ijcv manuscript No.
(will be inserted by the editor)

Operator Splittings, Bregman Methods and Frame Shrinkage in Image Processing

Simon Setzer

Received: date / Accepted: date

Abstract We examine the underlying structure of popular algorithms for variational methods used in image processing. We focus here on operator splittings and Bregman methods based on a unified approach via fixed point iterations and averaged operators. In particular, the recently proposed alternating split Bregman method can be interpreted from different points of view - as a Bregman, as an augmented Lagrangian and as a Douglas-Rachford splitting algorithm which is a classical operator splitting method. We also study similarities between this method and the forward-backward splitting method when applied to two frequently used models for image denoising which employ a Besov-norm and a total variation regularization term, respectively. In the first setting, we show that for a discretization 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. For the total variation regularizer, we also present a numerical comparison with multistep methods.

Keywords Douglas-Rachford splitting · forward-backward splitting · Bregman methods · augmented Lagrangian method · alternating split Bregman algorithm · image denoising

1 Introduction

Many problems in image restoration can be solved by means of variational methods, i.e., the resulting images are minimizers of appropriate energy functionals. The success of these models led to a great number of computational algorithms. A common idea is to derive iterative algorithms which consist in each iteration of subproblems which are easier to solve. Three important ways to do this will appear in this paper: operator splitting, Lagrangian and Bregman methods. In the minimization problems we consider here, the objective functions consist of the sum of two terms. Hence, *operator splitting methods* which make use of this special structure are a natural choice. *Lagrangian methods* allow one to consider primal and dual variable at the same time via related constrained problems. The idea of *Bregman methods*, on the other hand, is to introduce a term which penalizes the distance to the last iterate. This also gives rise to easier problems in each iteration.

In the first part of this paper, we will therefore underline the common roots of operator splitting and Bregman methods from the point of view of fixed point theory. In many cases, convergence can be guaranteed via the notion of *averaged operators*. We then consider a new Bregman technique, called the *alternating split Bregman algorithm*, proposed by Goldstein and Osher for image restoration and compressed sensing. We illustrate the three different perspectives for this method. In fact, in our setting the alternating split Bregman algorithm coincides with the *alternating direction method of multipliers* which is a special augmented Lagrangian method and it can also be interpreted as a classical operator splitting algorithm, namely, a *Douglas-Rachford splitting algorithm*, cf. (Esser 2009). This connection also clarifies the convergence of the alternating Split Bregman algorithm.

In the second part of this paper, we consider the application to image denoising. First, we consider the following image restoration model which uses an L_2 data-fitting term and a Besov-norm regularization term, see (DeVore and Lucier 1992),

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

S. Setzer
University of Mannheim, Germany
E-mail: ssetzer@kiwi.math.uni-mannheim.de

We present a discrete version of this problem involving Parseval frames. Interestingly, the corresponding alternating split Bregman algorithm is not only equivalent to the Douglas-Rachford splitting method but also to another popular operator splitting method: the *forward-backward splitting algorithm* which is for our particular problem simply a *gradient descent reprojection algorithm*, see (Chambolle 2005). Since our method is based on frame transformations and soft shrinkage, we also underline the relation to the classical wavelet shrinkage scheme which uses orthonormal wavelet transforms.
Finally, we consider the Rudin-Osher-Fatemi model

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

see (Rudin et al 1992), which uses a total variation regularization term. Note that for the infinite-dimensional setting the relation between problems (1) and (2) was studied in connection with so-called near minimizers in (Bechler et al 2006; Cohen et al 1999). Solving (2) is a successful image denoising method, especially, in terms of the preservation of edges and sharp object boundaries. Here, we first apply our findings to frame-based discretizations of the absolute value of the gradient. Similar to the Besov-norm setting, we derive a relation between the alternating split Bregman method, the forward-backward splitting algorithm and frame shrinkage. Second, we compare these algorithms numerically with a class of first-order methods that has attracted a lot of interest in image processing recently: the so-called multistep methods.

Remark 1 A shorter version of this paper has been published in the Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision 2009, see (Setzer 2009a).

2 Picard iterations for the solution of variational problems

Let us start with general minimization problems of the form

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

where H_1 and H_2 are Hilbert spaces, $D: H_1 \to H_2$ 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 lower semi-continuous (l.s.c.). The corresponding dual problem has the form

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

where g^* and Φ^* are the conjugate functions of g and Φ, respectively. In this paper we further assume that solutions $\hat{u}$ and $\hat{b}$ of the primal and dual problem, respectively, exist and that the duality gap is zero, i.e., (P) and (D) have the same value.

In other words, we suppose that there is a pair $(\hat{u},\hat{d})$ which satisfies the *Karush-Kuhn-Tucker conditions*, cf., e.g., (Bonnans and Shapiro 2000)

$$\begin{aligned} 0 &\in \partial g(\hat{u}) + D^*\hat{b}, \\ 0 &\in -D\hat{u} + \partial\Phi^*(\hat{b}), \end{aligned} \tag{3}$$

where the *subdifferentials* ∂g and $\partial\Phi^*$ are defined as follows: For any convex function $\mathscr{F}: H \to \mathbb{R}\cup\{+\infty\}$ on a Hilbert space H the subdifferential is defined as the following set-valued operator

$$\partial\mathscr{F}(\tilde{u}) = \{v\in H:\ \langle v, u-\tilde{u}\rangle \le \mathscr{F}(u)-\mathscr{F}(\tilde{u}),\ \forall u\in H\}. \tag{4}$$

By *Fermat's rule*, $\hat{u}$ is a solution of (P) if and only if $0\in\partial\mathscr{F}_P(\hat{u})$, and analogously for the dual problem. We further assume in this paper that the following so-called regularity conditions hold true

$$0 \in \operatorname{int}(D\operatorname{dom} g - \operatorname{dom}\Phi), \tag{5}$$

$$0 \in \operatorname{int}(\operatorname{dom} g^* + D^*\operatorname{dom}\Phi^*), \tag{6}$$

where int denotes the interior of the corresponding sets, see (Rockafellar 1970; Borwein and Zhu 2006). Note that in the finite-dimensional setting a weaker condition using the notion of the relative interior of a set can be found in (Rockafellar 1970, Theorem 31.1). Now we can write the primal and dual problem in the equivalent form

$$(P') \quad 0 \in \partial\mathscr{F}_P(\hat{u}) = \partial g(\hat{u}) + \partial(\Phi\circ D)(\hat{u})$$

and

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

Observe that for both the primal problem (P') and the dual problem (D') one has to solve a problem of the form

$$0 \in A(\hat{p}) + B(\hat{p}) \tag{7}$$

for set-valued operators A and B. The main idea of the methods we want to examine in this paper is to write our problems in terms of a fixed point equation, i.e.,

$$0 \in A(\hat{p}) + B(\hat{p}) \ \Leftrightarrow\ \hat{p} = T(\hat{p}) \tag{8}$$

for an appropriate operator $T: H \to H$. Here, $H = H_1$ or $H = H_2$, if we solve the primal problem (P) or the dual problem (D), respectively. Based on (8), we consider the following basic *Picard iterations* to compute such a fixed point $\hat{p}$:

$$p^{(k+1)} = T(p^{(k)}). \tag{9}$$

The operator T must be chosen according to (8) and such that the Picard iterations converge. Recall that an operator $T : H \to H$ on a Hilbert space H is *nonexpansive* if $\|T(u) - T(v)\| \leq \|u - v\|$ for all $u, v \in H$ and *contractive* if $\|T(u) - T(v)\| \leq \beta\|u - v\|$ for all $u, v \in H$ and some constant $\beta \in (0,1)$. In contrast to the property of being contractive, the nonexpansivity of an operator does not guarantee the convergence of the corresponding Picard iterations. Therefore, we use the stronger notion of an averaged operator, cf., e.g., (Bauschke and Borwein 1996; Byrne 2004; Combettes 2004). By definition, $T : H \to H$ is *averaged* if for a nonexpansive operator R and some $\alpha \in (0,1)$ we can write T as

$$T = \alpha I + (1-\alpha)R,$$

where I denotes the identity operator. Note that every contractive operator is averaged but in contrast to contractions, averaged operators can have more than one fixed point. For averaged operators we have the following convergence result:

Theorem 1 *Let H be a Hilbert space and let $T : H \to H$ be an averaged mapping which has at least one fixed point. Then, for every $p^{(0)} \in H$ the Picard sequence* (9) *converges weakly to a fixed point of T.*

This result has its origins in (Mann 1953; Krasnoselskii 1955; Schäfer 1957) and proofs can also be found, e.g., in (Browder and Petryshyn 1966; Opial 1967; Combettes 2004).

2.1 Proximation and the resolvent operator

Let us briefly recap some important elements of convex analysis which will appear as building blocks for the fixed point operators presented in the next sections.

Let H be a Hilbert space and $\mathscr{F} : H \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and l.s.c. function. Then, the *proximity operator* $\operatorname{prox}_{\gamma\mathscr{F}} : H \to H$, introduced in (Moreau 1965), is given by

$$\operatorname{prox}_{\gamma\mathscr{F}}(f) := \operatorname*{argmin}_{u \in H} \{\tfrac{1}{2\gamma}\|u - f\|^2 + \mathscr{F}(u)\}, \tag{10}$$

where $\gamma > 0$. Observe that this minimization problem is equivalent to (P) with $g := \frac{1}{2\gamma}\|\cdot - f\|^2$, $\Phi := \mathscr{F}$ and $D := I$. The objective function in (10) is strictly convex and coercive so that for any $f \in H$ the proximum $\operatorname{prox}_{\gamma\mathscr{F}}(f)$ exists and is unique, cf. (Ekeland and Temam 1976). By Fermat's rule, we have

$$\begin{aligned} \hat{u} = \operatorname{prox}_{\gamma\mathscr{F}}(f) &\Leftrightarrow 0 \in \tfrac{1}{\gamma}(\hat{u} - f) + \partial\mathscr{F}(\hat{u}) \\ &\Leftrightarrow \hat{u} = \underbrace{(I + \gamma\partial\mathscr{F})^{-1}}_{J_{\gamma\partial\mathscr{F}}}(f), \end{aligned}$$

where $J_{\gamma\partial\mathscr{F}}$ is called the *resolvent* of $\gamma\partial\mathscr{F}$.

2.2 Proximal point method

The common structure of the methods discussed in this paper is best seen by considering first the *classical proximal point algorithm*. We refer to (Eckstein and Bertsekas 1992) for references and historical background of this method. Applied to the primal problem (P), it has the form

$$u^{(k+1)} = \operatorname*{argmin}_{u \in H_1} \{\tfrac{1}{2\gamma}\|u - u^{(k)}\|^2 + \mathscr{F}_P(u)\} \tag{11}$$

for a step length $\gamma > 0$. Using our definition from Section 2.1 we can write the fixed point operator applied in (11) as

$$T = \operatorname{prox}_{\gamma\mathscr{F}_P} = J_{\gamma\partial\mathscr{F}_P}.$$

This also explains the name proximal point method.

Clearly, the fixed points of T are exactly the solutions of (P). The main idea of the proximal point method is that, instead of solving the (hard) original problem, we solve in each step a nicer problem which is constructed by adding a "cost-to-move" term to the original objective functional. This term penalizes the distance between two iterates. To show the convergence of (11) we need the following classical results from convex analysis, see, e.g., (Eckstein and Bertsekas 1992) and the references therein:

Since every subdifferential of a proper, convex and l.s.c. function is *maximal monotone* it follows that $\partial\mathscr{F}_P$ is maximal monotone. The resolvent of a maximal operator is *firmly nonexpansive* which means that it is averaged with parameter $1/2$. Hence, we can apply Theorem 1 to conclude weak convergence to a fixed point of T, i.e., to a solution of (P). In the same way, we can define the proximal point algorithm for (D)

$$\begin{aligned} b^{(k+1)} &= \operatorname*{argmin}_{b \in H_2} \left\{\tfrac{1}{2\gamma}\|b - b^{(k)}\|^2 + \mathscr{F}_D(b)\right\} \\ &= \operatorname{prox}_{\gamma\mathscr{F}_D}(b^{(k)}) = J_{\gamma\partial\mathscr{F}_D}(b^{(k)}) \end{aligned} \tag{12}$$

and the same convergence result holds true. It is well-known that this proximal point algorithm for (D) is equivalent to the *augmented Lagrangian method* of (Hestenes 1969; Powell 1969), see, e.g., (Rockafellar 1976; Iusem 1999; Frick 2008). To define this algorithm, we first transform (P) into the constrained minimization problem

$$\min_{u \in H_1, d \in H_2} E(u,d) \quad \text{subject to} \quad Du = d, \tag{13}$$

where $E(u,d) := g(u) + \Phi(d)$, see also (Wang et al 2008). The corresponding augmented Lagrangian algorithm is then defined as

$$\begin{aligned} (u^{(k+1)}, d^{(k+1)}) = \operatorname*{argmin}_{u \in H_1, d \in H_2} &\left\{E(u,d)\right. \\ &\left. + \langle b^{(k)}, Du - d\rangle + \tfrac{1}{2\gamma}\|Du - d\|^2\right\}, \\ b^{(k+1)} = b^{(k)} + \tfrac{1}{\gamma}(Du^{(k+1)} &- d^{(k+1)}). \end{aligned} \tag{14}$$

The first step of (14) the Lagrangian functional is "augmented" by a quadratic penalty term and for the same initial value $b^{(0)}$ the sequence $(b^{(k)})_{k\in\mathbb{N}}$ coincides with the one produced by the proximal point algorithm applied to (D). 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. (Iusem 1999).

2.3 Operator splitting methods

In the above proximal point method we have to compute the resolvent of the subdifferential of the complete objective function. However, the objective functions of (P) and (D) have an additive structure. Hence, we can exploit this to define operators T whose Picard iterations are easier to compute. In this paper, we restrict our attention to the following two operator splitting methods: the *forward-backward splitting* method (FBS), introduced in (Lions and Mercier 1979; Passty 1979), and the *Douglas-Rachford splitting* algorithm (DRS) whose origins lie in (Douglas and Rachford 1956) and which was first applied to image processing in (Combettes and Pesquet 2007).

Note that there exist other method like the Peaceman-Rachford and the backward-backward splitting algorithm, see (Lions and Mercier 1979; Combettes 2004) and the references therein. They are not treated here because of their inferior performance for the applications we want to consider in this paper.

To motivate the forward-backward splitting algorithm we rewrite the inclusion (7) as the fixed point relation

$$\begin{aligned} &\hat{p} - \eta B(\hat{p}) \in \hat{p} + \eta A(\hat{p}) \\ \Leftrightarrow\ &\hat{p} \in J_{\eta A}(I - \eta B)(\hat{p}), \quad \text{for } \eta > 0. \end{aligned} \tag{15}$$

The forward-backward splitting algorithm is then just the corresponding iterations with respect to the operator $T = J_{\eta A}(I - \eta B)$. Again it can be shown, see, e.g., (Lions and Mercier 1979; Combettes 2004; Combettes and Wajs 2005), that under the conditions stated in Theorem 2 below the operator T is averaged and convergence follows by Theorem 1. A somewhat different approach to the proof of the following theorem can be found in (Tseng 1991).

Theorem 2 (Forward-Backward Splitting)
Let $A : H \to 2^H$ be a maximal monotone operator and let $\beta B : H \to H$ be firmly nonexpansive for some $\beta > 0$. Furthermore, assume that a solution of (7) *exists. Then, for any $p^{(0)} \in H$ and $\eta \in (0, 2\beta)$ the following* forward-backward splitting algorithm

$$p^{(k+1)} = J_{\eta A}(I - \eta B)(p^{(k)}) \tag{16}$$

converges weakly to a solution of problem (7).

To introduce the Douglas-Rachford splitting algorithm, we first note that if B is single-valued we can rewrite the fixed point relation (15) as follows

$$\begin{aligned} &\hat{p} \in J_{\eta A}(I - \eta B)(\hat{p}) \\ \Leftrightarrow\ &\hat{p} + \eta B(\hat{p}) \in J_{\eta A}(I - \eta B)(\hat{p}) + \eta B(\hat{p}) && (17) \\ \Leftrightarrow\ &\hat{p} = J_{\eta B}\big(J_{\eta A}(I - \eta B)(\hat{p}) + \eta B(\hat{p})\big). && (18) \end{aligned}$$

If B is set-valued the Picard iterations

$$p^{(k+1)} \in J_{\eta B}\big(J_{\eta A}(I - \eta B)(p^{(k)}) + \eta B(p^{(k)})\big) \tag{19}$$

corresponding to (18) are called the "loose" Douglas-Rachford splitting algorithm, cf. (Eckstein 1989). In general, the algorithm (19) does not converge to a solution of (7). However, if we choose the element of $\eta B(p^{(k)})$ in a special way we do obtain a convergent algorithm. To this end, consider the fixed point equation for the operator $Q : H \to H$ given by

$$\hat{t} = Q(\hat{t}) = J_{\eta A}(2J_{\eta B}(\hat{t}) - \hat{t}) - J_{\eta B}(\hat{t}) + \hat{t}.$$

For such a fixed point $\hat{t}$ we define $\hat{p} := J_{\eta B}(\hat{t})$ and thus $\hat{\xi} := \hat{t} - \hat{p}$ lies in $\eta B(\hat{p})$. With this choice $\hat{\xi} \in \eta B(\hat{p})$ the element $\hat{p}$ is indeed a solution of (7), cf. (17):

$$\begin{aligned} \hat{p} + \hat{\xi} &= J_{\eta A}(2\hat{p} - (\hat{p} + \hat{\xi})) - \hat{p} + \hat{p} + \hat{\xi} \\ &= J_{\eta A}(\hat{p} - \hat{\xi}) + \hat{\xi} \\ \Rightarrow \quad &\hat{p} \in J_{\eta A}(I - \eta B)(\hat{p}) \Leftrightarrow 0 \in A(\hat{p}) + B(\hat{p}). \end{aligned}$$

This gives rise to the iterative algorithm of Theorem 3. Its convergence holds true since under suitable conditions on A and B the operator Q is averaged, see (Lions and Mercier 1979; Combettes 2004), so that we can apply Theorem 1.

Theorem 3 (Douglas-Rachford Splitting)
Let $A, B : H \to 2^H$ be maximal monotone operators and assume that a solution of (7) *exists. Then, for any initial elements $t^{(0)}$ and $p^{(0)}$ and any $\eta > 0$, the following* Douglas-Rachford splitting 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 the sequence $\left(p^{(k)}\right)_{k\in\mathbb{N}}$ converges to a solution $\hat{p}$.

Observe that in contrast to the forward-backward splitting algorithm the operator B is now allowed to be set-valued and we make use of its resolvent. Another difference is that there are no restrictions on the step length.

2.4 Bregman methods

In the preceding section we did not apply the resolvent of the full objective function as in the proximal point algorithm. Instead, we made use of the additive structure of the problem and worked with the resolvents of the individual terms. Another approach to modify the proximal point method consists in generalizing the definition of the resolvent or, equivalently, of the proximation operator. More precisely, we consider the following more general distances in the cost-to-move term of (11) and (12).

Let a Gâteaux differentiable *Bregman function* $\varphi : H \to \mathbb{R} \cup \{+\infty\}$ with Gâteaux derivative $v \mapsto \nabla\varphi(v)$ be given. We define the *Bregman distance* D_φ as follows, cf. (Bregman 1967; Censor and Lent 1981):

$$D_\varphi(u,v) = \varphi(u) - \varphi(v) - \langle \nabla\varphi(v), u - v\rangle.$$

Based on this notion of a Bregman distance the *Bregman proximal point algorithm* was introduced in (Censor and Zenios 1992). We refer to (Eckstein 1993; Censor and Zenios 1997; Kiwiel 1997; Frick 2008) for additional conditions necessary to guarantee convergence. Applied to (P), the Bregman proximal point algorithm is defined as follows for an initial value $u^{(0)}$ and a parameter $\gamma > 0$:

$$u^{(k+1)} = \operatorname*{argmin}_{u\in H_1}\{\tfrac{1}{\gamma} D_\varphi(u, u^{(k)}) + \mathscr{F}_P(u)\}. \tag{20}$$

Note that we use the Bregman distance to the last iterate $u^{(k)}$ as a cost-to-move term now. The classical proximal point method for (P) is just a special case of the Bregman proximal point algorithm for $\varphi := \frac{1}{2}\|\cdot\|^2$. We can also write (20) in the form

$$\begin{aligned} u^{(k+1)} &= T(u^{(k)}) \\ &= (\nabla\varphi + \gamma\partial\mathscr{F}_P)^{-1}\big(\nabla\varphi(u^{(k)})\big) \end{aligned} \tag{21}$$

and the fixed points of T are exactly the solution of the primal problem. Under certain assumptions, we can drop the condition of φ being differentiable, e.g., in finite-dimensional spaces for strictly convex Bregman function with full domain, cf. (Kiwiel 1997). The Bregman distance is then defined as

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

with $p \in \partial\varphi(v)$ and the corresponding Bregman proximal point algorithm is given by

$$\begin{aligned} u^{(k+1)} &= \operatorname*{argmin}_{u\in H_1}\{\tfrac{1}{\gamma} D_\varphi^{(p^{(k)})}(u, u^{(k)}) + \mathscr{F}_P(u)\}, \\ p^{(k+1)} &\in \partial\varphi(u^{(k+1)}), \end{aligned} \tag{22}$$

see also (Eckstein 1993; Frick 2008; Kiwiel 1997; Osher et al 2005) and the references therein.

To solve the constrained optimization problem (13), Goldstein and Osher (2009) proposed to use the Bregman proximal point algorithm (22) in the following way:

As the Bregman function φ the function E defined in (13) is chosen. This results in the Bregman distance

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

Instead of $\mathscr{F}_P$ in (22) we use the term $\frac{1}{2\gamma}\|Du - d\|^2$ so that the constraint in (13) is satisfied if we have convergence. The resulting algorithm is given by

$$\begin{aligned} (u^{(k+1)}, d^{(k+1)}) = \operatorname*{argmin}_{u\in H_1, d\in H_2} \big\{ & D_E^{(p^{(k)})}(u,d,u^{(k)},d^{(k)}) \\ & + \tfrac{1}{2\gamma}\|Du - d\|^2\big\}, \end{aligned} \tag{23}$$

$$p_u^{(k+1)} = p_u^{(k)} - \tfrac{1}{\gamma} D^*(Du^{(k+1)} - d^{(k+1)}), \tag{24}$$

$$p_d^{(k+1)} = p_d^{(k)} + \tfrac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)}), \tag{25}$$

where it is used that (23) implies

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

so that $\big(p_u^{(k)}, p_d^{(k)}\big) \in \partial E(u^{(k)}, d^{(k)})$ for $k \geq 1$. As we will see now we can choose $p_u^{(0)} := -\frac{1}{\gamma} D^* b^{(0)}$ and $p_d^{(0)} := \frac{1}{\gamma} b^{(0)}$ for any element $b^{(0)} \in H_2$ to obtain a convergent algorithm: If we define in each iteration

$$b^{(k+1)} := b^{(k)} + Du^{(k+1)} - d^{(k+1)}$$

it holds by (24)-(25) that $p_u^{(k)} = -\frac{1}{\gamma} D^* b^{(k)}$ and $p_d^{(k)} = \frac{1}{\gamma} b^{(k)}$ for all $k \geq 0$. Using that D is a bounded linear operator, the objective function in (23) can thus be written as

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

Hence, Goldstein and Osher (2009) obtained the *split Bregman method*

$$\begin{aligned} & (u^{(k+1)}, d^{(k+1)}) = \\ & \quad \operatorname*{argmin}_{u\in H_1, d\in H_2} \big\{E(u,d) + \tfrac{1}{2\gamma}\|b^{(k)} + Du - d\|^2\big\}, \\ & b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \end{aligned} \tag{26}$$

As already discovered in (Yin et al 2008), see also (Tai and Wu 2009), the split Bregman algorithm (26) coincides with the augmented Lagrangian algorithm (14) with the only difference that in (26) the iterates $b^{(k)}$ are scaled by γ. By Section 2.2 it also equivalent the proximal point algorithm for (D).

Therefore, we can conclude from Theorem 1 that the sequence $(\frac{1}{\gamma} b^{(k)})_{k\in\mathbb{N}}$ generated by the split Bregman method

(26) converges weakly to a solution of the dual problem. To summarize:

Proximal Point Alg. for (D)	=	Augmented Lagrangian Alg. for (P)	=	Split Bregman Alg. for (P)

2.5 Alternating split Bregman algorithm

Recall that by definition $E(u,d) = g(u) + \Phi(d)$ and the minimization problem in (26) may be hard to solve. Hence, Goldstein and Osher (2009) proposed the following *alternating split Bregman algorithm* to solve (13) where they minimize with respect to u and d alternatingly:

$$u^{(k+1)} = \operatorname*{argmin}_{u\in H_1} \big\{ g(u) + \tfrac{1}{2\gamma}\|b^{(k)} + Du - d^{(k)}\|^2 \big\}, \tag{27}$$

$$d^{(k+1)} = \operatorname*{argmin}_{d\in H_2} \big\{ \Phi(d) + \tfrac{1}{2\gamma}\|b^{(k)} + Du^{(k+1)} - d\|^2 \big\}, \tag{28}$$

$$b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \tag{29}$$

As also pointed out in (Esser 2009) the same idea to minimize alternatingly with respect to the variables was presented for the augmented Lagrangian algorithm (14) in (Gabay and Mercier 1976; Glowinski and Marroco 1975). The resulting algorithm is called the alternating direction method of multipliers (ADMM), cf. (Gabay 1983). It is equivalent to the alternating split Bregman algorithm since as we have seen in the preceding subsection, the augmented Lagrangian algorithm and the split Bregman algorithm coincide in our setting. Here, too, we can take a third point of view and interpret the algorithm as an operator splitting algorithm, namely, a Douglas-Rachford splitting algorithm. We thus have:

DRS for (D)	=	ADMM for (P)	=	Alternating Split Bregman Alg. for (P)

For the sake of completeness, we include the next Theorem 4 which shows as it was done in the same way in (Gabay 1983; Eckstein 1989) for the ADMM that the alternating split Bregman method is indeed a Douglas-Rachford splitting algorithm. If H_1 and H_2 are finite-dimensional, this provides us with a convergence result for the sequences $(b^{(k)})_{k\in\mathbb{N}}$ and $(d^{(k)})_{k\in\mathbb{N}}$. A different proof of convergence of the alternating split Bregman algorithm can be found in (Cai et al 2009).

Theorem 4 *The alternating split Bregman algorithm coincides with the Douglas-Rachford splitting algorithm applied to (D) with $A := \partial(g^* \circ (-D^*))$ and $B := \partial\Phi^*$, where $\eta = 1/\gamma$ and*

$$t^{(k)} = \eta(b^{(k)} + d^{(k)}), \quad p^{(k)} = \eta b^{(k)}, \quad k \geq 0. \tag{30}$$

Proof: 1. First, we show that for a proper, convex, l.s.c. 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:

$$\begin{aligned} &\hat v = \operatorname*{argmin}_{v\in H_1} \big\{ \tfrac{\eta}{2}\|Kv - w\|^2 + h(v) \big\} \\ &\Rightarrow \eta(K\hat v - w) = J_{\eta\,\partial(h^*\circ(-K^*))}(-\eta w). \end{aligned} \tag{31}$$

The first equality in (31) is equivalent to

$$\begin{aligned} &0 \in \eta K^*(K\hat v - w) + \partial h(\hat v) \\ &\Leftrightarrow \hat v \in \partial h^*\big(-\eta K^*(K\hat v - w)\big). \end{aligned}$$

Applying the operator $-\eta K$ implies

$$\begin{aligned} -\eta K\hat v \in\ & -\eta K \partial h^*\big(-\eta K^*(K\hat v - w)\big) \\ &= \eta\,\partial\big(h^*\circ(-K^*)\big)\big(\eta(K\hat v - w)\big). \end{aligned}$$

We now add $-\eta q$ on both sides to get

$$-\eta w \in \big(I + \eta\,\partial(h^*\circ(-K^*))\big)\big(\eta(K\hat v - w)\big),$$

which is, by definition of the resolvent, equivalent to the second equality of (31).
2. Applying (31) to (27) with $h := g$, $K := D$ and $w := d^{(k)} - b^{(k)}$ yields

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

Assume that the alternating split Bregman iterates are related to those of the Douglas-Rachford splitting algorithm via the identification (30) up to some $k \in \mathbb{N}$. Using this induction hypothesis, it follows that

$$\begin{aligned} &\eta(b^{(k)} + Du^{(k+1)}) \\ &= J_{\eta A}(\underbrace{\eta(b^{(k)} - d^{(k)})}_{2p^{(k)} - t^{(k)}}) + \underbrace{\eta d^{(k)}}_{t^{(k)} - p^{(k)}} = t^{(k+1)}. \end{aligned} \tag{32}$$

By definition of $b^{(k+1)}$ in (29), we can conclude that $\eta(b^{(k+1)} + d^{(k+1)}) = t^{(k+1)}$. Next we apply (31) to (28) with $h := \Phi$, $K := -I$ and $w := b^{(k)} + Du^{(k+1)}$ Together with (32) this gives

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

Again by the formula (29) for $b^{(k+1)}$ we obtain $\eta b^{(k+1)} = p^{(k+1)}$ which completes the proof.

Finally, we study the convergence properties of $(u^{(k)})_{k\in\mathbb{N}}$ in the case where H_1 and H_2 are finite-dimensional.

Proposition 1 *Assume that H_1 and H_2 are finite-dimensional Hilbert spaces. Then, every cluster point of the sequence $(u^{(k)})_{k\in\mathbb{N}}$ generated by the alternating split Bregman algorithm is a solution of the primal problem (P).*

Proof: We rewrite (27)-(29) in the equivalent form

$$0 \in \partial g(u^{(k+1)}) + \tfrac{1}{\gamma} D^* b^{(k)} + \tfrac{1}{\gamma} D^*(Du^{(k+1)} - d^{(k)}), \quad (33)$$

$$0 \in \partial \Phi(d^{(k+1)}) - \tfrac{1}{\gamma} b^{(k)} - \tfrac{1}{\gamma}(Du^{(k+1)} - d^{(k+1)}), \quad (34)$$

$$b^{(k+1)} = b^{(k)} + Du^{(k+1)} - d^{(k+1)}. \quad (35)$$

We know from Theorem 4 and Theorem 3 that the sequences $(b^{(k)})_{k\in\mathbb{N}}$ and $(d^{(k)})_{k\in\mathbb{N}}$ converge. Let $\hat{b}$ and $\hat{d}$ denote the corresponding limits. Furthermore, let $\hat{u}$ be a cluster point of $(u^{(k)})_{k\in\mathbb{N}}$ with convergent subsequence $(u^{(k_l)})_{l\in\mathbb{N}}$. Because the subdifferentials of the functions g and Φ are maximal monotone we can pass to the limits with respect to the indices k_l in (33) and (34), see, e.g., (Aubin and Frankowska 2009, Proposition 3.5.6).

This yields the Karush-Kuhn-Tucker conditions

$$\begin{aligned} 0 &\in \partial g(\hat{u}) + \tfrac{1}{\gamma} D^* \hat{b}, \\ 0 &\in -\tfrac{1}{\gamma}\hat{b} + \partial \Phi(D\hat{u}), \end{aligned} \quad (36)$$

cf. (3). Thus, $\hat{u}$ and $\frac{1}{\gamma}\hat{b}$ solve (P) and (D), respectively.

Now the natural question arises under which condition we can guarantee that $(u^{(k)})_{k\in\mathbb{N}}$ converges.

Theorem 5 *Assume that H_1 and H_2 are finite-dimensional Hilbert spaces and let $\hat{b}$ and $\hat{d}$ be the limit points arising from the alternating split Bregman algorithm* (27)-(29)*. Then, a sufficient condition for $(u^{(k)})_{k\in\mathbb{N}}$ to converge to a solution of (P) is that the problem*

$$\operatorname*{argmin}_{u\in H_1} \left\{ g(u) + \tfrac{1}{2\gamma} \|\hat{b} + Du - \hat{d}\|^2 \right\} \quad (37)$$

has a unique solution.

Proof: Let us rewrite (33) as

$$u^{(k+1)} \in (\gamma \partial g + D^* D)^{-1}(D^*(d^{(k)} - b^{(k)})).$$

Since $\gamma\partial g + D^*D$ is maximal monotone as a subdifferential of a proper, convex and l.s.c. function we have that the mapping $(\gamma\partial g + D^*D)^{-1}$ is also maximal monotone, cf. (Aubin and Frankowska 2009, Ch. 3.5).

Consequently, $(\gamma\partial g + D^*D)^{-1}$ is upper semicontinuous, see (Borwein and Zhu 2005, Ch. 5.1). In our case where $D^*(d^{(k)} - b^{(k)}) \to D^*(\hat{d} - \hat{b})$ as $k \to \infty$ and $(\gamma\partial g + D^*D)^{-1}$ is single-valued at the limit point $D^*(\hat{d} - \hat{b})$ this implies that $(u^{(k)})_{k\in\mathbb{N}}$ converges to the unique solution $\hat{u}$ of (37). By Proposition 1, $\hat{u}$ must be a solution of (P).

In many examples the following special case of Theorem 5 can be used.

Corollary 1 *If H_1 and H_2 are finite-dimensional Hilbert spaces and if the primal problem (P) has a unique solution then $(u^{(k)})_{k\in\mathbb{N}}$, defined by the alternating split Bregman algorithm, converges to the solution of (P).*

Proof: See Appendix.

Remark 2 Assume the functional g has the form $g(u) := \frac{1}{2}\|Ku - f\|^2$. Then, Corollary 1 implies that convergence holds true if the matrix $(\gamma K^*K + D^*D)^{-1}$ is invertible, or in other words, if $\mathcal{N}(K^*K) \cap \mathcal{N}(D^*D) = \{0\}$. This is clearly fulfilled for the choice $K = I$ we will make in the next section.

3 Application to image denoising

We now restrict our attention to the discrete setting and consider digital images defined on $\{1,\dots,n\}\times\{1,\dots,n\}$ which are reshaped 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. Our goal is to apply the algorithms defined in the sections above to the discrete denoising problem

$$\operatorname*{argmin}_{u\in\mathbb{R}^N} \left\{ \tfrac{1}{2}\|u - f\|_2^2 + \Phi(Du) \right\}, \quad (38)$$

where $D \in \mathbb{R}^{M,N}$ with $M \geq N$ and Φ is one of the following functions on $\mathbb{R}^M$:

i) $\Phi_1(v) := \|\Lambda v\|_1$
with $\Lambda := \operatorname{diag}(\lambda_j)_{j=1}^M$, $\lambda_j \geq 0$,

ii) $\Phi_2(v) := \|\tilde{\Lambda}\,|v|\,\|_1$ with $\tilde{\Lambda} := \operatorname{diag}(\tilde{\lambda}_j)_{j=1}^N$, $\tilde{\lambda}_j \geq 0$ and $|v| := \left(\|\mathbf{v}_j\|_2\right)_{j=1}^N$ for $\mathbf{v}_j := (v_{j+kN})_{k=0}^{p-1}$ and $M = pN$.

The corresponding conjugate functions are given by

i) $\Phi_1^*(v) := \iota_C(v)$ with
$C := \{v \in \mathbb{R}^M : |v_j| \leq \lambda_j,\ j = 1,\dots,M\}$,

ii) $\Phi_2^*(v) := \iota_{\widetilde{C}}(v)$ with
$\widetilde{C} := \{v \in \mathbb{R}^M : \|\mathbf{v}_j\|_2 \leq \tilde{\lambda}_j,\ j = 1,\dots,N\}$,

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

$$\begin{aligned} \operatorname{prox}_{\Phi_1}(f) &= S_\Lambda(f), \\ \operatorname{prox}_{\Phi_2}(f) &= \tilde{S}_{\tilde{\Lambda}}(f), \end{aligned}$$

where S_Λ denotes the *soft shrinkage* function (also called soft thresholding) given componentwise by

$$S_{\lambda_j}(f_j) := \begin{cases} 0 & \text{if } |f_j| \leq \lambda_j, \\ f_j - \lambda_j \operatorname{sgn}(f_j) & \text{if } |f_j| > \lambda_j. \end{cases} \quad (39)$$

$\tilde{S}_{\tilde{\Lambda}}$ denotes the *coupled shrinkage* function

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

compare (Chambolle 2005; Mrázek and Weickert 2003; Welk et al 2008). Similarly, we obtain

$$\begin{aligned}\operatorname{prox}_{\Phi_1^*}(f) &= f - S_\Lambda(f),\\ \operatorname{prox}_{\Phi_2^*}(f) &= f - \tilde{S}_{\tilde{\Lambda}}(f).\end{aligned} \tag{40}$$

Consider the alternating split Bregman algorithm (27)-(29) with $g(u) := \frac{1}{2}\|u-f\|_2^2$. Theorem 4 and Corollary 1 imply the convergence of $\left(u^{(k)}\right)_{k\in\mathbb{N}}$ and $\left(\frac{1}{\gamma}b^{(k)}\right)_{k\in\mathbb{N}}$ to a solution of the primal and the dual problem, respectively. With the above choice of g, we have to solve the following quadratic problem in (27):

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

cf. Section 3.3. Applying (39), we see that for $\Phi = \Phi_1$ the solution of the proximation problem in (28) is given by

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

and similarly for $\Phi = \Phi_2$. This leads to the following alternating split Bregman shrinkage algorithm:

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

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

As we will see below, soft shrinkage $S_{\gamma\Lambda}$ can be used for $B^1_{1,1}$ regularization. In the case where $\Phi = \Phi_2$ which is appropriate, e.g., for total variation regularization, we have to replace it by the coupled shrinkage function $\tilde{S}_{\gamma\tilde{\Lambda}}$. 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 $d^{(k+1)}$, $b^{(k+1)}$ and $u^{(k+1)}$. This is allowed because there are no restrictions on the choice of the starting values. Indeed, if we start in (27)-(29) with $d^{(0)} := Df$ and $b^{(0)} := 0$, we get the same iterates as in Algorithm I.

Note that Algorithm I can also be used for the deblurring problem which differs from (38) in having a more general data-fitting term $g(u) := \frac{1}{2}\|Ku - f\|_2^2$ with some linear blur operator K, cf. Section 3.3.

Remark 3 Based on the splitting (13), a quadratic penalty approach to solve (38) and the corresponding deblurring problem was proposed in (Wang et al 2008).

The problem (38) can also be solved via its dual problem:

$$\hat{b} = \operatorname*{argmin}_{b\in\mathbb{R}^M}\{\tfrac{1}{2}\|f - D^*b\|_2^2 + \Phi^*(b)\}, \tag{41}$$

and $\hat{u} = f - D^*\hat{b}$, see, e.g., (Chambolle 2004). Applying the forward-backward splitting algorithm (16) to this problem gives

$$b^{(k+1)} = \operatorname{prox}_{\gamma\Phi^*}\left(b^{(k)} + \gamma D(f - D^*b^{(k)})\right),$$

where $0 < \gamma < 2/\|D^*D\|_2$. Using relation (40) we obtain for $\Phi = \Phi_1$:

$$b^{(k+1)} = b^{(k)} + \gamma D(f - D^*b^{(k)}) - S_\Lambda\left(b^{(k)} + \gamma D(f - D^*b^{(k)})\right),$$

which can be written as follows:

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

$$\begin{aligned}d^{(k+1)} &:= S_\Lambda\left(b^{(k)} + \gamma Du^{(k)}\right),\\ b^{(k+1)} &:= b^{(k)} + \gamma Du^{(k)} - d^{(k+1)},\\ u^{(k+1)} &:= f - D^*b^{(k+1)}.\end{aligned}$$

If we use Φ_2, we have to replace the shrinkage functional S_Λ by $\tilde{S}_{\tilde{\Lambda}}$. This algorithm can also be deduced as a simple *gradient descent reprojection algorithm* as it was done, e.g., in (Chambolle 2005). Note that this is *not* the often cited Chambolle algorithm of (Chambolle 2004). A relation of this method to the Bermúdez-Moreno algorithm which also turns out to be a forward-backward splitting algorithm was shown in (Aujol 2009).

3.1 Besov-norm regularization

For a sufficiently smooth orthonormal wavelet basis $(\psi_i)_{i\in\Xi}$ of $L_2(\Omega)$ with wavelets of more than one vanishing moment, we can solve problem (1) in the wavelet domain by finding

$$\hat{d} = \operatorname*{argmin}_{d\in\ell_2(\Xi)}\left\{\tfrac{1}{2}\|d - c\|_{\ell_2}^2 + \lambda\|d\|_{\ell_1}\right\}, \quad c := (\langle f, \psi_i\rangle)_{i\in\Xi} \tag{42}$$

and defining $\hat{u} = \sum_{i\in\Xi}\hat{d}_i\psi_i$. 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, using the notation $\Lambda := \lambda I_N$, problem (1) has the discrete counterpart

$$\begin{aligned}\hat{u} &= \operatorname*{argmin}_{u\in\mathbb{R}^N}\left\{\tfrac{1}{2}\|u - f\|_2^2 + \|\Lambda Wu\|_1\right\}\\ &= \operatorname*{argmin}_{u\in\mathbb{R}^N}\left\{\tfrac{1}{2}\|Wu - Wf\|_2^2 + \|\Lambda Wu\|_1\right\}.\end{aligned} \tag{43}$$

By the orthogonality of W we can solve this as in (42): We compute for $c := Wf$

$$\hat{d} = \operatorname*{argmin}_{d\in\mathbb{R}^N}\left\{\tfrac{1}{2}\|d - c\|_2^2 + \|\Lambda d\|_1\right\} \tag{44}$$

and set $\hat{u} = W^* \hat{d}$. Because of (39), we thus obtain $\hat{u}$ by the known *wavelet shrinkage procedure* $\hat{u} = W^* S_\Lambda (W f)$ consisting of a wavelet transform W followed by a soft shrinkage S_Λ of the wavelet coefficients and the application of the inverse wavelet transform W^*.

However, for image processing tasks like denoising or segmentation, ordinary orthonormal wavelets are not suited due to their lack of translational invariance leading to visible artifacts. Nevertheless, without the usual subsampling, the method becomes translationally invariant and the results can be improved. Then, however, $W \in \mathbb{R}^{M,N}$, $M = pN$, for p equal to three times the decomposition level plus one for the rows belonging to the scaling function filters on the coarsest scale. We still have $W^*W = I_N$ but of course $WW^* \neq I_M$, 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., (Ron and Shen 1997; Daubechies et al 2003; Dong and Shen 2007). Equality (43) is still true for Parseval frames, but the problem is no longer equivalent to (44). Instead of (44), we have to solve the constrained problem

$$\hat{d} = \operatorname*{argmin}_{d \in \mathbb{R}^M} \left\{ \tfrac{1}{2} \|c - d\|_2^2 + \|\Lambda d\|_1 \right\} \quad \text{subject to } d \in \mathscr{R}(W), \tag{45}$$

where $\mathscr{R}(W)$ denotes the range of W. The solution in the image space is then again given by $\hat{u} = W\hat{d}$. Note that the constraint in (45) is equivalent to $(I_M - WW^*)d = d$, i.e., d must coincide with its orthogonal projection onto $\mathscr{R}(W)$. One could also penalize this condition. In the context of inpainting, this was suggested by Cai et al (2008).

We will now show that FBS shrinkage and alternating split Bregman shrinkage with $D = W$, $\gamma = 1$ and $\Phi = \Phi_1$ applied to (43) or, equivalently, to (45) coincide with the following algorithm which underlines the relation to the wavelet shrinkage algorithm with orthonormal transforms.

Algorithm III (Iterated Frame Shrinkage)

Initialization: $u^{(0)} := f$, $b^{(0)} := 0$.

For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$$\begin{aligned} d^{(k+1)} &:= S_\Lambda (b^{(k)} + W u^{(k)}), \\ b^{(k+1)} &:= b^{(k)} + W u^{(k)} - d^{(k+1)}, \\ u^{(k+1)} &:= W^* d^{(k+1)}. \end{aligned} \tag{46}$$

The first step of the algorithm, i.e., $u^{(1)} = W^* S_\Lambda (W f)$ is an ordinary frame shrinkage step which also appears if we disregard the constraint in (45). In the following iterations, the algorithm differs from the usual iterated frame shrinkage in the summand $b^{(k)}$ we have to add before the shrinkage step yielding $d^{(k+1)}$.

Note that in order to use the forward-backward splitting algorithm for problem (43), γ has to fulfill $0 < \gamma < 2/\|W^*W\|_2$. Since $W^*W = I_N$ we have to choose a $\gamma \in (0,2)$ and $\gamma = 1$ is an admissible choice.

Proposition 2 *For $D := W$, where $W^*W = I_N$, and $\gamma := 1$ the FBS shrinkage and the alternating split Bregman shrinkage algorithm coincide with the iterated frame shrinkage algorithm.*

Proof: We start with $u^{(0)} := f$ and $b^{(0)} := 0$ in all three algorithms and in the way we have written them, they only differ in the third step, where we have

$$\begin{aligned} u^{(k+1)} &= \tfrac{1}{2}\big(f + W^*(d^{(k+1)} - b^{(k+1)})\big) && \text{in Algorithm I} \\ u^{(k+1)} &= f - W^* b^{(k+1)} && \text{in Algorithm II} \\ u^{(k+1)} &= W^* d^{(k+1)} && \text{in Algorithm III} \end{aligned}$$

We now use induction on k. Assume that $u^{(k)} = f - W^* b^{(k)}$. Then, we obtain by definition of $b^{(k+1)}$ that

$$\begin{aligned} u^{(k+1)} &= f - W^* b^{(k+1)} \\ &= f - W^* b^{(k)} - u^{(k)} + W^* d^{(k+1)} \\ &= W^* d^{(k+1)}, \end{aligned}$$

so that the elements $u^{(k+1)}$ are the same for the alternating split Bregman shrinkage algorithm and the iterated frame shrinkage algorithm. Further, we see

$$\begin{aligned} &\tfrac{1}{2}\big(f + W^*(d^{(k+1)} - b^{(k+1)})\big) \\ &= \tfrac{1}{2}\big(f + W^* d^{(k+1)} - W^* b^{(k)} - u^{(k)} + W^* d^{(k+1)}\big) \\ &= W^* d^{(k+1)} + \tfrac{1}{2}\big(f - W^* b^{(k)} - u^{(k)}\big) \\ &= W^* d^{(k+1)}, \end{aligned}$$

so that the alternating split Bregman shrinkage algorithm coincides with the iterated frame shrinkage algorithm, too.

Let us restate our result for our special setting in the following diagram:

Alternating Split Bregman Shrinkage	=	FBS Shrinkage	=	Iterated Frame Shrinkage

3.2 Total variation regularization

In this section, we apply the algorithms presented so far to the discrete Rudin-Osher-Fatemi denoising method, i.e., to (38) with $\Phi = \Phi_2$ and a special 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 here, concerning Neumann boundary conditions, see, e.g., (Chan et al 2008). 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 $W^*W = I_N$ but $WW^* \neq I_{4N}$:

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

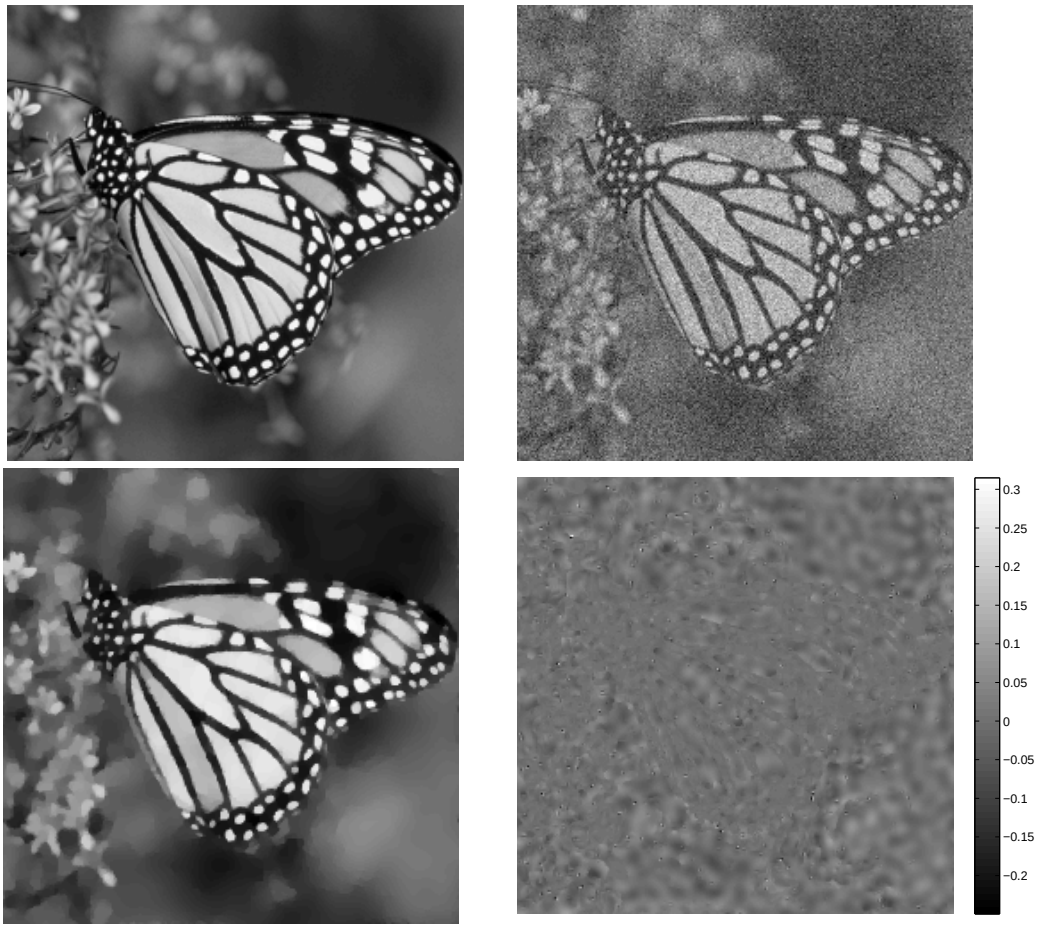

Fig. 1 Comparison of Algorithm IV and the alternating split Bregman method with $D := \mathscr{H}_1$. Stopping criterion: $\|u^{(k+1)} - u^{(k)}\|_\infty < 0.5$. *Top left*: Original image. *Top right*: Noisy image (white Gaussian noise with standard deviation 25). *Bottom left*: Algorithm IV, $\lambda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

Mrázek and Weickert (2003) and Welk et al (2008) showed 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 $|\nabla u|$. Hence, our discrete version of the Rudin-Osher-Fatemi functional (2) reads

$$\operatorname*{argmin}_{u \in \mathbb{R}^N} \big\{ \tfrac{1}{2}\|u - f\|_2^2 + \|\tilde{\Lambda}\,|\mathscr{H}_1 u|\,\|_1 \big\}, \quad \tilde{\Lambda} := \lambda I_N. \tag{47}$$

If we use the ordinary alternating split Bregman algorithm with $D = \mathscr{H}_1$ for this problem we have to solve a linear system of equations in each iteration. This can be avoided by using that $\mathscr{H}_1$ is part of the Parseval frame W. To this end, we define the proper, convex and l.s.c. functional $\widetilde{\Phi}_2$ which differs from Φ_2 in that the first part of the input vector is neglected, i.e.,

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

Now we can rewrite (47) as follows

$$\operatorname*{argmin}_{u \in \mathbb{R}^N} \big\{ \tfrac{1}{2}\|u - f\|_2^2 + \widetilde{\Phi}_2(Wu) \big\}. \tag{48}$$

Proposition 2 shows that the alternating split Bregman algorithm and the forward-backward splitting method with $\gamma := 1$ coincide for this problem and can be rewritten as:

Algorithm IV (Iterated Frame Shrinkage for (48)**)**
Initialization: $u^{(0)} := f$, $b^{(0)} := 0$.
For $k = 0, 1, \ldots$ repeat until a stopping criterion is reached

$$\begin{aligned}
d_0^{(k+1)} &:= (Wu^{(k)})_0, \\
d_1^{(k+1)} &:= \tilde{S}_{\tilde{\lambda}}\big(b^{(k)} + (Wu^{(k)})_1\big), \\
b^{(k+1)} &:= b^{(k)} + (Wu^{(k)})_1 - d_1^{(k+1)}, \\
u^{(k+1)} &:= W^* \begin{pmatrix} d_0^{(k+1)} \\ d_1^{(k+1)} \end{pmatrix},
\end{aligned}$$

where $(Wu)_0 := \mathscr{H}_0 u$ and $(Wu)_1 := \mathscr{H}_1 u$. Note that starting with $b_0^{(0)} := 0$ all iterates $b_0^{(k)}$ remain zero vectors. We also obtain Algorithm IV if we apply FBS shrinkage directly to (47) with $D := \mathscr{H}_1$ and $\gamma := 1$.
We now give a numerical example, where we denoise an image with Algorithm IV and with the usual alternating split

Bregman shrinkage algorithm applied to (47) with $D := \mathscr{H}_1$. Note that the latter requires the solution of a linear system of equations in each step. Here, we have used the diagonalization via the FFT to do this, cf. Section 3.3 for a more detailed discussion.

The results are depicted in Fig. 1. We only show the denoised image for Algorithm IV and its difference to the image obtained by the alternating split Bregman method with $D := \mathscr{H}_1$ here since the difference between the two result is marginal. Note that the two algorithms need nearly the same number of iterations which is also true for other regularization parameters λ and other stopping criteria.

3.3 Comparison with multistep algorithms

Minimizing the dual problem (41) corresponding to the discrete Rudin-Osher-Fatemi model means that we minimize a Lipschitz continuously differentiable function over a closed and convex set. Recently, there has been a lot of interest in solving problems of this form via *multistep methods*. The main idea is to make use of the history of preceding iterations. Two interesting multistep methods were proposed based on an algorithm of Nesterov (1983): The generalized fast iterative shrinkage thresholding algorithm (FISTA) of (Beck and Teboulle 2008, 2009) is a projected version of the algorithm in (Nesterov 1983) and can be seen as an improved gradient descent reprojection method. The method which is now widely known as Nesterov's algorithm, cf., (Nesterov 2005), is a modification of the algorithm in (Nesterov 1983) including projections.

We now compare the performance of these two multistep methods with those of the alternating split Bregman shrinkage (Algorithm I) and the FBS shrinkage (Algorithm II), see also (Goldstein and Osher 2009; Aujol 2009; Weiss et al 2009) for related numerical comparisons.

Our computations were performed on a dual core desktop (2.4 GHz processors, 3 GB memory) using MATLAB 7.6.0. For the sake of comparability, we do not use the frame-based discretization of D presented in Section 3.2 but the following widely used forward difference discretization of the gradient

$$D := \begin{pmatrix} I \otimes D_f \\ D_f \otimes I \end{pmatrix},$$

where

$$D_f := \begin{pmatrix} -1 & 1 & 0 \dots & 0 & 0 & 0 \\ 0 & -1 & 1 \dots & 0 & 0 & 0 \\ & & \ddots\ \ddots\ \ddots & & & \\ 0 & 0 & 0 \dots & -1 & 1 & 0 \\ 0 & 0 & 0 \dots & 0 & -1 & 1 \\ 0 & 0 & 0 \dots & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{n,n}.$$

Our test image is shown in Fig. 2. Note that we invert the matrix $\gamma I + D^*D$ in Algorithm II via multiplications with the cosine-II matrix. A speed-up might be possible by using fast DCT algorithms. Note that if we use periodic boundary conditions in the discretization the FFT can be used directly. Another possible speed improvement could be achieved by approximating the matrix inverse, e.g., via Gauß-Seidel iterations as proposed in (Goldstein and Osher 2009).

Note that these techniques can also be used in many cases for the corresponding deblurring algorithm where we have to invert a matrix of the form $\gamma K^*K + D^*D$.

The two tables in Fig. 3 contain our numerical results. We measure the performance of the algorithms in terms of the values of dual objective function in (41) and the primal objective function in (38) which are achieved after fixed running times. More specifically, we consider the logarithm of the distance between these values and the corresponding function values of a reference image obtained after many iterations.

With respect to the value of the dual objective function, we see that FISTA performs best, followed by Nesterov's algorithm and the alternating split Bregman shrinkage (ASB).

When we consider the primal objective function, however, alternating split Bregman shrinkage performs best and Nesterov's algorithm is now better than FISTA. It is important to note that the alternating split Bregman shrinkage is very sensitive to the choice of the parameter γ. Hence, in contrast to Section 3.2, we have optimized it numerically for each of the experiments, i.e., for the two error measures and the different computation times. For the forward-backward splitting algorithm we have chosen the parameter $\gamma = 0.249$.

It is not surprising that the multistep-methods perform well compared to the forward-backward splitting method: Let k denote the number of iterations. It is shown in (Nesterov 2005; Beck and Teboulle 2008, 2009) that the (non-logarithmic) error with respect to the dual functional is of the order $\mathscr{O}\left(\frac{1}{k^2}\right)$ compared to $\mathscr{O}\left(\frac{1}{k}\right)$ for the FBS shrinkage. Observe that these three algorithms are first-order method.

The good performance of the alternating split Bregman shrinkage algorithm, especially in terms of the dual variable, can be explained by its close relation to the Levenberg-Marquardt method, cf. (Levenberg 1944; Marquardt 1963), which is a special Newton method, i.e., a *second-order method*. Note that all the other algorithms considered here are first-order methods. For more details, we refer to (Setzer 2009b).

Observe that recently there is growing interest in improving the forward-backward splitting method via dynamic step length strategies, especially the so-called *Barzilai-Borwein techniques*, see, e.g., (Barzilai and Borwein 1988; Zhu 2008). This is not covered here.

Moreover, we want to mention that for the above denoising problem a special Lagrangian method, the *primal-dual hybrid gradient algorithm*, was introduced in (Zhu and Chan 2008), see also (Esser et al 2009).

Fig. 2 Left: Original image with values in $[0,255]$. Middle: Noisy version with additive white Gaussian noise of standard deviation 25. Right: Denoising result using the Rudin-Osher-Fatemi model with regularization parameter $\lambda = 25$.

Time	ASB	FISTA	Nesterov	FBS
5	4.62	2.73	3.28	7.15
10	2.45	0.10	1.28	5.80
15	1.33	-1.01	0.09	5.04
20	0.24	-2.05	-0.71	4.52
30	-0.92	-3.35	-1.62	3.93
40	-1.69	-4.47	-2.32	3.49
60	-3.33	-5.56	-3.68	2.68

Time	ASB	FISTA	Nesterov	FBS
5	6.47	7.02	5.38	8.56
10	4.50	5.14	3.73	7.40
15	3.48	4.25	2.78	6.71
20	2.48	3.32	2.14	6.26
30	1.27	2.28	1.36	5.73
40	0.40	1.52	0.73	5.34
60	-1.92	0.79	-0.47	4.61

Fig. 3 Corresponding to the experiment in Fig. 2: Logarithmic error with respect to the dual objective (left) function and the primal objective function (right) for different running times (in seconds).

4 Conclusions

In this paper, we described main ideas behind some important minimization techniques for image restoration, in particular split Bregman methods. It had previously been observed that the Split Bregman method can be interpreted as a classical proximal point method or, equivalently, as an augmented Lagrangian method. Since there is still a hard subproblem to solve in each iteration of the split Bregman algorithm, the alternating split Bregman method had been proposed. This algorithm is equivalent to other well-known methods, the Douglas-Rachford splitting and the alternating direction method of minimizers. We also studied the convergence properties in more detail, especially for the primal variable.

As an application, image denoising via two popular models which apply $B^1_{1,1}$ and total variation regularizers, respectively, was considered. In the $B^1_{1,1}$ case, alternating split Bregman and gradient descent reprojection methods led to the same algorithm in a special setting based on Parseval frames. This underlines the common roots of these techniques. A certain similarity could also been established for the total variation regularization functional. Finally, we have compared these methods numerically with multistep algorithms for a forward difference discretization of the total variation regularizer. We found that if the step length parameter is optimized the alternating split Bregman method performs well, especially with respect to the primal function value. This can be motivated by relating it to a special Newton method.

The minimization algorithms we considered here can be applied to other problems in image processing, e.g., to the minimization of nonlocal total variation functionals, cf., (Gilboa et al 2006; Gilboa and Osher 2008; Kindermann et al 2005). See also (Buades et al 2008) for more on the idea of using nonlocal operators in image processing. Moreover, Goldstein et al (2009) used the alternating split Bregman method for segmentation and surface reconstruction.

As pointed out in (Esser 2009), the alternating split Bregman method is especially useful for more complicated minimization problems, e.g., for image deblurring in the presence of noise. For Gaussian and impulse noise this was discussed in (Goldstein and Osher 2009; Esser 2009). The application to deblurring in the presence of Poisson noise and multiplicative noise can be found in (Figueiredo and Bioucas-Dias 2009; Bioucas-Dias and Figueiredo 2009; Setzer et al 2009). An alternating split Bregman algorithm for image denoising in the presence of multiplicative noise using a nonlocal total variation term was recently proposed by Steidl and Teuber (2009).

Appendix: Proof of Corollary 1

We will make use of the following lemma.

Lemma 1 *Let $\mathscr{F}$ be a l.s.c. and convex function on a finite-dimensional Hilbert space H. Assume that $\mathscr{F}$ has a unique minimizer, then $\mathscr{F}$ is coercive.*

Proof: Suppose that $\mathscr{F}$ is not coercive, i.e., there exists a sequence $(u^{(k)})_{k\in\mathbb{N}}$ with $\|u^{(k)}\| \to +\infty$ as $k \to +\infty$ and $|\mathscr{F}(u^{(k)})| \leq C < +\infty$ for

all $k \in \mathbb{N}$. W.l.o.g. assume that $\hat{u} = 0$ is the unique minimizer of $\mathscr{F}$ and that $\mathscr{F}(0) = 0$ (otherwise use similar arguments as in the proof of Corollary 1 below). We consider the sequence defined by

$$v^{(k)} = \frac{u^{(k)}}{\|u^{(k)}\|},$$

which is clearly bounded and thus has a cluster point $\hat{v}$. The convexity of $\mathscr{F}$ yields for k large enough that

$$\mathscr{F}(v^{(k)}) \leq \frac{1}{\|u^{(k)}\|}\mathscr{F}(u^{(k)}) \to 0 \quad \text{for } k \to \infty.$$

Since $\mathscr{F}$ is l.s.c. we obtain $\mathscr{F}(\hat{v}) \leq 0$ and thus $\mathscr{F}(\hat{v}) = 0$. This contradicts the uniqueness of the minimizer because by construction $\|\hat{v}\| = 1$.

Proof of Corollary 1: It is sufficient to show that $(u^{(k)})_{k\in\mathbb{N}}$ is bounded because we know from Proposition 1 that every cluster point is equivalent to the unique solution of (P). Taking (27) into account, we see that boundedness holds true if the functional $F : H_1 \to \mathbb{R} \cup \{+\infty\}$ defined by

$$F(u) = g(u) + \tfrac{1}{2}\|Du\|^2 \tag{A.1}$$

is coercive. From Lemma 1 we know that $g + \Phi \circ D$ is coercive. It remains to show that this implies the coercivity of F. Note that each element $u \in H_1$ has an orthogonal decomposition $u = u_1 + u_2$ with $u_1 \in \mathscr{N}(D)$, where $\mathscr{N}(D)$ is the null space of D. If $\mathscr{N}(D) = \{0\}$ we are done since g is convex and thus F is coercive. So, let $\mathscr{N}(D) \neq \{0\}$. If for a sequence $(u^{(k)})_{k\in\mathbb{N}}$ it holds that

$$\|u_2^{(k)}\| \leq C_1 < +\infty, \quad \forall k \in \mathbb{N}, \tag{A.2}$$

then $(u^{(k)})_{k\in\mathbb{N}}$ is unbounded by convexity of g.

So, assume that (A.2) holds true. We have to show that there cannot be an unbounded sequence $(u^{(k)})_{k\in\mathbb{N}}$ with $|g(u^{(k)})| \leq C_2 < +\infty$. Assume that such a sequence exists. For $v_0 \in H_1$ with $F(v_0) < +\infty$, we define for any $m \in \mathbb{N}$

$$v_m^{(k)} = v_0 + m\frac{u^{(k)} - v_0}{\|u^{(k)} - v_0\|}.$$

Then, $(v_m^{(k)})_{k\in\mathbb{N}}$ is a bounded sequence and thus has a cluster point v_m. Assumption (A.2) implies that $v_m \in v_0 + \mathscr{N}(D)$. Convexity of g yields for k large enough

$$\begin{aligned} g(v_m^{(k)}) &\leq \big(1 - \tfrac{m}{\|u^{(k)} - v_0\|}\big)g(v_0) + \tfrac{m}{\|u^{(k)} - v_0\|}g(u^{(k)}) \\ &\leq \underbrace{\big(1 - \tfrac{m}{\|u^{(k)} - v_0\|}\big)}_{\to 1}g(v_0) + \underbrace{\tfrac{m}{\|u^{(k)} - v_0\|}}_{\to 0}C_2. \end{aligned}$$

Since g is l.s.c. we obtain $g(v_m) \leq g(v_0)$. The sequence $(g(v_m))_{m\in\mathbb{N}}$ must be bounded from below because a solution of (P) exists. Hence, we have constructed an unbounded sequence $(v_m)_{m\in\mathbb{N}}$ for which both the corresponding values of g and $\Phi \circ D$ are bounded. This yields a contradiction since $g + \Phi \circ D$ is coercive by Lemma 1.

References

Aubin JP, Frankowska H (2009) Set-valued Analysis. Birkäuser, Boston

Aujol JF (2009) Some first-order algorithms for total variation based image restoration. Journal of Mathematical Imaging and Vision 34(3):307–327

Barzilai J, Borwein JM (1988) Two-point step size gradient methods. IMA Journal of Numerical Analysis 8:141–148

Bauschke HH, Borwein JM (1996) On projection algorithms for solving convex feasibility problems. SIAM Review 38(3):367–426

Bechler P, DeVore R, Kamont A, Petrova G, Wojtaszczyk P (2006) Greedy wavelet projections are bounded on BV. Transactions of the American Mathematical Society 359(2):619–635

Beck A, Teboulle M (2008) Fast iterative shrinkage-thresholding algorithm for linear inverse problems. IEEE Transactions on Image Processing Accepted for publication

Beck A, Teboulle M (2009) Fast gradient-based algorithms for constrained total variation image denoising and deblurring. SIAM Journal on Imaging Sciences 2:183–202

Bioucas-Dias J, Figueiredo M (2009) Total variation restoration of speckled images using a split-bregman algorithm. In: IEEE International Conference on Image Processing, Cairo, Egypt, submitted

Bonnans JF, Shapiro A (2000) Perturbation Analysis of Optimization Problems, Springer Series in Operations Research, vol 7. Springer, Berlin

Borwein JM, Zhu QJ (2005) Techniques of variational analysis. Springer, New York

Borwein JM, Zhu QJ (2006) Variational methods in convex analysis. Journal of Global Optimization 35(2):197–213

Bregman LM (1967) 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

Browder FE, Petryshyn WV (1966) The solution by iteration of nonlinear functional equations in Banach spaces. Bulletin of the American Mathematical Society 72:571–575

Buades A, Coll B, Morel JM (2008) Nonlocal image and movie denoising. International Journal of Computer Vision 76(2):123–139

Byrne C (2004) A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Problems 20:103–120

Cai JF, Chan RH, Shen Z (2008) A framelet-based image inpainting algorithm. Applied and Computational Harmonic Analysis 24:131–149

Cai JF, Osher S, Shen Z (2009) Split Bregman methods and frame based image restoration. Technical report, UCLA Computational and Applied Mathematics

Censor Y, Lent A (1981) An iterative row-action method for interval convex programming. Journal of Optimization Theory and Applications 34(3):321–353

Censor Y, Zenios SA (1992) Proximal minimization algorithm with D-functions. Journal of Optimization Theory and Applications 73(3):451–464

Censor Y, Zenios SA (1997) Parallel Optimization: Theory, Algorithms, and Applications. Oxford University Press, New York

Chambolle A (2004) An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20(1-2):89–97

Chambolle A (2005) Total variation minimization and a class of binary MRF models. In: Rangarajan A, Vemuri BC, Yuille AL (eds) Energy Minimization Methods in Computer Vision and Pattern Recognition, EMMCVPR, Springer, Lecture Notes in Computer Science, vol 3757, pp 136–152

Chan RH, Setzer S, Steidl G (2008) Inpainting by flexible Haar-wavelet shrinkage. SIAM Journal on Imaging Science 1:273–293

Cohen A, DeVore R, Petrushev P, Xu H (1999) Nonlinear approximation and the space BV($\mathbb{R}^2$). American Journal of Mathematics 121:587–628

Combettes PL (2004) Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization 53(5–6):475–504

Combettes PL, Pesquet JC (2007) A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE Journal of Selected Topics in Signal Processing 1(4):564–574

Combettes PL, Wajs VR (2005) Signal recovery by proximal forward-backward splitting. Multiscale Modelling & Simulation 4:1168–

1200
Daubechies I, Han B, Ron A, Shen Z (2003) Framelets: MRA-based construction of wavelet frames. Applied and Computational Harmonic Analysis 14:1–46
DeVore RA, Lucier BJ (1992) Fast wavelet techniques for near-optimal image processing. In: IEEE MILCOM '92 Conf. Rec., IEEE Press, San Diego, CA, vol 3, pp 1129–1135
Dong B, Shen Z (2007) Pseudo-splines, wavelets and framelets. Applied and Computational Harmonic Analysis 22:78–104
Douglas J, Rachford HH (1956) On the numerical solution of heat conduction problems in two and three space variables. Transactions of the American Mathematical Society 82(2):421–439
Eckstein J (1989) Splitting methods for monotone operators with applications to parallel optimization. PhD thesis, Massachusetts Institute of Technology
Eckstein J (1993) Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Mathematics of Operations Research 18(1):202–226
Eckstein J, Bertsekas DP (1992) On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming 55:293–318
Ekeland I, Temam R (1976) Convex Analysis and Variational Problems. North-Holland, Amsterdam
Esser E (2009) Applications of Lagrangian-based alternating direction methods and connections to split Bregman. Technical report, UCLA Computational and Applied Mathematics
Esser E, Zhang X, Chan TF (2009) A general framework for a class of first order primal-dual algorithms for TV minimization. Technical report, UCLA Computational and Applied Mathematics
Figueiredo M, Bioucas-Dias J (2009) Deconvolution of Poissonian images using variable splitting and augmented Lagrangian optimization. In: IEEE Workshop on Statistical Signal Processing, Cardiff
Frick K (2008) The Augmented Lagrangian Method and Associated Evolution Equations. PhD thesis
Gabay D (1983) 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, North–Holland, Amsterdam, chap 9, pp 299–331
Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & Mathematics with Applications 2:17–40
Gilboa G, Osher S (2008) Nonlocal operators with applications to image processing. Multiscale Modelling & Simulation 7(3):1005–1028
Gilboa G, Darbon J, Osher S, Chan TF (2006) Nonlocal convex functionals for image regularization. UCLA CAM Report 06-57
Glowinski R, Marroco A (1975) Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. Revue française d'automatique, informatique, recherche opérationnelle Analyse numérique 9(2):41–76
Goldstein T, Osher S (2009) The Split Bregman method for L1-regularized problems. SIAM Journal on Imaging Sciences 2(2):323–343
Goldstein T, Bresson X, Osher S (2009) Geometric applications of the split Bregman method: Segmentation and surface reconstruction. Technical report, UCLA Computational and Applied Mathematics
Hestenes MR (1969) Multiplier and gradient methods. Journal of Optimization Theory and Applications 4:303–320
Iusem AN (1999) Augmented Lagrangian methods and proximal point methods for convex optimization. Investigación Operativa 8:11–49
Kindermann S, Osher S, Jones PW (2005) Deblurring and denoising of images by nonlocal functionals. Multiscale Modelling & Simulation 4(4):1091–1115
Kiwiel KC (1997) Proximal minimization methods with generalized Bregman functions. SIAM Journal on Control and Optimization 35(4):1142–1168
Krasnoselskii MA (1955) Two observations about the method of successive approximations. Uspekhi Matematicheskikh Nauk 10:123–127, in Russian
Levenberg K (1944) A method for the solution of certain problems in least squares. Quarterly of Applied Mathematics 2:164–168
Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis 16(6):964–979
Mann WR (1953) Mean value methods in iteration. Proceedings of the American Mathematical Society 16(4):506–510
Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal of Applied Mathematics 11:431–441
Moreau JJ (1965) Proximité et dualité dans un espace hilbertien. Bulletin de la Societé Mathématique de France 93:273–299
Mrázek P, Weickert J (2003) Rotationally invariant wavelet shrinkage. In: Michaelis B, Krell G (eds) Pattern Recognition, Springer, Berlin, vol 2781, pp 156–163
Nesterov YE (1983) A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Soviet Mathematics Doklady 27(2):372–376
Nesterov YE (2005) Smooth minimization of non-smooth functions. Mathematical Programming 103:127–152
Opial Z (1967) Weak convergence of a sequence of successive approximations for nonexpansive mappings. Bulletin of the American Mathematical Society 73:591–597
Osher S, Burger M, Goldfarb D, Xu J, Yin W (2005) An iterative regularization method for the total variation based image restoration. Multiscale Modeling & Simulation 4:460–489
Passty GB (1979) Ergodic convergence to a zero of the sum of monotone operators in hilbert space. Journal of Mathematical Analysis and Applications 72:383–390
Powell MJD (1969) A method for nonlinear constraints in minimization problems. Academic Press, London
Rockafellar RT (1970) Convex Analysis. Princeton University Press, Princeton
Rockafellar RT (1976) Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research 1(2):97–116
Ron A, Shen Z (1997) Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator. Journal of Functional Analysis 148:408–447
Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D 60:259–268
Schäfer H (1957) Über die Methode sukzessiver Approximationen. Jahresbericht der Deutschen Mathematiker-Vereinigung 59:131–140
Setzer S (2009a) Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: Lie A, Lysaker M, Morken K, Tai XC (eds) Second International Conference on Scale Space Methods and Variational Methods in Computer Vision, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings, Springer, Lecture Notes in Computer Science, vol 5567, pp 464–476
Setzer S (2009b) Splitting methods in image processing. PhD thesis, University of Mannheim
Setzer S, Steidl G, Teuber T (2009) Deblurring Poissonian images by split Bregman methods. Journal of Visual Communication and Image Representation Accepted
Steidl G, Teuber T (2009) Removing multiplicative noise by Douglas-Rachford splitting methods. Journal of Mathematical Imaging and Vision Accepted
Tai XC, Wu C (2009) Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. In: Lie A, Lysaker M, Morken K, Tai XC (eds) Second International Conference on

Scale Space Methods and Variational Methods in Computer Vision, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings, Springer, Lecture Notes in Computer Science, vol 5567, pp 502–513

Tseng P (1991) Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM Journal on Control and Optimization 29:119–138

Wang Y, Yang J, Yin W, Zhang Y (2008) A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences 1(3):248–272

Weiss P, Blanc-Féraud L, Aubert G (2009) Efficient schemes for total variation minimization under constraints in image processing. SIAM journal on Scientific Computing 31(3):2047–2080

Welk M, Steidl G, Weickert J (2008) Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Applied and Computational Harmonic Analysis 24:195–224

Yin W, Osher S, Goldfarb D, Darbon J (2008) Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences 1(1):143–168

Zhu M (2008) Fast numerical algorithms for total variation based image restoration. PhD thesis, University of California, Los Angeles, USA

Zhu M, Chan TF (2008) An efficient primal-dual hybrid gradient algorithm for total variation image restauration. Technical report, UCLA Computational and Applied Mathematics