Split Bregman Algorithm, Douglas-Rachford Splitting and Frame Shrinkage

Simon Setzer

University of Mannheim, A5, 68131 Mannheim, Germany
ssetzer@kiwi.math.uni-mannheim.de
http://kiwi.math.uni-mannheim.de

Abstract. We examine relations between popular variational methods in image processing and classical operator splitting methods in convex analysis. We focus on a gradient descent reprojection algorithm for image denoising and the recently proposed Split Bregman and alternating Split Bregman methods. By identifying the latter with the so-called Douglas-Rachford splitting algorithm we can guarantee its convergence. We show that for a special setting based on Parseval frames the gradient descent reprojection and the alternating Split Bregman algorithm are equivalent and turn out to be a frame shrinkage method.

1 Introduction

In recent years variational models were successfully applied in image restoration. These methods came along with various computational algorithms. Interestingly, the roots of many restoration algorithms can be found in classical algorithms from convex analysis dating back more than 40 years. It is useful from different points of view to discover these relations: Classical convergence results carry over to the restoration algorithms at hand and ensure their convergence. On the other hand, earlier mathematical results have found new applications and should be acknowledged.

The present paper fits into this context. Our aim is twofold: First, we show that the *Alternating Split Bregman Algorithm* proposed by Goldstein and Osher for image restoration and compressed sensing can be interpreted as a *Douglas-Rachford Splitting Algorithm*. In particular, this clarifies the convergence of the algorithm. Second, we consider the following denoising problem which uses an L_2 data-fitting and a Besov-norm regularization term [1]

$$\operatorname*{argmin}_{u \in B^1_{1,1}(\Omega)} \{\frac{1}{2}\|u - f\|^2_{L_2(\Omega)} + \lambda \|u\|_{B^1_{1,1}(\Omega)}\}. \tag{1}$$

We show that for discrete versions of this problem involving Parseval frames the corresponding alternating Split Bregman Algorithm can be seen as an application of a *Forward-Backward Splitting Algorithm*. The latter is also related to the *Gradient Descent Reprojection Algorithm*, see Chambolle [2]. Since our methods are based on soft (coupled) frame shrinkage, we also establish the relation to the

X.-C. Tai et al. (Eds.): SSVM 2009, LNCS 5567, pp. 464–476, 2009.

classical wavelet shrinkage scheme. Finally, we consider the Rudin-Osher-Fatemi model [3]

$$\operatorname*{argmin}_{u \in BV(\Omega)} \frac{1}{2}\|u-f\|^2_{L_2(\Omega)} + \lambda \int_\Omega |\nabla u(x)|\, dx, \tag{2}$$

which is a successful edge-preserving image restoration method. We apply our findings to create an efficient frame-based minimization algorithm for the discrete version of this problem.

2 Operator Splitting Methods

Proximation and Soft Shrinkage. We start by considering the *proximity operator*

$$\operatorname{prox}_{\gamma\Phi}(f) := \operatorname*{argmin}_{u \in H}\{\frac{1}{2\gamma}\|u-f\|^2 + \Phi(u)\} \tag{3}$$

on a Hilbert space H. If $\Phi : H \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lower semi-continuous (lsc), then for any $f \in H$, there exists a unique minimizer $\hat{u} := \operatorname{prox}_{\gamma\Phi}(f)$ of (3). By Fermat's rule, this minimizer is determined by the inclusion

$$\begin{aligned} 0 &\in \frac{1}{\gamma}(\hat{u} - f) + \partial\Phi(\hat{u}) \\ \Leftrightarrow f &\in \hat{u} + \gamma\partial\Phi(\hat{u}) \quad \Leftrightarrow \quad \hat{u} = (I + \gamma\partial\Phi)^{-1} f, \end{aligned}$$

where the set-valued function $\partial\Phi : H \to 2^H$ is the *subdifferential* of Φ. If Φ is proper, convex and lsc, then $\partial\Phi$ is a maximal monotone operator. For a set-valued function $F : H \to 2^H$, the operator $J_F := (I + F)^{-1}$ is called the *resolvent* of F. If F is maximal monotone, then J_F is single-valued and firmly nonexpansive.
In this paper, we are mainly interested in the following two functions Φ_i, $i = 1, 2$, on $H := \mathbb{R}^M$:

i) $\Phi_1(u) := \|\Lambda u\|_1$ with $\Lambda := \operatorname{diag}(\lambda_j)_{j=1}^M$, $\lambda_j \geq 0$,
ii) $\Phi_2(u) := \|\tilde{\Lambda}\,|u|\,\|_1$ with $\tilde{\Lambda} := \operatorname{diag}(\tilde{\lambda}_j)_{j=1}^N$, $\tilde{\lambda}_j \geq 0$ and $|u| := \Big(\|\mathbf{u}_j\|_2\Big)_{j=1}^N$ for $\mathbf{u}_j := (u_{j+kN})_{k=0}^{p-1}$ and $M = pN$.

The corresponding Fenchel conjugate functions are given by

i) $\Phi_1^*(u) := \iota_C(u)$ with $C := \{u \in \mathbb{R}^M : |u_j| \leq \lambda_j,\ \ j = 1, \ldots, M\}$,
ii) $\Phi_2^*(u) := \iota_{\tilde{C}}(u)$ with $\tilde{C} := \{u \in \mathbb{R}^M : \|\mathbf{u}_j\|_2 \leq \tilde{\lambda}_j,\ \ j = 1, \ldots, N\}$,

where ι_C the indicator function of the set C (or $\tilde{C}$), i.e., $\iota_C(u) := 0$ for $u \in C$ and $\iota_C(u) := +\infty$ otherwise. A short calculation shows that for any $f \in \mathbb{R}^M$ we have

$$\operatorname{prox}_{\Phi_1}(f) = T_\Lambda(f), \quad \operatorname{prox}_{\Phi_2}(f) = \tilde{T}_{\tilde{\Lambda}}(f),$$

where T_Λ denotes the *soft shrinkage* function given componentwise by

$$T_{\lambda_j}(f_j) := \begin{cases} 0 & \text{if } |f_j| \le \lambda_j, \\ f_j - \lambda_j \operatorname{sgn}(f_j) & \text{if } |f_j| > \lambda_j, \end{cases} \tag{4}$$

and $\tilde{T}_{\tilde{\Lambda}}$ denotes the *coupled shrinkage* function, compare [2, 4, 5],

$$\tilde{T}_{\tilde{\lambda}_j}(\mathbf{f}_j) := \begin{cases} 0 & \text{if } \|\mathbf{f}_j\|_2 \le \tilde{\lambda}_j, \\ \mathbf{f}_j - \tilde{\lambda}_j \mathbf{f}_j / \|\mathbf{f}_j\|_2 & \text{if } \|\mathbf{f}_j\|_2 > \tilde{\lambda}_j. \end{cases}$$

Similarly, we obtain

$$\operatorname{prox}_{\Phi_1^*}(f) = f - T_\Lambda(f), \quad \operatorname{prox}_{\Phi_2^*}(f) = f - \tilde{T}_{\tilde{\Lambda}}(f). \tag{5}$$

Operator Splittings. Now we consider more general minimization problems of the form

$$(P) \qquad \min_{u \in H_1} \{ \underbrace{g(u) + \Phi(Du)}_{:=\mathcal{F}_P(u)} \},$$

where $D : H_1 \to H_2$ is a bounded linear operator and both functions $g : H_1 \to \mathbb{R} \cup \{+\infty\}$ and $\Phi : H_2 \to \mathbb{R} \cup \{+\infty\}$ are proper, convex and lsc. Furthermore, we assume that $0 \in \operatorname{int}(D \operatorname{dom}(g) - \operatorname{dom}(\Phi))$. For $g(u) := \frac{1}{2\gamma}\|u - f\|^2$ and $D = I$ this is again our proximation problem. The corresponding dual problem has the form

$$(D) \qquad -\min_{b \in H_2} \{ \underbrace{g^*(-D^*b) + \Phi^*(b)}_{:=\mathcal{F}_D(b)} \}.$$

We assume that solutions $\hat{u}$ and $\hat{b}$ of the primal and dual problems, respectively, exist and that the duality gap is zero. In other words, we suppose that there is a pair $(\hat{u}, \hat{d})$ which satisfies the *Karush-Kuhn-Tucker conditions* $0 \in \partial g(\hat{u}) + D^*\hat{b},\ 0 \in -D\hat{u} + \partial\Phi^*(\hat{b})$. Then $\hat{u}$ is a solution of (P) if and only if

$$0 \in \partial \mathcal{F}_P(\hat{u}) = \partial g(\hat{u}) + \partial(\Phi \circ D)(\hat{u}).$$

Similarly, a solution $\hat{b}$ of the dual problem is characterized by

$$0 \in \partial \mathcal{F}_D(\hat{b}) = \partial(g^* \circ (-D^*))(\hat{b}) + \partial\Phi^*(\hat{b}).$$

In both primal and dual problem, one finally has to solve an inclusion of the form

$$0 \in A(\hat{p}) + B(\hat{p}). \tag{6}$$

Various splitting techniques make use of this additive structure. In this paper, we restrict our attention to the *forward-backward splitting* (FBS) and the *Douglas-Rachford splitting* (DRS). The inclusion (6) can be rewritten as fixed point equation

$$\hat{p} - \eta B(\hat{p}) \in \hat{p} + \eta A(\hat{p}) \quad \Leftrightarrow \quad \hat{p} \in J_{\eta A}(I - \eta B)\hat{p}, \quad \eta > 0 \tag{7}$$

and the FBS algorithm is just the corresponding iteration. For the following convergence result and generalizations of the algorithm we refer to [6, 7, 8, 9].

Theorem 1 (**FBS**). *Let* $A : H \to 2^H$ *be a maximal monotone and* $\beta B : H \to H$ *be firmly nonexpansive for some* $\beta > 0$. *Furthermore, assume that a solution of* (6) *exists. Then, for any* $p^{(0)}$ *and any* $\eta \in (0, 2\beta)$ *the following* FBS algorithm *converges weakly to such a solution of* (6)

$$p^{(k+1)} = J_{\eta A}(I - \eta B)p^{(k)}. \tag{8}$$

To introduce the DRS, we rewrite the right-hand side of (7) as

$$\hat{p} + \eta B\hat{p} \in J_{\eta A}(I - \eta B)\hat{p} + \eta B\hat{p} \quad \Leftrightarrow \quad \hat{p} \in J_{\eta B}\big(\underbrace{J_{\eta A}(I - \eta B)\hat{p} + \eta B\hat{p}}_{:=\hat{t}}\big)$$

The DRS algorithm [10] is the corresponding iteration, where we use $t^{(k)} := p^{(k)} + \eta B p^{(k)}$. For the following convergence result, which in contrast to the FBS algorithm holds also for set-valued operators B, see [6, 8].

Theorem 2 (**DRS**). *Let* $A, B : H \to 2^H$ *be maximal monotone operators and assume that a solution of* (6) *exists. Then, for any initial elements* $t^{(0)}$ *and* $p^{(0)}$ *and any* $\eta > 0$, *the following* DRS algorithm *converges weakly to an element* $\hat{t}$:

$$\begin{aligned} t^{(k+1)} &= J_{\eta A}(2p^{(k)} - t^{(k)}) + t^{(k)} - p^{(k)}, \\ p^{(k+1)} &= J_{\eta B}(t^{(k+1)}). \end{aligned}$$

Furthermore, it holds that $\hat{p} := J_{\eta B}(\hat{t})$ *satisfies* $0 \in A(\hat{p}) + B(\hat{p})$. *If* H *is finite-dimensional, then the sequence* $\big(p^{(k)}\big)_{k \in \mathbb{N}}$ *converges to* $\hat{p}$.

3 Bregman Methods

For a function $\varphi : H \to \mathbb{R} \cup \{+\infty\}$, the *Bregman distance* $D_\varphi^{(p)}$ is defined as

$$D_\varphi^{(p)}(u, v) = \varphi(u) - \varphi(v) - \langle p, u - v \rangle,$$

with $p \in \partial\varphi(v)$, cp. [11]. Given an arbitrary initial value $u^{(0)}$ and a parameter $\gamma > 0$, the *Bregman proximal point algorithm* (BPP) *applied to* (P) has the form [12, 13, 14]

$$u^{(k+1)} = \operatorname*{argmin}_{u \in H_1} \{ \frac{1}{\gamma} D_\varphi^{(p^{(k)})}(u, u^{(k)}) + \mathcal{F}_P(u) \}, \quad p^{(k+1)} \in \partial\varphi(u^{(k+1)}). \tag{9}$$

For conditions on φ such that $(u^{(k)})_{k \in \mathbb{N}}$ converges to a minimizer of (P), see [13] and the references therein. For $\varphi := \frac{1}{2}\|\cdot\|_2^2$, we recover the classical *proximal point algorithm* (PP) *for* (P) which can be written as follows, compare [15],

$$u^{(k+1)} = \operatorname{prox}_{\gamma\mathcal{F}_P}(u^{(k)}) = \operatorname*{argmin}_{u \in H_1} \big\{ \frac{1}{2\gamma}\|u - u^{(k)}\|_2^2 + \mathcal{F}_P(u) \big\} = J_{\gamma\partial\mathcal{F}_P}(u^{(k)}).$$

Under our assumptions on g, Φ and D, the weak convergence of the PP algorithm is guaranteed for any initial point $u^{(0)}$, see [16]. In the same way, we can define the PP *algorithm for* (D)

$$b^{(k+1)} = \mathrm{prox}_{\gamma \partial \mathcal{F}_D}(b^{(k)}) = \operatorname*{argmin}_{b \in H_2} \big\{ \frac{1}{2\gamma} \|b - b^{(k)}\|_2^2 + \mathcal{F}_D(b) \big\} = J_{\gamma \partial \mathcal{F}_D}(b^{(k)})$$

and the same convergence result holds true. It is well-known that the PP algorithm applied to (D) is equivalent to the *augmented Lagrangian method* (AL) for the primal problem, see, e.g., [15,14]. To define this algorithm we first transform (P) into the constrained minimization problem

$$\min_{u \in H_1, d \in H_2} E(u,d) \quad \text{s.t.} \quad Du = d, \tag{10}$$

where $E(u,d) := g(u) + \Phi(d)$. This problem was introduced in [29]. The corresponding AL *algorithm for* (P) is then defined as

$$\begin{aligned} (u^{(k+1)}, d^{(k+1)}) &= \operatorname*{argmin}_{u \in H_1, d \in H_2} \big\{ E(u,d) + \langle b^{(k)}, Du - d \rangle + \frac{1}{2\gamma} \|Du - d\|_2^2 \big\} \\ b^{(k+1)} &= b^{(k)} + \frac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)}). \end{aligned} \tag{11}$$

Indeed, it has been shown that for the same initial value $b^{(0)}$ the sequence $(b^{(k)})_{k \in \mathbb{N}}$ coincides with the one produced by the PP algorithm applied to (D), see [15]. Moreover, if $(b^{(k)})_{k \in \mathbb{N}}$ converges strongly then every strong cluster point of $(u^{(k)})_{k \in \mathbb{N}}$ is a solution of (P), cf. [17]. To solve the constrained optimization problem (10), Goldstein and Osher [18] proposed to use the Bregman distance

$$D_E^{(p^{(k)})}(u,d,u^{(k)},d^{(k)}) = E(u,d) - E(u^{(k)},d^{(k)}) - \langle p_u^{(k)}, u - u^{(k)} \rangle - \langle p_d^{(k)}, d - d^{(k)} \rangle$$

and the term $\frac{1}{2\gamma}\|Du - d\|_2^2$ instead of $\mathcal{F}_P$ in (9). This results in the algorithm

$$(u^{(k+1)}, d^{(k+1)}) = \operatorname*{argmin}_{u \in H_1, d \in H_2} \big\{ D_E^{(p^{(k)})}(u,d,u^{(k)},d^{(k)}) + \frac{1}{2\gamma}\|Du - d\|_2^2 \big\}, \tag{12}$$

$$p_u^{(k+1)} = p_u^{(k)} - \frac{1}{\gamma} D^*(Du^{(k+1)} - d^{(k+1)}),\ p_d^{(k+1)} = p_d^{(k)} + \frac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)}),$$

where we have used that (12) implies

$$\begin{aligned} 0 \in\ & \partial E(u^{(k+1)}, d^{(k+1)}) - \big(p_u^{(k)}, p_d^{(k)}\big) \\ & + \big(\frac{1}{\gamma} D^*(Du^{(k+1)} - d^{(k+1)}), -\frac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)})\big), \\ =\ & \partial E(u^{(k+1)}, d^{(k+1)}) - \big(p_u^{(k+1)}, p_d^{(k+1)}\big), \end{aligned}$$

so that $\big(p_u^{(k)}, p_d^{(k)}\big) \in \partial E(u^{(k)}, d^{(k)})$. Setting $p_u^{(k)} = -\frac{1}{\gamma} D^* b^{(k)}$ and $p_d^{(k)} = \frac{1}{\gamma} b^{(k)}$ for all $k \geq 0$ and regarding that for a bounded linear operator D,

$$\begin{aligned} D_E^{(p^{(k)})}(u,d,u^{(k)},d^{(k)}) + \frac{1}{2\gamma}\|Du - d\|_2^2 &= E(u,d) - E(u^{(k)},d^{(k)}) \\ -\frac{1}{\gamma}\langle b^{(k)}, Du - Du^{(k)} \rangle - \frac{1}{\gamma}\langle b^{(k)}, d - d^{(k)} \rangle &+ \frac{1}{2\gamma}\|Du - d\|_2^2, \end{aligned}$$

Goldstein and Osher obtained the *Split Bregman method* [18]

$$\begin{aligned}(u^{(k+1)}, d^{(k+1)}) &= \underset{u\in H_1, d\in H_2}{\operatorname{argmin}} \left\{E(u,d) + \frac{1}{2\gamma}\|b^{(k)} + Du - d\|_2^2\right\},\\ b^{(k+1)} &= b^{(k)} + Du^{(k+1)} - d^{(k+1)}.\end{aligned} \tag{13}$$

As already discovered in [19], the Split Bregman algorithm (13) is just the AL algorithm (11) with the only difference that in (13) the iterates $b^{(k)}$ are scaled by γ. Hence, we can conclude that the sequence $(\frac{1}{\gamma}b^{(k)})_{k\in\mathbb{N}}$ generated by the Split Bregman method (13) converges to solutions of the dual problem. The same holds true for the sequence $(p_d^{(k)})_{k\in\mathbb{N}}$ we get from (12). To summarize:

$$\boxed{\text{PP for } (D)} = \boxed{\text{AL for } (P)} = \boxed{\text{Split Bregman Alg.}}$$

Since the minimization problem in (13) is hard to solve, Goldstein and Osher [18] proposed the following *alternating Split Bregman algorithm* without a convergence proof:

$$u^{(k+1)} = \underset{u\in H_1}{\operatorname{argmin}} \left\{g(u) + \frac{1}{2\gamma}\|b^{(k)} + Du - d^{(k)}\|_2^2\right\}, \tag{14}$$

$$d^{(k+1)} = \underset{d\in H_2}{\operatorname{argmin}} \left\{\Phi(d) + \frac{1}{2\gamma}\|b^{(k)} + Du^{(k+1)} - d\|_2^2\right\}, \tag{15}$$

$$b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \tag{16}$$

The next theorem identifies this alternating Split Bregman method as a special case of a DRS.

$$\boxed{\text{DRS for } (D)} = \boxed{\text{Alternating Split Bregman Alg.}}$$

If H_1 and H_2 are finite-dimensional it therefore provides us with a convergence result for the sequence $(b^{(k)})_{k\in\mathbb{N}}$ of this algorithm.

Theorem 3. *The alternating Split Bregman algorithm coincides with the DRS algorithm applied to* (D) *with* $A := \partial(g^* \circ (-D^*))$ *and* $B := \partial\Phi^*$*, where* $\eta = 1/\gamma$ *and*

$$t^{(k)} = \eta(b^{(k)} + d^{(k)}), \quad p^{(k)} = \eta b^{(k)}, \quad k \geq 0. \tag{17}$$

Proof: 1. First, we show that for a proper, convex, lsc function $h : H_1 \to \mathbb{R}\cup\{+\infty\}$ and a bounded linear operator $K : H_1 \to H_2$ the following relation holds true:

$$\hat{p} = \underset{p\in H_1}{\operatorname{argmin}} \left\{\frac{\eta}{2}\|Kp - q\|^2 + h(p)\right\} \quad \Rightarrow \quad \eta(K\hat{p} - q) = J_{\eta\,\partial(h^*\circ(-K^*))}(-\eta q). \tag{18}$$

The first equality is equivalent to

$$0 \in \eta K^*(K\hat{p} - q) + \partial h(\hat{p}) \quad \Leftrightarrow \quad \hat{p} \in \partial h^*\big(-\eta K^*(K\hat{p} - q)\big).$$

Applying $-\eta K$ on both sides and adding $-\eta q$ implies

$$\begin{aligned} -\eta K\hat{p} &\in -\eta K \partial h^*\big(-\eta K^*(K\hat{p}-q)\big) \;=\; \eta\,\partial\big(h^*\circ(-K^*)\big)\big(\eta(K\hat{p}-q)\big) \\ -\eta q &\in \big(I+\eta\,\partial(h^*\circ(-K^*))\big)\big(\eta(K\hat{p}-q)\big) \end{aligned}$$

which is by the definition of the resolvent equivalent to the right equality in (18).
2. Applying (18) to (14) with $h := g$, $K := D$ and $q := d^{(k)} - b^{(k)}$ we get

$$\eta(b^{(k)} + Du^{(k+1)} - d^{(k)}) = J_{\eta A}(\eta(b^{(k)} - d^{(k)})).$$

Assume that the alternating Split Bregman iterates and the DRS iterates coincide with the identification (17) up to some $k \in \mathbb{N}$. Using this induction hypothesis it follows that

$$\eta(b^{(k)} + Du^{(k+1)}) = J_{\eta A}(\underbrace{\eta(b^{(k)} - d^{(k)})}_{2p^{(k)}-t^{(k)}}) + \underbrace{\eta d^{(k)}}_{t^{(k)}-p^{(k)}} = t^{(k+1)}. \tag{19}$$

By definition of $b^{(k+1)}$ in (16) we see that $\eta(b^{(k+1)} + d^{(k+1)}) = t^{(k+1)}$. Next we apply (18) to (15) with $h := \Phi$, $K := I$ and $q := b^{(k)} + Du^{(k+1)}$ which gives together with (19),

$$\eta(b^{(k)} + Du^{(k+1)} - d^{(k+1)}) = J_{\eta B}(\underbrace{\eta(b^{(k)} + Du^{(k+1)})}_{t^{(k+1)}}) = p^{(k+1)}.$$

Again by the formula (16) for $b^{(k+1)}$ we obtain $\eta b^{(k+1)} = p^{(k+1)}$ which completes the proof. □

A similar result was shown in [20, 21].

4 Application to Image Denoising

In the following, we restrict our attention to a discrete setting. We consider digital images defined on $\{1,\dots,n\} \times \{1,\dots,n\}$ and reshape them columnwise into vectors $f \in \mathbb{R}^N$ with $N = n^2$. If not stated otherwise the multiplication of vectors, their square root etc. are meant componentwise.

We will now apply the algorithms defined in Sections 2 and 3 to the discrete denoising problem of the form

$$\operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|u-f\|_2^2 + \Phi(Du)\Big\}, \quad D \in \mathbb{R}^{M,N}, \quad M \geq N, \tag{20}$$

where Φ is defined as in Section 2. Consider the alternating Split Bregman algorithm (14)-(16) with $g(u) := \frac{1}{2}\|u-f\|_2^2$. Theorem 3 implies the convergence of $\big(b^{(k)}\big)_{k\in\mathbb{N}}$ and it is not hard to show that for this special choice of g, the sequence $\big(u^{(k)}\big)_{k\in\mathbb{N}}$ converges to a solution of the primal problem. The quadratic functional in (14) with the above choice of g can simply be minimized by setting its gradient to zero which results in

$$u^{(k+1)} = (\gamma I + D^{\mathrm{T}}D)^{-1}\big(\gamma f + D^{\mathrm{T}}(d^{(k)} - b^{(k)})\big).$$

Goldstein and Osher proposed to calculate the inverse $(\gamma I + D^{\mathrm{T}} D)^{-1}$ by Gauß-Seidel iterations. Applying (4) we see that for $\Phi = \Phi_1$ the solution of the proximation problem in (15) is given by

$$d^{(k+1)} = T_{\gamma \Lambda}(b^{(k)} + Du^{(k+1)}).$$

The following algorithm shows the case $\Phi = \Phi_1$. Observe that in order to better compare this method to the other algorithms in this section, we have changed the order in which we compute $u^{(k+1)}$. This is allowed because there are no restrictions on the choice of the starting values.

Algorithm (Alternating Split Bregman Shrinkage)
Initialization: $u^{(0)} := f$, $b^{(0)} := 0$.
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$$\begin{aligned} d^{(k+1)} &:= T_{\gamma \Lambda}(b^{(k)} + Du^{(k)}), \\ b^{(k+1)} &:= b^{(k)} + Du^{(k)} - d^{(k+1)}, \\ u^{(k+1)} &:= (\gamma I + D^{\mathrm{T}} D)^{-1}\big(\gamma f + D^{\mathrm{T}}(d^{(k+1)} - b^{(k+1)})\big). \end{aligned}$$

For $\Phi = \Phi_2$ we have to replace the soft shrinkage $T_{\gamma \Lambda}$ by the coupled shrinkage $\tilde{T}_{\gamma \tilde{\Lambda}}$. Note that this algorithm can also be used for the deblurring problem which differs from (20) in having a more general data-fitting term $g(u) := \frac{1}{2}\|Ku - f\|_2^2$ with some linear operator K. In this case one has to invert the matrix $\gamma K^{\mathrm{T}} K + D^{\mathrm{T}} D$ which can be diagonalized in many applications by FFT or DCT techniques, e.g., if it is circulant.

The problem (20) can also be solved via its dual problem by $\hat{u} = f - D^{\mathrm{T}} \hat{b}$, where

$$\hat{b} = \operatorname*{argmin}_{b \in \mathbb{R}^M} \{\frac{1}{2}\|f - D^{\mathrm{T}} b\|_2^2 + \Phi_i^*(b)\}, \quad i = 1, 2 \tag{21}$$

see, e.g., [22]. Applying the FBS algorithm (8) to the dual problem (21) gives

$$b^{(k+1)} = \mathrm{prox}_{\gamma \Phi_i^*}\left(b^{(k)} + \gamma D(f - D^{\mathrm{T}} b^{(k)})\right), \qquad i = 1, 2,$$

where $0 < \gamma < 2/\|D^{\mathrm{T}} D\|_2$. Using the relation (5) we obtain for $\Phi = \Phi_1$

$$b^{(k+1)} = b^{(k)} + \gamma D(f - D^{\mathrm{T}} b^{(k)}) - T_{\Lambda}\big(b^{(k)} + \gamma D(f - D^{\mathrm{T}} b^{(k)})\big).$$

This yields the following algorithm to compute the minimizer of (20) for $\Phi = \Phi_1$:

Algorithm (FBS Shrinkage)
Initialization: $u^{(0)} := f$, $b^{(0)} := 0$
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$$\begin{aligned} d^{(k+1)} &:= T_{\Lambda}\big(b^{(k)} + \gamma Du^{(k)}\big), \\ b^{(k+1)} &:= b^{(k)} + \gamma Du^{(k)} - d^{(k+1)}, \\ u^{(k+1)} &:= f - D^{\mathrm{T}} b^{(k+1)}. \end{aligned}$$

For the functional Φ_2 we have to replace the shrinkage functional by $\tilde{T}_{\tilde{\Lambda}}$. This algorithm can also be deduced as a simple *gradient descent reprojection algorithm* as it was done, e.g., by Chambolle [2]. Note that this is *not* the often cited Chambolle algorithm in [22]. A relation of this method to the Bermúdez-Moreno algorithm which also turns out to be an FBS algorithm was shown in [23]. A connection to min-max duality was established in [24].

4.1 Besov-Norm Regularization

For a sufficiently smooth orthogonal wavelet basis $\{\psi_i\}_{i\in I}$ of $L_2(\Omega)$ with wavelets of more than one vanishing moment, problem (1) can be rewritten as

$$\frac{1}{2}\|d-c\|_{\ell_2}^2 + \lambda\|d\|_{\ell_1},$$

where $c := (\langle f, \psi_i\rangle)_i$ and $d := (\langle u, \psi_i\rangle)_i$. In the discrete setting, consider the *orthogonal* matrix $W \in \mathbb{R}^{N,N}$ having as rows the filters of orthogonal wavelets (and scaling functions) up to a certain level. Then the minimization problem corresponding to (1) is given by

$$\begin{aligned}\hat{u} &= \operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|u-f\|_2^2 + \|\Lambda W u\|_1\Big\} \\ &= \operatorname*{argmin}_{u\in\mathbb{R}^N} \Big\{\frac{1}{2}\|Wu-Wf\|_2^2 + \|\Lambda W u\|_1\Big\}. \end{aligned} \tag{22}$$

The orthogonality of W yields further $\hat{u} = W^{\mathrm{T}}\hat{d}$, where

$$\hat{d} = \operatorname*{argmin}_{d\in\mathbb{R}^N} \Big\{\frac{1}{2}\|d-c\|_2^2 + \|\Lambda d\|_1\Big\}, \qquad c := Wf,\ \Lambda := \lambda I_N \tag{23}$$

and by (4) we obtain the known *wavelet shrinkage procedure* $\hat{u} = W^{\mathrm{T}}T_\Lambda(Wf)$ consisting of a wavelet transform W followed by soft shrinkage T_Λ of the wavelet coefficients and the inverse wavelet transform W^{T}.

However, for image processing tasks like denoising or segmentation, ordinary orthogonal wavelets are not suited due to their lack of translational invariance which leads to visible artefacts. Nevertheless, without the usual subsampling, the method becomes translationally invariant and the results can be improved. But then $W \in \mathbb{R}^{M,N}$, $M = pN$, where p is three times the decomposition level plus one for the rows belonging to the scaling function filters on the coarsest scale. We still have $W^{\mathrm{T}}W = I_N$, but of course $WW^{\mathrm{T}} \neq I_M$, i.e., the rows of W form a discrete *Parseval frame* on $\mathbb{R}^N$ but not a basis. For the design of such frames see, e.g., [25, 26]. Equality (22) is still true for Parseval frames, but the problem is no longer equivalent to (23). Instead we can apply FBS shrinkage or alternating Split Bregman shrinkage with $D = W$ and $\Phi = \Phi_1$. Note that in order to use the FBS algorithm, γ has to fulfill $0 < \gamma < 2/\|W^{\mathrm{T}}W\|_2$. Now $W^{\mathrm{T}}W = I_N$, thus we have to choose γ in $(0, 2)$ and $\gamma = 1$ is an admissible choice. It was shown in [27] that both algorithms coincide for $D = W$ with $W^{\mathrm{T}}W = I_N$ and $\gamma = 1$:

Alternating Split Bregman Shrinkage	=	FBS Shrinkage

Moreover, the third step of both algorithms can be simplified to the frame synthesis step

$$u^{(k+1)} = W^{\mathrm{T}} d^{(k+1)}. \tag{24}$$

4.2 ROF Regularization

In this section, we apply the algorithms presented so far to the discrete ROF denoising method. We use an appropriate discretization of the absolute value of the gradient. Let $h_0 := \frac{1}{2}[1\ 1]$ and $h_1 := \frac{1}{2}[1\ -1]$ be the filters of the Haar wavelet. For convenience of notation, we use periodic boundary conditions and the corresponding circulant matrices are denoted by $H_0 \in \mathbb{R}^{n,n}$ and $H_1 \in \mathbb{R}^{n,n}$. Then the following matrix fulfills $W^{\mathrm{T}}W = I_N$ but $W^{\mathrm{T}}W \neq I_{4N}$

$$W := \begin{pmatrix} H_0 \otimes H_0 \\ H_0 \otimes H_1 \\ H_1 \otimes H_0 \\ H_1 \otimes H_1 \end{pmatrix} = \begin{pmatrix} \mathcal{H}_0 \\ \\ \mathcal{H}_1 \\ \end{pmatrix}.$$

In [4,5] it was shown that $\Big(\big((H_0 \otimes H_1)\, u\big)^2 + \big((H_1 \otimes H_0)\, u\big)^2 + \big((H_1 \otimes H_1)\, u\big)^2 \Big)^{\frac{1}{2}}$ is a consistent finite difference discretization of $|\nabla u|$. Using this gradient discretization, the discrete version of the ROF functional in (2) reads

$$\operatorname*{argmin}_{u \in \mathbb{R}^N} \big\{ \frac{1}{2} \|u - f\|_2^2 + \|\tilde{\Lambda}\, |\mathcal{H}_1 u|\, \|_1 \big\}, \qquad \tilde{\Lambda} := \lambda I_N. \tag{25}$$

Observe that if we use the alternating Split Bregman algorithm with $D = \mathcal{H}_1$ for this problem we have to solve a linear system of equations in the third step of each iteration. This problem can be avoided by using that $\mathcal{H}_1$ is part of a Parseval frame, cp. [27]. To this end we define the proper, convex and lsc functional $\tilde{\Phi}_2$ which differs from Φ_2 in that the first part of the input vector is neglected, i.e.,

$$\tilde{\Phi}_2(c) = \|\tilde{\Lambda}\, |c_1|\, \|_1, \qquad \text{for } c = (c_0, c_1) \in \mathbb{R}^N \times \mathbb{R}^{3N}.$$

Now we can rewrite (25) as follows

$$\operatorname*{argmin}_{u \in \mathbb{R}^N} \big\{ \frac{1}{2} \|u - f\|_2^2 + \tilde{\Phi}_2(Wu) \big\}.$$

Applying the alternating Split Bregman algorithm, or equivalently the FBS method, with $\gamma = 1$ and (24) we obtain the following algorithm.

Initialization: $u^{(0)} := f$, $b^{(0)} := 0$.
For $k = 0, 1, \dots$ repeat until a stopping criterion is reached

$$
\begin{aligned}
d_0^{(k+1)} &:= (Wu^{(k)})_0, \\
d_1^{(k+1)} &:= \tilde{T}_{\tilde{\Lambda}}\big(b^{(k)} + (Wu^{(k)})_1\big), \\
b^{(k+1)} &:= b^{(k)} + (Wu^{(k)})_1 - d_1^{(k+1)}, \\
u^{(k+1)} &:= W^{\mathrm{T}} \begin{pmatrix} d_0^{(k+1)} \\ d_1^{(k+1)} \end{pmatrix},
\end{aligned}
\tag{26}
$$

where $(Wu)_0 := \mathcal{H}_0 u$ and $(Wu)_1 := \mathcal{H}_1 u$. Note that starting with $b_0^{(0)} := 0$ all iterates $b_0^{(k)}$ remain zero vectors. We also obtain algorithm (26) if we apply FBS shrinkage directly to (25) with $D = \mathcal{H}_1$ and $\gamma = 1$.

We now give a numerical example for these two algorithms. The computations were performed in MATLAB. In Fig. 1 we see the result of applying the two algorithms to a noisy image. Note that we only show the resulting image for algorithm (26) here, since the difference to the alternating Split Bregman

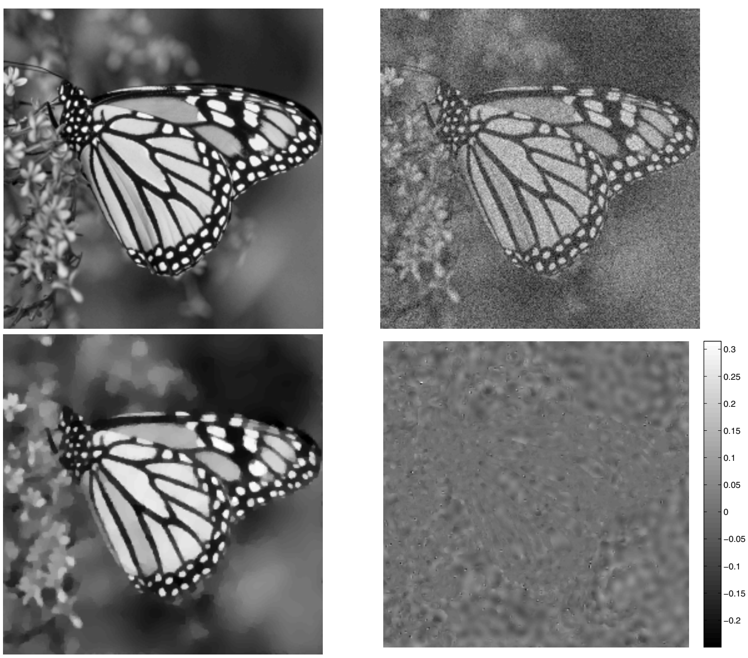

Fig. 1. Comparison of algorithm (26) and the alternating Split Bregman method with $D = \mathcal{H}_1$. Stopping criterion: $\|u^{(k+1)} - u^{(k)}\|_\infty < 0.5$. *Top left*: Original image. *Top right*: Noisy image (white Gaussian noise with standard deviation 25). *Bottom left*: Algorithm (26), $\lambda = 70$, (53 iterations). *Bottom right*: Difference to alternating Split Bregman shrinkage with $D = \mathcal{H}_1$, (53 iterations).

method with $D = \mathcal{H}_1$ is marginal. We also found that the two algorithms need nearly the same number of iterations. However, algorithm (26) is extremely fast and does not require solving a linear system of equations as the alternating Split Bregman shrinkage does. Moreover, $\gamma = 1$ seems to be a very good parameter choice. For the above numerical experiment we used periodic boundary conditions, concerning Neumann boundary conditions, see, e.g., [28].

References

1. DeVore, R.A., Lucier, B.J.: Fast wavelet techniques for near-optimal image processing. In: IEEE MILCOM 1992 Conf. Rec., vol. 3, pp. 1129–1135. IEEE Press, San Diego (1992)
2. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 136–152. Springer, Heidelberg (2005)
3. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
4. Mrázek, P., Weickert, J.: Rotationally invariant wavelet shrinkage. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 156–163. Springer, Heidelberg (2003)
5. Welk, M., Steidl, G., Weickert, J.: Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Applied and Computational Harmonic Analysis 24, 195–224 (2008)
6. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis 16(6), 964–979 (1979)
7. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM Journal on Control and Optimization 29, 119–138 (1991)
8. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization 53(5–6), 475–504 (2004)
9. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Modeling and Simulation 4, 1168–1200 (2005)
10. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Transactions of the American Mathematical Society 82(2), 421–439 (1956)
11. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Computational Mathematics and Mathematical Physics 7(3), 200–217 (1967)
12. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Mathematics of Operations Research 18(1), 202–226 (1993)
13. Kiwiel, K.C.: Proximal minimization methods with generalized Bregman functions. SIAM Journal on Control and Optimization 35(4), 1142–1168 (1997)
14. Frick, K.: The Augmented Lagrangian Method and Associated Evolution Equations, Dissertation, University of Innsbruck (2008)
15. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research 1(2), 97–116 (1976)

16. Browder, F.E., Petryshyn, W.V.: The solution by iteration of nonlinear functional equations in Banach spaces. Bulletin of the American Mathematical Society 72, 571–575 (1966)
17. Iusem, A.N.: Augmented Lagrangian methods and proximal point methods for convex optimization. Investigación Operativa 8, 11–49 (1999)
18. Goldstein, D., Osher, S.: The Split Bregman method for l_1 regularized problems. UCLA CAM Report (2008)
19. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences 1(1), 143–168 (2008)
20. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming 55, 293–318 (1992)
21. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary–Value Problems. Studies in Mathematics and its Applications, vol. 15, pp. 299–331. North–Holland, Amsterdam (1983)
22. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20, 89–97 (2004)
23. Aujol, J.F.: Some first-order algorithms for total variation based image restoration. Preprint ENS Cachan (2008)
24. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restauration. UCLA CAM Report (2008)
25. Daubechies, I., Han, B., Ron, A., Shen, Z.: Framelets: MRA-based construction of wavelet frames. Applied and Computational Harmonic Analysis 14, 1–46 (2003)
26. Dong, B., Shen, Z.: Pseudo-splines, wavelets and framelets. Applied and Computational Harmonic Analysis 22, 78–104 (2007)
27. Setzer, S., Steidl, G.: Split Bregman method, gradient descent reprojection method and Parseval frames. Preprint Univ. Mannheim (2008)
28. Chan, R.H., Setzer, S., Steidl, G.: Inpainting by flexible Haar-wavelet shrinkage. SIAM Journal on Imaging Science 1, 273–293 (2008)
29. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences 1(3), 248–272 (2008)