$$J(u) = \int_{\Omega} (u(x) \log u(x) + 1 - u(x)) \, dx,$$

resulting in the Bregman distance

$$D_J^p(v, u) = \int_{\Omega} \left(v(x) \log \frac{v(x)}{u(x)} + u(x) - v(x) \right) dx. \quad (7)$$

2.2 Bregman Distances and Duality

Duality is a basic ingredient in convex optimization (cf. [32]) and hence it is also interesting to understand some connections of duality and Bregman distances. For this sake we employ the convex conjugate (also called Legendre–Fenchel transform) of a functional J given by $J^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfying

$$J^*(p) = \sup_{u \in X} (\langle p, u \rangle - J(u)). \quad (8)$$

Noticing that for $p \in \partial J(u)$ we have $J^*(p) = \langle p, u \rangle - J(u)$ one can immediately rewrite the Bregman distance as

$$D_J^p(v, u) = J(v) + J^*(p) - \langle p, v \rangle, \quad (9)$$

which can be interpreted as measuring the deviation of p from being a subgradient in $\partial J(v)$ or the deviation of v from being a subgradient in $\partial J^*(p)$.

A key identity relates Bregman distances with respect to J to those with respect to the convex conjugate J^* :

Proposition 5. *Let $p \in \partial J(u)$ and $q \in \partial J(v)$. Then*

$$D_J^p(v, u) = D_{J^*}^q(p, q). \quad (10)$$

Proof. By simple reordering we find

$$\begin{aligned} D_J^p(v, u) &= J(v) - \langle p, v \rangle + \langle p, u \rangle - J(u) \\ &= J(v) - \langle p, v \rangle + J^*(p), \end{aligned}$$

where we have used the maximality relation for the convex conjugate, which is equivalent to $p \in \partial J(u)$. With analogous reasoning we find $J^*(q) = \langle q, v \rangle - J(v)$ and hence

$$D_J^p(v, u) = J(v) + J^*(p) - J^*(q) - \langle p - q, v \rangle = D_{J^*}^q(p, q),$$

noticing that $q \in \partial J(v)$ implies $v \in \partial J^*(q)$.

A second aspect of duality related to Bregman distance is the convex conjugate of the latter, which shows that Bregman distances are dual to measuring differences via a functional:

Proposition 6. *Let $q \in \partial J(v)$ and H be defined by*

$$H(u) = D_J^q(u, v). \quad (11)$$

Then

$$H^*(p) = J^*(p+q) - J^*(q). \quad (12)$$

Proof. We have

$$\begin{aligned} H^*(p) &= \sup_u [\langle p, u \rangle - J(u) + J(v) - \langle q, v - u \rangle] \\ &= \sup_u [\langle p+q, u \rangle - J(u)] - [\langle q, v \rangle - J(v)]. \end{aligned}$$

The first term equals $J^*(p+q)$ by definition and the second equals $J^*(q)$ since $q \in \partial J(v)$.

2.3 Bregman Distances and Fenchel Duality

In the following we further investigate some properties of Bregman distances for a combination of two convex functionals $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$, $G : Y \rightarrow \mathbb{R} \cup \{+\infty\}$. The classical setting is related the Fenchel duality theorem (cf. [32]), where

$$J(u) := F(u) + G(Ku) \quad (13)$$

with $K : X \rightarrow Y$ a bounded linear operator between Banach spaces. The Fenchel duality theorem shows that under suitable conditions

$$\inf_u J(u) = \sup_w [F^*(-K^*w) + G^*(w)], \quad (14)$$

together with equations relating optimal solutions \hat{u} and \hat{w} via subdifferentials of the involved functionals

$$-K^*\hat{w} \in \partial F(\hat{u}), \quad K\hat{u} \in \partial G^*(\hat{w}). \quad (15)$$

The above duality opens the possibility to employ Bregman distances on the dual problem as well as on the primal, which is nicely complemented by the duality relations for Bregman distances of a functional and its convex conjugate.

In the following we derive a basic estimates for the variational problem (13), which clarifies the relation of perturbations of one functional with duality and Bregman distances. We shall assume that the regularity of F and G is such that

$$\partial J(u) = \partial F(u) + K^* \partial G(Ku)$$

and the Fenchel duality theorem holds (cf. [32] for details).

Then we obtain the following estimate for perturbations of J :

Theorem 7. *Let F , G and K be as above, and let \tilde{G} be a perturbation of G satisfying the same assumptions. Let $u \in X$ be a minimizer of J with $-K^*w \in \partial F(u)$ and \tilde{u} be a minimizer of $F(\cdot) + \tilde{G}(K\cdot)$ with $-K^*\tilde{w} \in \partial F(\tilde{u})$. Then*

$$D_F^{-K^*w, -K^*\tilde{w}}(u, \tilde{u}) \leq G^*(\tilde{w}) - G^*(w) + \tilde{G}^*(w) - \tilde{G}^*(\tilde{w}). \quad (16)$$

Proof. We have

$$\begin{aligned} D_F^{-K^*w, -K^*\tilde{w}}(u, \tilde{u}) &= \langle K^*\tilde{w} - K^*w, u - \tilde{u} \rangle \\ &= \langle Ku, \tilde{w} - w \rangle + \langle K\tilde{u}, w - \tilde{w} \rangle. \end{aligned}$$

By the Fenchel duality theorem we have $Ku \in \partial G^*(w)$ and $K\tilde{u} \in \partial \tilde{G}^*(\tilde{w})$, which implies the assertion by inserting the subgradient inequality.

2.4 Bregman Distances for One-Homogeneous Functionals

The case of convex one-homogeneous functionals J , i.e.,

$$J(tu) = |t|J(u) \quad \forall t \in \mathbb{R}, \quad (17)$$

received strong attention recently, and also appears to be a particularly interesting one with respect to Bregman distances. In the one-homogeneous case one has

$$J(u) = \langle p, u \rangle \quad (18)$$

for $p \in \partial J(u)$. Thus, the Bregman distance simply reduces to

$$D_J^p(v, u) = J(v) - \langle p, v \rangle. \quad (19)$$

An interesting property in the one-homogeneous case is the fact that the convex conjugate is the indicator function of a convex set C , i.e.,

$$J^*(p) = \begin{cases} 0 & \text{if } p \in C \\ +\infty & \text{else.} \end{cases} \quad (20)$$

This sheds interesting light on (10), noticing that $p \in \partial J(u)$ implies $p \in C$. Hence,

$$D_J^p(v, u) = D_{J^*}^v(p, q) = \langle q - p, v \rangle.$$

An alternative way to see this property is (19) combined with $\langle q, v \rangle = J(v)$.

In the one-homogeneous case we immediately find an example of Bregman distances vanishing for $v \neq u$. Let $t > 0$ and $v = tu$, then $\partial J(v) = \partial J(u)$ implies $D_J^p(v, u) = 0$. On the other hand, we observe that the Bregman distance distinguishes different orientation. Choosing $v = tu$ for $t < 0$ we have $\partial J(v) = -\partial J(u)$, hence $D_J^p(v, u) = 2J(v)$.

3 Applications in Inverse Problems and Imaging

In the last decade, Bregman distances have become an important tool in inverse problems and image processing. Their main use is twofold: On the one hand, they are of particular importance for all kinds of error estimates as already sketched above and in particular they are quite useful for the analysis of variational regularization techniques with non-differentiable regularization functionals. This route has been initiated in [14] and subsequently expanded, e.g., in [8, 18, 34, 36–38, 52, 54, 58]. On the other hand, Bregman distances can be used to construct novel iterative techniques with superior properties compared to classical variational regularization. This route goes back to [50] and was developed further, e.g., in [12, 16, 22, 35, 47, 60, 61, 63], the methods also had a huge impact on various applications (cf., e.g., [12, 28, 49]).

The basic setup we consider is the solution of a problem of the form $Ku = f$, where $K : X \rightarrow Y$ is a bounded linear operator between Banach spaces and f are given data. Since in most cases K does not have a closed range (or is even a compact operator) and data contain measurement errors, this problem can be ill-posed. To cure this issue variational regularization methods employ a convex regularization functional $R : X \rightarrow \mathbb{R} \cup \{+\infty\}$, which introduces the a-priori knowledge that reasonable approximations of the solution u have small (minimal) values $R(u)$. Variational regularization methods make a compromise between approximating the data f and minimizing R and solve a problem of the form

$$D(Ku, f) + \alpha R(u) \rightarrow \min_{u \in X}, \quad (21)$$

where $D : Y \times Y \rightarrow \mathbb{R}$ is an appropriate distance measure and $\alpha > 0$ is a regularization parameter to be chosen appropriately in dependence of the measurement error (often referred to as data noise). Specific forms of the distance measure are derived, e.g., via statistical modelling as the negative log-likelihood of the data noise. Frequently D is simply a least-squares term, i.e., Y is a Hilbert space and

$$D(Ku, f) = \frac{1}{2} \|Ku - f\|_Y^2. \quad (22)$$

A classical example is the ROF-model for image denoising [55], where R is the total variation seminorm, K is an embedding from $BV(\Omega) \cap L^2(\Omega)$ into $L^2(\Omega)$, and D the squared L^2 -norm. For the whole section we shall assume that D is convex with respect to the first variable, which is the case for almost all commonly used examples.

3.1 Error Estimates

Error estimates for solutions of (21) are of interest with respect to two quantities: First of all, the distance of the data f to the ideal data Ku^* , where u^* is the unknown ideal solution. This part is referred to as data error or noise. Second, the regularization parameter α , which should equal zero in the case of ideal data and introduces a systematic error in the case of perturbed data (when it needs to be positive). In the setting of (13) we thus need to choose

$$F(u) = \alpha R(u), \quad G(Ku) = D(Ku, f). \quad (23)$$

The optimality conditions for a minimizer u_α are then of the form

$$p_\alpha = K^* w_\alpha, \quad p_\alpha \in \partial R(u_\alpha) \quad - \alpha K^* w_\alpha \in \partial D(Ku_\alpha, f), \quad (24)$$

where the subgradient of D is meant to be computed with respect to the first argument for fixed f .

In order to obtain error estimates for some different data \tilde{f} we choose $\tilde{G}(Ku) = D(Ku, \tilde{f})$ and denote by \tilde{u}_α its corresponding regularized solution with

$$\tilde{p}_\alpha = K^* \tilde{w}_\alpha, \quad \tilde{p}_\alpha \in \partial R(\tilde{u}_\alpha).$$

Then (16) yields

$$\alpha D_R^{K^* w_\alpha, K^* \tilde{w}_\alpha}(u, \tilde{u}) \leq G^*(\tilde{w}_\alpha) - G^*(w_\alpha) + \tilde{G}^*(w_\alpha) - \tilde{G}^*(\tilde{w}_\alpha). \quad (25)$$

To further illustrate the behaviour consider the case of a quadratic data fidelity

$$G(Ku) = D(Ku, f) = \frac{1}{2} \|Ku - f\|^2, \quad (26)$$

for some squared Hilbert space norm, which yields $G^*(w) = \frac{1}{2} \|w\|^2 + \langle w, f \rangle$. Hence,

$$\alpha D_R^{K^* w_\alpha, K^* \tilde{w}_\alpha}(u, \tilde{u}) \leq \langle f - \tilde{f}, \tilde{w}_\alpha - w_\alpha \rangle. \quad (27)$$

In the case (26) one can see quite immediately why the (symmetric) Bregman distance is an appropriate error measure for the estimates. Starting with the optimality conditions

$$\begin{aligned} Ku_\alpha - f + \alpha w_\alpha &= 0, & p_\alpha &= K^* w_\alpha \in R(u_\alpha), \\ K\tilde{u}_\alpha - f + \alpha \tilde{w}_\alpha &= 0, & \tilde{p}_\alpha &= K^* \tilde{w}_\alpha \in R(\tilde{u}_\alpha), \end{aligned}$$

we find

$$K(u_\alpha - \tilde{u}_\alpha) + \alpha(w_\alpha - \tilde{w}_\alpha) = f - f^*. \quad (28)$$

The right-hand side is exactly the perturbation of the data, whose norm we want to use to estimate errors in the solution u_α . Hence we simply take the squared norm on both sides and obtain by expanding on the left-hand side

$$\|K(u_\alpha - \tilde{u}_\alpha)\|^2 + 2\alpha \langle w_\alpha - \tilde{w}_\alpha, K(u_\alpha - \tilde{u}_\alpha) \rangle + \alpha^2 \|w_\alpha - \tilde{w}_\alpha\|^2 = \|f - \tilde{f}\|^2.$$

Finally using $K^* w_\alpha = p_\alpha$ we arrive at

$$\|K(u_\alpha - \tilde{u}_\alpha)\|^2 + 2\alpha D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) + \alpha^2 \|w_\alpha - \tilde{w}_\alpha\|^2 = \|f - \tilde{f}\|^2, \quad (29)$$

which implies (by the nonnegativity of all involved terms) the immediate estimate

$$D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) \leq \frac{1}{2\alpha} \|f - \tilde{f}\|^2 \quad (30)$$

for the Bregman distance. Note that (29) is not just an estimate, but indeed an equality for three error terms—the error in the image of the operator K (somehow the residual), the error in the dual variables w , and the Bregman distance of solutions. Here Ku and w are elements of a Hilbert space and it is of course natural to measure their deviations in the corresponding norm, so (29) yields the Bregman distance as the naturally induced error measure in the Banach space X .

Having obtained (29) it is interesting to note that one can alternatively obtain estimates for two parts of the right-hand side by taking scalar products of (28) with appropriate elements and subsequent application of the Cauchy–Schwarz, respectively, Young’s inequality. The first is obtained by a scalar product with $Ku_\alpha - Ku^*$, which yields

$$\|K(u_\alpha - \tilde{u}_\alpha)\|^2 + \alpha D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) = \langle f - \tilde{f}, K(u_\alpha - \tilde{u}_\alpha) \rangle \leq \frac{1}{2} \|f - \tilde{f}\|^2 + \frac{1}{2} \|K(u_\alpha - \tilde{u}_\alpha)\|^2,$$

hence

$$\|K(u_\alpha - \tilde{u}_\alpha)\|^2 + 2\alpha D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) \leq \|f - \tilde{f}\|^2. \quad (31)$$

Using analogous reasoning, a scalar product of (28) with $w_\alpha - \tilde{w}_\alpha$ leads to

$$2\alpha D_R^{p_\alpha, \tilde{p}_\alpha}(u_\alpha, \tilde{u}_\alpha) + \alpha^2 \|w_\alpha - \tilde{w}_\alpha\|^2 \leq \|f - \tilde{f}\|^2. \quad (32)$$

3.2 Asymptotics

A key question in inverse problems is the behaviour of the regularized solution u_α as $\alpha \rightarrow 0$, which only makes sense if the noise in the data vanishes, i.e., $f = Ku^*$ for some desired solution $u^* \in X$. It is well-known that for ill-posed problems the convergence can be arbitrarily slow as $\alpha \rightarrow 0$ without further conditions on the desired solution u^* . For a further characterization it is important to note that under appropriate choice of α a limiting solution u^* of the variational model (21) satisfies

$$R(u) \rightarrow \min_{u \in X} \quad \text{subject to } Ku = Ku^*. \quad (33)$$

This can be seen from the estimate

$$D(Ku_\alpha, f) + \alpha R(u_\alpha) \leq D(Ku^*, f) + \alpha R(u^*).$$

Using $\alpha \rightarrow 0$ and $D(Ku^*, f) \rightarrow 0$ we see that $D(Ku_\alpha, f) \rightarrow 0$, hence the limit is a solution of $Ku^* = f$. Dividing by α and using nonnegativity of D , we find

$$R(u_\alpha) \leq R(u^*) + \frac{D(Ku^*, f)}{\alpha},$$

and under the standard condition on the parameter choice

$$\frac{D(Ku^*, f)}{\alpha} \rightarrow 0,$$

we observe that the limit of u_α cannot have a larger value of R than any other solution of $Ku = f$, i.e., it solves (33).

The key observation in [14, 27] is that appropriate conditions in the case of variational regularization is related to the existence of a Lagrange multiplier for (33). The Lagrange functional is given by $L(u, w) = R(u) - \langle w, Ku - Ku^* \rangle$, hence the existence of a Lagrange multiplier is the so-called *source condition*

$$p^* = K^* w^* \in \partial R(u^*). \quad (34)$$

Let us again detail the arguments in the case (26), where we can indeed use the above error estimates like (26) with $\tilde{u}_\alpha = u^*$ and $\tilde{w}_\alpha = w^*$. In order to obtain u_α as the solution of a variational problem we can indeed choose $\tilde{f} = Ku^* + \alpha w^*$ (note that (34) is equivalent to the existence of some \tilde{f} such that u^* solves the variational problem with data \tilde{f} , cf. [14]). Hence, (29) becomes

$$\|K(u_\alpha - u^*)\|^2 + 2\alpha D_R^{p^*, p^*}(u_\alpha, u^*) + \alpha^2 \|w_\alpha - w^*\|^2 = \|f - Ku^* - \alpha w^*\|^2. \quad (35)$$

Again with Young's inequality we end up at

$$D_R^{p_\alpha, p^*}(u_\alpha, u^*) \leq \frac{\|f - Ku^*\|^2}{\alpha} + \alpha \|w^*\|^2, \quad (36)$$

which gives the usual optimal choice $\alpha \sim \|f - Ku^*\|$ of regularization parameter in terms of the noise level, exactly as in the linear Hilbert space case (cf. [33]).

3.3 Bregman Iterations and Inverse Scale Space Methods

A frequent observation made for variational methods as discussed above is a systematic bias, in particular the methods yield solutions u_α with $R(u_\alpha)$ too small, which, e.g., results into a local loss of contrast in the case of total variation regularization (the contrast loss is larger for smaller structures cf. [17]). In order to cure such systematic errors in particular in the case of one-homogeneous regularization it turned out that the well-known Bregman iteration is a perfect tool. Instead of solving the variational problem only once one usually starts at u_0 being a minimizer of the regularization functional R , i.e., at the coarsest scale (if one agrees that scale is defined by R). Then of course $p_0 = 0 \in \partial R(u_0)$ and one can subsequently iterate

$$u_{k+1} \in \arg \min_{u \in X} (D(Ku, f) + \alpha D_R^{p_k}(u, u_k)), \quad (37)$$

where the subgradient p_k is updated via the optimality condition

$$p_{k+1} - p_k \in -\frac{1}{\alpha} K^* \partial D(Ku_k, f). \quad (38)$$

Noticing that we can again write $p_k = K^* w_k$ one can also construct an iteration

$$w_{k+1} - w_k \in -\frac{1}{\alpha} \partial D(Ku_k, f), \quad (39)$$

from which one can derive the well-known equivalence to augmented Lagrangian methods for minimizing R subject to $Ku = f$.

The convergence analysis in the case $f = Ku^*$ follows the well-known route for the Bregman iteration, but due to the ill-posedness of $Ku = f$ there is a particularly interesting aspect in the case of noisy data f differing from the ideal Ku^* . If the range of K is not closed, one has to take care of the situation where neither a solution $Ku = f$ nor some kind of least-squares solution (a minimizer of $D(Ku, f)$) exists in X . Hence, the Bregman iteration has the role of an iterative regularization method and needs to be stopped appropriately before the noise effects start to deteriorate the quality of the solution. Indeed one can show that the Bregman distance $D^{p_k}(u^*, u_k)$ is decreasing during the first iterations up to a certain point when the residual $D(Ku_k, f)$ becomes too small (i.e. one approximates the noisy data stronger than

Ku^*). Successful stopping criteria as the discrepancy principle are indeed based on comparing the residual with an estimate of the noise $D(Ku^*, f)$ and stop when $D(u_k, f)$ drops below this estimate.

In imaging a particularly interesting and quite related aspect of Bregman iterations is the scale behaviour. As mentioned above, with scale defined as above by properties of the regularization functional R , the Bregman iteration inserts finer and finer scales during its progress. In order not to miss certain scales it is obviously interesting to make small enough steps, which amounts to choosing α sufficiently large. For the limit of $\alpha \rightarrow \infty$ one can interpret the iteration as a backward Euler discretization (with timestep $\frac{1}{\alpha}$) of a flow, which has been called inverse scale space method by a reminiscence to so-called scale space methods in image processing, which exhibit the opposite scale behaviour (cf. [57, 59]). The inverse scale space flow is a solution of the differential inclusion

$$\partial_t p(t) \in -K^* \partial D(Ku(t), f), \quad p(t) \in \partial J(u(t)), \quad (40)$$

with initial value $u(0) = u_0$ such that $p(0) = 0 \in \partial R(u_0)$. It can be interpreted by a gradient flow for the subgradient p on a dual functional (cf. [17]) or as a doubly nonlinear evolution equation. For the latter we will give an explanation on the analysis in terms of Bregman distances related to the involved functionals in the next section, which is also the appropriate way to analyse the inverse scale space method.

An unexpected result is the behaviour of the inverse scale space flow for polyhedral functions such as the ℓ^1 -norm. Roughly speaking the polyhedral case means that for any $u \in X$ a subdifferential $\partial R(u)$ can be obtained via convex combinations of a finite number of elements (independent of u). It has been shown (cf. [20, 48]) that in such cases and $D(Ku, f) = \frac{1}{2} \|Ku - f\|^2$ the dynamics of the solution $u(t)$ is piecewise constant in time, i.e., quite far from a continuous flow, while the dynamics of the subgradients $p(t)$ is piecewise linear in time. Interestingly, the time steps t_k at which the solution changes can be computed explicitly, and the value of $u(t_k)$ is obtained by minimizing

$$\|Ku - f\|^2 \quad \text{subject to } p(t_k) \in \partial R(u).$$

This is particularly attractive in the case of sparse optimization with R being the ℓ^1 -norm, since the condition $p(t_k) \in \partial R(u)$ defines the sign of u and in particular the set of zeros. This means that the least-squares problems have to be solved on a rather small support, which is highly attractive for computational purposes (cf. [20]). Let us briefly explain the behaviour for $R : \mathbb{R}^N \rightarrow \mathbb{R}^+$ being the ℓ^1 -norm and some arbitrary differentiable functional G on the right-hand side, i.e.,

$$\partial_t p_i(t) = -\partial_{u_i} G(u(t)). \quad (41)$$

In this case the subdifferential is the multivalued sign of $u_i(t)$ and for $u_0 = p_0 = 0$ we obviously find $u_i(t) = 0$ for sufficiently small time since $|p_i(t)| < 1$, which holds for all i . Hence for $t < t_1$ with t_1 to be determined we find

$$\partial_t p_i(t) = -\partial_{u_i} G(0), \quad (42)$$

which can be integrated easily to

$$p_i(t_1) = -t_1 \partial_{u_i} G(0). \quad (43)$$

The key observation is that $u_i \neq 0$ for some i is only possible if $|p_i(t_1)| = 1$. This implies that the first time with possibly nonzero u is

$$t_1 = \frac{1}{\|\partial G(0)\|_\infty}. \quad (44)$$

At time t_1 the sign of all u_i is determined by $p_i(t_1)$ and one can check that a solution is obtained by minimizing

$$u(t_1) \in \arg \min_{u \in \mathbb{R}^N} G(u) \quad \text{subject to } p_i(t_1) \in \partial |u_i(t_1)|, \quad (45)$$

or in other words

$$u(t_1) \in \arg \min_{u \in \mathbb{R}^N} G(u) \quad \text{subject to } p_i(t_1) u_i(t_1) \geq |u_i(t_1)| \quad \forall i.$$

The optimality condition for the latter problem can be written as

$$\partial_{u_i} G(u(t_1)) + \lambda_i (q_i - p_i(t_1)) = 0, \quad q_i \in \partial |u_i(t_1)|. \quad (46)$$

for some $\lambda \in \mathbb{R}^N$ satisfying the complementarity conditions

$$\lambda_i \geq 0, \quad \lambda_i (p_i(t_1) u_i(t_1) - |u_i(t_1)|) = 0.$$

This implies $\partial_{u_i} G(u(t_1)) = 0$ if $u_i(t_1) \neq 0$, $\partial_{u_i} G(u(t_1)) \geq 0$ if $u_i(t_1) = 0$ and $p_i(t_1) = 1$, and $\partial_{u_i} G(u(t_1)) \leq 0$ if $u_i(t_1) = 0$ and $p_i(t_1) = -1$. This implies that we can find a time interval (t_1, t_2) such that

$$u(t) = u(t_1), \quad p(t) = p(t_1) - (t - t_1) \partial G(u(t_1))$$

is a solution, and t_2 is again defined as the minimal time where there exists i such that $|p_i(t_2)| = 1$ and $|p_i(t)| < 1$. Again, the solution at time t_2 is defined by a solution of the variational problem

$$u(t_2) \in \arg \min_{u \in \mathbb{R}^N} G(u) \quad \text{subject to } p_i(t_2) u_i(t_2) \geq |u_i(t_2)| \quad \forall i.$$

By an inductive procedure one obtains that the same kind of dynamics goes on for all t until it stops after finite time steps t_n at a minimizer of G .

As mentioned above the scale behaviour of the inverse scale space flow is highly attractive in image processing. In the polyhedral case there is a somehow exact decomposition into different scales by the steps made at times t_k . Indeed $\partial_t u$ is a sum of concentrated measures in time, and one may eliminate certain scales by leaving out the corresponding jump $u(t_k + \tau) - u(t_k - \tau)$. This observation leads the way to a much more general definition of filters from the inverse scale space method, which was discussed in [21]

$$\partial_t p(t) = f - u(t), \quad p(t) \in \partial R(u(t)). \quad (47)$$

A certain scale filter is defined by

$$F(f) = u_0 + \int_0^\infty w(t) d\partial_t u(t), \quad (48)$$

with measurable weights $w(t) \in [0, 1]$. In the case $w \equiv 1$ one simply obtains f , while certain scales can be damped out choosing $w(t) = 0$ for t in an appropriate interval. The design of filters for certain purpose is an ongoing subject of research.

4 Applications in Partial Differential Equations

In the following, we provide an overview of different aspects of partial differential equations, where Bregman distances are a useful tool. Unlike the case of inverse problems and image processing discussed above the notion of Bregman distance is not used widely in this field, and indeed most applications do not refer to this term or use it in a very hidden way. Our goal in the following section is to work out the basic ideas related to Bregman distances in a structured way, which sheds new light on many established techniques and hopefully also opens routes towards novel results. For this sake we employ a formal approach and avoid technicalities such as detailed function spaces, which of course can be worked out from existing literature.

4.1 Entropy Dissipation Methods for Gradient Systems

Entropy dissipation methods are a frequently used tool in partial differential equations (cf. [4, 41]), which is often based on using the logarithmic entropy

$$E(u) = \int_{\Omega} u(x) \log u(x) dx \quad (49)$$

as a Lyapunov functional, e.g., in diffusion equations (cf., e.g., [2–5, 24]), kinetic equations (cf., e.g., [4]), or fluid mechanics (cf., e.g., [56]). In particular in gradient

systems also different convex functionals are used regularly and in a structured way. The abstract form of a gradient system is

$$\partial_t u(t) = -L(u(t))E'(u(t)), \quad (50)$$

where $L(u)$ is a linear symmetric positive semi-definite operator on appropriate spaces and E a convex energy functional, which we assume differentiable for simplicity (similar treatment for non-differentiable convex functionals is possible by using subgradients, but beyond our scope). The entropy dissipation property can be verified by the straightforward computation

$$\frac{d}{dt}E(u(t)) = E'(u(t))\partial_t u(t) = -\langle E'(u(t)), L(u(t))E'(u(t)) \rangle \leq 0. \quad (51)$$

The negative of the right-hand side is frequently called entropy dissipation functional $D(u(t))$ and can be used to derive further quantitative information about the decay to equilibrium. A standard example (cf. [4, 24]) are nonlinear Fokker–Planck equations of the form

$$\partial_t u = \nabla \cdot (m(u)\nabla(e'(u) + V)) \quad (52)$$

on a domain $\Omega \subset \mathbb{R}^d$ with no-flux boundary conditions. Here, $e : \mathbb{R}^+ \rightarrow \mathbb{R}$ is a convex function, $m : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ a (potentially nonlinear) mobility function, and $V : \Omega \rightarrow \mathbb{R}$ an external potential. Recently also systems of Fokker–Planck equations as well as certain reaction–diffusion systems of the form

$$\partial_t u_i = D_i \Delta u_i + F_i(u_1, \dots, u_M), \quad i = 1, \dots, M \quad (53)$$

have been investigated with entropy dissipation techniques (cf. [29, 42, 45]).

The major purpose of entropy dissipation techniques is to obtain qualitative or ideally quantitative results about the decay to equilibrium for transient solutions. An equilibrium solution u_∞ is a minimizer of E on a convex set K , to which also the transient solution $u(t)$ belongs for all t . An example is the Fokker–Planck equation with linear mobility $m(u) = u$, where K is the set of nonnegative integrable functions with prescribed mean value. Hence, u_∞ satisfies

$$E'(u_\infty)(u - u_\infty) \geq 0 \quad \forall u \in K. \quad (54)$$

If further the operator $L(u)$ is such that

$$L(u)E'(u_\infty) = 0 \quad \forall u \in K, \quad (55)$$

which is indeed the case for the typical examples, then one can rewrite the gradient system as

$$\partial_t u(t) = -L(u(t))(E'(u(t)) - E'(u_\infty)). \quad (56)$$

Hence, the right-hand side is expressed in a difference of energy gradients for the transient and equilibrium solution. In a similar way, the entropy dissipation can be rewritten in terms of a distance between those and the Bregman distance (usually called relative entropy) plays a key role for this purpose. One observes that

$$\begin{aligned} \frac{d}{dt} D_E^{E'(u_\infty)}(u(t), u_\infty) &= E'(u(t)) \partial_t u(t) - E'(u_\infty) \partial_t u(t) \\ &= -\langle E'(u(t)) - E'(u_\infty), L(u(t))(E'(u(t)) - E'(u_\infty)) \rangle \\ &=: -F(u(t), u_\infty). \end{aligned}$$

Of course, the above computation holds for smooth solutions only, for weak solutions one can usually derive the time-integrated version

$$D_E^{E'(u_\infty)}(u(t), u_\infty) + \int_s^t F(u(\tau)) d\tau \leq D_E^{E'(u_\infty)}(u(s), u_\infty). \quad (57)$$

The above computation shows that entropy dissipation can be rephrased as the decrease of the Bregman distance between stationary and transient solution. We notice that the use of the Bregman distance is not essential in this computation, but the understanding of this structure can be quite beneficial, in particular if one wants to use dual variables, the so-called entropy variables

$$\varphi(t) = E'(u(t)), \quad \varphi_\infty = E'(u_\infty). \quad (58)$$

The entropy variable φ solves the system

$$\partial_t (E'(\varphi(t))) = -L((E^*)'(\varphi(t)))\varphi(t), \quad (59)$$

where E^* is the convex conjugate of E . When analysing the dual flow (59) a dissipation property can now be derived immediately using relation (10). Thus, we obtain a dual entropy dissipation of the form

$$\frac{d}{dt} D_{E^*}^{u(t)}(\varphi_\infty, \varphi(t)) = -\langle \varphi(t) - \varphi_\infty, L((E^*)'(\varphi(t))) (\varphi(t) - \varphi_\infty) \rangle. \quad (60)$$

The duality relation is particularly interesting for constructing approximations in terms of the entropy variables, as, e.g., carried out for degenerate cross-diffusion systems in (cf. [19, 40, 62]).

4.2 Lyapunov Functionals for Gradient Systems Out of Equilibrium

The appropriate use of Bregman distances seems to be less explored, but maybe even more crucial for the derivation of Lyapunov functionals if gradient systems are perturbed out of equilibrium. The simplest example is the linear Fokker–Planck equation with non-potential force as investigated in [5]

$$\partial_t u = \nabla \cdot (\nabla u + uF) \quad \text{in } \Omega \times \mathbb{R}^+, \quad (61)$$

supplemented by no-flux boundary conditions

$$(\nabla u + uF) \cdot n = 0 \quad \text{on } \partial\Omega \times \mathbb{R}^+. \quad (62)$$

If the vector field F is not the gradient of some potential function, then a stationary solution cannot be constructed as the minimizer of an entropy functional. However, the existence and uniqueness of a stationary solution can be shown under quite general assumptions on F (cf. [31]). In a form similar to gradient flows (cf. [1, 53]) we write (61) as

$$\partial_t u = \nabla \cdot (u(\nabla e'(u) + F)), \quad e(u) = u \log u + 1 - u, \quad (63)$$

which suggests to further investigate distances based on the entropy functional

$$E(u) = \int_{\Omega} e(u) \, dx = \int_{\Omega} (u \log u - u + 1) \, dx. \quad (64)$$

The dissipation of the relative entropy can be computed via

$$\begin{aligned} \frac{d}{dt} D_E^{E'(u_{\infty})}(u(t), u_{\infty}) &= \int_{\Omega} (e'(u(t)) - e'(u_{\infty})) \partial_t u(t) \, dx \\ &= \int_{\Omega} (e'(u(t)) - e'(u_{\infty})) \nabla \cdot u(\nabla(e'(u(t)) \\ &\quad - e'(u_{\infty})) + \nabla e'(u_{\infty}) + F) \, dx \\ &= - \int_{\Omega} u |\nabla(e'(u(t)) - e'(u_{\infty}))|^2 \, dx \\ &\quad + \int_{\Omega} (e'(u(t)) - e'(u_{\infty})) \nabla \cdot u(t) (\nabla e'(u_{\infty}) + F) \, dx, \end{aligned}$$

where we have used the no-flux boundary conditions

$$(\nabla e(u(t)) + F) \cdot n = (\nabla e(u_{\infty}) + F) \cdot n = 0 \quad \text{on } \partial\Omega \times \mathbb{R}^+$$

in order to apply integration by parts in the first term on the right-hand side. The second term is simplified via

$$\begin{aligned}
 & \nabla \cdot (u(\nabla e'(u_\infty) + F)) \\
 &= \nabla \cdot \left(\frac{u(t)}{u_\infty} u_\infty (\nabla e'(u_\infty) + F) \right) \\
 &= u_\infty \nabla \left(\frac{u(t)}{u_\infty} \right) \cdot (\nabla e'(u_\infty) + F) \\
 &= u_\infty \nabla \exp(e'(u(t)) - e'(u_\infty)) \cdot (\nabla e'(u_\infty) + F) \\
 &= u_\infty \exp(e'(u(t)) - e'(u_\infty)) \nabla(e'(u(t)) - e'(u_\infty)) \cdot (\nabla e'(u_\infty) + F).
 \end{aligned}$$

With Ψ satisfying $\Psi'(z) = z \exp(z)$ we can further write

$$\begin{aligned}
 & \int_{\Omega} (e'(u(t)) - e'(u_\infty)) \nabla \cdot u(t) (\nabla e'(u_\infty) + F) \, dx = \\
 & \int_{\Omega} \nabla \Psi(e'(u(t)) - e'(u_\infty)) \cdot u_\infty (\nabla e'(u_\infty) + F) \, dx = 0,
 \end{aligned}$$

which can be seen again through integration by parts. Hence, we finally obtain the decrease of the Bregman distance via

$$\frac{d}{dt} D_E^{E'(u_\infty)}(u(t), u_\infty) = - \int_{\Omega} u |\nabla(e'(u(t)) - e'(u_\infty))|^2 \, dx, \tag{65}$$

and the logarithmic Sobolev inequality (cf. [3]) implies exponential convergence to equilibrium.

Another example are boundary-driven nonlinear Fokker–Planck equations

$$\partial_t u = \nabla \cdot (\nabla m(u) + m(u)F) \quad \text{in } \Omega \times \mathbb{R}^+, \tag{66}$$

considered in [10] with Dirichlet boundary conditions

$$u = g \quad \text{on } \partial\Omega \times \mathbb{R}^+. \tag{67}$$

We mention that an analogous analysis holds in the case of no-flux boundary conditions (in which case we have a direct generalization of the nonsymmetric Fokker–Planck equation above) or mixed Dirichlet and no-flux boundary conditions. Bodineau et al. [10] construct Lyapunov functionals of the form

$$H(u, u_\infty) = \int_{\Omega} \int_{u_\infty(x)}^{u(x,t)} \Phi' \left(\frac{m(s)}{m(u_\infty(x))} \right) \, ds \, dx, \tag{68}$$

where Φ is a nonnegative function with unique minimum at zero. Such a construction seems far from being intuitive, but it becomes much more clear for Φ being the logarithmic entropy, i.e., $\Phi'(t) = \log t$. In this case the Lyapunov functional becomes

$$H(u, u_\infty) = \int_{\Omega} \int_{u_\infty(x)}^{u(x,t)} \log m(s) - \log m(u_\infty(x)) \, ds \, dx, \quad (69)$$

and with a function e such that $e'(s) = \log m(s)$, we further obtain

$$H(u, u_\infty) = \int_{\Omega} (e(u(x,t)) - e(u_\infty(x)) - e'(u_\infty(x))(u(x,t) - u_\infty(x))) \, ds \, dx, \quad (70)$$

which is nothing but the Bregman distance for the entropy functional

$$E(u) = \int_{\Omega} e(u) \, dx, \quad \text{with } e'(u) = \log m(u). \quad (71)$$

Since Eq. (66) can be written as

$$\partial_t u = \nabla \cdot (m(u)(\nabla \log m(u) + F)), \quad \text{in } \Omega \times \mathbb{R}^+, \quad (72)$$

the above form of E is also a natural choice. The detailed computations for the entropy dissipation are indeed completely analogous to the case of the linear Fokker–Planck equation, the crucial point appears to be the logarithmic relation between entropy derivatives $e'(u)$ and mobilities $m(u)$.

4.3 Doubly Nonlinear Evolution Equations

A generalization of gradient systems are doubly nonlinear evolution equations with a gradient structure either of the form

$$\partial_t p(t) \in -\partial G(u(t)), \quad p(t) \in \partial F(u(t)) \quad (73)$$

or as

$$\partial F(\partial_t u) + \partial G(u(t)) \ni 0. \quad (74)$$

The best studied case, which is also the one where both coincide, corresponds to $F(u) = \frac{1}{2} \|u\|^2$ for a norm in a Hilbert space, which yields the classical gradient flow

$$\partial_t u(t) \in -\partial G(u(t)). \quad (75)$$

We have seen a system in the form (73) already above in the inverse scale space method, while the form (74) appears frequently in mechanical problems (cf., e.g., [46] and the references therein). There is indeed a duality relation for (73) and (74). Starting from (73) we obtain $u(t) \in \partial G^*(-\partial_t p(t)) \cap \partial F^*(p(t))$, respectively $-u(t) \in \partial G^*(\partial_t p(t))$ if G satisfies a symmetry condition around zero. This yields

$$\partial G^*(\partial_t p(t)) + \partial F^*(p(t)) \neq 0,$$

the analogue of (74).

Doubly nonlinear evolution equations have recently been investigated extensively, and in particular tools from convex analysis have been employed (cf. [46]). Here we add our Bregman distance point of view to derive estimates for such equations. Let us start with a straightforward computation on the change of the time derivative of the Bregman distance:

Lemma 1. *Let F be differentiable and u a solution of (73). Then*

$$\frac{d}{dt} D_F^{p(t)}(v, u(t)) = -\langle \partial_t p(t), v - u(t) \rangle \leq G(v) - G(u(t)).$$

This can be used to quantify the distance of $u(t)$ to a minimizer of G :

Corollary 1. *Let F be differentiable, u_∞ a minimizer of G , and u a solution of (73). Then*

$$\frac{d}{dt} D_F^{p(t)}(u_\infty, u(t)) + D_G^0(u(t), u_\infty) \leq 0. \quad (76)$$

Since it is straightforward to see

$$\frac{d}{dt} D_G^0(u(t), u_\infty) = \frac{d}{dt} G(u(t)) \leq 0 \quad (77)$$

we see after integrating (76) in time

$$\begin{aligned} D_F^{p(t)}(u_\infty, u(t)) + t D_G^0(u(t), u_\infty) &\leq D_F^{p(t)}(u_\infty, u(t)) \\ &+ \int_0^t D_G^0(u(s), u_\infty) ds \leq D_F^{p(0)}(u_\infty, u(0)), \end{aligned}$$

leading to linear decay of the Bregman distance:

Theorem 2. *Let F be differentiable, u_∞ a minimizer of G , and u a solution of (73). Then*

$$D_G^0(u(t), u_\infty) \leq \frac{1}{t} D_F^{p(0)}(u_\infty, u(0)). \quad (78)$$

4.4 Error Estimates for Nonlinear Elliptic Problems

We finally turn our attention to the analysis of discretization methods for nonlinear elliptic problems such as the p -Laplace equation. Such elliptic problems are optimality conditions of some energy functional of the form

$$E(u) = J(u) - \langle f, u \rangle, \quad (79)$$

where J is a convex functional on a Banach space X , typically a Sobolev space of first order derivatives. The elliptic differential equation (or more general differential inclusion) is the optimality condition

$$p = f, \quad p \in \partial J(u) \quad (80)$$

A canonical example is the p -Laplace equation

$$-\nabla \cdot (|\nabla u|^{p-2} \nabla u) = f, \quad (81)$$

which is related to the functional

$$J(u) = \frac{1}{p} \int_{\Omega} |\nabla u(x)|^p dx. \quad (82)$$

For variational discretizations of such problems the Bregman distance appears to be a quite useful tool, which is still not fully exploited. In many approaches the Bregman distance is used in a hidden way and strict convexity is used to obtain an estimate in terms of the underlying norms (with potentially suboptimal constants, however). For the p -Laplace equation such an approach is carried out in [30]. Again in the limiting case $p = 1$ related to total variation minimization the Bregman distance is even more crucial and appears, e.g., in [6]. Here we briefly sketch the obvious role of Bregman distances in Galerkin discretizations of the form

$$E(u) \rightarrow \min_{u \in X_h}, \quad (83)$$

where X_h is a finite-dimensional subspace of X , e.g., constructed by finite elements.

Let us start by pointing out the basic structure of error estimates for Galerkin methods in the linear case related to the minimization of a positive definite quadratic form

$$J(u) = B(u, u), \quad (84)$$

where $B : X \times X \rightarrow \mathbb{R}$ is a bounded and coercive bilinear form. The optimality condition in weak form is given by

$$B(u, v) = \langle f, v \rangle \quad \forall v \in X, \quad (85)$$

and the Galerkin discretization yields a solution $u_h \in X_h$ of

$$B(u_h, v) = \langle f, v \rangle \quad \forall v \in X_h. \quad (86)$$

Error estimates for such discretizations are obtained in two steps: first the error between u and u_h is estimated by the projection error to the subspace X_h and then the projection error is estimated, e.g., via the interpolation error. The crucial property for the first step is the so-called *Galerkin orthogonality*

$$B(u - u_h, v) = 0 \quad \forall v \in X_h, \quad (87)$$

which implies

$$B(u - u_h, u - u_h) = B(u - u_h, u - v) \quad \forall v \in X_h, \quad (88)$$

and by the Cauchy–Schwarz inequality for the positive definite bilinear form B

$$B(u - u_h, u - u_h) \leq B(u - v, u - v) \quad \forall v \in X_h. \quad (89)$$

In other words u_h is the projection of u on the subspace X_h , when the (squared) norm induced by B is used as a distance measure.

Since the term $B(u - v, u - v)$ above is just the Bregman distance related to quadratic functional J one might think of an analogous property in the case of nonquadratic J , when the Bregman projection is used. Indeed, we can derive such a relation in the case of arbitrary convex J . For this sake let again u be a minimizer of E and u_h a minimizer of E constrained to the subspace X_h . Then we have $f \in \partial J(u)$ and thus, since u_h minimizes E on X_h , we have for all $v \in X_h$

$$\begin{aligned} D_J^f(u_h, u) &= J(u_h) - J(u) - \langle f, u_h - u \rangle \\ &= E(u_h) - J(u) + \langle f, u \rangle \\ &\leq E(v) - J(u) + \langle f, u \rangle. \end{aligned}$$

Rewriting the last term we hence obtain the Bregman projection property

$$D_J^f(u_h, u) \leq D_J^f(v, u), \quad \forall v \in X_h. \quad (90)$$

This observation opens a way to analyse Galerkin methods for such nonlinear problems in the same way as in the linear case, the key step to be developed for specific problems and specific discretizations (X_h) is the estimation of the Bregman projection error.

Note again the role of the Bregman distance for error estimation: The one-sided distance $D_J^f(u_h, u)$ is particularly suitable for the estimation of a-priori errors as above, while a-posteriori error estimation should rather be based on the distance $D_J^{p_h}(u, u_h)$ with $p_h \in \partial J(u_h)$. We have by the minimizing property of u

$$\begin{aligned} D_J^{p_h}(u, u_h) &= J(u) - J(u_h) - \langle p_h, u - u_h \rangle \\ &= E(u) - E(u_h) + \langle p_h - f, u_h - u \rangle \\ &\leq \langle p_h - f, u_h - u \rangle. \end{aligned}$$

Using the duality relation $u \in \partial J^*(f)$, this could be further estimated to the full a-posteriori estimate

$$D_J^{p_h}(u, u_h) \leq \langle p_h - f, u_h \rangle + J^*(2f - p_h) - J^*(f). \quad (91)$$

For practical purposes the above abstract estimate is not useful in most cases, since computing the adjoint J^* means to solve a nonlinear partial differential equation as well, which might be as difficult as the original one. However, the general strategy can be exploited together with specific properties of the functional J and the subspace X_h . In particular for gradient energies of the form

$$J(u) = \int_{\Omega} j(\nabla u) \, dx \quad (92)$$

one can derive alternative versions using only the convex conjugate j^* , which is significantly easier to compute.

5 Further Developments

In this final section we discuss some aspects of Bregman distances that came up recently and will potentially have strong further impact, in particular we will explore some developments related to probability.

5.1 Uncertainty Quantification in Inverse Problems

Since Bregman distances appear to be a suitable tool for estimates in certain nonlinear deterministic problems, it seems natural to exploit them also in the stochastic counterparts of such problems. The obvious measure for error estimates is then the expected value of the Bregman distance with respect to the stochastic quantity. Such approaches have been used successfully in particular in statistical

inverse problems (cf., e.g., [58]), which we also want to discuss in the following. In order to avoid technicalities we restrict ourselves to a purely finite-dimensional setup.

Consider the inverse problem $Ku = f$, where $K : \mathbb{R}^N \rightarrow \mathbb{R}^M$ and the data are generated from a true solution u^* with additive Gaussian noise, i.e.,

$$f = Ku^* + \sigma n, \quad (93)$$

with n a Gaussian random variable with zero mean and covariance matrix I_M . Let again R be a convex regularization functional and u_α a solution of the variational problem

$$J(u) = \frac{1}{2\sigma^2} \|(Ku - f)\|^2 + \alpha J(u) \rightarrow \min_{u \in \mathbb{R}^N}. \quad (94)$$

Then u_α satisfies the optimality condition

$$\frac{1}{\sigma^2} K^* K(u_\alpha - u^*) + \alpha p_\alpha = \frac{1}{\sigma^2} K^* n, \quad p_\alpha \in \partial R(u_\alpha), \quad (95)$$

which implies $p_\alpha = K^* w_\alpha$. Now assume u^* satisfies the source condition (34), then we have

$$K(u_\alpha - u^*) + \alpha \sigma^2 (w_\alpha - w^*) = n - \alpha \sigma^2 w^*.$$

Taking the squared norm and subsequently expectation with respect to w in this identity we obtain

$$\begin{aligned} 2\alpha\sigma^2 E[D_R^{p_\alpha, p^*}(u_\alpha, u^*)] &\leq E[\|K(u_\alpha - u^*)\|^2 + 2\alpha\sigma^2 D_R^{p_\alpha, p^*}(u_\alpha, u^*) + \alpha^2 \sigma^4 \|w_\alpha - w^*\|^2] \\ &= E[\|n - \alpha\sigma^2 w^*\|^2] \\ &= E[\|n\|^2] + \alpha^2 \sigma^4 \|w^*\|^2 = \sigma^2 M + \alpha^2 \sigma^4 \|w^*\|^2. \end{aligned}$$

Thus, the expected error in the Bregman distance is estimated by

$$E[D_R^{p_\alpha, p^*}(u_\alpha, u^*)] \leq \frac{M}{2\alpha} + \frac{\alpha\sigma^2}{2} \|w^*\|^2. \quad (96)$$

We notice that the above approach not only yields an estimate of the Bregman distance, but also indeed an exact value for the sum of three error measures, in addition to the Bregman distance also the residual error as well as the error in the source space (related to $w_\alpha - w^*$). Usually the latter is the largest of the three, so one needs to expect a blow up of this term as $M \rightarrow \infty$ if α is not increasing as M . If one is interested in the first two terms only, one can simply use a duality product with

$u_\alpha - u^*$ in (95) and subsequently estimate the expected value of the right-hand side in a different way, which may lead to robust estimates in terms of M , respectively, estimates that can be carried out for infinite-dimensional white noise.

An application of Bregman distances in Bayesian modelling was recently investigated in [13], considering frequently used posterior densities of the form

$$\pi(u|f) \sim e^{-\frac{\|Ku-f\|^2}{2\sigma^2} - \alpha R(u)}, \quad (97)$$

where again R is a convex and Lipschitz continuous functional on \mathbb{R}^N (generalizations to posterior distributions in infinite-dimensional spaces where further studied in [39]). It has been shown that the posterior can be centred around the so-called maximum a-posteriori probability (MAP) estimate \hat{u} , which maximizes $p(u|f)$, in the form

$$\pi(u|f) \sim e^{-\frac{\|Ku-K\hat{u}\|^2}{2\sigma^2} - \alpha D_R^\hat{p}(u, \hat{u})}. \quad (98)$$

Based on the observation

$$\langle s, u - \hat{u} \rangle = \frac{\|Ku - K\hat{u}\|^2}{\sigma^2} + \alpha \langle p - \hat{p}, u - \hat{u} \rangle. \quad (99)$$

for $p \in \partial R(u)$ and

$$s = \frac{1}{\sigma^2} K^*(Ku - f) + \alpha p \in \partial(-\log \pi(u|f)),$$

a Bayes cost of the form

$$\Gamma(v) = \mathbf{E}_{p(u|f)} \left[\frac{\|Ku - Kv\|^2}{\sigma^2} + \alpha \langle q - \hat{p}, v - \hat{u} \rangle \right] \quad (100)$$

has been introduced for $q \in \partial R(v)$ (note that selection of $p \in R(u)$ is only needed on a set of zero measure due to Rademacher's theorem). A simple integration by parts argument then shows that the MAP-estimate \hat{u} is a minimizer of the Bayes cost, which is a quite natural choice compared to the highly degenerate cost usually used to characterize MAP estimates (cf. [43]). A direct consequence is the fact that the MAP estimate has smaller Bregman distance in expectation than the frequently used conditional mean estimate, hence one obtains a theoretical argument explaining the success of MAP estimates in practice.

5.2 Bregman Distances and Optimal Transport

Bregman distances can be used also as a cost in optimal transport, which has been investigated in [23] for a convex and differentiable functional J on \mathbb{R}^N . Given two probability measures μ and ν , an optimal transport plan is a probability measure γ on $\mathbb{R}^N \times \mathbb{R}^N$ with marginals μ and ν minimizing the functional

$$F(\gamma) = \int_{\mathbb{R}^N \times \mathbb{R}^N} D_J^{(u)}(v, u) d\gamma(v, u). \quad (101)$$

The resulting optimal value of F can be interpreted as a transport distance between the measures μ and ν .

Besides the important question of well-posedness solved in (cf. [23]) there are several interesting problems such as the existence of transport maps under certain condition (i.e. concentration of γ on a set described by the graph of a map $T: \mathbb{R}^N \rightarrow \mathbb{R}^N$) as well as relations to uncertainty quantification. A first example is the Bayes cost approach described in the previous section, which can indeed be interpreted as the transport distance between the posterior distribution and a measure concentrated at the MAP estimate. This motivates further research in the future, an obvious next step might be to estimate distances between different posterior distributions in transport distances related to Bregman distances.

A different use of Bregman distances in optimal transport was recently made in [7] for the solution of Monge–Kantorovich formulations in optimal transport. They consider entropic regularizations of the problem, i.e., for $\epsilon > 0$ they minimize a discrete version of

$$F_\epsilon(\gamma) = \int_{\mathbb{R}^N \times \mathbb{R}^N} C(v, u) d\gamma(v, u) + \epsilon E(\gamma), \quad (102)$$

where E is the entropy

$$E(\gamma) = \int_{\mathbb{R}^N \times \mathbb{R}^N} \log \left(\frac{d\gamma}{d\mathcal{L}} \right) d\gamma(v, u), \quad (103)$$

where $\frac{d\gamma}{d\mathcal{L}}$ is the Radon–Nikodym derivative with respect to the Lebesgue measure. The key observation is that the minimization of F_ϵ can be rewritten equivalently as the minimization of the Kullback–Leibler divergence, i.e., the Bregman distance related to E , between γ and the Gibbs measure φ_ϵ with density $e^{-C/\epsilon}$

$$D_E(\gamma, \varphi_\epsilon) \rightarrow \min_\gamma, \quad (104)$$

which transforms the problem into a Bregman projection problem of the Gibbs density onto the set of plans with given marginals, which can be computed much more efficiently than the original transport control problem. Note that the general procedure can be carried out as well with an arbitrary convex functional whose

domain are positive densities, the corresponding Gibbs density is then to be defined as $\varphi_\epsilon = (E^*)'(-C/\epsilon)$. A particular computational advantage of the logarithmic entropy is the fact that iterative Bregman projections can be computed explicitly and realized with low complexity, in the discrete sets it only needs multiplications and scalar products of diagonal matrices with the matrix discretizing the Gibbs measure (cf. [7] for further details).

5.3 Infimal Convolution of Bregman Distances

Infimal convolution of convex functionals become popular recently in image processing in order to combine favourable properties of certain regularization functionals, e.g., total variation and higher-order versions thereof. A quite unexplored topic is the infimal convolution of Bregman distances, however. Since they are convex functionals of the first variable one may consider the infimal convolution

$$[D_R^{p_1}(\cdot, u_1) \square D_R^{p_2}(\cdot, u_2)](u) = \inf_{v \in X} [D_R^{p_1}(u - v, u_1) + D_R^{p_2}(v, u_2)], \quad (105)$$

with an obvious extension to more than two values.

Of particular interest in imaging applications appears to be the case of $p_2 = -p_1$ and $u_2 = -u_1$ for a one-homogeneous functional such as total variation. The latter was used to obtain a regularization functional enforcing partly equal edge sets (REF colorbregman). While minimizing the Bregman distance for total variation strongly favours edge sets with jumps of equal sign (see also the discussion related to orientation for one-homogeneous functionals in Sect. 2.4), the infimal convolution of Bregman distances eliminates this part and hence measures differences in edge sets rather than jumps of the same sign. A further study of theoretical properties as well as applications of such kind of infimal convolution of Bregman distances remains an interesting property for future research. Obvious candidates are problems in compressed sensing where one first of all aims at obtaining the correct support of the solution rather than the sign.

Acknowledgements This work was partially supported by ERC via Grant EU FP 7 - ERC Consolidator Grant 615216 LifeInverse and by the German Science Foundation DFG via BU 2327/6-1 and EXC 1003 Cells in Motion Cluster of Excellence, Münster, Germany

References

1. Ambrosio, L., Gigli, N., Savare, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhaeuser, Basel/Boston/Berlin (2005)
2. Arnold, A., Unterreiter, A.: Entropy decay of discretized Fokker-Planck equations I: temporal semidiscretization. *Comput. Math. Appl.* **46**, 1683–1690 (2003)

3. Arnold, A., Markowich, P., Toscani, G., Unterreiter, A.: On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Commun. Partial Differ. Equ.* **26**, 43–100 (2001)
4. Arnold, A., Carrillo, J.A., Desvillettes, L., Dolbeault, J., Jüngel, A., Lederman, C., Markowich, P., Toscani, G., Villani, C.: Entropies and equilibria of many-particle systems: an essay on recent research. *Mon. Hefte Math.* **142**, 35–43 (2004)
5. Arnold, A., Carlen, E., Ju, Q.: Large-time behavior of non-symmetric Fokker-Planck type equations. *Commun. Stoch. Anal.* **2**, 153–175 (2008)
6. Bartels, S.: Error control and adaptivity for a variational model problem defined on functions of bounded variation. *Math. Comput.* **84**, 1217–1240 (2015)
7. Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., Peyre, G.: Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.* **37**, A1111–A1138 (2015)
8. Benning, M., Burger, M.: Error estimates for general fidelities. *Electron. Trans. Numer. Anal.* **38**, 44–68 (2011)
9. Benning, M., Burger, M.: Ground states and singular vectors of convex variational regularization methods. *Methods Appl. Anal.* **20**, 295–334 (2013)
10. Bodineau, T., Lebowitz, J., Mouhot, C., Villani, C.: Lyapunov functionals for boundary-driven nonlinear drift-diffusion equations. *Nonlinearity* **27**, 2111 (2014)
11. Bregman, L.M.: The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **7**, 200–217 (1967)
12. Brune, C., Sawatzky, A., Burger, M.: Primal and dual Bregman methods with application to optical nanoscopy. *Int. J. Comput. Vis.* **92**, 211–229 (2011)
13. Burger, M., Lucka, F.: Maximum-a-posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. *Inverse Prob.* **30**, 114004 (2014)
14. Burger, M., Osher, S.: Convergence rates of convex variational regularization. *Inverse Prob.* **20**, 1411–1421 (2004)
15. Burger, M., Osher, S.: A guide to the TV zoo. In: Burger, M., Osher, S. (eds.) *Level Set and PDE-Based Reconstruction Methods in Imaging*. Springer, Berlin (2013)
16. Burger, M., Gilboa, G., Osher, S., Xu, J.: Nonlinear inverse scale space methods. *Commun. Math. Sci.* **4**, 179–212 (2006)
17. Burger, M., Frick, K., Osher, S., Scherzer, O.: Inverse total variation flow. *Multiscale Model. Simul.* **6**, 366–395 (2007)
18. Burger, M., Resmerita, E., He, L.: Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing* **81**, 109–135 (2007)
19. Burger, M., Di Francesco, M., Pietschmann, J.F., Schlake, B.: Nonlinear cross-diffusion with size exclusion. *SIAM J. Math. Anal.* **42**, 2842–2871 (2010)
20. Burger, M., Moeller, M., Benning, M., Osher, S.: An adaptive inverse scale space method for compressed sensing. *Math. Comput.* **82**, 269–299 (2013)
21. Burger, M., Eckardt, L., Gilboa, G., Moeller, M.: Spectral representation of one-homogeneous functionals. In: Aujol, J.F., Niholovq, M., Pzpadikis, N. (eds.) *Scale Space and Variation Methods in Computer Vision*. Springer, Berlin, Heidelberg, 16–27
22. Cai, J., Osher, S., Shen, Z.: Linearized Bregman iterations for compressed sensing. *Math. Comput.* **78**, 1515–1536 (2009)
23. Carlier, G., Jimenez, C.: On Monge’s problem for Bregman-like cost functions. *J. Convex Anal.* **14**, 647–655 (2007)
24. Carrillo, J.A., Jüngel, A., Markowich, P., Toscani, G., Unterreiter, A.: Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. *Mon. Hefte Math.* **133**, 1–82 (2001)
25. Censor, Y., Lent, A.: An iterative row-action method for interval convex programming. *J. Optim. Theory Appl.* **34**, 321–353 (1981)
26. Censor, Y., Zenios, S.: *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, Oxford (1998)

27. Chavent, G., Kunisch, K.: Regularization of linear least squares problems by total bounded variation. *ESAIM: Control Optim. Calc. Var.* **2**, 359–376 (1997)
28. Choi, K., Fahimian, B.P., Li, T., Suh, T.S., Lei, X.: Enhancement of four-dimensional cone-beam computed tomography by compressed sensing with Bregman iteration. *J. X-Ray Sci. Technol.* **21**, 177–192 (2013)
29. Di Francesco, M., Fellner, K., Markowich, P.: The entropy dissipation method for inhomogeneous reaction-diffusion systems. *Proc. R. Soc. A* **464**, 3272–3300 (2008)
30. Diening, L., Kreuzer, C.: Linear convergence of an adaptive finite element method for the p-Laplacian equation. *SIAM J. Numer. Anal.* **46**, 614–638 (2008)
31. Droniou, J., Vazquez, J.L.: Noncoercive convection-diffusion elliptic problems with Neumann boundary conditions. *Calc. Var. Partial Differ. Equ.* **34**, 413–434 (2009)
32. Ekeland, I., Temam, R.: *Convex Analysis and Variational Problems*. SIAM, Philadelphia (1999)
33. Engl, H., Hanke-Bourgeois, M., Neubauer, A.: *Regularization of Inverse Problems*. Kluwer, Dordrecht (1996)
34. Flemming, J.: Theory and examples of variational regularization with non-metric fitting functionals. *J. Inverse III-Posed Probl.* **18**, 677–699 (2010)
35. Goldstein, T., Osher, S.: The split Bregman method for l_1 regularized problems. *SIAM J. Imaging Sci.* **2**, 323–343 (2008)
36. Grasmair, M.: Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems* **11**, 115014 (2010)
37. Grasmair, M.: Variational inequalities and higher order convergence rates for Tikhonov regularisation on Banach spaces. *J. Inverse III-Posed Probl.* **21**, 379–394 (2013)
38. Hein, T.: Tikhonov regularization in Banach spaces - improved convergence rates results. *Inverse Prob.* **25**, 035002 (2009)
39. Helin, T., Burger, M.: Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. Preprint arXiv:1412.5816 (2015)
40. Jüngel, A.: The boundedness-by-entropy method for cross-diffusion systems. *Nonlinearity* **28**, 1963–2001 (2015)
41. Jüngel, A., Matthes, D.: Entropiemethoden für nichtlineare partielle Differentialgleichungen. *Int. Math. Nachr.* **209**, 1–14 (2008)
42. Jüngel, A., Stelzer, I.V.: Entropy structure of a cross-diffusion tumor-growth model. *Math. Mod. Meth. Appl. Sci.* **22**, 1250009, (2012)
43. Kaipio, J., Somersalo, E.: *Statistical and Computational Inverse Problems*. Springer, New York (2005)
44. Kiwiell, K.C.: Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.* **35**, 1142–1168 (1997)
45. Mielke, A., Haskovec, J., Markowich, P.A.: On uniform decay of the entropy for reaction-diffusion systems. *J. Dyn. Differ. Equ.* **27**, 897–928 (2015)
46. Mielke, A., Rossi, R., Savare, G.: Nonsmooth analysis of doubly nonlinear evolution equations. *Calc. Var. Partial Differ. Equ.* **46**, 253–310 (2013)
47. Moeller, M.: Multiscale methods for polyhedral regularizations and applications in high dimensional imaging. Ph.D. thesis, University of Muenster (2012)
48. Moeller, M., Burger, M.: Multiscale methods for polyhedral regularizations. *SIAM J. Optim.* **23**, 1424–1456 (2013)
49. Müller, J., Brune, C., Sawatzky, A., Kösters, T., Schäfers, K.P., Burger, M.: Reconstruction of short time PET scans using Bregman iterations. In: *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 2383–2385. IEEE, New York (2011)
50. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. *SIAM Multiscale Model. Simul.* **4**, 460–489 (2005)
51. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. *Commun. Partial Differ. Equ.* **26**, 101–174 (2001)
52. Pöschl, C.: An overview on convergence rates for Tikhonov regularization methods for non-linear operators. *J. Inverse III-posed Probl.* **17**, 77–83 (2009)

53. Reid, M.: Meet the Bregman divergences. <http://mark.reid.name/blog/meet-the-bregman-divergences.html> (2013)
54. Resmerita, E.: Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems* **21**, 1303 (2005)
55. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenom.* **60**, 259–268 (1992)
56. Saint-Raymond, L.: Convergence of solutions to the Boltzmann equation in the incompressible Euler limit. *Arch. Ration. Mech. Anal.* **166**, 47–80 (2003)
57. Scherzer, O., Groetsch, C.: Inverse scale space theory for inverse problems. In: *Scale-Space and Morphology in Computer Vision*, pp. 317–325. Springer, Berlin/Heidelberg (2001)
58. Werner, F., Hohage, T.: Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems* **28**, 104004 (2012)
59. Witkin, A.P.: Scale-space filtering: a new approach to multi-scale description. In: *IEEE International Conference on ICASSP'84 Acoustics, Speech, and Signal Processing*, vol. 9, pp. 150–153 (1984)
60. Xu J., Osher, S.: Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. *IEEE Trans. Image Process.* **16**, 534–544 (2007)
61. Yin, W.: Analysis and generalizations of the linearized Bregman method. *SIAM J. Imaging Sci.* **3**, 856–877 (2010)
62. Zamponi, N., Jüngel, A.: Analysis of degenerate cross-diffusion population models with volume filling, TU Vienna. Preprint (2015)
63. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM J. Imaging Sci.* **3**, 253–276 (2010)