

Color Bregman TV

Michael Moeller^{*1}, Eva-Maria Brinkmann^{1,2}, Martin Burger², and Tamara Seybold¹

¹Arnold & Richter Cine Technik GmbH & Co. Betriebs KG, Image Science
Research and Development, Munich
²Westfälische Wilhelms-Universität Münster, Department for Computational
and Applied Mathematics

October 11, 2013

Abstract

In this paper we present a novel iterative procedure for multichannel image and data reconstruction using Bregman distances. With the motivation that multiple channels sharing a common subgradient with respect to a suitable regularization implies desirable properties such as a common edge set (and a common direction of the normals to the level lines) in the case of the total variation (TV), we propose to determine each iterate by regularizing each channel with a weighted linear combination of Bregman distances to all other image channels from the previous iteration. In this sense we generalize the Bregman iteration proposed by Osher et al. in [32] to multichannel images. We prove the convergence of the proposed scheme, analyze stationary points and present numerical experiments on color image denoising, which show the superior behavior of our approach in comparison to TV, TV with Bregman iterations on each channel separately, and vectorial TV. Additionally, we propose to use the infimal convolution of Bregman distances to different channels from the previous iteration to obtain the independence of the sign and hence the independence of the direction of the edge. While this work focuses on TV regularization, the proposed scheme can potentially improve any variational multichannel reconstruction method with a 1-homogeneous regularization.

Key words: Bregman iteration, total variation, multichannel image reconstruction, color image reconstruction, ℓ^1 regularization, inverse scale space methods

AMS subject classification: 47A52, 65J22, 68U10, 49M30

*Email: m.moeller@gmx.net

1 Introduction

Variational methods for image restoration have become one of the dominant approaches due to their flexibility and wide range of applicability. A majority of image processing and reconstruction tasks have been modeled as energy minimization problems in which the reconstructed image \hat{u} is determined by

$$\hat{u} = \arg \min_u H_f(u) + J(u), \quad (1)$$

for a data fidelity term $H_f(u)$ measuring the closeness to the data f and a regularization term $J(u)$ imposing some kind of smoothness or prior information (cf. [3, 29, 17, 30]). Classical choices for the fidelity term are squared norms between f and some operator applied to the image u or log-likelihoods derived from different noise models (cf. [38, 2, 27, 39]). The most frequently used regularization terms are total variation (cf. [35, 29]), variants thereof (cf. e.g. [10, 23]), or other ℓ^1 -type regularizations yielding sparse reconstructions in certain dictionaries (cf. [19, 26, 41]). Further improvements in variational image restoration have been made by considering Bregman iterations related to (1), iteratively improving the reconstruction via

$$u^{k+1} = \arg \min_u H_f(u) + J(u) - J(u^k) - \langle p^k, u - u^k \rangle, \quad (2)$$

where $p^k \in \partial J(u^k)$. Here the regularization term is a generalized Bregman distance between u and u^k . Starting with an initial value such that $p^0 = 0 \in \partial J(u^0)$, (1) is exactly the first step of the Bregman iteration, in the further step the regularization is replaced from J (reinterpreted as the Bregman distance to the minimizer of J) to the Bregman distance to the last image. A regularizing effect in case of noisy data can be obtained by appropriately terminating the iterative procedure (cf. [32]). Both experimental (cf. [32, 5, 13, 28, 45]) as well as analytical results (cf. [32, 15, 5, 6, 14, 43, 31]) confirm the superior properties of Bregman iterations, in particular the systematic error compared to (1) is reduced.

The majority of the investigations with methods mentioned above was related to gray-scale images, in particular Bregman iterations have been investigated exclusively to that case (with the trivial exception of separately applying Bregman iterations to different color channels). In the case of color - or, more generally, multi channel - image restoration, one crucial question is what kind of relation between the different image channels can be assumed. Even among methods considering the total variation (TV) as a regularization functional, many different variants have been proposed, e.g. Vectorial TV [8, 36, 24] or TV in different color spaces, e.g. chrominance and luminance based TV in [18]. While these regularizations work well in many applications, they share the property of coupling the intensities of different color channels. Our novel idea is to use Bregman distance regularization in an iterative setting to make use of the fact that the edge sets of different image channels should coincide.

Consider an underdetermined multichannel image restoration problem, which requires the solution of

$$K_i u_i = f_i \quad (3)$$

for M different channels $i \in \{1, \dots, M\}$ with measurements f_i and bounded linear operators K_i . Our motivation is to look for a solution $u = (u_1, \dots, u_M)$ with reasonably small total variation in each channel and additionally require the channels u_i to share a common edge set. This way, no assumption has to be made on the sizes of the jumps across the edges in each channel. As we will see in the course of this paper, a common edge set is strongly correlated with a common subgradient of the total variation such that - in its strictest form - one could consider minimizing

$$\|K_i u_i - f_i\|^2 \quad (4)$$

under the condition that there exists a $p \in \partial J(u_i)$ for all $i = 1, \dots, M$, i.e. that the channels share a common subgradient and consequently the edge set in the case $J = TV$. While the constraint of a common subgradient is difficult to impose directly, we will use the Bregman distances between different color channels in an iterative procedure to measure the differences between edge sets of different channels. More precisely, we shall compute updates from using an average over Bregman distances to different channels, i.e.

$$u_i^{k+1} \in \arg \min_{u_i \in \mathcal{U}} \left(\frac{\lambda}{2} \|K_i u_i - f_i\|^2 + \sum_{j=1}^M w_{i,j} (J(u_i) - \langle p_j^k, u_i \rangle) \right) \quad (5)$$

with positive weights $w_{i,j}$, $p_j^k \in \partial J(u_j^k)$, and a data fidelity weight λ . Clearly this makes sense in particular for one-homogeneous J , and we will focus on the particular case of $J = TV$. Note that for a weight matrix being identical we obtain the channel-wise Bregman iteration. In this sense, our approach generalizes and extends the work of Osher et al. [32] for color image restoration.

While a common subgradient of different image channels forces the edges to - loosely speaking - 'go into the same direction', e.g. in 1d either all be a 'step up' or all be a 'step down', we also consider the case of different channels sharing an edge set but not (necessarily) the edge direction. This will be particularly useful for image channels coming from different image modalities, e.g. in combined medical imaging devices such as SPECT/CT or PET/MR scanners. Mathematically, we use the infimal convolution of Bregman distances to achieve the independence of the edge direction.

Besides TV, collaborative sparse reconstruction using ℓ^1 regularization is another possible application and motivation for our approach. In case multiple sparse quantities have to be reconstructed from multiple measurements and one can expect the solutions to have the same sign, one can use the Bregman distances between the quantities as a penalty, thus encouraging a common subgradient. The infimal convolution of Bregman distances allows to relax this constraint to the u_i having the same support, but not necessarily the same sign.

The idea of using edge information from different channels or images has been considered in literature before, see for instance in [42, 21, 4, 33] and the references therein, but to the best knowledge of the authors no work has been done in the context of jointly minimizing the Bregman distances and hence encouraging common subgradients of different images.

The remainder of the paper is organized as follows: In section 2 we illustrate our general idea using the example of discrete one dimensional TV denoising. We continue in section 3 by discussing Bregman distance priors with a particular focus on the relation between the TV Bregman distance and common edge sets as well as the ℓ^1 Bregman distance and a common support. Our novel idea for iterative color image restoration using Bregman distances is presented in section 4 along with an analysis of stationary points, the primal-dual and dual formulation, as well as the proof of the well-definedness of the proposed iteration. Section 5 analyzes the convergence of the proposed scheme. The numerical implementation as well as color image denoising results are presented in section 6. Finally, we draw conclusions and outline possible extensions of the proposed scheme in section 7.

Concerning notation, we shall use the following setting uniformly throughout the paper

- A subscript i will denote the scalar image at channel i , $u = (u_1, \dots, u_M)$ is the complete multi-channel image.
- A superscript k will always denote the iteration number, i.e. u^k is the k -th Bregman iteration, u_i^k its i -th channel.
- The argument will always denote the spatial location, either continuously as $u(x)$ or referring to discrete pixel / voxel number $u(j)$, which will be clear from the context.

2 The Idea in a Nutshell: Discrete 1d TV

As a first motivation, let us consider discrete total variation regularization in 1d, which is easy to analyze and clearly motivates our approach. Let $f_i \in \mathbb{R}^N$ be measured signals (later: channels of an image) with $i \in \{1, \dots, M\}$. Let D denote the finite difference matrix with differences in x and y direction. The discrete 1d total variation is defined as $J(u) = \|Du\|_1$. The corresponding subdifferential can be characterized as

$$\partial J(u) = \{p \mid p = D^T q, q \in \text{sign}(Du)\} \quad (6)$$

In other words, the j -th component of q is given as

$$q(j) \begin{cases} = 1 & \text{if } (Du)(j) > 0 \\ = -1 & \text{if } (Du)(j) < 0 \\ \in [-1, 1] & \text{if } (Du)(j) = 0 \end{cases}$$

Now consider the regular Bregman iteration on each signal f_i separately. In the $(k+1)$ -th iteration one solves

$$\begin{aligned} u_i^{k+1} &= \arg \min_u \left(\frac{\lambda}{2} \|u_i - f_i\|^2 + \|Du_i\|_1 - \langle q_i^k, Du_i \rangle \right) \\ &= \arg \min_u \left(\frac{\lambda}{2} \|u_i - f_i\|^2 + \sum_j (\text{sign}(Du_i(j)) - q_i^k(j)) Du_i(j) \right) \end{aligned}$$

We can see that the Bregman iteration basically weights the penalization of each component of Du_i , or, in other words, weights the penalization of the approximated derivative. If the j -th component of u_i^k had a 'jump up', i.e. $Du_i^k(j) > 0$, i.e. $q_i^k(j) = 1$, then a 'jump up' of u_i^{k+1} at the j -th component is not penalized anymore and the size of the jump is determined exclusively by the minimization of $\|u_i - f_i\|^2$. Notice that this property has been studied in the literature and it has been shown theoretically as well as experimentally, that this iterative procedure restores the contrast of the reconstructed image and often leads to higher quality results than plain TV minimization.

In the above Bregman iteration the $q_i^k(j)$ could be interpreted as the likelihood of the true image having an edge in the corresponding direction at this component. The higher the value of $q_i^k(j)$, the smaller is the weight for an edge with $Du_i^{k+1}(j) > 0$ or, in other words, the more likely is an edge in this direction. Notice that once $q_i^k(j) = 1$ or $q_i^k(j) = -1$ the magnitude of $Du_i^{k+1}(j)$ does not matter anymore.

Viewing the $q_i^k(j)$ as likelihoods of the true image having an edge, motivates our idea for color Bregman iteration: If we have a signal (e.g. an image) with multiple channels,

we consider it to be likely that all channels share the position of the true edges as well as their direction. If $q_i^k(j) > 0$ for all channels i we want to reduce the penalty on all u_i^{k+1} for having a 'jump up' at the j -th component. On the other hand, if some $q_i^k(j) > 0$ and some $q_i^k(j) < 0$ this behavior is likely due to noise and one should not change the penalization of the corresponding component. We therefore propose to compute the 'likelihood' of an edge in each color channel as a linear combination of all q_i^k , i.e. computing

$$u^{k+1} = \arg \min_u \left(\frac{\lambda}{2} \|u - f\|^2 + \|Du\|_1 - \langle \tilde{p}_i^k, u \rangle \right) \quad (7)$$

with $\tilde{p}_i^k = \sum_l w_{i,l} p_l^k$ and non-negative weights $w_{i,l}$. We will demonstrate in this paper, that this kind of color Bregman iteration has significant advantages over the channel by channel reconstruction of color images.

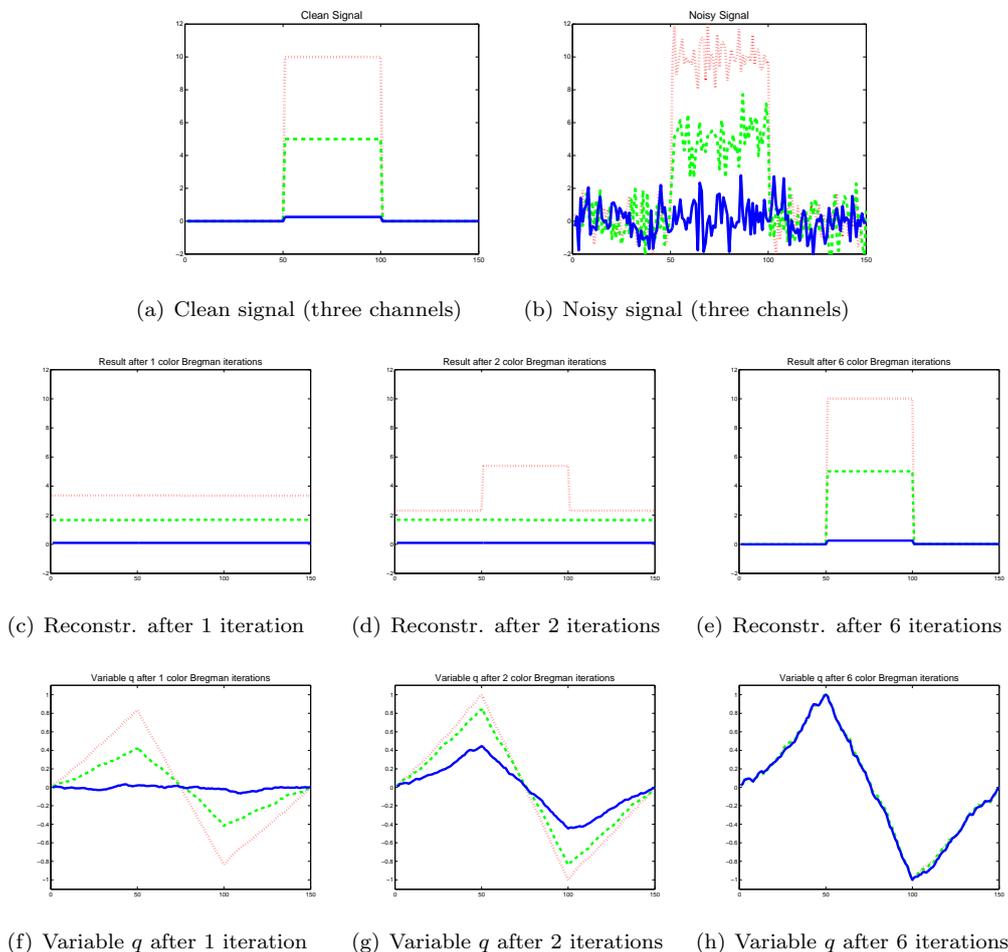


Figure 1: One dimensional example of color Bregman iteration

As a first example, consider the simple, one dimensional case shown in figure 1: The first row shows the clean data in the form of three one dimensional channels shown in red, green

and blue, as well as a noisy version of the signal. Notice that the heavy noise in comparison to the small signal in the blue channel makes the reconstruction of the blue signal without additional information basically impossible. This claim is underlined by looking at the variable q after the first iteration: The variable q reaching an absolute value of 1 indicates an edge at the corresponding position. More precisely, $q(i) = 1$ means a step up and $q(i) = -1$ means a step down at position i . We can see that the true steps are visible in the red and green but not in the blue channel. Even worse, the minimum and maximum values of q in the blue channel are not at positions 50 and 100 such that the first edges introduced in a single channel reconstruction of blue will not be in the right place. As we can see in figure 1, color Bregman iteration allows the blue channel to use the edge information of red and green to obtain a much better estimate for the true variable q . Finally, after six iterations, we reconstruct the true signal almost perfectly despite the heavy noise.

Additional to the color Bregman iteration described above, we look at the case where different images or channels are likely to share an edge set but not necessarily the sign of $Du(j)$. In this case it would be natural to consider

$$u_i^{k+1} = \arg \min \frac{\lambda}{2} \|u_i - f_i\|^2 + \sum_j \left(1 - \sum_l w_{i,l} |q_l^k(j)| \right) |Du_i(j)| \quad (8)$$

such that the magnitude but not the sign of $\tilde{q}_l^k(j)$ matter. Again, we can combine edge information of multiple channels by using a weighted average of the absolute values of the $q_l^k(j)$. In a more general setting, we propose to use the infimal convolution of Bregman distances to achieve the independence of the sign.

In the following sections we will analyze the above regularization procedure in detail in the continuous case and look at different regularizations $J(u)$. We will start by establishing the connection between the Bregman distance and common subgradients of different elements.

3 Bregman Distance Priors

The Bregman distance as first considered in [11] and later generalized to arbitrary proper convex functions $J(u)$ in [32] is given by

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

for $p \in \partial J(u)$. In the last years the Bregman distance has been used for various tasks including regularization [32], optimization [25, 16, 44] as well as for error estimates [15].

For one-homogeneous functionals, the pairs (u, v) with a Bregman distance of zero can be characterized as follows:

Proposition 3.1. *Let J be one-homogeneous and convex. Then*

$$p \in \partial J(u) \cap \partial J(v) \Leftrightarrow D_J^p(v, u) = 0. \quad (10)$$

Proof. Notice that the subdifferential of one-homogeneous functionals $J(u)$ can be classified as

$$\partial J(u) = \{p \mid \langle p, u \rangle = J(u), J(v) - \langle p, v \rangle \geq 0 \forall v\} \quad (11)$$

such that the Bregman distance simplifies to

$$D_J^p(v, u) = J(v) - \langle p, v \rangle. \quad (12)$$

For $p \in \partial J(v)$ we can use (11) in (12) to see that $D_J^p(v, u) = 0$. Vice versa, $D_J^p(v, u) = 0$ means that $J(v) = \langle p, v \rangle$ and since $p \in \partial J(u)$ we have $J(z) - \langle p, z \rangle \geq 0 \forall z$ such that $p \in \partial J(v)$ based on (11). \square

Proposition 3.1 gives a first indication that the Bregman distance measures the distance between the subdifferentials of two elements rather than measuring a direct distance. The motivation for our new, extended Bregman iteration comes from considering the total variation as a regularization functional, for which the subgradients encode information about the edges of an element without considering the size of the jump. As we will see in the next subsection, the corresponding Bregman distance measures the difference between the edge sets of two functions. To our minds the idea of assuming common edge sets for instance for the color channels of an image is very reasonable and allows us to propose a regularization that does not rely on any relation between the values of the color channels.

We also mention that the assumption of merely equal discontinuity sets can be derived from the usual model of color images as projections of a hyperspectral image U , i.e.

$$u_i(x) = \int U(x, \omega) \phi_i(\omega) d\omega, \quad (13)$$

where ϕ_i is an appropriate kernel for each channel. Under the natural condition that U is smooth in the spectral dimension ω , but changes between the characteristic spectra of different materials at some discontinuity set Γ it is straight-forward to see that the discontinuity set of all images u_i is contained in Γ and it is even very likely to be the equal to Γ for all i .

3.1 TV Bregman Distance and Edges

Throughout this subsection, we will assume that J is the total variation defined as

$$J(u) = \sup_{q \in C_0^\infty(\Omega, \mathbb{R}^n), \|q\|_\infty \leq 1} \int_\Omega \operatorname{div}(q) u \, dx$$

for an image domain Ω .

A well-known result (cf. [34]) serves as a first indication that a common edge set can be encoded by joint subgradients respectively by zero Bregman distance:

Lemma 3.2. *Let $v = f(u)$ for some monotone function f and $J = TV$. Then*

$$D_J^p(v, u) = 0 \quad \forall p \in \partial J(u). \quad (14)$$

Another relevant result is related to images having a joint edge set and some disjoint parts at nonzero distance:

Theorem 3.3. *Let u and v be piecewise constant with C^1 discontinuity sets*

$$D(u) = \Gamma \cup \Sigma \quad (15)$$

$$D(v) = \Gamma \cup \Upsilon, \quad (16)$$

such that Γ , Σ , Υ , and $\partial\Omega$ have pairwise positive distance and are boundaries of open sets, respectively. Moreover, let

$$[u][v] > 0 \quad \text{on } \Gamma, \quad (17)$$

where $[u]$ denotes the jump across Γ in normal direction. Then there exists $p \in \partial J(u) \cap \partial J(v)$.

Proof. Let ϵ be smaller than the minimal distance between any of the sets Γ , Σ , Υ , $\partial\Omega$. Let d denote the signed distance function to $D(u) \cup D(v)$. Moreover, let f_ϵ be a nonnegative function with support in $(-\epsilon, \epsilon)$ attaining its maximum value one at argument zero. Then

$$p = \nabla \cdot (s f_\epsilon(d) \nabla d),$$

is a subgradient if s equals the sign of $[u]$ on $D(u)$ and the sign of $[v]$ on $D(v)$. Note that the latter choice is possible since $[u][v] > 0$ on the intersection of the jump sets. We only verify that $p \in \partial J(u)$, the assertion for v is completely analogous. First of all we know that for a piecewise constant function with smooth subset the total variation is given by (cf. [1, 22])

$$J(u) = \int_{D(u)} |[u]| \, d\sigma.$$

Now we have that p is the divergence of a vector field with maximal value of the Euclidean norm equal to one and compact support, which implies

$$\langle p, \varphi \rangle \leq J(\varphi) \quad \forall \varphi \in BV(\Omega).$$

It remains to verify $\langle p, u \rangle = J(u)$. We have

$$\langle p, u \rangle = \int_{\Omega} u \nabla \cdot (s f_{\epsilon}(d) \nabla d) \, dx.$$

Integrating by parts on the open sets separated by $D(u)$ and using that ∇u vanishes away from the discontinuity set we obtain

$$\langle p, u \rangle = \int_{D(u)} [u] s f_{\epsilon}(d) \nabla d \cdot n \, d\sigma.$$

Since on $D(u)$ we have $d = 0$, $\nabla d = n$, and $[u]s = |[u]|$ we conclude that the right-hand side coincides with the above formula for the total variation of u . \square

Theorem 3.3 confirms that a joint subgradient indeed exists (respectively, two elements have a zero Bregman distance) if certain conditions on the edge sets are met. We can see that one image having an edge at a place where the other one is constant does not influence the existence of a joint subgradient. This is an important piece of information for using the Bregman distance as a regularization, because it tells us that no edges are introduced artificially if one channel has an edge at a place where the other one does not.

Theorem 3.3, however, also tells us that a common edge alone is not enough. Condition (17) shows that on the common edge set, the jumps across the discontinuity sets have to point in the same direction. In the following we will speak of the jump across the discontinuity set as the 'direction' of an edge. In this sense, a zero Bregman distance encodes that the edges two images (or channels) have in common 'point in the same direction'. Notice that the condition (17) is the two dimensional analogon of the $Du_i(j)$ having the same sign for all i in section 2. Also notice, that the edge-aligning property of the Bregman distance, or more precisely the alignment of the level lines, was one of the motivations for the iterative Bregman regularization in [32].

In the next subsection we will look at ℓ^1 regularization, where the analogous property of the 'edges pointing in the same direction' is easy to understand.

3.2 ℓ^1 Bregman Distance and Support

It is well known that the subdifferential of the ℓ^1 norm can be characterized as follows:

$$p \in \partial \|u\|_1 \Leftrightarrow p(j) \begin{cases} = 1 & \text{if } u(j) > 0, \\ = -1 & \text{if } u(j) < 0, \\ \in [-1, 1] & \text{else.} \end{cases} \quad (18)$$

With the help of this representation it is easy state the analogous version of Theorem 3.3 in the ℓ^1 case.

Proposition 3.4. *Let $u, v \in \mathbb{R}^n$ and let $J = \|\cdot\|_1$. Let I be the index set where u and v are nonzero, $I = \{i \mid u_i \neq 0, v_i \neq 0\}$. Then $\text{sign}(u_i) = \text{sign}(v_i)$ for all $i \in I$ is equivalent to the existence of a $p \in \partial J(u) \cap \partial J(v)$.*

As we can see the signs of u and v on the common support have to coincide. For relating a common subgradient and finally an equal support with equal signs to the Bregman distance, it helps to write the Bregman distance as the following sum:

$$D_{\|\cdot\|_1}^p(v, u) = \sum_j (\text{sign}(v(j)) - p(j))v(j). \quad (19)$$

The following lemma gives the conditions under which the support as well as the signs of nonzero entries of two elements u and v coincide:

Lemma 3.5. *Let $u, v \in \mathbb{R}^n$ and let $J = \|\cdot\|_1$. Then*

$$D_J^p(v, u) = 0 \quad \forall p \in \partial J(u). \quad (20)$$

implies that $\text{sign}_0(u)v \geq 0$, i.e. $v(i)$ cannot have opposite signs as $u(i)$. Moreover $\text{sign}_0(u) = \text{sign}_0(v)$ if and only if

$$D_J^p(v, u) = 0, \quad D_J^q(u, v) = 0 \quad \forall p \in \partial J(u), q \in \partial J(v). \quad (21)$$

Proof. Let $I = \{j \mid v(j) \neq 0, \text{sign}_0(u(j)) \neq \text{sign}_0(v(j))\}$. Using (19) with the subgradient $p = \text{sign}_0(u)$, it is easy to see that

$$D_{\|\cdot\|_1}^p(v, u) = \sum_{j \in I} (\text{sign}_0(v(j)) - \text{sign}_0(u(j)))v(j).$$

Since the signs of $v(j)$ and $u(j)$ in each summand are different, any summand would be strictly positive. Due to $D_J^p(v, u) = 0$, I has to be empty. Particularly, the support of v is contained in the support of u . If additionally $D_J^q(u, v) = 0$ for all q , the support of u is contained in the support of v and the $u(i)$ cannot have opposite signs as the $v(i)$, which leads to $\text{sign}_0(u) = \text{sign}_0(v)$. The reverse implication follows straight-forward using equation (19). \square

Keeping the idea of regularizing with the Bregman distances between several elements in mind, the natural question arising from lemma 3.5, proposition 3.4, and theorem 3.3 is if a zero Bregman distance is too restrictive considering the 'same sign' (respectively the 'edges pointing in the same direction') conditions.

In the context of the recent literature (c.f. [40] and the references therein), a zero Bregman distance with respect to the ℓ^1 norm can be seen as a special case of collaborative sparsity, where not only the support, but also the signs should coincide. In case the equality of signs is too restrictive, the independence of signs can be achieved by considering the infimal convolution of Bregman distances as a regularization, as Lemma 3.7 shows.

The infimal convolution of two convex functionals $J(\cdot)$ and $R(\cdot)$ is defined as follows

$$J(v) \square R(v) = \inf_{\phi, \psi, v = \phi + \psi} J(\phi) + R(\psi). \quad (22)$$

In the following we are going to further investigate the infimal convolution of Bregman distances, not to be confused with the Bregman distance for the infimal convolution of functionals. More precisely, the infimal convolution is carried out for different subgradients (related to the second argument) considering the Bregman distance as a functional of the first argument, which is indeed convex as soon as J is.

Lemma 3.6. *The infimal convolution between the ℓ^1 Bregman distance $D_J^p(v, u)$ and $D_J^{-p}(v, -u)$ is given by*

$$D_J^p(v, u) \square D_J^{-p}(v, -u) = \sum_{i=1}^N |v(i)|(1 - |p(i)|). \quad (23)$$

Proof. We compute

$$\begin{aligned} D_J^p(v, u) \square D_J^{-p}(v, -u) &= \inf_{\phi, \psi, v=\phi+\psi} J(\phi) - \langle \phi, p \rangle + J(\psi) + \langle \psi, p \rangle \\ &= \inf_{\phi, \psi, v=\phi+\psi} \sum_i (|\phi(i)| - p(i)\phi(i) + |\psi(i)| + p(i)\psi(i)) \\ &= \inf_{\phi, \psi, v=\phi+\psi} \sum_i |\phi(i)|(1 - p(i)\text{sign}(\phi(i))) + \sum_i |\psi(i)|(1 + p(i)\text{sign}(\psi(i))), \\ &\geq \inf_{\phi, \psi, v=\phi+\psi} \sum_i |\phi(i)|(1 - |p(i)|) + \sum_i |\psi(i)|(1 - |p(i)|) \geq \sum_i |v(i)|(1 - |p(i)|) \end{aligned}$$

On the other hand, $D_J^p(v, u) \square D_J^{-p}(v, -u) \leq \sum_i |v(i)|(1 - |p(i)|)$ holds for $\phi(i) = v(i)$ and $\psi(i) = 0$ if $\text{sign}(v(i)) = \text{sign}(p(i))$ and $\phi(i) = 0$ and $\psi(i) = v(i)$ if $\text{sign}(v(i)) = -\text{sign}(p(i))$. \square

As we can see $D_J^p(v, u) \square D_J^{-p}(v, -u)$ does not penalize $v(i)$ if $|p(i)| = 1$ and thus is independent of the sign of $u(i)$. We can now formulate the property of two elements having the same support in terms of the infimal convolution of Bregman distances:

Lemma 3.7. *Let $u, v \in \mathbb{R}^n$ and let $J = \|\cdot\|_1$. Then*

$$D_J^p(v, u) \square D_J^{-p}(v, -u) = 0 \quad \forall p \in \partial J(u) \quad (24)$$

implies that the support of v is contained in the support of u . Moreover u and v have equal support if and only if (24) and

$$D_J^q(u, v) \square D_J^{-q}(u, -v) = 0 \quad \forall q \in \partial J(v) \quad (25)$$

hold.

Proof. Using the representation (23) as well as the fact that the sum is zero if and only if each summand is zero, the lemma simplifies to stating that u and v have the same support if and only if

$$|v(i)|(1 - |p(i)|) = 0 \quad \forall p(i) \in \text{sign}(u(i)). \quad (26)$$

for all i . If there exists an i for which $u(i) = 0$ but $v(i) \neq 0$, $p(i) = 0$ is a valid choice for the i -th component of the subgradient and equation (26) is violated. On the other hand, if there exists a $p(i)$ such that $p(i) \in \text{sign}(u(i))$ but $|v(i)|(1 - |p(i)|) \neq 0$, then $|v(i)| > 0$ and $|p(i)| < 1$, which, however, means that $u(i) = 0$. \square

This motivates us to consider the infimal convolution of Bregman distances in the TV case as well. Our goal is to show that the infimal convolution $D_J^p(v, u) \square D_J^{-p}(v, -u)$ with respect to TV can be zero independent of the direction of the edge set.

3.3 Infimal Convolution of Bregman Distances and Total Variation

As a first auxiliary result and as a reasoning for also considering the Bregman distance using $-p$, the following Lemma states how the subdifferential of a one-homogeneous functional behaves with respect to a change of sign.

Lemma 3.8. *Let J be convex and one-homogeneous. Then $\partial J(-u) = -\partial J(u)$.*

Proof. Let $p \in \partial J(-u)$. By definition $J(-u) = \langle p, -u \rangle$ such that $J(u) = \langle -p, u \rangle$ and also by definition $J(v) - \langle p, v \rangle \geq 0$ for all v (particularly for all $-v$) such that $J(v) - \langle -p, v \rangle \geq 0$ for all v . Together, this shows $p \in -\partial J(u)$. Similar arguments show the reverse also holds. \square

As a consequence it indeed makes sense to consider the infimal convolution of the Bregman distances with subgradients of opposite sign. An explicit formula seems out of reach currently, but we can provide upper and lower bounds, from which we can gain further insight.

Proposition 3.9. *Let $v \in C^1$, $p = \nabla \cdot g \in \partial J(u)$ for $J = TV$. Then*

$$\int_{\Omega} (1 - |g|) |\nabla v| \, dx \leq D_J^p(v, u) \square D_J^{-p}(v, -u) \leq \int_{\Omega} |\nabla v| \, dx - \left| \int_{\Omega} \nabla v \cdot g \, dx \right|. \quad (27)$$

Proof. We use the definition of the infimal convolution as the infimum of the functional

$$F(w) = J(v - w) + J(w) - \langle p, v - w \rangle + \langle p, w \rangle.$$

By a density argument we can compute the infimal convolution as the infimum over all C^1 functions w of

$$F(w) = \int_{\Omega} |\nabla(v - w)| + |\nabla w| + g \cdot \nabla(v - 2w) \, dx.$$

With the Cauchy-Schwarz inequalities on the gradient fields we obtain

$$F(w) \geq \int_{\Omega} (1 - |g|) (|\nabla(v - w)| + |\nabla w|) \, dx,$$

and the triangle inequality yields the lower bound.

Choosing $w = 0$ and $w = v$ and using the regularity of v to integrate by parts we see

$$\inf_w F(w) \leq F(0) = \int_{\Omega} |\nabla v| + g \cdot \nabla v \, dx$$

and

$$\inf_w F(w) \leq F(v) = \int_{\Omega} |\nabla v| - g \cdot \nabla v \, dx$$

we obtain the upper bound. \square

With the infimal convolution of Bregman distances we obtain a more general version of the morphological invariance, the infimal convolution also vanishes for nonmonotone transformations of the intensities:

Lemma 3.10. *Let $v = f(u)$ for some function $f : \mathbb{R} \rightarrow \mathbb{R} \in BV$ and $J = TV$. Then*

$$D_J^p(v, u) \square D_J^{-p}(v, -u) = 0 \quad \forall p \in \partial J(u). \quad (28)$$

Proof. According to the Jordan decomposition (c.f. [7] for a nice extension to higher dimensions), any function $f : \mathbb{R} \rightarrow \mathbb{R} \in BV$ can be written as $f = f_1 - f_2$ for monotonically increasing functions f_1 and f_2 . Now lemma 3.2 tells us that $J(f_1(u)) - \langle p, f_1(u) \rangle = 0$ and $J(f_2(u)) - \langle p, f_2(u) \rangle = 0$ for all p , which immediately yields $D_J^p(v, u) \square D_J^{-p}(v, -u) = 0$ for all p . \square

Another generalization concerns the case of piecewise constants, now we obtain a version independent of the signs of jumps as desired:

Theorem 3.11. *Let u and v be piecewise constant with C^1 discontinuity sets*

$$D(u) = \Gamma \cup \Sigma \tag{29}$$

$$D(v) = \Gamma \cup \Upsilon, \tag{30}$$

such that Γ , Σ , and Υ have pairwise positive distance. Then there exists $p \in \partial J(u)$ such that

$$D_J^p(v, u) \square D_J^{-p}(v, -u) = 0.$$

Proof. Due to the definition of the infimal convolution and the nonnegativity it suffices to find $w \in BV(\Omega)$ and $p = \nabla \cdot g \in \partial J(u)$ such that

$$F(w) = J(v - w) + J(w) - \langle p, v - w \rangle + \langle p, w \rangle$$

vanishes. For this sake we construct w with jump set $\Gamma \cup \Upsilon \cup \Theta$ such that Θ has positive distance from Γ , Υ , and Σ , such that $w = 0$ in a neighbourhood of Υ and of the part of Γ where u and v have the same jump sign, and $w = v$ in a neighbourhood of the part of Γ where u and v have the same opposite jump sign. It is well-known that for a piecewise constant u the vector field g needs to satisfy $g \cdot n = \text{sign}([u])$ on $\Gamma \cup \Sigma$, but can be chosen arbitrarily (up to the constraint on the supremum norm) on $\Upsilon \cup \Theta$ to obtain a subgradient $p = \nabla \cdot g \in \partial J(u)$. Hence, let

$$g \cdot n = -\text{sign}([w]) \quad \text{on } \Theta \quad \text{and} \quad g \cdot n = \text{sign}([v]) \quad \text{on } \Upsilon.$$

With the above choices, we find

$$F(w) = \int_{\Gamma \cup \Upsilon} (|[v - w]| + |[w]| - g \cdot n[v - 2w]) \, d\sigma + 2 \int_{\Theta} (|[w]| + g \cdot n[w]).$$

The second integral vanishes due to the above properties of $g \cdot n$. To show that the first integral vanishes we can inspect the integrand on the different parts. On Υ the integrand equals $|[v]| - g \cdot n[v]$, which vanishes because of the above choice of g . On the part of Γ where u and v have opposite jump sign the same applies. On the part of Γ with opposite jump sign the integrand equals

$$|[v]| + g \cdot n[v] = |[v]| - |[v]| = 0.$$

Hence, we conclude $F(w) = 0$. \square

4 Color Bregman Iterations

Our above analysis shows that in the case of TV regularization, having the same edge set and direction of the edges means having a common subgradient. Let us consider the case of reconstructing signals u_i from noisy measurements f_i . Considering a regularization which encourages multiple channels to share the edge sets and their direction could therefore lead to a model of the form

$$\sum_{i=1}^M \left(\frac{1}{2} \|K_i u_i - f_i\|^2 + \alpha (J(u_i) - \langle p_0, u_i \rangle) \right) \rightarrow \min_{u \in \mathcal{U}^M, p_0 \in \cap_i \partial J(u_i)} \quad (31)$$

where K_i denote linear operators which are specific to the measurements or the reconstruction task and α is a weight for the regularization term. Unfortunately, the constraint of a common subgradient leads to a highly degenerate set of constraints. Additionally, the above problem is nonconvex, such that we suggest to instead use an iterative procedure where each iteration is the solution of a convex minimization problem using the Bregman distances to all u_i of the previous iteration as a regularization.

4.1 Iteration Schemes

Let $K : \mathcal{U}^M \rightarrow \mathcal{Y}^M$ be a block diagonal operator given by $u \mapsto (K_i u_i)$. We propose to find the iterate u^{k+1} as the vector of minimizers

$$u_i^{k+1} \in \arg \min_{u_i \in \mathcal{U}} \left(\frac{1}{2} \|K_i u_i - f_i\|^2 + \alpha \sum_{j=1}^M w_{i,j} D_J^{p_j^k}(u_i, u_j^k) \right) \quad (32)$$

for non-negative weights $w_{i,j}$ with $\sum_j w_{i,j} = 1$. The optimality condition to the minimization problems (32) can be written in vector form as

$$p^{k+1} = W p^k + \lambda K^*(K u^{k+1} - f), \quad p_i^{k+1} \in \partial J(u_i^{k+1}). \quad (33)$$

where we used $\lambda = \frac{1}{\alpha}$ and the entries of the matrix W are $w_{i,j}$. Therefore, the rows of the weight matrix W sum to one

$$\sum_{j=1}^M w_{i,j} = 1 \quad \forall i = 1, \dots, M, \quad (34)$$

such that $(1, 1, \dots, 1)^T$ is an eigenvector for W with eigenvalue 1.

The alternative scheme using infimal convolution of Bregman distances is given by

$$u_i^{k+1} \in \arg \min_{u_i \in \mathcal{U}} \left(\frac{1}{2} \|K_i u_i - f_i\|^2 + \alpha w_{i,i} D_J^{p_i^k}(u_i, u_i^k) + \alpha \sum_{j=1, j \neq i}^M w_{i,j} (D_J^{p_j^k}(u_i, u_j^k) \square D_J^{-p_j^k}(u_i, -u_j^k)) \right) \quad (35)$$

Note that we treat the diagonal term differently, namely without infimal convolution. This is natural however, since the different sign of jumps should only be relevant for different channels, while we want to use the information on the i -th channel itself as in the original Bregman iteration. Optimality conditions can be stated as above, and one observes that it

is also important to use $D_J^{p_i^k}$ only in the diagonal term in order to really obtain an update equation for the subgradients.

Since the subgradients of the infimal-convolution are not very intuitive, we derive an alternative formulation using new auxiliary variables as in the definition of the infimal convolution

$$u_i^{k+1} \in \arg \min_{u_i \in \mathcal{U}} \min_{z_{ij} \in \mathcal{U}} \left(\frac{1}{2} \|K_i u_i - f_i\|^2 + \alpha w_{i,i} D_J^{p_i^k}(u_i, u_i^k) + \alpha \sum_{j=1, j \neq i}^M w_{i,j} (J(u_i - z_{ij}) + J(z_{ij}) - \langle p_j^k, u_i - z_{ij} \rangle + \langle p_j^k, z_{ij} \rangle) \right). \quad (36)$$

Now we can easily derive optimality with respect to u_i and z_{ij}

$$K_i^*(K_i u_i^{k+1} - f_i) + \alpha w_{i,i} (p_i^{k+1} - p_i^k) + \alpha \sum_{j \neq i} w_{i,j} (r_{ij}^{k+1} - p_j^k) = 0, \quad (37)$$

$$2p_j^k - r_{ij}^{k+1} + g_{ij}^{k+1} = 0, \quad (38)$$

$$p_i^{k+1} \in \partial J(u_i^{k+1}), \quad g_{ij}^{k+1} \in \partial J(z_{ij}^{k+1}), \quad r_{ij}^{k+1} \in \partial J(u_i^{k+1} - z_{ij}^{k+1}). \quad (39)$$

4.2 Stationary Solutions

The following subsection investigates possible stationary points of the two iterative schemes described above. Additionally to the results presented below, we include an explicit example for the behavior of color Bregman iteration in the case where $f_i = c_i K \nu$ for ν being a singular vector in the supplementary material to this manuscript. We refer the reader to [6] for details regarding the concept of singular values for nonlinear variational methods.

4.2.1 Color Bregman Iteration

In the following we study potential stationary solutions of the iteration scheme, i.e. $(u, p) \in \mathcal{U}^M \times (\mathcal{U}^*)^M$ such that

$$p = Wp - \lambda K^*(Ku - f), \quad p_i \in \partial J(u_i). \quad (40)$$

Of particular interest are of course solutions with zero residual $K^*(Ku - f) = 0$, for which one needs $p = Wp$. This implies $p = p_1 e_1$, where e_1 is an eigenvector of W with eigenvalue 1. We are interested in the case of $e_1 = (1, 1, \dots, 1)^T$ being the unique positive normalized eigenvector, which is guaranteed e.g. by the Perron-Frobenius theorem if all w_{ij} are positive. Then the above property would imply $p_i = p_1$ for all i and we would have found a solution for which all channels share a common subgradient.

In the general case there might however be different stationary solutions. If W is symmetric, which appears to be the natural choice, then we have that e_1^T is a right eigenvector of W , which implies

$$e_1^T p = e_1^T Wp - \lambda e_1^T K^*(Ku - f) = e_1^T p - \lambda e_1^T K^*(Ku - f),$$

i.e. $e_1^T K^*(Ku - f) = 0$. This means that the sum of the residuals is zero for any stationary solution, but not necessarily each residual itself.

For W symmetric being such that $WK^* = K^*W$ one can derive more detailed arguments. Note that the latter property is satisfied automatically in the prominent case of all operators

K_i being identical. In this case we obtain that for a stationary solution $(I - W)p$ is in the range of K^* , i.e. there exists $q \in \mathcal{Y}$ such that

$$\alpha(I - W)q = (f - Ku), \quad K_i^* q_i \in \partial J(u_i).$$

This relation is a saddle point condition for the functional

$$\mathcal{S}(u, q) = \sum_{i=1}^M J(u_i) - \langle q, Ku - f \rangle - \frac{\alpha}{2} \langle (I - W)q, q \rangle. \quad (41)$$

Note that \mathcal{S} is convex with respect to u , hence a natural condition would be concavity with respect to q . The latter is satisfied since under the above conditions (symmetry, nonnegative entries, row sums equal to one) the matrix $I - W$ is positive semidefinite. Note that e_1 is always in the nullspace of $I - W$, hence it will not be positive definite. This property of W , together with the saddle-point interpretation of the stationary solution, will also be crucial in the convergence analysis below. Notice that in case that there exists a solution of $Ku = f$ with $q = Wq$ and $p_i = K_i^* q_i \in \partial J(u_i)$, then (u, q) is a saddle point of \mathcal{S} (independent of α).

4.2.2 Infimal Convolution Bregman Iteration

Let us now look at possible stationary solutions of the iteration using infimal convolutions of Bregman distances, i.e. at solutions of the equations

$$K_i^*(K_i u_i - f_i) + \alpha \sum_{j \neq i} w_{i,j} (r_{ij} - p_j) = 0, \quad (42)$$

$$2p_j - r_{ij} + g_{ij} = 0, \quad (43)$$

$$p_i \in \partial J(u_i), \quad g_{ij} \in \partial J(z_{ij}), \quad r_{ij} \in \partial J(u_i - z_{ij}). \quad (44)$$

Again we are particularly interested in the case of a zero residual $K^*(Ku - f) = 0$. We can state

Proposition 4.1. *A necessary and sufficient condition for a solution to (42), (43) and (44) to meet $K^*(Ku - f) = 0$ is that there exist z_{ij} which meet*

$$D_J^{p_j}(u_i - z_{ij}, u_j) = 0, \quad D_J^{-p_j}(z_{ij}, -u_j) = 0.$$

Proof. The above condition is sufficient since it allows us to choose $r_{ij} = p_j \in \partial J(z_{ij})$ and $g_{ij} = -p_j \in \partial J(u_i - z_{ij})$, such that (42) immediately yields $K_i^*(K_i u_i - f_i) = 0$ for all i . To see that the above condition is necessary, we consider the dual product

$$\begin{aligned} \frac{1}{\alpha} \langle K_i^*(f_i - K_i u_i), u_i \rangle &= \left\langle \sum_{j \neq i} w_{i,j} (r_{ij} - p_j), u_i \right\rangle = \sum_{j \neq i} w_{i,j} \langle r_{ij} - p_j, u_i - z_{ij} + z_{i,j} \rangle, \\ &= \sum_{j \neq i} w_{i,j} \left(\underbrace{\langle r_{ij}, u_i - z_{ij} \rangle}_{=D_J^{p_j}(u_i - z_{i,j}, u_j)} - \underbrace{\langle p_j, u_i - z_{ij} \rangle}_{=g_{ij} + p_j} + \langle r_{ij} - p_j, z_{i,j} \rangle \right), \\ &= \sum_{j \neq i} w_{i,j} \left(D_J^{p_j}(u_i - z_{i,j}, u_j) + \underbrace{\langle g_{ij}, z_{i,j} \rangle - \langle -p_j, z_{i,j} \rangle}_{=D_J^{-p_j}(z_{i,j}, -u_j)} \right). \end{aligned}$$

Obviously, the above being equal to zero is necessary for $K^*(Ku - f) = 0$. Due to the non-negativity of the Bregman distance, the above sum can only be equal to zero if each summand is zero, which yields the assertion. \square

As we can see any stationary point with zero residual of the color Bregman iteration also is a stationary point with zero residual of the infimal convolution Bregman iteration. Additionally, we gain all points that can be decomposed into a part with a zero Bregman distance to u_j and a part with zero Bregman distance to $-u_j$. Particularly, we can see the independence of the sign of the p_j . By choosing $z_{ij} = u_i$ or $z_{ij} = 0$, we see that any point with zero residual and a subgradient p such that for each i either $p \in \partial J(u_i)$ or $-p \in \partial J(u_i)$ holds, is a stationary point of the infimal convolution Bregman iteration.

4.3 Primal-Dual and Dual Formulation

Under the condition introduced in the last section, i.e. $K^*W = WK^*$, we can employ the saddle-point structure to provide alternative formulations of the iterative scheme (33). Such a reformulation is also well-known in the case of the original Bregman iteration and can be interpreted as an Augmented Lagrangian method. For initial values of p in the range of K^* it is straight-forward to see that $p^k = K^*q^k$ and we can rephrase (33) as

$$q^{k+1} = Wq^k - \lambda(Ku^{k+1} - f), \quad K_i^*q_i^{k+1} \in \partial J(u_i^{k+1}) \quad (45)$$

With \mathcal{S} defined as above, these conditions can be reinterpreted as

$$u^{k+1} \in \arg \min_u \mathcal{S}(u, q^{k+1}), \quad (46)$$

$$q^{k+1} = q^k + \partial_q \mathcal{S}(u^{k+1}, q^k). \quad (47)$$

Hence, the scheme is minimizing with respect to u at fixed q , while q^{k+1} is obtained with an ascent step with respect to q .

We observe that the primal-dual iteration (45) can be considered as an alternative in the case $K^*W \neq WK^*$. Eliminating q^{k+1} it can be rephrased as

$$K_i^*(Wq^k)_i - \lambda K_i^*(Ku^{k+1} - f)_i \in \partial J(u_i^{k+1}),$$

which is the optimality condition for the variational problem

$$u^{k+1} \in \arg \min_{u \in \mathcal{U}^M} \left(\frac{\lambda}{2} \|Ku - f\|^2 + \sum_i J(u_i) - \langle Wq^k, Ku - f \rangle \right). \quad (48)$$

Finally we shall also state a dual form. We will give a purely formal derivation here and not justify the duality, since we will use the dual formulation as a motivation of the convergence analysis below only, where we are interested in the convergence of u^k as well. Under appropriate regularity $p_i \in \partial J(u_i)$ can be reformulated to $u_i \in \partial J^*(p_i)$ with J^* being the convex conjugate of J . Note that $u_i^{k+1} \in \partial J^*(K_i^*q_i^{k+1})$ implies $Ku_i^{k+1} \in \partial_q J^*(K_i^*q_i^{k+1})$. Hence, the iteration

$$q^{k+1} + Ku^{k+1} - f = q^k - \alpha(I - W)q^k$$

can also be interpreted as a forward-backward splitting method on the dual functional

$$D(q) = \frac{\alpha}{2} \langle (I - W)q, q \rangle - \langle f, q \rangle + \sum_i J^*(K_i^*q_i), \quad (49)$$

where the forward splitting is applied to the first term and the backward splitting to the other two.

4.4 Well-Definedness

We assume that \mathcal{U} is a Banach space being the dual of some Banach space \mathcal{V} and that $\mathcal{R}(K_i^*) \subset \mathcal{V}$ for all $i = 1, \dots, M$. This is satisfied e.g. if K_i is the adjoint of some operator L_i mapping to \mathcal{V} .

Lemma 4.2. *Let the sublevel sets of $u \mapsto \|K_i u\|^2 + J(u)$ be bounded in \mathcal{U} and let $H : \mathcal{U} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper nonnegative convex functional. Then for each $q \in \mathcal{Y}$ and each $\epsilon > 0$ there exists a solution of the variational problem*

$$\frac{1}{2} \|K_i u_i - f_i\|^2 + \epsilon J(u_i) - \langle K_i^* q, u_i \rangle + H(u_i) \rightarrow \min_{u_i \in \mathcal{U}}. \quad (50)$$

Moreover, the solution is unique if K_i is injective.

Proof. By adding a constant independent of u_i we can rewrite the minimization equivalently as

$$\frac{1}{2} \|K_i u_i - f_i - q\|^2 + \epsilon J(u_i) + H(u_i) \rightarrow \min_{u_i \in \mathcal{U}}. \quad (51)$$

This functional is convex and hence weak-* lower semicontinuous on bounded sets. Moreover, the nonnegativity of H implies an upper bound on $\|K_i u_i\|^2 + J(u)$ on each sublevel set of the functional. Thus, the latter are bounded and thus weak-* compact by the Banach-Alaoglu theorem. These two properties yield the existence of a minimizer by a standard argument. \square

Theorem 4.3. *Let the sublevel sets of $u \mapsto \|K_i u\|^2 + J(u)$ be bounded in \mathcal{U} , let $p_i^0 = K_i^* q_i^0 \in \partial J(u_i^0)$. Then, the sequence of iterates (u_i^k, p_i^k) defined by (32) is well-defined and satisfies (33).*

Proof. Let $p_i^k = K_i^* q_i^k \in \partial J(u_i^k)$ be met (holds for $k = 0$ by the conditions of the theorem). We determine the next iterates (u_i^{k+1}, p_i^{k+1}) by (32). By defining $H(u_i) = \alpha \sum_{j \neq i} w_{i,j} D_J^{p_j^k}(u_i, u_j^k)$, which is proper, convex and nonnegative, we see that each minimization problem for the u_i has the form of (50) with $\epsilon = \alpha$ such that lemma 4.2 guarantees the existence of a solution. The optimality conditions yields

$$K_i^*(K_i u_i^{k+1} - f_i) + \alpha p_i^{k+1} - \alpha \sum_j w_{i,j} p_j^k = 0 \quad (52)$$

for $p_i^{k+1} \in \partial J(u_i^{k+1})$, such that (33) holds for $\lambda = 1/\alpha$. Induction yields the result for all k . \square

In a similar way the Bregman iteration with infimal convolutions can be analyzed, noticing that by the properties of the infimal convolutions the variational problem in each iteration can still be cast in the form of Lemma 4.2 if $\epsilon := \alpha w_{i,i} > 0$.

Theorem 4.4. *Let the sublevel sets of $u \mapsto \|K_i u\|^2 + J(u)$ be bounded in \mathcal{U} , let $p_i^0 = K_i^* q_i^0 \in \partial J(u_i^0)$, and $w_{i,i} > 0$ for all i . Then, the sequence of iterates (u_i^k, p_i^k) defined by (35) is well-defined.*

5 Convergence Analysis

In the following we discuss the convergence analysis of the iteration scheme (setting $\lambda = 1$ for simpler notation)

$$q^{k+1} = Wq^k + f - Ku^{k+1}, \quad K_i^* q_i^{k+1} \in \partial J(u_i^{k+1}). \quad (53)$$

Throughout the whole section we shall assume that

- W is symmetric with nonnegative entries and row sums equal to one.
- W has a simple maximal eigenvalue $\lambda_1 = 1$.
- The initial values satisfy $p_i^0 = K^* q_i^0 \in \partial J(u_i^0)$.

We shall denote by $e_1 = \frac{1}{\sqrt{M}}(1, 1, \dots, 1)^T \in \mathbb{R}^M$ the normalized positive eigenvector corresponding to λ_1 and by $Q_1 \in \mathbb{R}^{M \times M}$ the projection matrix on the orthogonal space to e_1 .

5.1 Auxiliary Estimates

We start with some estimates on quantities potentially dissipated by the iteration. Those are norms of q^k as well as Bregman distances between u^k and some potential final state \bar{u} .

Proposition 5.1. *Let (u^k, q^k) be a sequence generated by (53). Then the estimate*

$$\begin{aligned} & \|Q_1(q^{k+1} - q^k)\|^2 + \|e_1^T(q^{k+1} - q^k)\|^2 + 2D_{symm}^{k,k+1} + \|K(u^{k+1} - u^k)\|^2 \\ & \leq \lambda_2^2 \|Q_1(q^k - q^{k-1})\|^2 + \|e_1^T(q^k - q^{k-1})\|^2 \end{aligned} \quad (54)$$

holds, where $D_{symm}^{k,k+1}$ denotes the symmetric Bregman distance

$$D_{symm}^{k,k+1} = \langle p^{k+1} - p^k, u^{k+1} - u^k \rangle. \quad (55)$$

and λ_2 is the second largest eigenvalue of W .

Proof. Subtracting (53) for index k and $k-1$ we obtain

$$q^{k+1} - q^k + K(u^{k+1} - u^k) = W(q^k - q^{k-1}).$$

Taking the squared norm in the Hilbert space \mathcal{Y} and expanding the square on the left-hand side yields

$$\|q^{k+1} - q^k\|^2 + 2D_{symm}^{k,k+1} + \|K(u^{k+1} - u^k)\|^2 = \|W(q^k - q^{k-1})\|^2. \quad (56)$$

Now we can further expand the squared norms in q into the parts in the subspaces spanned by e_1 and orthogonal to e_1 , respectively. Finally, the standard spectral estimate

$$\|Q_1 W(q^k - q^{k-1})\| \leq \lambda_2 \|Q_1(q^k - q^{k-1})\|$$

yields (54). □

As a direct consequence of (54) we conclude after summation

$$\|q^{k+1} - q^k\|^2 + \sum_{j=1}^k \|K(u^{j+1} - u^j)\|^2 + (1 - \lambda_2^2) \sum_{j=1}^k \|Q_1(q^j - q^{j-1})\|^2 \leq \|q^1 - q^0\|^2 \quad (57)$$

From the finiteness of the second sum and equation (53) we can immediately conclude

$$Q_1((W - I)q^k + f - K u^{k+1}) \rightarrow 0. \quad (58)$$

Let us define the residual at the k -th iteration as

$$r^k = (W - I)q^k + f - K u^k. \quad (59)$$

We can state:

Corollary 5.2. *The quantities $\|r^k\|$ and $\|q^{k+1} - q^k\|$ are monotonically decreasing.*

Proof. With (56) and $\|W\| = 1$, which follows from the above properties, we have

$$\|W(q^{k+1} - q^k)\| \leq \|q^{k+1} - q^k\| \leq \|W(q^k - q^{k-1})\| \leq \|q^k - q^{k-1}\|.$$

The decrease of $\|r^k\|$ follows from

$$\begin{aligned} r^k &= (I - W)q^k + (f - Ku^k) \\ &= \underbrace{q^k - Wq^{k-1} + (f - Ku^k)}_{=0} - W(q^k - q^{k-1}), \end{aligned}$$

□

Another auxiliary estimate is the following:

Proposition 5.3. *Let (u^k, q^k) be a sequence generated by (53) and let $\bar{u} \in \mathcal{U}^M$, $\bar{q} \in \mathcal{Y}^M$. Then the estimates*

$$\|q^{k+1} - \bar{q}\|^2 + 2\langle p^{k+1} - K^*\bar{q}, u^{k+1} - \bar{u} \rangle \leq \|W(q^k - \bar{q})\|^2 + 2\langle q^{k+1} - \bar{q}, r \rangle \quad (60)$$

and

$$\|q^{k+1} - \bar{q}\|^2 + 2\langle p^{k+1} - K^*\bar{q}, u^{k+1} - \bar{u} \rangle + \|Ku^{k+1} - K\bar{u}\|^2 \leq \|W(q^k - \bar{q})\|^2 + 2\langle W(q^k - \bar{q}), r \rangle + \|r\|^2 \quad (61)$$

hold, where r is the residual

$$r = (W - I)\bar{q} + f - K\bar{u}. \quad (62)$$

Proof. Adding and subtracting \bar{q} and $K\bar{u}$ as well as some rearrangement in (53) yields

$$q^{k+1} - \bar{q} + K(u^{k+1} - \bar{u}) = W(q^k - \bar{q}) + r. \quad (63)$$

Taking now a squared norm in \mathcal{Y} and expanding squares on both sides yields (61). Taking the scalar product with $q^{k+1} - \bar{q}$ and an application of Young's inequality yields (60). □

The above result is particularly interesting if $K_i^*\bar{q}_i \in \partial J(\bar{u}_i)$, since then the second term on the left-hand side of (60) and (61) is a symmetric Bregman distance between u^{k+1} and \bar{u} and in particular nonnegative.

A straight-forward computation further shows:

Proposition 5.4. *Let (u^k, q^k) with $p^k = K^*q^k$ be a sequence generated by (53) and let $\bar{u} \in \mathcal{U}^M$, $\bar{q} \in \mathcal{Y}^M$. Then the identity*

$$\begin{aligned} D_J^{p^{k+1}}(\bar{u}, u^{k+1}) - D_J^{p^k}(\bar{u}, u^k) &= -D_J^{p^k}(u^{k+1}, u^k) - \|q^{k+1} - q^k\|^2 \\ &\quad + \langle q^{k+1} - q^k, r - (I - W)(q^k - \bar{q}) \rangle. \end{aligned} \quad (64)$$

holds, where r is the residual defined by (62).

Proof. We rearrange equation (63) to

$$\begin{aligned} q^{k+1} - q^k + q^k - \bar{q} + K(u^{k+1} - \bar{u}) &= W(q^k - \bar{q}) + r. \\ \Rightarrow K(u^{k+1} - \bar{u}) &= -(I - W)(q^k - \bar{q}) + r - (q^{k+1} - q^k). \end{aligned}$$

We take the dual product with $q^{k+1} - q^k$ to obtain

$$\langle K(u^{k+1} - \bar{u}), q^{k+1} - q^k \rangle = \langle r - (I - W)(q^k - \bar{q}), q^{k+1} - q^k \rangle - \|q^{k+1} - q^k\|^2.$$

The left hand side can be rewritten as

$$\begin{aligned} \langle K(u^{k+1} - \bar{u}), q^{k+1} - q^k \rangle &= \langle u^{k+1} - \bar{u}, p^{k+1} - p^k \rangle, \\ &= -\langle p^k, u^{k+1} \rangle + J(\bar{u}) - \langle p^{k+1}, \bar{u} - u^{k+1} \rangle - (J(\bar{u}) - \langle p^k, \bar{u} \rangle), \\ &= J(u^{k+1}) - \langle p^k, u^{k+1} \rangle + D_J^{p^{k+1}}(\bar{u}, u^{k+1}) - (J(\bar{u}) - \langle p^k, \bar{u} \rangle), \\ &= D_J^{p^k}(u^{k+1}, u^k) + D_J^{p^{k+1}}(\bar{u}, u^{k+1}) - D_J^{p^k}(\bar{u}, u^k), \end{aligned}$$

which proves the proposition. \square

5.2 Clean Data

In the following we consider the case of clean data, i.e. we assume there exists $\bar{u} \in \mathcal{U}^M$, $\bar{q} \in \mathcal{Y}^M$ such that

$$0 = r = (W - I)\bar{q} + f - K\bar{u}, \quad K_i^* \bar{q}_i \in \partial J(\bar{u}_i). \quad (65)$$

Note that (65) is always satisfied in our original setting, i.e. if we have a solution of $Ku = f$ with joint subgradient, if $K_i = K_1$ for all i , and the source-type condition $\bar{p}_1 = K_1^* \bar{q}_1$. In this case we automatically have

$$(W - I)\bar{q} = (W - I)e_1 \bar{p}_1 = 0.$$

Using the fact that the terms depending on r drop out in the auxiliary estimates, we can state the monotonic behavior of q^k approaching \bar{q} :

Corollary 5.5. *The quantities $\|q^k - \bar{q}\|$ and $\|W(q^k - \bar{q})\|$ are monotonically decreasing.*

Proof. Equation (60) with $r = 0$ and $\langle p^{k+1} - K^* \bar{q}, u^{k+1} - \bar{u} \rangle = D_{\text{symm}}(u^{k+1}, \bar{u}) \geq 0$ leads to $\|q^{k+1} - \bar{q}\| \leq \|W(q^k - \bar{q})\|$, which together with $\|W\| = 1$ yields the assertion. \square

Using $r = 0$ in the auxiliary estimates from the previous section, we can obtain a convergence result for the iterative scheme:

Theorem 5.6. *Let $r = 0$ and let (u^k, q^k) with $p^k = K^* q^k$ be a sequence generated by (53). Then there exists a subsequence (u^{k_ℓ}, q_{k_ℓ}) converging in the weak-star respectively weak topology. The limit (u, q) of each such subsequence satisfies*

$$0 = (W - I)q + f - Ku. \quad (66)$$

Moreover, $Q_1 q^k \rightarrow Q_1 \bar{q}$, and if K_i^* is compact then $K_i^* q_i \in \partial J(u_i)$.

Proof. Summation of (54) yields

$$\|q^{k+1} - \bar{q}\|^2 + 2 \sum_{j=1}^{k+1} \langle p^j - K^* \bar{q}, u^j - \bar{u} \rangle + (1 - \lambda_2^2) \sum_{j=1}^k \|Q_1(q^j - \bar{q})\|^2 \leq \|q^0 - \bar{q}\|^2 \quad (67)$$

This implies $Q_1 q^k \rightarrow Q_1 \bar{q}$

From (64) we deduce by applying Young's inequality

$$2D_J^{p^{k+1}}(\bar{u}, u^{k+1}) - 2D_J^{p^k}(\bar{u}, u^k) \leq -2D_J^{p^k}(u^{k+1}, u^k) - \|q^{k+1} - q^k\|^2 + \|(I - W)(q^k - \bar{q})\|^2.$$

Adding this estimate and (60) we further deduce

$$\begin{aligned} & 2D_J^{p^{k+1}}(\bar{u}, u^{k+1}) + 2D_J^{p^k}(u^{k+1}, u^k) + \|q^{k+1} - \bar{q}\|^2 + \\ & 2\langle p^{k+1} - K^*\bar{q}, u^{k+1} - \bar{u} \rangle + \|q^{k+1} - q^k\|^2 \\ & \leq 2D_J^{p^k}(\bar{u}, u^k) + \|W(q^k - \bar{q})\|^2 + \|(I - W)(q^k - \bar{q})\|^2. \end{aligned}$$

Since $I - W$ is positive semidefinite it is straightforward to see

$$\|Wq\|^2 + \|(I - W)q\|^2 \leq \|q\|^2.$$

Using this on the right-hand side of the previous estimate together with summation yields

$$\begin{aligned} & 2D_J^{p^{k+1}}(\bar{u}, u^{k+1}) + \|q^{k+1} - \bar{q}\|^2 + 2\sum_{j=0}^k D_J^{p^j}(u^{j+1}, u^j) + \\ & 2\sum_{j=0}^k \langle p^{j+1} - K^*\bar{q}, u^{j+1} - \bar{u} \rangle + \sum_{j=0}^k \|q^{j+1} - q^j\|^2 \\ & \leq 2D_J^0(\bar{u}, u^0) + \|q^0 - \bar{q}\|^2. \end{aligned}$$

With the monotonicity of $\|q^{k+1} - q^k\|$ from (54) we hence deduce

$$\|q^{k+1} - q^k\| \leq \frac{C}{\sqrt{k+1}} \quad (68)$$

for some constant C .

Moreover, we deduce that q^k is bounded, which implies the existence of a subsequence q^{k_ℓ} converging weakly to some \hat{q} . Moreover, from (61) we conclude that Ku^k is uniformly bounded, thus

$$J(u^k) = \langle K^*q^k, u^k \rangle = \langle q^k, Ku^k \rangle \leq \|q^k\| \|Ku^k\|$$

is uniformly bounded as well. From the boundedness of the sublevel sets of $\frac{1}{2}\|K \cdot\|^2 + J$ we obtain the existence of a weak-star convergent subsequence of u^k , again denoted by k_ℓ .

Finally, since $q^{k+1} - q^k \rightarrow 0$, we conclude that the limits q and u of such converging subsequences satisfy $q = Wq + f - Ku$. Moreover, the weak convergence of q^{k_ℓ} yields

$$\langle K_i^* q_i, v \rangle = \lim \langle K_i^* q_i^{k_\ell}, v \rangle \leq J(v) \quad \forall v \in \mathcal{U}.$$

If K_i^* is compact, then the $K_i^* q_i^{k_\ell}$ converges strongly and hence

$$\langle K_i^* q_i^{k_\ell}, u_i^{k_\ell} \rangle \rightarrow \langle K_i^* q_i, u_i \rangle.$$

Now the lower semicontinuity of J implies for any $v_i \in \mathcal{U}$

$$J(u_i) + \langle K_i^* q_i, v_i - u_i \rangle \leq \liminf J(u_i^{k_\ell}) + \langle K_i^* q_i^{k_\ell}, v_i - u_i^{k_\ell} \rangle \leq J(v_i),$$

hence $K_i^* q_i \in \partial J(u_i)$. □

From (68) and $\|W\| = 1$ we can immediately conclude

Corollary 5.7. *Let the assumptions of Theorem 5.6 be satisfied. Then the color Bregman iteration converges with*

$$\|r^k\| \leq \frac{C}{\sqrt{k}} \quad (69)$$

for some constant C .

Under the stronger condition that there is indeed a solution with subgradient $\bar{p} = \bar{p}_1 e_1$ one can easily give a refined statement:

Corollary 5.8. *Let in addition to the assumptions of Theorem 5.6 \bar{u} satisfy $K\bar{u} = f$. Then $Ku^k \rightarrow f$ and p^k converges weakly to $p_1 e_1$.*

Proof. The assertion follows from Theorem 5.6 combined with a summation of (61). \square

Example 1. *For regular Bregman iteration without any coupling of the subgradients we have $W = I$. Therefore, corollary 5.2 tells us that $\|r^k\| = \|Ku^k - f\|$ is monotonically decreasing and corollary 5.7 tells us that $\|Ku^k - f\| \leq \frac{C}{\sqrt{k}}$, which both is in agreement with the results of [32].*

5.3 Noisy Data

A key property of Bregman iterations for noisy data is the decrease of the Bregman distance to the clean image at least until the residual reaches the noise level. This yields the stability of the iteration and a strategy to choose a stopping criterion in dependence of the noise.

In the case of the color Bregman iteration (64) this will not work by comparing r^{k+1} and r , but rather shifted versions, namely

$$\sigma^k = r^{k+1} - (I - W)(q^k - q^{k+1}) = q^{k+1} - q^k, \quad (70)$$

and

$$\bar{\sigma}^k = r - (I - W)(q^k - \bar{q}). \quad (71)$$

Indeed, from (64) we can conclude

$$D_J^{p^{k+1}}(\bar{u}, u^{k+1}) \leq D_J^{p^k}(\bar{u}, u^k) \quad (72)$$

as long as

$$\|\sigma^k\| \geq \|\bar{\sigma}^k\|. \quad (73)$$

At the initial iteration the comparison in (73) is the same as in the classical discrepancy, since

$$\sigma^0 = f - Ku^1, \quad \bar{\sigma}^0 = f - K\bar{u}.$$

Moreover, we have shown above that $\|\sigma^k\|$ is decreasing during the iteration, which further indicates that there is an appropriate stopping index at which (73) is violated for the first time.

6 Computational Results

In this section we will demonstrate the general behavior of the color Bregman TV as well as of the infimal convolution Bregman iteration model. In our numerical experiments of denoising 24 test images and a comparison to TV, TV with separate channel Bregman iteration, and vectorial TV, our proposed methods, show very promising results. Additional to the denoising results, we include detailed numerical experiments on different artificial test images showing the behavior of color Bregman iteration for images with edges pointing in different directions in the supplementary material.

Before jumping into the numerical experiments, let us very briefly discuss the numerical implementation for both approaches, color Bregman iteration and infimal convolution Bregman iteration.

6.1 Numerical Implementation

6.1.1 Color Bregman TV Denoising

In the case of TV denoising (using the simplified notation $J(u) = \|\nabla u\|_1$ for the total variation) we rewrite (32) by defining the sum of all residual f_i^k with $f_i^0 = f_i$ and obtain

$$u_i^{k+1} \in \arg \min_{u_i \in \mathcal{U}} \left(\frac{1}{2} \|u_i - f_i^k\|^2 + \alpha \|\nabla u_i\|_1 \right) \quad (74)$$

$$\tilde{f}_i^{k+1} = \sum_j w_{i,j} (f_j^k - u_j^{k+1}) \quad (75)$$

which is similar to the 'adding-back-the-noise' formulation from [32].

For the above minimization in u , we apply the split Bregman method [25] also known as the alternating directions method of multipliers (ADMM), see [9]: A new variable d and an additional constraint $d = \nabla u$ is introduced, which is enforced using Bregman iteration. The minimization between u and d is done in an alternating fashion. This leads to an inner loop (ADMM algorithm to minimize for u) and an outer loop for the color Bregman iteration given by (75). The inner loop becomes

$$\begin{aligned} u_i &\leftarrow \arg \min_{u_i} \left(\frac{1}{2} \|u_i - f_i^k\|^2 + \frac{\mu}{2} \|\nabla u_i - d_i + b_i\|^2 \right) \\ d_i &\leftarrow \arg \min_{d_i} \left(\frac{\mu}{2} \|\nabla u_i - d_i + b_i\|^2 + \alpha \|d_i\|_1 \right) \\ b_i &\leftarrow b_i + \nabla u_i - d_i \end{aligned}$$

Or, using the optimality conditions for u and d

$$\begin{aligned} u_i &\leftarrow (I - \mu\Delta)^{-1} (f_i^k + \mu \operatorname{div}(d_i - b_i)) \\ d_i &\leftarrow \operatorname{shrink} \left(\nabla u_i + b_i, \frac{\alpha}{\mu} \right) \\ b_i &\leftarrow b_i + \nabla u_i - d_i \end{aligned}$$

In our implementation we use an adaptive μ as well as a stopping criterion in the inner loop based on the primal and dual residual, see [9] for details.

6.1.2 Infimal-Convolution Color Bregman TV

To compute the minimizers needed for each infimal convolution iteration, we again apply the ADMM algorithm. For the sake of avoiding too many indices, we omit additional indices indicating the iteration number of the minimization algorithm. We rewrite the minimization with respect to u_i as follows

$$\begin{aligned}
& E(u_i, (\phi_j, d_i, d_{j,+}, d_{j,-})_{j=1,\dots,c}) \\
&= \frac{\lambda}{2} \|u_i - f_i\|^2 + w_{i,i} (\|d_i\|_1 - \langle q_i^k, d_i \rangle) + \frac{\mu}{2} \|\nabla u_i - d_i + b_i\|^2 \\
&+ \sum_{j \neq i} w_{i,j} (\|d_{j,+}\|_1 - \langle q_j^k, d_{j,+} \rangle + \|d_{j,-}\|_1 + \langle q_j^k, d_{j,-} \rangle) \\
&+ \frac{\mu}{2} \sum_{j \neq i} (\|\nabla(u_i - \phi_j) - d_{j,+} + b_{j,+}\|^2 + \|\nabla \phi_j - d_{j,-} + b_{j,-}\|^2),
\end{aligned}$$

which corresponds to a splitting with the constraints $\nabla u_i = d_i$, $\nabla(u_i - \phi_j) = d_{j,+}$ and $\nabla \phi_j = d_{j,-}$. We solve for $(u_i, \{\phi_j\})$, and $(d_i, \{d_{j,+}\}, \{d_{j,-}\})$ in an alternating fashion and obtain

$$0 = \lambda(u_i - f_i) - \mu \operatorname{div}(\nabla u_i - d_i + b_i + \sum_{j \neq i} (\nabla u_i - \nabla \phi_j - d_{j,+} + b_{j,+}))$$

$$0 = \mu \operatorname{div}(\nabla u_i - \nabla \phi_j - d_{j,+} + b_{j,+} - \nabla \phi_j + d_{j,-} - b_{j,-}) \quad \forall j \neq i$$

as the optimality conditions for $(u_i, \{\phi_j\})$, which we can rewrite as

$$(\lambda - \mu c \Delta) u_i + \sum_{j \neq i} \mu \Delta \phi_j = \lambda f_i + \mu \div (b_i - d_i + \sum_{j \neq i} (b_{j,+} - d_{j,+})) \quad (76)$$

$$\Delta u_i - 2 \Delta \phi_j = \operatorname{div}(d_{j,+} - b_{j,+} + b_{j,-} - d_{j,-}) \quad \forall j \neq i \quad (77)$$

Notice that we can use the equations (77) in (76) to eliminate all ϕ_j such that solving for u_i reduces to the linear equation

$$(\lambda - \frac{\mu c + 1}{2} \Delta) u_i = \lambda f_i + \mu \operatorname{div}(b_i - d_i + \frac{1}{2} \sum_{j \neq i} (b_{j,+} - d_{j,+} + b_{j,-} - d_{j,-}))$$

After finding u_i , equations (77) yield a Poisson equation for each ϕ_j . Both, the solving for u_i and solving for the ϕ_j can be done very efficiently using the discrete cosine transform. The minimization with respect to $(d_i, \{d_{j,+}\}, \{d_{j,-}\})$ yields simple shrinkage formulas for each variable:

$$\begin{aligned}
d_i &= \operatorname{shrink} \left(\nabla u_i + b_i + \frac{w_{i,i}}{\mu} q_i^k, \frac{w_{i,i}}{\mu} \right) \\
d_{j,+} &= \operatorname{shrink} \left(\nabla(u_i - \phi_j) + b_{j,+} + \frac{w_{i,j}}{\mu} q_j^k, \frac{w_{i,j}}{\mu} \right) \\
d_{j,-} &= \operatorname{shrink} \left(\nabla \phi_j + b_{j,-} - \frac{w_{i,j}}{\mu} q_j^k, \frac{w_{i,j}}{\mu} \right)
\end{aligned}$$

Finally, we update b_i , $b_{j,+}$ and $b_{j,-}$. The formulas for the updates are the typical ADMM updates and are left out here for the sake of brevity. After convergence we use

$$q_i^{k+1} = q_i^k + \frac{\mu}{w_{i,i}} b_i$$

to compute the q^{k+1} for the next iteration.

6.2 Denoising Results

6.2.1 Artificial Images

We presented detailed discussions about the behavior of the color Bregman as well as the infimal convolution Bregman approach regarding their properties in the cases where the data channels do or do not have a common subgradient. As shown above the property of having a common subgradient translates to the edges being at the same positions and pointing into the same direction when using TV regularization. To closely investigate the effect of this property on color Bregman and infimal convolution iteration, we generated several simple testimages with two nested, differently colored squares, such that the jumps sometimes point in the same and sometimes point in opposite directions. We ran both of the proposed methods for 20 iterations and illustrate the temporal evolution of the iteration in a sequence of images in the accompanying supplementary material. A more detailed description of the corresponding experimental setup can also be found in the supplementary material.

6.2.2 Natural Color Images

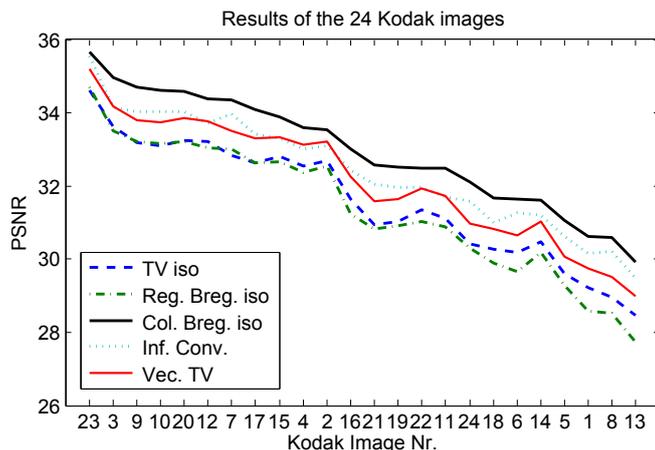


Figure 2: PSNR results for the 24 images of the Kodak image set ($\sigma = 0.05$), sorted in descending order.

Despite the illustration of the behavior of the proposed methods on the artificial test images, the success of the proposed methods in real applications will depend on how well the assumption of common edge sets and edges pointing in the same direction is met in practice. Therefore, we will evaluate the denoising results on a set of real and realistic images.

For our numerical experiments we used the 24 images of the Kodak image data set (c.f. <http://r0k.us/graphics/kodak/>), a representative set of color images. We added Gaussian noise with a standard deviation of $\sigma = 0.05$ (for images on a scale from 0 to 1) to obtain noisy color images and applied a channel-by-channel TV, regular Bregman iteration, infimal convolution Bregman iteration and color Bregman iteration to the noisy images and additionally compare our results to the vectorial total variation (VTV) approach proposed by Bresson and Chan in [12], for which the Matlab code is available on the author's website <http://www.cs.cityu.edu.hk/~xbresson/codes.html>.

The mean peak signal-to-noise ratio (PSNR) values over all 24 Kodak images are shown in Tab. 1. We included the TV and Bregman iteration based TV results using the anisotropic

Method	PSNR
Anisotropic TV	31.60
Isotropic TV	31.59
Anisotropic channel-by-channel Bregman iteration	31.07
Isotropic channel-by-channel Bregman iteration	31.38
Anisotropic color Bregman iteration	32.63
Isotropic color Bregman iteration	32.95
Isotropic Infimal convolution Bregman iteration	32.41
Vectorial TV	32.16

Table 1: Mean PSNR results of the 24 Kodak images with AWGN ($\sigma = 0.05$).

as well as the isotropic definition of the TV. As we can see, isotropic color Bregman iteration yields the highest and infimal convolution Bregman iteration the second highest PSNR values. The mean PSNR using isotropic color Bregman iteration is about 1.35 dB higher than the PSNR using classical TV. In comparison to VTV, the infimal convolution approach yields a small improvement while color Bregman iteration can further improve the infimal convolution results by more than 0.5db.

Figure 2 shows the PSNR results of all methods on all 24 Kodak images as a line plot. For the sake of readability the results are sorted in descending order based on the color Bregman PSNR value. As we can see the color Bregman iteration results show the highest PSNR for all the 24 images. The infimal convolution approach outperforms VTV in most cases but remains clearly behind the color Bregman iteration results. Thus, for our test data, the stricter assumption of the edges pointing into the same direction seems to help much more with the suppression of the noise than it harms at parts where the assumption might be violated.

A visual comparison of the results obtained on Kodak image 23 (see figure 3) shows the advantage of obtaining higher contrast images with color Bregman iteration. Notice that of the proposed method: The contrast and detail recovery is highest in the color Bregman results. Notice that a higher noise level of $\sigma = 0.15$ was chosen in this example to better show the differences of the method results. The reference and the noisy image crop of the Kodak image 23 are shown in Fig. 3(a) and 3(b) while Fig. 3(c) - 3(f) show the results of classical channel-by-channel TV, vectorial TV and the two proposed methods.

In Fig. 3(e) and Fig. 3(f) the results of the proposed infimal convolution and the color Bregman iteration are shown. Around the eye of the bird we observe higher contrast in the infimal convolution and the color Bregman iteration results compared to the TV results. The black color in the circle around the eye can be recovered completely using the color Bregman iteration. The contrast using infimal convolution Bregman iteration is still higher than the contrast of VTV (Fig. 3(d)), thus the mentioned circle is darker in the infimal convolution result than in the VTV result. The color Bregman iteration is the only method that brings back details (dark spots) on the blue wing of the bird.

A visually unpleasant effect of TV denoising is the staircasing effect, which is especially visible in the smooth white-to-green transition in the right part of the image. In future research it might therefore be interesting to combine improvements and extension of the total variation (such as the total generalized variation proposed in [10]) with color Bregman iteration.

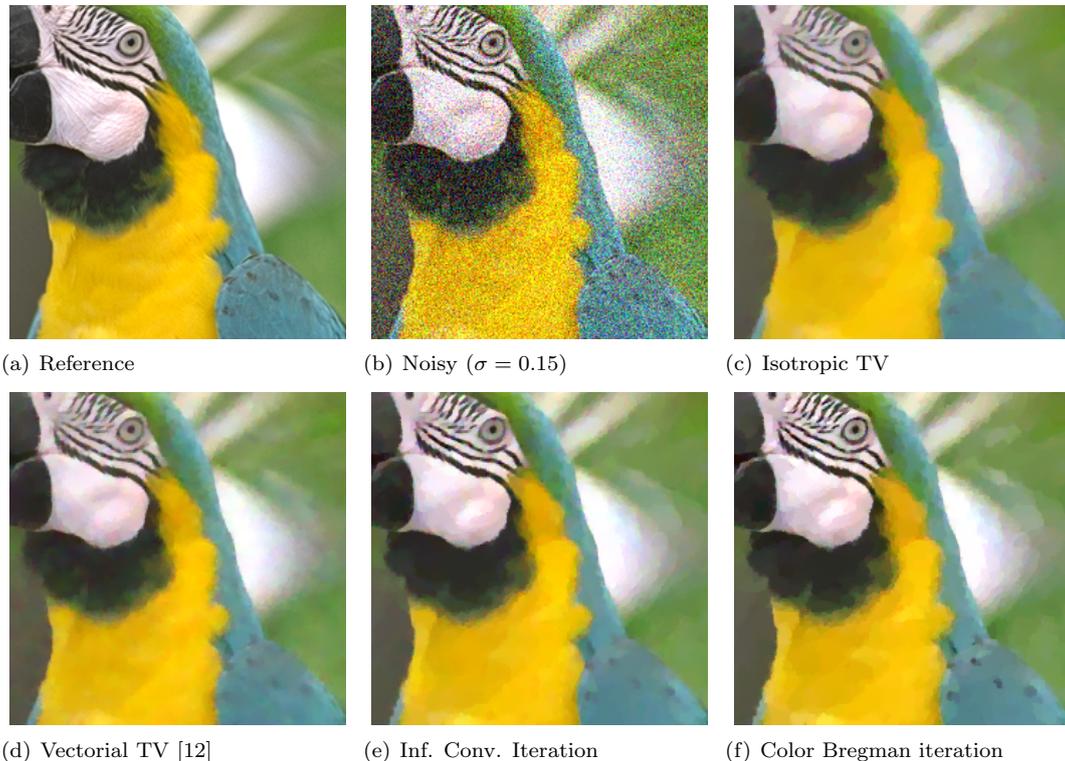


Figure 3: Crop of the Kodak image nr. 23.

7 Conclusions and Further Extensions

In this paper we proposed an iterative procedure for the reconstruction of multichannel data using Bregman distances, which encourages a common subgradient of all channels. In the case of TV regularization it leads to the image channels sharing a common edge set as well as a common edge direction. The latter effect can be avoided by using the infimal convolution of Bregman distances. We proved the convergence of the new iterative procedure, analyzed stationary points and demonstrated its effectiveness with numerical experiments on color image denoising.

Since the idea of coupling subgradients with a weight matrix W in a Bregman iteration is a completely novel idea, there are a number of open questions which go beyond the scope of this paper and will be subject to future research. Nevertheless, we would like to point out which open questions and further ideas we find particularly interesting.

- First of all, the infimal convolution iteration is lacking a rigorous analysis such as a convergence proof, convergence speed, or stationary points of the iteration.
- Secondly, the above color Bregman iteration needs to be tested in detail in different settings (e.g. denoising, demosaicking, inpainting, or deconvolution) and compared to state of the art methods not related to Bregman iteration, such as variants of non-local means or dictionary learning approaches. As we have pointed out in section 4.3 there are two different types of color Bregman iterations in the cases where K^* and W do not commute. While in the prominent case of denoising the commutation is obvious

since $K = I$, an operator acting differently on different color channels, e.g. the operator arising in demosaicking, will not have this property. It will be interesting to investigate the differences between the two different Bregman iterations in this case.

- Thirdly, the application of our approach to multimodal imaging data will be interesting. Notice that the proposed Bregman iteration only assumes a relation between the edge sets of the image channels and not between the image values such that joint reconstruction of very different types of images is possible with each channel profiting from the information in the other channels.
- Finally, notice that we formulated the above Bregman iteration as $q^{k+1} = Wq^k + \lambda(f - Ku^{k+1})$, where the most important parts of the convergence analysis are based on W and $I - W$ being positive semi-definite and $\|W\| \leq 1$. At least in the discrete setting, one could consider to not only couple different image channels, but also couple patches within each channel in a non-local fashion. Similar to non-local means (or non-local TV, c.f. [23, 37]) one could determine similarity weights between image patches and use these weights in the operator/matrix W to couple the parts of the subgradient of different patches. Notice that this technique could find significantly more similar patches than non-local means since not the patches themselves, but only their edge sets have to be similar: A red patch with a jump could help to denoise a green patch with a jump. This idea will also be subject to future research.

8 Acknowledgements

The authors would like to thank Martin Benning for providing efficient MATLAB code for the inversion of the operator $I + \gamma\Delta$. This work was partly supported by the German Ministry for Science and Education (BMBF) via the project HYPERMATH. MB further acknowledges support from the German Science Foundation (DFG) via projects BU 2327/6-1 and BU 2327/8-1.

References

- [1] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, Oxford University Press, New York, 2000.
- [2] G. AUBERT AND J.-F. AUJOL, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946.
- [3] G. AUBERT AND P. KORNPBST, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer, 2006.
- [4] C. BALLESTER AND J. VERDERA, *A variational model for P+XS image fusion*, in Workshop on Variational and Level Set Methods, 2003.
- [5] M. BENNING, C. BRUNE, M. BURGER, AND J. MÜLLER, *Higher-order TV methods—enhancement via Bregman iteration*, J. Sci. Comput., 54 (2013), pp. 269–310.
- [6] M. BENNING AND M. BURGER, *Ground states and singular vectors of convex variational regularization methods*. To appear in Methods and Applications of Analysis.

- [7] S. BIANCHINI AND D. TONON, *A decomposition theorem for BV functions*, Communications on Pure and Applied Analysis, 10 (2011), pp. 1549–1566.
- [8] P. BLOMGREN AND T. F. CHAN, *Color TV: Total variation methods for restoration of vector valued images*, IEEE Trans. Image Processing, 7 (1996), pp. 304–309.
- [9] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
- [10] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM J. Img. Sci., 3 (2010), pp. 492–526.
- [11] L. BREGMAN, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. Math. Phys., 7 (1967), pp. 200–217.
- [12] X. BRESSON AND T. F. CHAN, *Fast dual minimization of the vectorial total variation norm and applications to color image processing*, Inverse Problems and Imaging, 2 (2008), pp. 255–284.
- [13] C. BRUNE, A. SAWATZKY, AND M. BURGER, *Primal and dual Bregman methods with application to optical nanoscopy*, Int. J. Comput. Vis., 92 (2011), pp. 211–229.
- [14] M. BURGER, G. GILBOA, S. OSHER, AND J. XU, *Nonlinear inverse scale space methods*, Comm. Math. Sci., 4 (2006), pp. 179–212.
- [15] M. BURGER, E. RESMERITA, AND L. HE, *Error estimation for bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135.
- [16] J. CAI, S. OSHER, AND Z. SHEN, *Linearized bregman iterations for compressed sensing*, Math. Comp., 78 (2009), pp. 1515–1536.
- [17] T. CHAN AND J. SHEN, *Image Processing and Analysis*, SIAM, Philadelphia, 2005.
- [18] C. CONDAT, *Joint demosaicking and denoising by total variation minimization*, in IEEE Conference on Image Processing, 2012, pp. 2781–2784.
- [19] I. DAUBECHIES, B. HAN, A. RON, AND Z. SHEN, *Framelets: MRA-based constructions of wavelet frames*, Applied and Computational Harmonic Analysis, 14 (2003), pp. 1 – 46.
- [20] K. DEIMLING, *Nonlinear Functional Analysis*, Springer, Berlin, 1985.
- [21] M. J. EHRHARDT AND S. ARRIDGE, *Vector-valued image processing by parallel level sets*, IEEE Trans. on Image Processing, (2012).
- [22] L. EVANS AND R. GARIEPY, *Measure theory and fine properties of functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.
- [23] G. GILBOA AND S. OSHER, *Nonlocal operators with applications to image processing*, Multiscale Modeling and Simulation, 7 (2009), pp. 1005–1028.
- [24] B. GOLDLUECKE AND D. CREMERS, *An approach to vectorial total variation based on geometric measure theory*, in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 327–333.

- [25] T. GOLDSTEIN AND S. OSHER, *The split Bregman method for l_1 regularized problems*, SIAM J. on Imaging Sci., 2 (2008), pp. 323–343.
- [26] G. KUTYNIOK AND D. LABATE, *Shearlets: Multiscale Analysis for Multivariate Data*, Applied and numerical harmonic analysis, Springer, 2012.
- [27] T. LE, R. CHANRTRAND, AND T. J. ASAKI, *A variational approach to reconstructing images corrupted by poisson noise*, Journal of Mathematical Imaging and Vision, 27 (2007), pp. 257–263.
- [28] A. MARQUINA AND S. OSHER, *Image super-resolution by TV-regularization and Bregman iteration*, J. Sci. Comput., 37 (2008), pp. 367–382.
- [29] M. BURGER AND S. OSHER, *A guide to the tv zoo*, in Level Set and PDE -based Reconstruction Methods in Imaging, M. Burger and S. Osher, eds., Springer, Berlin, 2013.
- [30] J. MODERSITZKI, *FAIR: Flexible Algorithms for Image Registration*, SIAM, Philadelphia, 2009.
- [31] M. MOELLER, *Multiscale Methods for Polyhedral Regularizations and Applications in High Dimensional Imaging*, PhD thesis, University of Muenster, Germany, 2012.
- [32] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, SIAM Multiscale Model. Simul., 4 (2005), pp. 460–489.
- [33] V. S. PRASATH, J. C. MORENO, AND K. PALANIAPPAN, *Color image denoising by chromatic edges based vector valued diffusion*. Submitted to IEEE Signal Processing Letters, arXiv:1304.5587v2.
- [34] E. RESMERITA AND O. SCHERZER, *Error estimates for non-quadratic regularization and the relation to enhancement*, Inverse Problems, 22 (2006), pp. 801–814.
- [35] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
- [36] G. SAPIRO AND D. RINGACH, *Anisotropic diffusion of multivalued images with applications to color filtering*, IEEE Trans. Img. Proc., 5 (1996), pp. 1582–1586.
- [37] A. SAWATZKY, *(Nonlocal) Total Variation in Medical Imaging*, PhD thesis, University of Muenster, 2011.
- [38] A. SAWATZKY, C. BRUNE, J. MUELLER, AND M. BURGER, *Total variation processing of images with poisson statistics*, in Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns, E. X. Jiang and N. Petkov, eds., vol. 5702 of LNCS, Springer, july 2009, pp. 533–540.
- [39] A. SAWATZKY, D. TENBRINCK, X. JIANG, AND M. BURGER, *A variational framework for region-based segmentation incorporating physical noise models*, Journal of Mathematical Imaging and Vision, (2013).
- [40] P. SPRECHMANN, I. RAMIREZ, G. SAPIRO, AND Y. C. ELДАР, *C-hilasso: A collaborative hierarchical sparse modeling framework*, IEEE Trans. Sig. Proc., 59 (2011), pp. 4183–4198.

- [41] J.-L. STARCK, E. CANDES, AND D. DONOHO, *The curvelet transform for image denoising*, Image Processing, IEEE Transactions on, 11 (2002), pp. 670–684.
- [42] D. TSCHUMPERLE AND R. DERICHE, *Vector-valued image regularization with PDEs: A common framework for different applications*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, pp. 506–517.
- [43] J. XU AND S. OSHER, *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising*, IEEE Trans. on Image Processing, 16 (2007), pp. 534–544.
- [44] W. YIN, *Analysis and generalizations of the linearized Bregman method*, SIAM J. Imaging Sci., 3 (2010), pp. 856–877.
- [45] X. ZHANG, M. BURGER, X. BRESSON, AND S. OSHER, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, SIAM J. Imaging Sci., 3 (2010), pp. 253–276.