Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing

Gerlind Plonka [1], Jianwei Ma [2,3]

[1] Department of Mathematics, University of Duisburg-Essen, Campus Duisburg, 47048 Duisburg, Germany

[2] School of Aerospace, Tsinghua University, Beijing 100084, China

[3] Centre of Geoscience, Ecole des Mines de Paris, 35 rue Saint-Honore 77305, Fontainebleau, France

Abstract

Compressed sensing is a new concept in signal processing. Assuming that a signal can be represented or approximated by only a few suitably chosen terms in a frame expansion, compressed sensing allows to recover this signal from much fewer samples than the Shannon-Nyquist theory requires. Many images can be sparsely approximated in expansions of suitable frames as wavelets, curvelets, wave atoms and others. Generally, wavelets represent point-like features while curvelets represent line-like features well. For a suitable recovery of images, we propose models that contain weighted sparsity constraints in two different frames. Given the incomplete measurements $f = \Phi u + \epsilon$ with the measurement matrix $\Phi \in \mathbb{R}^{K\times N}$, $K << N$, we consider a jointly sparsity-constrained optimization problem of the form $\underset{u}{\operatorname{argmin}}\{\|\Lambda_c\Psi_c u\|_1 + \|\Lambda_w\Psi_w u\|_1 + \frac{1}{2}\|f - \Phi u\|_2^2\}$. Here Ψ_c and Ψ_w are the transform matrices corresponding to the two frames, and the diagonal matrices Λ_c, Λ_w contain the weights for the frame coefficients. We present efficient iteration methods to solve the optimization problem, based on Alternating Split Bregman algorithms. The convergence of the proposed iteration schemes will be proved by showing that they can be understood as special cases of the Douglas-Rachford Split algorithm. Numerical experiments for compressed sensing based Fourier-domain random imaging show good performances of the proposed curvelet-wavelet regularized split Bregman (CWSpB) methods, where we particularly use a combination of wavelet and curvelet coefficients as sparsity constraints.

Keywords: Compressed sensing, compressive sampling, Alternating Split Bregman iteration, Douglas-Rachford split algorithm, iterative shrinkage/thresholding (IST), curvelets, remote sensing CS imaging

1 Introduction

1.1 Background

In many engineering fields, e.g., geophysical exploration, medical magnetic resonance imaging (MRI) or remote sensing, we have to deal with problems on incomplete and inaccurate measurements limited by physical constraints or extreme expenses. Compressed sensing (CS), see [7, 8, 9, 21], is a new concept to solve these problems. The CS theory says that

a compressible unknown signal can be recovered from a small number of random measurements by sparsity-promoting nonlinear recovery algorithms. Instead of taking usual samples, the measurements are inner products of signals with random vectors. The data acquisition now depends on its sparsity rather than its bandwidth. Compressed sensing might have an important impact for measurement mechanisms and design of devices in various engineering fields.

Over the last few years, there have been increased interests in applications of the compressed sensing in compressive optical imaging [23], medical MRI [35, 36], analog-to-information conversion [26], DNA biosensing [48], seismic exploration [34], astronomy [4], machine learning [54], surface metrology [38], and remote sensing [39, 40, 41], etc.

One crucial step for the CS is its nonlinear recovery, the so-called decoding. Different recovery algorithms have been proposed in the last couple of years, as e.g. linear programming [7], reweighted linear programming [10], gradient projection sparse reconstruction [27], orthogonal matching pursuit (OMP) [52], stagewise OMP [22], Bregman iteration [55], fixed-point continuation [30], inverse scale space nonlinear diffusion [31, 37], Bayesian statistical method [32], spectral projected gradient method [53], iterative shrinkage/thresholding (IST) algorithms [18, 2, 3, 24, 45, 28, 40], model-based recovery algorithm [1], and nonconvex ℓ_p-norm optimization with $p \in (0, 1)$ [13, 46].

Among the existing methods, the iterative shrinkage methods are quite universal, robust and simple to implement by engineers. Another advantage is that existing transforms can be incorporated easily into the IST framework. Therefore, it became one of the most popular tools for solving of linear inverse problems.

In this paper, we propose a numerical algorithm based on the Alternating Split Bregman method. The Bregman iteration is a concept from functional analysis for finding extrema of convex functionals. The splitting technique is used to decouple the ℓ_1- and ℓ_2-components in the functional that we wish to minimize.

The split Bregman method was recently proposed by Goldstein and Osher [29]. It can effectively solve general ℓ_1-regularized optimization problems with multiple ℓ_1-regularized terms while the linearized Bregman algorithm [5, 55] and fixed point continuation methods [30] fail. If the considered optimization problem is uniquely solvable, then the convergence of the split Bregman iteration for one ℓ_1-regularization term has been proved by Cai, Osher and Shen [6]. Recently, Setzer [47] showed very close relations between the split Bregman iteration, the augmented Lagrangian method, the Douglas-Rachford splitting and frame shrinkage. Another advantage of the split Bregman method is that it has a relatively small memory footprint and is easy to program by users [29, 6]. These properties are significant for large scale problems.

In this paper, we propose a two-frame regularized alternating split Bregman algorithm for compressed sensing, and prove its convergence. For the numerical examples, we apply a wavelet basis [17] and the curvelet frame [11, 12]. Our paper is motivated by the observation that images consist of different components (point-like features and line-like features) that can not be sparsely represented by one frame expansion equally well. Therefore, we want to allow weighted sparsity constraints in two different frames. For example, wavelets and curvelets are complementary well-suited for point-like features and line-like features, respectively. Another motivation is to use the split Bregman method to accelerate previous iterative curvelet thresholding methods for real applications in large-scale remote sensing.

1.2 The model for compressed sensing and sparse frame expansions

Let us consider the compressed sensing problem with incomplete measurements [7, 8, 9, 21]

$$f = \Phi u + \epsilon, \tag{1.1}$$

where $\Phi \in \mathbb{R}^{K\times N}\,(K < N)$ is a given so-called CS measurement matrix or a lens-based optical imaging architecture. The recovery of the signal $u \in \mathbb{R}^N$ from the observation $f \in \mathbb{R}^K$ is an underdetermined linear system that leads to a seriously ill-posed problem because there are much fewer rows than columns of Φ. However, let us assume that the signal u possesses a sparse representation in a certain basis or a frame, i.e., there is a transform matrix $\Psi \in \mathbb{R}^{M\times N}$ (with $M = N$ for a basis and $M > N$ for a frame), such that Ψu contains only a small set of significant entries. Further, let the measurement matrix Φ not be correlated with Ψ. Usually, one assumes that Φ satisfies the so-called Restricted Isometry Property (RIP), see [7, 9]. Then u can be reconstructed with high accuracy from the incomplete measurements f. Frequently used measurement matrices Φ are sparse binary random matrices, randomly downsampled Fourier matrices and Toeplitz block matrices. Especially, pseudo-orthogonal measurement matrices Φ satisfying $\Phi\,\Phi^T = I_K$, where I_K denotes the identity matrix of size K, have been shown to result in faster convergence in nonlinear decoding schemes, see e.g. [23] for single-pixel CS imaging.

Applying a transform Ψ to u, the CS-problem reads in coefficient domain

$$f = \Phi\Psi^{-1}\Psi u + \epsilon = \tilde{\Phi}\vartheta + \epsilon \tag{1.2}$$

with $\vartheta = \Psi u$ and $\tilde{\Phi} = \Phi\Psi^{-1}$, where the matrix Ψ denotes a forward transform (e.g., a curvelet transform, a wavelet transform or a trigonometric transform) and Ψ^{-1} is its inversion. (In the case of frames, $\Psi^{-1} = \Psi^{\dagger}$ denotes the generalized inverse.)

One classical method to find a solution of the CS-problem (1.1) (resp.(1.2)) is the sparsity-promoting basis pursuit denoising method which solves the convex optimization problem

$$BP_\zeta:\ \underset{u}{\operatorname{argmin}}\left\{\|\Psi u\|_1:\ \ \|f-\Phi u\|\le\zeta\right\}, \tag{1.3}$$

where the positive parameter ζ is an estimate of the noise level. Alternatively, one considers the unconstrained problem

$$QP_\lambda:\ \underset{u}{\operatorname{argmin}}\left\{\frac{1}{2}\|f-\Phi u\|_2^2+\lambda\|\Psi u\|_1\right\} \tag{1.4}$$

that can be solved by quadratic programming, and the Lasso (LS) method

$$LS_\xi:\ \underset{u}{\operatorname{argmin}}\left\{\frac{1}{2}\|f-\Phi u\|_2^2:\ \|\Psi u\|_1<\xi\right\}. \tag{1.5}$$

It has been verified that these three problems (BP_ζ, QP_λ, and LS_ξ) are closely related by using appropriate parameters ζ, λ, and ξ, see e.g. [53]. These models are called "analysis models" [25], where one seeks the u whose coefficients in a certain frame possess an ℓ_1 minimum.

The above problems can be also solved in coefficient domain, replacing u by its frame coefficients ϑ and the matrix Φ by $\Phi\Psi^{-1}$, e.g.,

$$QP_\lambda:\ \underset{\vartheta}{\operatorname{argmin}}\left\{\frac{1}{2}\|f-\Phi\Psi^{-1}\vartheta\|_2^2+\lambda\|\vartheta\|_1\right\}. \tag{1.6}$$

This model is called "synthesis model" [25], where one directly seeks the ℓ_1-minimization coefficients.

Whether to take a synthesis or an analysis based model depends on the real applications. The synthesis based approach is only applicable if $\Psi \in \mathbb{R}^{M\times N}$ with $M \geq N$ possesses a left inverse. If one uses an orthogonal transform Ψ, both models are equivalent. In this paper we will focus on the analysis based approach.

1.3 Two-frame sparsity promoting models for compressed sensing

Images usually consist of different components, they may contain smooth structures and textures. Therefore, while parts of the image can be represented in a highly sparse manner by a wavelet-like transform, others do not. Within the last years, many sophisticated methods were developed to design efficient representations of two-dimensional data. Curvelets [11, 12], contourlets [20] and wave atoms [19] are examples for non-adaptive function frames with strong anisotropic directional selectivity. These transforms behave differently for smooth parts and textures. For oriented textures, wave atoms lead to a significantly sparser expansion than Garbor filters or tensor product wavelets. Therefore, we propose to solve an optimization problem that involves two different basis/frame transforms of the signal u. These transforms need to be chosen suitably using a priori information of the signal u.

Let us consider two different transform matrices Ψ_c and Ψ_w, where we assume that these matrices correspond to orthogonal bases or Parseval frames, i.e., let

$$\Psi_c \in \mathbb{R}^{N_1\times N}, \quad \Psi_w \in \mathbb{R}^{N_2\times N} \quad \text{with} \quad \Psi_c^T\Psi_c = \Psi_w^T\Psi_w = I_N, \quad N_1 \geq N,\, N_2 \geq N,$$

where I_N denotes the identity matrix of size N, and Ψ^T is the transpose of Ψ.

In coefficient domain, we are interested in minimizing the ℓ_1-norm of suitably scaled (transform) coefficients. The scaling can be used for emphasizing (or reducing) of coefficients in different levels or different regions of the image, and also for weighting of the coefficients of the two different frames. Therefore, we allow diagonal matrices Λ_c and Λ_w, and propose the following two models.

Denoising model. If the measurements f contain significant noise, we consider the generalized optimization problem

$$\operatorname*{argmin}_u\{J(u) + \frac{1}{2}\|f - \Phi u\|_2^2\}, \tag{1.7}$$

with a smoothing term of the form

$$J(u) = \|\Lambda_c\Psi_c u\|_1 + \|\Lambda_w\Psi_w u\|_1.$$

Observe that for a wavelet transform, the ℓ_1-norm of (suitably scaled) wavelet coefficients is equivalent with a Besov norm of u, while for other frames, we do not have such a direct interpretation of the functional $J(u)$.

The corresponding constrained problem that is equivalent with (1.7) reads

$$\operatorname*{argmin}_{u,\vartheta_c,\vartheta_w}\{\|\Lambda_c\vartheta_c\|_1 + \|\Lambda_w\vartheta_w\|_1 + \frac{1}{2}\|f - \Phi u\|_2^2\} \quad \text{s.t.} \quad \vartheta_c = \Psi_c u, \quad \vartheta_w = \Psi_w u. \tag{1.8}$$

Reconstruction model. If the measurements f contain only a negligible amount of noise, we consider the optimization problem

$$\underset{u}{\operatorname{argmin}}\{J(u)\} \quad s.t. \quad \Phi u = f$$

with $J(u)$ as above. The corresponding constrained problem in coefficient domain now reads

$$\begin{aligned} &\underset{\vartheta_c,\vartheta_w}{\operatorname{argmin}}\{\|\Lambda_c\vartheta_c\|_1 + \|\Lambda_w\vartheta_w\|_1\} \\ &\text{s.t. there exists an } u \in \mathbb{R}^N \text{ with} \quad f = \Phi u, \quad \vartheta_c = \Psi_c u, \quad \vartheta_w = \Psi_w u. \end{aligned} \tag{1.9}$$

In Section 2, we derive fast algorithms for the optimization problems (1.8) and (1.9), respectively. For this purpose, we use the Alternating Split Bregman method that has been recently proposed in [29]. In the framework of the Split Bregman method, the variables ϑ_c, ϑ_w, and u are decoupled so that they can be found separately and flexibly. In particular, in contrast to [6, 29], using a CS matrix $\Phi \in \mathbb{R}^{K\times N}$ that is obtained by random downsampling of the rows of an orthogonal matrix, we need not to solve a linear system of equations in each iteration step of the algorithm.

In Section 3, we prove the convergence of the proposed Alternating Split Bregman algorithm for the denoising model. Using the observations in [47], our iteration method is shown to be equivalent with the Douglas-Rachford splitting algorithm [33, 14]. In this way we can guarantee its convergence. Finally, in Section 4, the methods are applied to compressed sensing with a curvelet frame and a wavelet basis. Wavelets and curvelets are complementary well-suited for point-like features and line-like features, respectively. A short description of curvelets can be found in the Appendix.

Let us explicitly outline the difference of our paper from some related work. Our method can be interpreted as an extension of Goldstein and Osher's split Bregman iteration [29], where the TV regularization and Haar wavelet regularization are considered. Compared with [29], we use a more general scaling coefficient approach. Our method is closely related to Cai et al's work [6], where the framelets are considered. As in [6], we apply the so-called analysis-based approach. But our proof of convergence for the derived split Bregman iteration is completely different from that in [6]. In particular, we need not to assume that the considered optimization problem is uniquely solvable. This assumption is a key ingredient in the convergence proof in [6], but it will be hardly satisfied for optimization problems in compressed sensing. Our proof is based on the observations by Setzer [47], but it does not use the dual formulation of the optimization problem as in [47].

Observe that our model is different from the models for image separation (see [50, 49]). In [50], Starck et al. proposed a MCA (morphological component analysis) method for image separation, in which a model $\underset{u_c,u_w}{\operatorname{argmin}}\{\|\Psi_c u_c\|_1 + \|\Psi_w u_w\|_1 + \frac{1}{2}\|f - u_c - u_w\|_2^2\}$ uses two frame ℓ_1 sparsity constraints to decompose the image into cartoon and texture components $u_c + u_w$. Our model is similar to [49], where Starck et al. used the curvelet thresholding and wavelet thresholding repeatedly to gain the image deconvolution. In contrast to [49], however, our method is implemented with Split Bregman iterations and applied to compressed sensing.

2 The Alternating Split Bregman method

In order to be able to solve the constrained optimization problems (1.8) and (1.9), we wish to convert them into unconstrained optimization problems. This can be done using the concept of "Bregman distance".

Let H be a Hilbert space, and let $E : H \to \mathbb{R} \cup \{\infty\}$ be a convex functional. The Bregman distance associated with E at $v \in H$ is defined by

$$D_E^p(u, v) := E(u) - E(v) - \langle p, u - v\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in H, and where p is a subgradient of E at v, i.e. $p \in \partial E(v)$, see e.g. [44, 29]. The Bregman distance is usually not symmetric, but we have $D_E^p(u, v) \geq 0$ and $D_E^p(u, v) \geq D_E^p(w, v)$ if $w = \lambda u + (1 - \lambda)v$ with $\lambda \in [0, 1]$. This concept will be used now to derive the new iteration methods.

2.1 The Denoising Model

The classical method to convert (1.8) into an unconstrained problem is to consider a sequence of problems of the form

$$\operatorname*{argmin}_{u,\vartheta_c,\vartheta_w} \left\{ E(u, \vartheta_c, \vartheta_w) + \frac{\mu_k}{2} \left(\|\Psi_c u - \vartheta_c\|_2^2 + \|\Psi_w u - \vartheta_w\|_2^2 \right) \right\}, \tag{2.1}$$

where

$$E(u, \vartheta_c, \vartheta_w) := \|\Lambda_c \vartheta_c\|_1 + \|\Lambda_w \vartheta_w\|_1 + \frac{1}{2}\|f - \Phi u\|_2^2$$

and with an increasing sequence $\mu_1 < \mu_2 < \ldots$. In contrast to this approach, Goldstein and Osher [29] suggest to apply the Bregman distance in order to obtain an iterative method to solve the unconstrained optimization problem. In our case, let

$$\begin{aligned} D_E^p((u, \vartheta_c, \vartheta_w), (u^k, \vartheta_c^k, \vartheta_w^k)) \;&:=\; E(u, \vartheta_c, \vartheta_w) - E(u^k, \vartheta_c^k, \vartheta_w^k) \\ &\quad - \langle p_u^k, u - u^k\rangle - \langle p_c^k, \vartheta_c - \vartheta_c^k\rangle - \langle p_w^k, \vartheta_w - \vartheta_w^k\rangle, \end{aligned}$$

where (p_u^k, p_c^k, p_w^k) is the subgradient of E at $(u^k, \vartheta_c^k, \vartheta_w^k)$. Instead of (2.1), we consider the iteration

$$\begin{aligned} &(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) \\ =\;& \operatorname*{argmin}_{u,\vartheta_c,\vartheta_w} \left\{ D_E^p((u, \vartheta_c, \vartheta_w), (u^k, \vartheta_c^k, \vartheta_w^k)) + \frac{\mu}{2} \left(\|\Psi_c u - \vartheta_c\|_2^2 + \|\Psi_w u - \vartheta_w\|_2^2 \right) \right\} \\ =\;& \operatorname*{argmin}_{u,\vartheta_c,\vartheta_w} \Big\{ E(u, \vartheta_c, \vartheta_w) - \langle p_u^k, u\rangle - \langle p_c^k, \vartheta_c\rangle - \langle p_w^k, \vartheta_w\rangle \\ &\qquad + \frac{\mu}{2} \left(\|\Psi_c u - \vartheta_c\|_2^2 + \|\Psi_w u - \vartheta_w\|_2^2 \right) \Big\}, \end{aligned} \tag{2.2}$$

where $\mu > 0$ can be taken as a fixed parameter. This yields the necessary condition

$$\begin{aligned} 0 \;\in\; \partial \Big[& E(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) - \langle p_u^k, u^{k+1}\rangle - \langle p_c^k, \vartheta_c^{k+1}\rangle - \langle p_w^k, \vartheta_w^{k+1}\rangle \\ & + \frac{\mu}{2} \left(\|\Psi_c u^{k+1} - \vartheta_c^{k+1}\|_2^2 + \|\Psi_w u^{k+1} - \vartheta_w^{k+1}\|_2^2 \right) \Big]. \end{aligned}$$

Calculating this subdifferential we find by $(p_u^{k+1}, p_c^{k+1}, p_w^{k+1}) \in \partial E(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1})$ the recursion relations

$$\left\{\begin{array}{lcl} p_u^{k+1} &=& p_u^k - \mu\left(\Psi_c^T(\Psi_c u^{k+1} - \vartheta_c^{k+1}) + \Psi_w^T(\Psi_w u^{k+1} - \vartheta_w^{k+1})\right), \\ p_c^{k+1} &=& p_c^k - \mu\left(\vartheta_c^{k+1} - \Psi_c u^{k+1}\right), \\ p_w^{k+1} &=& p_w^k - \mu\left(\vartheta_w^{k+1} - \Psi_w u^{k+1}\right). \end{array}\right. \tag{2.3}$$

Applying the simplification idea in [29], we obtain from (2.2)–(2.3) the split Bregman iteration

$$\left\{\begin{array}{lcl} (u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) &=& \underset{u,\vartheta_c,\vartheta_w}{\operatorname{argmin}} \big\{ E(u, \vartheta_c, \vartheta_w) + \frac{\mu}{2}\|\vartheta_c - \Psi_c u - b_c^k\|_2^2 \\ & & \qquad + \frac{\mu}{2}\|\vartheta_w - \Psi_w u - b_w^k\|_2^2 \big\}, \\ b_c^{k+1} &=& b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1}, \\ b_w^{k+1} &=& b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}, \end{array}\right. \tag{2.4}$$

where we can choose an arbitrary start vector $(u^0, \vartheta_c^0, \vartheta_w^0)$ and $(b_c^0, b_w^0) = (0, 0)$. We show that the iteration in (2.4) indeed coincides with the previous algorithm.

Theorem 2.1 *The sequence $\{(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1})\}_{k\geq 0}$ obtained by the iteration rules (2.2)–(2.3) with an arbitrary fixed starting vector $(u^0, \vartheta_c^0, \vartheta_w^0)$ and with $(p_u^0, p_c^0, p_w^0) = (0, 0, 0)$ coincides with the corresponding sequence $\{(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1})\}_{k\geq 0}$ obtained by (2.4) with $(b_c^0, b_w^0) = (0, 0)$.*

Proof. We apply the iteration rules (2.2)–(2.3) and determine the sequences $\{b_c^k\}_{k\geq 0}$ and $\{b_w^k\}_{k\geq 0}$ by

$$b_c^k := \frac{1}{\mu} p_c^k, \qquad b_w^k := \frac{1}{\mu} p_w^k. \tag{2.5}$$

Then from $(p_u^0, p_c^0, p_w^0) = (0, 0, 0)$ we find $(b_c^0, b_w^0) = (0, 0)$, and for $k = 0$, we simply observe that equation (2.2) and the first equation in (2.4) coincide. Further, b_c^k and b_w^k in (2.5) satisfy by (2.3) the recursion formulas

$$b_c^{k+1} = \frac{1}{\mu} p_c^{k+1} = \frac{1}{\mu} p_c^k - \vartheta_c^{k+1} + \Psi_c u^{k+1} = b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1},$$

and analogously, $b_w^{k+1} = b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}$. Using these recursions for b_c^k and b_w^k, we find

$$\begin{aligned} \frac{\mu}{2}\|\vartheta_c - \Psi_c u - b_c^k\|_2^2 &= \frac{\mu}{2}\|\vartheta_c - \Psi_c u\|_2^2 - \mu\langle \vartheta_c - \Psi_c u, b_c^k\rangle + \frac{\mu}{2}\|b_c^k\|_2^2 \\ &= \frac{\mu}{2}\|\vartheta_c - \Psi_c u\|_2^2 - \langle \vartheta_c - \Psi_c u, p_c^k\rangle + \frac{\mu}{2}\|b_c^k\|_2^2, \end{aligned}$$

and

$$\frac{\mu}{2}\|\vartheta_w - \Psi_w u - b_w^k\|_2^2 = \frac{\mu}{2}\|\vartheta_w - \Psi_w u\|_2^2 - \langle \vartheta_w - \Psi_w u, p_w^k\rangle + \frac{\mu}{2}\|b_w^k\|_2^2.$$

Hence (2.2) can be rewritten as

$$\begin{aligned} (u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) = \underset{u,\vartheta_c,\vartheta_w}{\operatorname{argmin}} \Big\{ E(u, \vartheta_c, \vartheta_w) + \frac{\mu}{2}\left(\|\Psi_c u - \vartheta_c - b_c^k\|_2^2 + \|\Psi_w u - \vartheta_w - b_w^k\|_2^2\right) \\ - \langle p_c^k, \Psi_c u\rangle - \langle p_w^k, \Psi_w u\rangle - \langle p_u^k, u\rangle \Big\}. \end{aligned}$$

This relation is equal to the first equation in (2.4) since from (2.3) we observe that

$$p_u^{k+1} - p_u^k = -\Psi_c^T(p_c^{k+1} - p_c^k) - \Psi_w^T(p_w^{k+1} - p_w^k),$$

and iterative application yields $p_u^{k+1} + \Psi_c^T p_c^{k+1} + \Psi_w^T p_w^{k+1} = p_u^0 + \Psi_c^T p_c^0 + \Psi_w^T p_w^0 = 0$ for all $k \geq 0$. □

In order to solve the minimization problem (2.4), an Alternating Split Bregman algorithm is applied. We finally obtain the following scheme.

Alternating Split Bregman Algorithm for the Denoising Model.

$$\begin{cases} u^{k+1} &= \underset{u}{\operatorname{argmin}} \left\{ \frac{1}{2}\|f - \Phi u\|_2^2 + \frac{\mu}{2}\left(\|\vartheta_c^k - \Psi_c u - b_c^k\|_2^2 + \|\vartheta_w^k - \Psi_w u - b_w^k\|_2^2\right)\right\}, \\ \vartheta_c^{k+1} &= \underset{\vartheta_c}{\operatorname{argmin}} \left\{ \|\Lambda_c \vartheta_c\|_1 + \frac{\mu}{2}\|\vartheta_c - \Psi_c u^{k+1} - b_c^k\|_2^2\right\}, \\ \vartheta_w^{k+1} &= \underset{\vartheta_w}{\operatorname{argmin}} \left\{ \|\Lambda_w \vartheta_w\|_1 + \frac{\mu}{2}\|\vartheta_w - \Psi_w u^{k+1} - b_w^k\|_2^2\right\}, \\ b_c^{k+1} &= b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1}, \\ b_w^{k+1} &= b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}. \end{cases} \tag{2.6}$$

At the end of this subsection, we will show that a fixed point of this iteration process indeed solves the original optimization problem (1.8). The convergence proof is given in Section 3.

Let us now turn to the numerical treatment of the iteration rules in (2.6). The second and the third equation can be simply solved by a componentwise shrinkage. Indeed, with $\Lambda_c = \operatorname{diag}\left((\lambda_{c,l})_{l=0}^{N_1-1}\right)$ we have for the l-th component of ϑ_c^{k+1},

$$\vartheta_{c,l}^{k+1} = \underset{\vartheta_{c,l}}{\operatorname{argmin}} \left\{ |\lambda_{c,l}\,\vartheta_{c,l}| + \frac{\mu}{2}|\vartheta_{c,l} - (\Psi_c u^{k+1})_l - b_{c,l}^k|^2 \right\},$$

i.e.,

$$0 \in |\lambda_{c,l}| \frac{\vartheta_{c,l}^{k+1}}{|\vartheta_{c,l}^{k+1}|} + \mu\left(\vartheta_{c,l}^{k+1} - (\Psi_c u^{k+1})_l - b_{c,l}^k\right), \tag{2.7}$$

where $\frac{\vartheta_{c,l}^{k+1}}{|\vartheta_{c,l}^{k+1}|}$ denotes the set $[-1,1]$ for $\vartheta_{c,l}^{k+1} = 0$. Hence we find a solution by soft shrinkage, i.e., for $l = 0, \ldots, N_1 - 1$,

$$\vartheta_{c,l}^{k+1} = \begin{cases} (\Psi_c u^{k+1})_l + b_{c,l}^k - \frac{|\lambda_{c,l}|}{\mu} & \text{if } (\Psi_c u^{k+1})_l + b_{c,l}^k \geq \frac{|\lambda_{c,l}|}{\mu}, \\ (\Psi_c u^{k+1})_l + b_{c,l}^k + \frac{|\lambda_{c,l}|}{\mu} & \text{if } (\Psi_c u^{k+1})_l + b_{c,l}^k \leq -\frac{|\lambda_{c,l}|}{\mu}, \\ 0 & \text{if } |(\Psi_c u^{k+1})_l + b_{c,l}^k| < \frac{|\lambda_{c,l}|}{\mu}. \end{cases}$$

Analogously, we compute ϑ_w^{k+1} by a componentwise shrinkage procedure. In vector form, we write this shrinkage shortly as

$$\vartheta_c^{k+1} = \mathcal{T}_{\mu^{-1}|\Lambda_c|}(\Psi_c u^{k+1} + b_c^k), \qquad \vartheta_w^{k+1} = \mathcal{T}_{\mu^{-1}|\Lambda_w|}(\Psi_w u^{k+1} + b_w^k), \tag{2.8}$$

where $|\Lambda_c| := \operatorname{diag}\left(|\lambda_{c,l}|\right)_{l=0}^{N_1-1}$, $|\Lambda_w| := \operatorname{diag}\left(|\lambda_{w,l}|\right)_{l=0}^{N_1-1}$ have only nonnegative entries at the diagonal. Using the last two equations in (2.6), the results for ϑ_c^{k+1} and ϑ_w^{k+1} in (2.8) directly imply that

$$b_{c,l}^{k+1} \in \left[-\tfrac{|\lambda_{c,l}|}{\mu}, \tfrac{|\lambda_{c,l}|}{\mu}\right], \qquad b_{w,l}^{k+1} \in \left[-\tfrac{|\lambda_{w,l}|}{\mu}, \tfrac{|\lambda_{w,l}|}{\mu}\right]. \tag{2.9}$$

The functional in the first equation of (2.6) is differentiable. We find

$$\Phi^T(\Phi u^{k+1} - f) - \mu\Psi_c^T(\vartheta_c^k - \Psi_c u^{k+1} - b_c^k) - \mu\Psi_w^T(\vartheta_w^k - \Psi_w u^{k+1} - b_w^k) = 0, \qquad (2.10)$$

and with

$$\begin{aligned} \Phi_{temp} &:= \Phi^T\Phi + \mu\,(\Psi_c^T\Psi_c + \Psi_w^T\Psi_w) \\ f_{temp} &:= \Phi^T f + \mu\left(\Psi_c^T(\vartheta_c^k - b_c^k) + \Psi_w^T(\vartheta_w^k - b_w^k)\right) \end{aligned}$$

it follows that

$$\Phi_{temp}\, u^{k+1} = f_{temp}.$$

Remember that $\Psi_c^T\Psi_c = \Psi_w^T\Psi_w = I_N$, such that $\Phi_{temp} = \Phi^T\Phi + 2\mu\, I_N$ is positive definite for any $\mu > 0$. Further, since $\Phi \in \mathbb{R}^{K\times N}$ $(K < N)$ is assumed to be obtained by random downsampling of the rows of an orthogonal $N \times N$-matrix, we have $\Phi\Phi^T = I_K$. Therefore, we obtain $(\Phi^T\Phi + 2\mu I_N)^{-1} = \frac{1}{2\mu}(I_N - \frac{1}{2\mu+1}\Phi^T\Phi)$ since

$$\begin{aligned} \tfrac{1}{2\mu}(I_N - \tfrac{1}{2\mu+1}\Phi^T\Phi)(\Phi^T\Phi + 2\mu I_N) &= \tfrac{1}{2\mu}(\Phi^T\Phi + 2\mu I_N - \tfrac{1}{2\mu+1}\Phi^T\Phi - \tfrac{2\mu}{2\mu+1}\Phi^T\Phi) \\ &= I_N. \end{aligned}$$

Hence, by $\frac{1}{2\mu}(I_N - \frac{1}{2\mu+1}\Phi^T\Phi)\Phi^T f = \frac{1}{2\mu}(\Phi^T f - \frac{1}{2\mu+1}\Phi^T f) = \frac{1}{2\mu+1}\Phi^T f$ we obtain

$$u^{k+1} = \tfrac{1}{2\mu+1}\Phi^T f + \tfrac{1}{2}(I_N - \tfrac{1}{2\mu+1}\Phi^T\Phi)(\Psi_c^T(\vartheta_c^k - b_c^k) + \Psi_w^T(\vartheta_w^k - b_w^k)).$$

We outline the obtained Alternating Split Bregman algorithm for two frames (e.g., curvelets and wavelets) in Table 1.

Input: $(u^0, \vartheta_w^0, \vartheta_c^0, b_c^0, b_w^0) = (0,0,0,0,0)$.
While a stop criterion is not satisfied
1) $u^{k+1} = \frac{1}{2\mu+1}\Phi^T f + \frac{1}{2}(I_N - \frac{1}{2\mu+1}\Phi^T\Phi)(\Psi_c^T(\vartheta_c^k - b_c^k) + \Psi_w^T(\vartheta_w^k - b_w^k))$
2) $\vartheta_c^{k+1} = \mathcal{T}_{\mu^{-1}\lvert\Lambda_c\rvert}(\Psi_c u^{k+1} + b_c^k)$
$\vartheta_w^{k+1} = \mathcal{T}_{\mu^{-1}\lvert\Lambda_w\rvert}(\Psi_w u^{k+1} + b_w^k)$
3) $b_c^{k+1} = b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1}$
$b_w^{k+1} = b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}$
end

Table 1: Alternating Split Bregman algorithm for two frames for solving (1.8).

Observe that the obtained algorithm is much simpler than the split Bregman algorithms proposed in [29] and [6] (for image restoration), since we need not to solve a linear system of equations in the first step but can compute u^{k+1} directly.

Let us remark that if the measurement matrix Φ is orthogonal, i.e., $\Phi^T\Phi = I_N$, we obtain $\Phi_{temp} = (2\mu+1)I_N$ and hence

$$u^{k+1} = \tfrac{1}{2\mu+1} f_{temp}.$$

This special case may occur if f is a vector of noisy measurements of a rotated image u.

In the algorithm in Table 1, steps 2 and 3 are carried out in coefficient domain.

Input: $(u^0, u_w^0, u_c^0, \tilde{b}_c^0, \tilde{b}_w^0) = (0, 0, 0, 0, 0)$.

While a stop criterion is not satisfied

1) $u^{k+1} = \frac{1}{2\mu+1}\Phi^T f + \frac{1}{2}(I_N - \frac{1}{2\mu+1}\Phi^T\Phi)((u_c^k - \tilde{b}_c^k) + (u_w^k - \tilde{b}_w^k))$

2) $u_c^{k+1} = \mathcal{S}_{\mu^{-1}|\Lambda_c|}(u^{k+1} + \tilde{b}_c^k)$
$u_w^{k+1} = \mathcal{S}_{\mu^{-1}|\Lambda_w|}(u^{k+1} + \tilde{b}_w^k)$

3) $\tilde{b}_c^{k+1} = \tilde{b}_c^k + u^{k+1} - u_c^{k+1}$
$\tilde{b}_w^{k+1} = \tilde{b}_w^k + u^{k+1} - u_w^{k+1}$

end

Table 2: Alternating Split Bregman algorithm in spatial domain for two frames.

Alternatively, one can derive a similar algorithm directly in spatial domain (purely analysis-based approach). Observe that, by assumption $\Psi_c^T\Psi_c = \Psi_w^T\Psi_w = I_N$, i.e., the matrices Ψ_c^T and Ψ_w^T are the (generalized) inverse transform matrices. Let now

$$u_c^k := \Psi_c^T\vartheta_c^k, \quad u_w^k := \Psi_w^T\vartheta_w^k, \quad \tilde{b}_c^k = \Psi_c^T b_c^k, \quad \tilde{b}_w^k = \Psi_w^T b_w^k.$$

Then we obtain the algorithm in spatial domain in Table 2.

Step 2 now contains a frame shrinkage procedure (e.g. a curvelet and a wavelet shrinkage), i.e.,

$$\begin{aligned}\mathcal{S}_{\mu^{-1}|\Lambda_c|}(u^{k+1} + \tilde{b}_c^k) &:= \Psi_c^T\mathcal{T}_{\mu^{-1}|\Lambda_c|}(\Psi_c(u^{k+1} - \tilde{b}_c^k)),\\ \mathcal{S}_{\mu^{-1}|\Lambda_w|}(u^{k+1} + \tilde{b}_w^k) &:= \Psi_w^T\mathcal{T}_{\mu^{-1}|\Lambda_w|}(\Psi_w(u^{k+1} - \tilde{b}_w^k)),\end{aligned}$$

with the shrinkage operator $\mathcal{T}_{\mu^{-1}|\Lambda|}$ defined in (2.8). Note that this frame shrinkage procedure is only equivalent with step 2 in Table 1, if $\Psi_c\,\Psi_c^T = \Psi_w\,\Psi_w^T = I_N$, i.e., if we apply orthonormal bases and no frames.

We prefer to use the analysis-based method in Table 2, since, due to the redundancy of curvelet transform, the number of curvelet coefficients is much higher than the number of pixels, so that the analysis-based method can reduce the computational cost to some extent. Further, it has been observed in numerical experiments that the analysis-based method is often superior over the synthesis-based algorithm (see e.g. [25, 10] for some descriptions and insights).

It should be noted that the algorithm given in Table 1 is not a rigorous synthesis-based model. However, using the constraints in (1.8), we can take for example $u = \frac{1}{2}\left(\Psi_c^T\vartheta_c + \Psi_w^T\vartheta_w\right)$, and a purely synthesis-based optimization model then reads

$$\underset{\vartheta_c,\vartheta_w}{\operatorname{argmin}}\{\|\Lambda_c\vartheta_c\|_1 + \|\Lambda_w\vartheta_w\|_1 + \tfrac{1}{2}\|f - \tfrac{1}{2}\Phi(\Psi_c^T\vartheta_c + \Psi_w^T\vartheta_w)\|_2^2\}.$$

The matrix $(\Psi_c^T, \Psi_w^T)^T$ denotes a combination of wavelet and curvelet transform and has also been used by Starck et al. in [50].

We show in the next theorem that any fixed point of the Alternating Split Bregman algorithm in (2.6) is indeed a minimizer of the original constrained problem (1.8). The convergence of this algorithm will be studied in Section 3.

Theorem 2.2 *Assume that $(u^*, \vartheta_c^*, \vartheta_w^*, b_c^*, b_w^*)$ is a fixed point of the Alternating Split Bregman iteration (2.6). Then $(u^*, \vartheta_c^*, \vartheta_w^*)$ solves the constrained optimization problem (1.8).*

Proof. From $b_c^* = b_c^* + \Psi_c u^* - \vartheta_c^*$ and $b_w^* = b_w^* + \Psi_w u^* - \vartheta_w^*$ it follows that the constraints $\vartheta_c^* = \Psi_c u^*$ and $\vartheta_w^* = \Psi_w u^*$ are satisfied. Let now

$$\Lambda := \begin{pmatrix} \Lambda_c & 0 \\ 0 & \Lambda_w \end{pmatrix}, \quad \vartheta := \begin{pmatrix} \vartheta_c \\ \vartheta_w \end{pmatrix}, \quad \Psi := \begin{pmatrix} \Psi_c \\ \Psi_w \end{pmatrix}.$$

Then from $\vartheta = \Psi u$ it follows that $\Psi^T \vartheta = \Psi^T \Psi u = 2u$, and the functional from (1.8) can be written as

$$\begin{aligned} E(\vartheta, u) &= \|\Lambda\vartheta\|_1 + \tfrac{1}{2}\|f - \Phi u\|_2^2 = \|\Lambda\vartheta\|_1 + \tfrac{1}{2}\|f - \tfrac{1}{2}\Phi\Psi^T\vartheta\|_2^2 = E(\vartheta) \\ &= \|\Lambda\Psi u\|_1 + \tfrac{1}{2}\|f - \Phi u\|_2^2 = E(u). \end{aligned}$$

In particular, this convex functional is minimized by $u^* = \frac{1}{2}\Psi^T\vartheta^*$ if $0 \in \Psi^T(\partial E(\vartheta^*))$, i.e., if

$$0 \in \Psi^T \left[|\Lambda| \left(\tfrac{\vartheta_l^*}{|\vartheta_l^*|} \right)_{l=0}^{N_1+N_2-1} - \tfrac{1}{2}\Psi\,\Phi^T(f - \tfrac{1}{2}\Phi\Psi^T\vartheta^*) \right] = \psi^T|\Lambda| \left(\tfrac{\vartheta_l^*}{|\vartheta_l^*|} \right)_{l=0}^{N_1+N_2-1} - \Phi^T(f - \Phi u^*) \tag{2.11}$$

with $|\Lambda| = \text{diag}\,(|\Lambda_c|, |\Lambda_w|)$, and where $\vartheta^* = (\vartheta_l^*)_{l=0}^{N_1+N_2-1}$, and as before, $\frac{\vartheta_l^*}{|\vartheta_l^*|}$ denotes the set $[-1, 1]$ for $\vartheta_l^* = 0$. But from the second and third formula in (2.6) it follows by (2.7) and $\vartheta^* = \Psi u^*$ that

$$0 \in |\Lambda| \left(\tfrac{\vartheta_l^*}{|\vartheta_l^*|} \right)_{l=0}^{N_1+N_2-1} + \mu\,(\vartheta^* - \Psi u^* - b^*),$$

and hence

$$0 \in \Psi^T|\Lambda| \left(\frac{\vartheta_l^*}{|\vartheta_l^*|} \right)_{l=0}^{N_1+N_2-1} - \mu\Psi^T b^*. \tag{2.12}$$

Furthermore, the first formula in (2.6) yields by (2.10)

$$\Phi^T(\Phi u^* - f) - \mu\Psi^T(-b^*) = \Phi^T(\Phi u^* - f) + \mu\Psi^T b^* = 0,$$

i.e., $\mu\Psi^T b^* = \Phi^T(f - \Phi u^*)$, and hence the functional E is indeed minimized by $(u^*, \vartheta^*) = (u^*, \vartheta_c^*, \vartheta_w^*)$. By (2.12), a fixpoint of the spatial domain algorithm in Table 2 is also a solution of (1.8). □

2.2 The Reconstruction Model

Let us now consider the constrained problem in (1.9). We shortly derive the Alternating Split Bregman Algorithm also for this model using a similar procedure as in Subsection 2.1. Let now

$$E(\vartheta_c, \vartheta_w) := \|\Lambda_c\vartheta_c\|_1 + \|\Lambda_w\vartheta_w\|_1.$$

Instead of (2.2), we consider the minimization problem

$$\begin{aligned} &(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) \\ &= \underset{u,\vartheta_c,\vartheta_w}{\text{argmin}} \left\{ D_E^p((\vartheta_c, \vartheta_w), (\vartheta_c^k, \vartheta_w^k)) + \tfrac{1}{2}\|\Phi u - f\|_2^2 + \tfrac{\mu}{2}\left(\|\Psi_c u - \vartheta_c\|_2^2 + \|\Psi_w u - \vartheta_w\|_2^2 \right) \right\} \\ &= \underset{u,\vartheta_c,\vartheta_w}{\text{argmin}} \left\{ E(\vartheta_c, \vartheta_w) - \langle p_c^k, \vartheta_c\rangle - \langle p_w^k, \vartheta_w\rangle + \tfrac{1}{2}\|\Phi u - f\|_2^2 \right. \\ &\qquad \left. + \tfrac{\mu}{2}\left(\|\Psi_c u - \vartheta_c\|_2^2 + \|\Psi_w u - \vartheta_w\|_2^2 \right) \right\}, \end{aligned} \tag{2.13}$$

where (p_c^k, p_w^k) is a subgradient of E at $(\vartheta_c^k, \vartheta_w^k)$. Using that $(p_c^{k+1}, p_w^{k+1}) \in \partial E(\vartheta_c^{k+1}, \vartheta_w^{k+1})$, we obtain the necessary conditions

$$\begin{aligned} p_c^{k+1} &= p_c^k - \mu(\vartheta_c^{k+1} - \Psi_c u^{k+1}), \\ p_w^{k+1} &= p_w^k - \mu(\vartheta_w^{k+1} - \Psi_w u^{k+1}), \end{aligned} \tag{2.14}$$

and

$$\Phi^T(\Phi u^{k+1} - f) + \mu\left(\Psi_c^T(\Psi_c u^{k+1} - \vartheta_c^{k+1}) + \Psi_w^T(\Psi_w u^{k+1} - \vartheta_w^{k+1})\right) = 0. \tag{2.15}$$

The simplified split Bregman iteration now reads

$$\left\{\begin{array}{lcl} (u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) & = & \underset{u,\vartheta_c,\vartheta_w}{\operatorname{argmin}} \big\{ E(\vartheta_c, \vartheta_w) + \frac{1}{2}\|\Phi u - f^k\|_2^2 \\ & & \qquad + \frac{\mu}{2}\left(\|\vartheta_c - \Psi_c u - b_c^k\|_2^2 + \|\vartheta_w - \Psi_w u - b_w^k\|_2^2\right)\big\}, \\ b_c^{k+1} & = & b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1}, \\ b_w^{k+1} & = & b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}, \\ f^{k+1} & = & f^k + f - \Phi u^{k+1}. \end{array}\right. \tag{2.16}$$

Observe that there is an essential difference to formula (2.4) in the denoising model, namely the feedback step for f^k, where f^k replaces f in the first equation.

As before, assuming that $(p_c^0, p_w^0) = (0,0)$ in (2.14), and $(b_c^0, b_w^0) = (0,0)$ as well as $f^0 = f$ in (2.16), we can verify that the iteration rules in (2.13)-(2.14) and in (2.16) provide the same sequence $(u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1})_{k\geq 0}$:

With $b_c^k := \frac{1}{\mu} p_c^k$ and $b_w^k := \frac{1}{\mu} p_w^k$, the recursions for b_c^{k+1} and b_w^{k+1} in (2.16) follow from (2.14). Further, with

$$\begin{aligned} \frac{\mu}{2}\|\vartheta_c - \Psi_c u - b_c^k\|_2^2 &= \frac{\mu}{2}\|\vartheta_c - \Psi_c u\|_2^2 - \mu\langle \vartheta_c - \Psi_c u, b_c^k\rangle + \frac{\mu}{2}\|b_c^k\|_2^2, \\ \frac{\mu}{2}\|\vartheta_w - \Psi_w u - b_w^k\|_2^2 &= \frac{\mu}{2}\|\vartheta_w - \Psi_w u\|_2^2 - \mu\langle \vartheta_w - \Psi_w u, b_w^k\rangle + \frac{\mu}{2}\|b_w^k\|_2^2, \\ \frac{1}{2}\|\Phi u - f^k\|_2^2 &= \frac{1}{2}\|\Phi u - f\|_2^2 + \langle \Phi u - f, f - f^k\rangle + \frac{1}{2}\|f - f^k\|_2^2, \end{aligned}$$

the optimization problem (2.13) can be rewritten as

$$\begin{aligned} (u^{k+1}, \vartheta_c^{k+1}, \vartheta_w^{k+1}) = \underset{u,\vartheta_c,\vartheta_w}{\operatorname{argmin}} \big\{ & E(\vartheta_c, \vartheta_w) + \tfrac{1}{2}\|\Phi u - f^k\|_2^2 + \tfrac{\mu}{2}\|\vartheta_c - \Psi_c u - b_c^k\|_2^2 \\ & + \tfrac{\mu}{2}\|\vartheta_w - \Psi_w u - b_w^k\|_2^2 - \mu\langle \Psi_c u, b_c^k\rangle - \mu\langle \Psi_w u, b_w^k\rangle - \langle \Phi u, f - f^k\rangle\big\} \end{aligned}$$

and is equivalent with the first equation in (2.16) if

$$\langle \Phi u, f - f^k\rangle + \mu\langle \Psi_c u, b_c^k\rangle + \mu\langle \Psi_w u, b_w^k\rangle = \langle u, \Phi^T(f - f^k)\rangle + \langle u, \Psi_c^T p_c^k\rangle + \langle u, \Psi_w^T p_w^k\rangle = 0.$$

Using (2.14), (2.15) and the iteration $f^k = f^{k-1} + f - \Phi u^k$, it follows indeed that

$$\begin{aligned} & \Phi^T(f - f^k) + \Psi_c^T p_c^k + \Psi_w^T p_w^k \\ = \; & \Phi^T(f - f^{k-1} - f + \Phi u^k) + \Psi_c^T(p_c^{k-1} - \mu\vartheta_c^k + \mu\Psi_c u^k) + \Psi_w^T(p_w^{k-1} - \mu\vartheta_w^k + \mu\Psi_w u^k) \\ = \; & \Phi^T(\Phi u^k - f) + \mu\Psi_c^T(\Psi u^k - \vartheta_c^k) + \mu\Psi_w^T(\Psi_w u^k - \vartheta_w^k) \\ & + \Phi^T(f - f^{k-1}) + \Psi_c^T p_c^{k-1} + \Psi_w^T p_w^{k-1} \\ = \; & \Phi^T(f - f^{k-1}) + \Psi_c^T p_c^{k-1} + \Psi_w^T p_w^{k-1} = \ldots = \Phi^T(f - f^0) + \Psi_c^T p_c^0 + \Psi_w^T p_w^0 = 0. \end{aligned}$$

Iterative minimization now yields the following scheme

Alternating Split Bregman Algorithm for the Reconstruction Model.

$$\begin{cases} u^{k+1} = \underset{u}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\Phi u - f^k\|_2^2 + \frac{\mu}{2}\left(\|\vartheta_c^k - \Psi_c u - b_c^k\|_2^2 + \|\vartheta_w^k - \Psi_w u - b_w^k\|_2^2\right)\right\}, \\ \vartheta_c^{k+1} = \underset{\vartheta_c}{\operatorname{argmin}} \left\{ \|\Lambda_c \vartheta_c\|_1 + \frac{\mu}{2}\|\vartheta_c - \Psi_c u^{k+1} - b_c^k\|_2^2\right\}, \\ \vartheta_w^{k+1} = \underset{\vartheta_w}{\operatorname{argmin}} \left\{ \|\Lambda_w \vartheta_w\|_1 + \frac{\mu}{2}\|\vartheta_w - \Psi_w u^{k+1} - b_w^k\|_2^2\right\}, \\ b_c^{k+1} = b_c^k + \Psi_c u^{k+1} - \vartheta_c^{k+1}, \\ b_w^{k+1} = b_w^k + \Psi_w u^{k+1} - \vartheta_w^{k+1}, \\ f^{k+1} = f^k + f - \Phi u^{k+1}. \end{cases} \tag{2.17}$$

Further, analogously as in Theorem 2.2, it can be shown that for a fixed point $(u^*, \vartheta_c^*, \vartheta_w^*, b_c^*, b_w^*, f^*)$ of the Alternating Split Bregman iteration (2.17) the vector $(\vartheta_c^*, \vartheta_w^*)$ solves the constrained optimization problem (1.9).

We now summarize the numerical algorithm in Table 3, using the same abbreviations as in Subsection 2.1. Optionally, as suggested in [29], one can also internally apply two or three iterations from step 1 to step 3 of the algorithm, i.e., one may execute step 4 not in each iteration step. As for the denoising model, we can derive the analysis-based method (spatial domain), similarly as in Table 2, but using f^k in step 1 and adding the "residual feedback" step $f^{k+1} = f^k + f - \Phi u^{k+1}$.

Input: $(u^0, \vartheta_w^0, \vartheta_c^0, b_c^0, b_w^0) = (0,0,0,0,0)$ and $f^0 = f$.
While a stop criterion is not satisfied
1) $u^{k+1} = \frac{1}{2\mu+1}\Phi^T f^k + \frac{1}{2}(I_N - \frac{1}{2\mu+1}\Phi^T\Phi)((u_c^k - \tilde{b}_c^k) + (u_w^k - \tilde{b}_w^k))$
2) $u_c^{k+1} = \mathcal{S}_{\mu^{-1}\lvert\Lambda_c\rvert}(u^{k+1} + \tilde{b}_c^k)$
$u_w^{k+1} = \mathcal{S}_{\mu^{-1}\lvert\Lambda_w\rvert}(u^{k+1} + \tilde{b}_w^k)$
3) $\tilde{b}_c^{k+1} = \tilde{b}_c^k + u^{k+1} - u_c^{k+1}$
$\tilde{b}_w^{k+1} = \tilde{b}_w^k + u^{k+1} - u_w^{k+1}$
4) $f^{k+1} = f^k + f - \Phi u^{k+1}$
end

Table 3: Alternating Split Bregman algorithm for two frames for solving (1.9).

3 Convergence of the Alternating Split Bregman method

In this section we study the convergence of the new algorithms for two frames proposed in Section 2. We focus here on the denoising model in (2.6). The reconstruction model can be treated similarly.

Our considerations may be even interesting for the simple case with only one frame, i.e., if one considers

$$\underset{u,\vartheta_c}{\operatorname{argmin}}\{\|\Lambda_c\vartheta_c\|_1 + \frac{1}{2}\|f - \Phi u\|_2^2\} \quad \text{s.t.} \quad \vartheta_c = \Psi_c u$$

instead of (1.8). Observe that Goldstein and Osher did not consider the convergence of their algorithm in [29]. In [47], Setzer showed that the Alternating Split Bregman

algorithm for a similar optimization problem can be understood as a special case of the so-called Douglas-Rachford splitting algorithm, whose convergence has been provided (under suitable assumptions) in [33] and in [14]. In Subsection 3.1, we shortly summarize the convergence result for the Douglas-Rachford splitting method. Afterwards, we follow the idea of [47] and apply this result to our alternating split Bregman algorithm by showing that the two algorithms coincide for suitably chosen operators.

3.1 The Douglas-Rachford splitting

Let H be a real Hilbert space, and let $A : H \to 2^H$ and $B : H \to 2^H$ be two set-valued mappings. We assume that A and B are maximal monotone, i.e., their resolvents $J_A := (Id + A)^{-1}$ and $J_B := (Id + B)^{-1}$ exist and are firmly nonexpansive. Here, an operator $J_A : H \to H$ is called *firmly nonexpansive*, if for all $x \in H$ and $y \in H$ the condition

$$\|J_A x - J_A y\| \leq \|x - y\|$$

is satisfied. In case of linear or affine operators, this relation is satisfied if the spectral radius of J_A is bounded by one, i.e., $\rho(J_A) := \lim_{n\to\infty} \|J_A^n\|^{1/n} \leq 1$.

We are interested in solving the inclusion

$$0 \in A(\hat{r}) + B(\hat{r}). \tag{3.1}$$

Let $\hat{r}$ be a solution of (3.1), i.e., assume that there are vectors $\sigma_1 \in A(\hat{r})$ and $\sigma_2 \in B(\hat{r})$ such that $\sigma_1 + \sigma_2 = 0$. The inclusion in (3.1) can with $\mu > 0$ be rewritten as

$$\hat{r} - \mu\sigma_2 = \hat{r} + \mu\sigma_1 \in (Id + \mu A)(\hat{r}),$$

i.e., $\hat{r} = (Id + \mu A)^{-1}(\hat{r} - \mu\sigma_2) = J_{\mu A}(\hat{r} - \mu\sigma_2)$. Addition of $\mu\sigma_2$ gives

$$\hat{r} + \mu\sigma_2 = J_{\mu A}(\hat{r} - \mu\sigma_2) + \mu\sigma_2 \in (Id + \mu B)(\hat{r}),$$

and finally

$$\hat{r} = J_{\mu B}\left(J_{\mu A}(\hat{r} - \mu\sigma_2) + \mu\sigma_2\right). \tag{3.2}$$

This fixed point equation can be solved iteratively using the Douglas-Rachford split algorithm

$$\begin{cases} s^{k+1} &= J_{\mu A}(2r^k - s^k) + s^k - r^k, \\ r^{k+1} &= J_{\mu B} s^{k+1}. \end{cases} \tag{3.3}$$

Indeed, if (s^*, r^*) is a fixed point of this iteration, then we obtain from $r^* = J_{\mu B} s^*$ that $\sigma_2 = \frac{1}{\mu}(s^* - r^*) \in B(r^*)$, and hence

$$\begin{aligned} r^* &= J_{\mu B}(J_{\mu A}(2r^* - s^*) + s^* - r^*) \\ &= J_{\mu B}(J_{\mu A}(r^* - \mu\sigma_2) + \mu\sigma_2), \end{aligned}$$

i.e., r^* satisfies the equation (3.2). We wish to apply the following result that has been shown in [16], Corollary 5.2.

Theorem 3.1 *Let $A, B : H \to 2^H$ two maximal monotone operators on a Hilbert space H, and let $\mu \in (0, \infty)$. Assume that a solution of (3.1) exists. Then for any initial elements s^0 and r^0, the Douglas-Rachford split algorithm in (3.3) converges weakly to $(\hat{s}, \hat{r})$, and we have $0 \in A(\hat{r}) + B(\hat{r})$.*

Note that for a finite-dimensional space H, it follows convergence of $(r^k)_{k\geq 0}$ to $\hat{r}$. Recent applications of the Douglas-Rachford splitting method for image denoising can be found in [15, 51].

3.2 Convergence of the Alternating Split Bregman algorithm

Using the above results we can show the following theorem.

Theorem 3.2 *The Alternating Split Bregman algorithm in (2.6) for the denoising model converges with arbitrary starting point $(u^0, \vartheta_c^0, \vartheta_w^0, b_c^0, b_w^0)$ to a fixed point $(u^*, \vartheta_c^*, \vartheta_w^*, b_c^*, b_w^*)$ of this iteration, and the obtained vector $(u^*, \vartheta_c^*, \vartheta_w^*)$ solves the constrained optimization problem (1.8).*

Proof. We show that the iteration scheme in (2.6) coincides with the Douglas-Rachford split algorithm in (3.3) with suitably chosen operators A and B. The convergence of the scheme then follows from Theorem 3.1, and the remaining assertion is a consequence of Theorem 2.2.

1. We start with the first equation in (2.6),

$$u^{k+1} = \operatorname*{argmin}_u \left\{ \frac{1}{2}\|f - \Phi u\|_2^2 + \frac{\mu}{2}\left(\|\vartheta_c^k - \Psi_c u - b_c^k\|_2^2 + \|\vartheta_w^k - \Psi_w u - b_w^k\|_2^2 \right) \right\},$$

which is equivalent to

$$0 \in \Phi^T(\Phi u^{k+1} - f) + \mu \left[-\Psi_c^T(\vartheta_c^k - \Psi_c u^{k+1} - b_c^k) - \Psi_w^T(\vartheta_w^k - \Psi_w u^{k+1} - b_w^k) \right].$$

The functional is strictly convex, and we find

$$\begin{aligned} u^{k+1} &= (\Phi^T\Phi)^\dagger \left[\Phi^T f - \mu(\Psi_c^T\Psi_c + \Psi_w^T\Psi_w)u^{k+1} + \mu\Psi_c^T(\vartheta_c^k - b_c^k) \right. \\ &\quad \left. + \mu\Psi_w^T(\vartheta_w^k - b_w^k) \right], \end{aligned}$$

where $(\Phi^T\Phi)^\dagger$ denotes the generalized inverse of $\Phi^T\Phi$. Let now

$$\Psi := \begin{pmatrix} \Psi_c \\ \Psi_w \end{pmatrix} \in \mathbb{R}^{(N_1+N_2)\times N}, \quad \vartheta^k := \begin{pmatrix} \vartheta_c^k \\ \vartheta_w^k \end{pmatrix} \in \mathbb{R}^{N_1+N_2}, \quad b^k := \begin{pmatrix} b_c^k \\ b_w^k \end{pmatrix} \in \mathbb{R}^{N_1+N_2}.$$

Then we obtain by multiplication with $-\mu\Psi$

$$-\mu\Psi u^{k+1} = -\mu\Psi(\Phi^T\Phi)^\dagger \left[\Phi^T f - \mu\Psi^T\Psi u^{k+1} + \mu\Psi^T(\vartheta^k - b^k) \right],$$

and choosing $\tilde{f}$ with $\Phi^T f = \Psi^T\tilde{f}$, it follows that

$$-\mu\Psi u^{k+1} = \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T \left[-\tilde{f} + \mu\Psi u^{k+1} - \mu(\vartheta^k - b^k) \right].$$

Addition of $-\tilde{f} + \mu\Psi u^{k+1} - \mu(\vartheta^k - b^k)$ yields

$$-\tilde{f} - \mu(\vartheta^k - b^k) = (I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T) \left[-\tilde{f} + \mu\Psi u^{k+1} - \mu(\vartheta^k - b^k) \right], \tag{3.4}$$

where I is the identity matrix of size N_1+N_2. Let now the operator $A : \mathbb{R}^{N_1+N_2} \to \mathbb{R}^{N_1+N_2}$ be given by

$$A(y) := \Psi(\Phi^T\Phi)^\dagger\Psi^T(y - \tilde{f}) = \Psi(\Phi^T\Phi)^\dagger(\Psi^T y - \Phi^T f).$$

Then from $z = (I + \mu A)(y) = (I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)(y - \tilde{f}) + \tilde{f}$, it follows for the resolvent

$$J_{\mu A} z = (I + \mu A)^{-1} z = (I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}(z - \tilde{f}) + \tilde{f}.$$

For $\mu > 0$, the matrix $(I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}$ is obviously positive definite, and by

$$\|J_{\mu A}x - J_{\mu A}y\|_2 \le \|(I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}\|_2\,\|x - y\|_2 \qquad x, y \in \mathbb{R}^{N_1+N_2},$$

the resolvent $J_{\mu A}$ is firmly nonexpansive for all $\mu > 0$ since the spectral norm of $(I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}$ is bounded by 1. We obtain from (3.4)

$$(I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}(-\tilde{f} - \mu(\vartheta^k - b^k)) = -\tilde{f} + \mu\Psi u^{k+1} - \mu(\vartheta^k - b^k),$$

i.e.,

$$\begin{aligned}\mu(\Psi u^{k+1} + b^k) &= (I + \mu\Psi(\Phi^T\Phi)^\dagger\Psi^T)^{-1}(-\tilde{f} - \mu(\vartheta^k - b^k)) + \tilde{f} + \mu\vartheta^k \\ &= J_{\mu A}(-\mu(\vartheta^k - b^k)) + \mu\vartheta^k. \end{aligned} \tag{3.5}$$

Let now $r^k := \mu\, b^k$ and $s^k := \mu(\vartheta^k + b^k)$. Then (3.5) is of the form

$$\mu(\Psi u^{k+1} + b^k) = J_{\mu A}(2r^k - s^k) + (s^k - r^k).$$

This is the first equation in the Douglas-Rachford iteration, since by definition of b^{k+1} in the last two equations of (2.6), we have $b^{k+1} = b^k + \Psi u^{k+1} - \vartheta^{k+1}$ and hence

$$s^{k+1} = \mu(b^{k+1} + \vartheta^{k+1}) = \mu(\Psi u^{k+1} + b^k) = J_{\mu A}(2r^k - s^k) + (s^k - r^k).$$

2. It remains to show that $r^{k+1} = \mu b^{k+1}$ satisfies the second relation in (3.3) with a suitably chosen operator B. For that purpose we consider the second and third equation in (2.6),

$$\begin{aligned}\vartheta_c^{k+1} &= \operatorname*{argmin}_{\vartheta_c}\left\{\|\Lambda_c\vartheta_c\|_1 + \frac{\mu}{2}\|\vartheta_c - \Psi_c u^{k+1} - b_c^k\|_2^2\right\}, \\ \vartheta_w^{k+1} &= \operatorname*{argmin}_{\vartheta_w}\left\{\|\Lambda_w\vartheta_w\|_1 + \frac{\mu}{2}\|\vartheta_w - \Psi_w u^{k+1} - b_w^k\|_2^2\right\},\end{aligned}$$

yielding

$$\begin{aligned}0 &\in |\Lambda_c|(\operatorname{sign}\vartheta_c^{k+1}) + \mu(\vartheta_c^{k+1} - \Psi_c u^{k+1} - b_c^k), \\ 0 &\in |\Lambda_w|(\operatorname{sign}\vartheta_w^{k+1}) + \mu(\vartheta_w^{k+1} - \Psi_w u^{k+1} - b_w^k),\end{aligned}$$

where $|\Lambda_c| := \operatorname{diag}\left(|\lambda_{c,l}|\right)_{l=0}^{N_1-1}$, $|\Lambda_w| := \operatorname{diag}\left(|\lambda_{w,l}|\right)_{l=0}^{N_1-1}$ have only nonnegative entries at the diagonal, and where sign is a set-valued operator,

$$\operatorname{sign}\vartheta_c^{k+1} := (\operatorname{sign}\vartheta_{c,l}^{k+1})_{l=0}^{N_1-1} \quad \text{with} \quad \operatorname{sign}\vartheta_{c,l}^{k+1} = \begin{cases} 1 & \vartheta_{c,l}^{k+1} > 0, \\ -1 & \vartheta_{c,l}^{k+1} > 0, \\ [-1,1] & \vartheta_{c,l}^{k+1} = 0. \end{cases}$$

Analogously the operator sign ϑ_w^{k+1} is defined. Then, with

$$|\Lambda| := \begin{pmatrix} |\Lambda_c| & 0 \\ 0 & |\Lambda_w| \end{pmatrix}$$

we have

$$0 \in (|\Lambda|\operatorname{sign})(\vartheta^{k+1}) + \mu(\vartheta^{k+1} - \Psi u^{k+1} - b^k).$$

We now define $B := (|\Lambda| \operatorname{sign})^{-1}$, i.e., B is a set-valued mapping with

$$B(y) = B\left((y_l)_{l=0}^{N_1+N_2-1}\right) = (z_l)_{l=0}^{N_1+N_2-1} \quad \text{with} \quad z_l := \begin{cases} 0 & |y_l| < |\lambda_l|, \\ [0,\infty) & y_l = |\lambda_l|, \\ (-\infty,0] & y_l = -|\lambda_l|, \end{cases}$$

where $|\lambda_l|$ is the l-th diagonal entry in $|\Lambda|$. Observe that $B(y)$ is only defined for $|y_l| \leq |\lambda_l|$, $l = 0, \ldots, N_1 + N_2 - 1$. It follows that

$$\vartheta^{k+1} \in B\left(\mu(-\vartheta^{k+1} + \Psi u^{k+1} + b^k)\right) = B(\mu b^{k+1}),$$

and by (2.9), this inclusion is well-defined. It is equivalent with

$$\begin{aligned} \mu\vartheta^{k+1} \quad &\in \quad \mu B\left(\mu(-\vartheta^{k+1} + \Psi u^{k+1} + b^k)\right) \\ &= (I + \mu B)\left(\mu(-\vartheta^{k+1} + \Psi u^{k+1} + b^k)\right) + \mu(\vartheta^{k+1} - \Psi u^{k+1} - b^k), \end{aligned}$$

and addition of $\mu(-\vartheta^{k+1} + \Psi u^{k+1} + b^k) = \mu b^{k+1}$ yields

$$\mu(\Psi u^{k+1} + b^k) \in (I + \mu B)\,(\mu b^{k+1}).$$

Now, the definition of B implies for its resolvent

$$J_{\mu B}(z) = (I + \mu B)^{-1}\left((z_l)_{l=0}^{N_1+N_2-1}\right) = (y_l)_{l=0}^{N_1+N_2-1} \quad \text{with} \quad y_l = \begin{cases} |\lambda_l| & z_l > |\lambda_l|, \\ -|\lambda_l| & z_l < -|\lambda_l|, \\ z_l & |z_l| < |\lambda_l|, \end{cases}$$

and hence $J_{\mu B} : \mathbb{R}^{N_1+N_2} \to \mathbb{R}^{N_1+N_2}$ is a single-valued, firmly nonexpansive operator. Thus,

$$\mu b^{k+1} = (I + \mu B)^{-1}\,(\mu(\Psi u^{k+1} + b^k)),$$

and by $s^{k+1} = \mu(\Psi u^{k+1} + b^k)$ and $r^{k+1} = \mu b^{k+1}$, it follows the second equation,

$$r^{k+1} = J_{\mu B}(s^{k+1}),$$

of the Douglas-Rachford iteration (3.3). Hence, the Alternating Split Bregman algorithm in (2.6) coincides with the Douglas-Rachford algorithm, and its convergence is ensured by Theorem 3.1, since the operators A and B, defined in the proof, are maximal monotone. □

Remark. The convergence of the Alternating Split Bregman algorithm can be shown in a similar manner if the parameter $\mu > 0$ is not fixed but changes in the different iteration steps.

4 Numerical experiments

In this section, we consider pseudo-random Fourier-domain CS imaging using 30% measurements. The pseudo-random downsampling is a polynomial variable density random sampling, which obeys a probability density function with dense sampling in low-frequency components and spare sampling in high-frequency components. One can find more details about the pseudo-random Fourier downsampling in [35]. Figure 2 (a) displays the sampling pattern with 30% measurements that will be used for the CS measurement matrix

in our numerical experiments. Here, the white points denote the sampling points. In Figure 2 (b), we show the 2D probability density function (upper subfigure) and its 1D center crossed profile for the sampling pattern. In CS imaging, we obtain the vector f with $K = 3N/10$ measurements at the same time by using an encoded random mask for Fourier coefficients of objects u.

In the first test, we show the performance of the proposed curvelet-wavelet regularized split Bregman (CWSpB) method (reconstruction model) for a standard Lena image of size 256×256, in comparison with iterative wavelet thresholding (IWT) and iterative curvelet thresholding (ICT) methods. The computational parameters are provided as follows. We take Daubechies' DB6 wavelet transform with a 4-level decomposition as one basis in our model, and the discrete curvelet transform as the second frame. The DB6 wavelet transform is also taken for the IWT. The soft shrinkage function is used for all iterative thresholding methods. For the proposed reconstruction model, we choose the increasing regularization parameters $\mu = \mu_0(1 + k/N_{numb})^4$ with an initial value $\mu_0 = 0.2$, where $N_{numb} = 20$ is the number of total iterations, and k denotes the iteration index. In particular, we use decreasing thresholds $\sigma = 0.02\frac{1}{\mu}$ for both iterative wavelet thresholding and curvelet thresholding. Further, supposing that we have no apriori information about the image to recover, we simply use the matrices $\Lambda_c = 0.02\, I_{N_1}$ and $\Lambda_w = 0.02\, I_N$ as weighted identity matrices.

Figure 3 (a) shows the original unknown image. Figure 3 (b) shows the decoded result by zero-filling recovery $u = \Phi^T f$. Observe that there exist serious noise-like artifacts. In the next rows, we display the decoding results (left column) and their recovery errors (right column) by using IWT, ICT and the proposed CWSpB method, respectively. We notice that our new method can recover especially texture-like feature more effectively.

In Figure 5 (a), we display the changes of the signal-to-noise ratio (SNR) values when the number of iterations increases. It can be seen that our proposed method converges much faster than the IWT and ICT methods. The CWSpB method combines the advantages of curvelets and wavelets to some extent, so that it achieves a higher SNR value. Figure 5 (b) shows the so-called pseudo-Pareto curves [53], i.e., the residual $\|f - \Phi u^k\|_2^2$ (vertical coordinate) versus the norm $\|u^k\|_1$ (horizontal coordinate).

In Figure 4, we apply the proposed method to a real scene of the moon surface. Figure 4 (a) is the original scene. Figure 4 (b) is the direct $u = \Phi^T f$ reconstruction. Figures 4 (c) - (e) are obtained by the IWT, the wavelet-total variation (TV) regularization [35], and the ICT, respectively. We notice that the TV regularization makes the results more smooth. Figure 4 (f) is the result obtained by the new proposed CWSpB method. In this example, we have tested the decay threshold values $\sigma = \sigma_0(1 - k/N_{numb})$ with an initial value $\sigma_0 = 0.06$ for all iterative methods. In Figure 6 (a) we show the changes of SNR values obtained respectively by IWT, ICT, and CWSpB methods, if the number of iterations increases. Figure 6 (b) shows their pseudo-Pareto curves. Again, we find that the new CWSpB algorithm, based on Bregman split technique obtains a good recovery result already after $5-7$ iterations and converges much faster than the iterative threshold methods.

Finally, Figure 7 shows the performance of the proposed CWSpB method for a real cloud scene by remote sensing. The proposed method achieves higher SNR values indeed.

5 Conclusions

In this paper, we have proposed two new models for CS remote sensing using the alternating split Bregman algorithm. The new iteration schemes have been shown to solve suitable optimization problems, where two different frames are allowed for regularization. These frames can be chosen according to possible a-priori knowledge on the data that have to be recovered. We have proved the convergence of the derived algorithms by showing equivalence with the Douglas-Rachford algorithm. In particular, we presented this connection by a concrete description of the operators A and B that have to be taken in the Douglas-Rachford iteration.

In our numerical experiments we have used an orthogonal wavelet basis and the curvelet frame, since these function families can present point-like and line-like features of images differently well. It comes out that the used split Bregman technique converges very fast. Without assuming any a-priori knowledge about the images, the weight matrices Λ_c and Λ_w for scaling of wavelet and curvelet coefficients have been chosen as (scaled) identity matrices. We have found that our method is quite sensitive to the regularization parameter μ. One may allow also different parameters μ_c and μ_w (instead of μ) for the terms $\Psi_c^T(\vartheta_c^k - b_c^k)$ and $\Psi_w^T(\vartheta_w^k - b_w^k)$ in the first step of the algorithms. An optimal choice of these parameters as well as of the matrices Λ_c and Λ_w will be subject of further research.

Acknowledgment

The research in this paper is funded by the National Natural Science Foundation of China under Grand No. 40704019 and 40674061, Tsinghua Basic Research Fund under Grand No. JC 2007030, and the projects PL 170/11-2 and PL 170/13-1 of the Deutsche Forschungsgemeinschaft (DFG). This is gratefully acknowledged. Furthermore, J. Ma would like to thank the Invited Professorship from the Ecole des Mines de Paris.

A Appendix: Curvelets and Iterative Curvelet Thresholding

The curvelets [11, 12] constitute a directional frame of $L^2(\mathbb{R}^2)$, which has anisotropic needle-shaped elements and allows an optimal sparse representation of objects with discontinuities along smooth curves. Unlike wavelets, curvelets are indexed by three parameters: a scale 2^{-j}, $j \in \mathbb{N}_0$; an equispaced sequence of rotation angles $\theta_{j,l} = 2\pi l \cdot 2^{-\lfloor j/2 \rfloor}$, $0 \le l \le 2^{\lfloor j/2 \rfloor} - 1$; and a position $x_k^{(j,l)} = R_{\theta_{j,l}}^{-1}(k_1\, 2^{-j}, k_2\, 2^{-\lfloor j/2 \rfloor})^T$, $(k_1, k_2) \in \mathbb{Z}^2$, where $R_{\theta_{j,l}}$ denotes the rotation matrix with angle $\theta_{j,l}$. The curvelet elements are defined by

$$\Psi_{j,l,k}(x) := \Psi_j(R_{\theta_{j,l}}(x - x_k^{(j,l)})), \qquad x = (x_1, x_2) \in \mathbb{R}^2, \tag{1.6}$$

where Ψ_j are smooth functions with compact support on wedges in Fourier domain. For a smooth object f with discontinuities along C^2-continuous curves, the best m-term approximation $\tilde{f}_m$ by curvelet thresholding obeys $\|f - \tilde{f}_m\|_2^2 \le C m^{-2} (\log m)^3$, while for wavelets the decay rate is only m^{-1} and for the Fourier transform the rate is $m^{-1/2}$. Unlike the isotropic elements of wavelet bases, the needle-shaped elements of this transform possess very high directional sensitivity and anisotropy (see Figure 1). Such elements are very efficient in representing line-like edges.

The family of curvelet functions forms a tight frame of $L_2(\mathbb{R}^2)$. That means, each function $f \in L_2(\mathbb{R}^2)$ has a representation

$$f = \sum_{j,l,k} \langle f, \Psi_{j,l,k} \rangle \, \Psi_{j,l,k}$$

where $\langle f, \Psi_{j,l,k} \rangle$ denotes the L_2-scalar product of f and $\Psi_{j,l,k}$. The coefficients $\langle f, \Psi_{j,l,k} \rangle$ are called curvelet coefficients of function f. The algorithm of the second-generation curvelet transform by Candes, Donoho, Demanet and Ying [12], used in this paper, can be found in www.curvelet.org. For more details on curvelets and recent applications we refer to [42, 43].

The simple iterative curvelet thresholding (see e.g., [34, 40, 41]) can be written as follows

$$u^{k+1} = \mathcal{S}_{\sigma,\mathcal{T}}(u^k + \Phi^T(f - \Phi u^k)). \tag{1.7}$$

Here the shrinkage operator $S_{\sigma,\mathcal{T}}$ is given by

$$\mathcal{S}_{\sigma,\mathcal{T}}(f) = \sum_{j,l,k} \mathcal{T}_\sigma(\langle f, \Psi_{j,l,k} \rangle) \, \Psi_{j,l,k}, \tag{1.8}$$

where $\mathcal{T}_\sigma$ can e.g. be taken as a soft shrinkage function defined by a fixed threshold $\sigma > 0$,

$$\mathcal{T}_{s,\sigma}(x) = \begin{cases} x - \sigma, & x \geq \sigma, \\ 0, & |x| < \sigma, \\ x + \sigma & x \leq -\sigma, \end{cases}$$

or a hard shrinkage function

$$\mathcal{T}_{h,\sigma}(x) = \begin{cases} x, & |x| \geq \sigma, \\ 0, & |x| < \sigma. \end{cases}$$

The iterative curvelet/wavelet thresholding has been used in the numerical experiments in Section 4 for comparisons.

References

[1] R. Baraniuk, V. Cevher, M. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Information Theory, submitted, 2008.

[2] J. Bioucas-Dias, M. Figueiredo, A new TwIST: two step iterative shrinkage/thresholding algorithms for image restoration, IEEE Trans. Image Process. **16** (12), 2992-3004 (2007).

[3] T. Blumensath, M. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. **14** (5), 629-654 (2008).

[4] J. Bobin, J. Starck, R. Ottensamer, Compressed Sensing in Astronomy, IEEE J. Selected Topics in Signal Process. **2** (5), 718-726 (2008).

[5] J. Cai, S. Osher, Z. Shen, Linear Bregman iteration for compressed sensing, Math. Comput. **78**, 1515-1536 (2009).

[6] J. Cai, S. Osher, Z. Shen, Split Bregman methods and frame based image restoration, submitted, 2009.

[7] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Infor. Theory **51** (12), 4203-4215 (2005).

[8] E. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. **59**, 1207-1223 (2006).

[9] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory **52** (2), 489-590 (2006).

[10] E. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, J. Fourier Anal. Appl. **14** (5), 877-905 (2008).

[11] E. Candès, D. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise singularities, Commun. Pure Appl. Math. **57**, 219-266 (2004).

[12] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. **5** (3), 861-889 (2006).

[13] R. Chartrand, Exact reconstructions of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. **14**, 707-710 (2007).

[14] P. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, Optimization **53** (5-6), 475-504 (2004).

[15] P. Combettes, J. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, IEEE J. Selected Topics in Signal Processing **1** (4), 564-574 (2007).

[16] P. Combettes, V. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Model. Simul. **4**, 1168-1200 (2005).

[17] I. Daubechies, Ten lectures on wavelets, Philadelphia, PA: SIAM, 1992.

[18] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. **57** (11), 1413-1457 (2004).

[19] L. Demanet, L. Ying, Wave atoms and sparsity of oscillatory patterns, Appl. Comput. Harmon. Anal. **23** (3), 368-387 (2007).

[20] M. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. **14** (12), 2091-2106 (2005).

[21] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory **52** (4), 1289-1306 (2006).

[22] D. Donoho, Y. Tsaig, I. Drori, J. Starck, Sparse solution of underdetermined linear equations by stagewise Orthogonal Matching Pursuit, submitted, 2008.

[23] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, Single-pixel imaging via compressive sampling, IEEE Signal Process. Mag. **25** (2), 83-91 (2008).

[24] M. Elad, B. Matalon, J. Shtok, M. Zibulevsky, A wide-angle view at iterated shrinkage algorithms, Proc. of SPIE, vol. 6701, pp. 670102: 1-19, 2007.

[25] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, Inverse Problems **23**, 947-968 (2007).

[26] Y. Eldar, Compressed sensing of analog signals in shift-invariant spaces, IEEE Trans. Signal Process., 2009, to appear.

[27] M. Figueiredo, R. Nowak, S. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J. Select Topic in Signal Process. **1** (4), 586-597 (2007).

[28] M. Fornasier, H. Rauhut, Iterative thresholding algorithms, Appl. Comput. Harmon. Anal. **25** (2), 187-208 (2008).

[29] T. Goldstein, S. Osher, The split Bregman method for ℓ_1 regularized problems, SIAM J. Imaging Sci. **2** (2), 323-343 (2009).

[30] E. Hale, W. Yin, Y. Zhang, Fixed-point continuation for ℓ_1-minimization: methodology and convergence, SIAM J. Optimization **19** (3), 1107-1130 (2008).

[31] L. He, T. Chang, S. Osher, T. Fang, P. Speier, MR image reconstruction by using the iterative refinement method and nonlinear inverse scale space methods, UCLA CAM Report 06-35, 2006.

[32] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. **56** (6), 2346-2356 (2008).

[33] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. **16** (6), 964-979 (1979).

[34] T. Lin, F. J. Herrmann, Compressed wavefield extrapolation, Geophysics **72** (5), SM77-SM93 (2007).

[35] M. Lustig, D. Donoho, J. Pauly, Sparse MRI: the application of compressed sensing for rapid MR imaging, Magnetic Resonance in Medicine **58** (6), 1182-1195 (2007).

[36] M. Lustig, D. Donoho, J. Santos, J. Pauly, Compressed sensing MRI, IEEE Signal Processing Magazine **25** (2), 72-82 (2008).

[37] J. Ma, Compressed sensing by inverse scale space and curvelet thresholding, Appl. Math. Comput. **206**, 980-988 (2008).

[38] J. Ma, Compressed sensing for surface characterization and metrology, IEEE Trans. Instrument and Measurement, 2009, to appear.

[39] J. Ma, Single-pixel remote sensing, IEEE Geosci. Remote Sensing Lett. **6** (2), 199-203 (2009).

[40] J. Ma, F.-X Le Dimet, Deblurring from highly incomplete measurements for remote sensing, IEEE Trans. Geosci. Remote Sensing **47** (3), 792-802 (2009).

[41] J. Ma, A single-pixel imaging system for remote sensing by two-step iterative curvelet thresholding, IEEE Geosci. Remote Sensing Lett., 2009, to appear.

[42] J. Ma, G. Plonka, Combined curvelet shrinkage and nonlinear anisotropic diffusion, IEEE Trans. Image Process. **16** (9), 2198-2206 (2007).

[43] J. Ma, G. Plonka, A review of curvelets and recent applications, IEEE Signal Processing Magazine, 2010, to appear.

[44] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model. Simul. **4**, 460-489 (2005).

[45] G. Peyré, Best basis compressed sensing, Proceedings of SSVM07, pp. 80-91, June 2007.

[46] R. Saab, O. Yilmaz, Sparse recovery by non-convex optimization– instance optimality, submitted, 2008.

[47] S. Setzer, Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage, in A. Lie, M. Lysaker, K. Mørken, and X. Tai, eds, Scale Space and Variational Methods, LNCS, Voss, Norway, 2009.

[48] M. Sheikh, O. Milenkovic, R. Baraniuk, Designing compressive sensing DNA microarrays, IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), St. Thomas, U.S. Virgin Islands, December 2007.

[49] J. Starck, M. Nguyen, F. Murtagh, Wavelets and curvelets for image deconvolution: a combined approach, Signal Processing, **83** (10), 2279-2283 (2003).

[50] J. Starck, M. Elad, D. Donoho, Image decomposition via the combination of sparse representations and a variational approach, IEEE Trans. Image Process. **14**, 1570-1582 (2005).

[51] G. Steidl, T. Teuber, Removing multiplicative noise by Douglas-Rachford splitting methods, preprint, 2009.

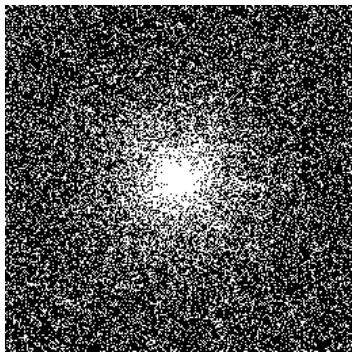
[52] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory **53** (12), 4655-4666 (2008).

[53] E. van den Berg, M. Friedlander, Probing the pareto frontier for basis pursuit solutions, SIAM J. Scientific Computing **31** (2), 890-912 (2008).

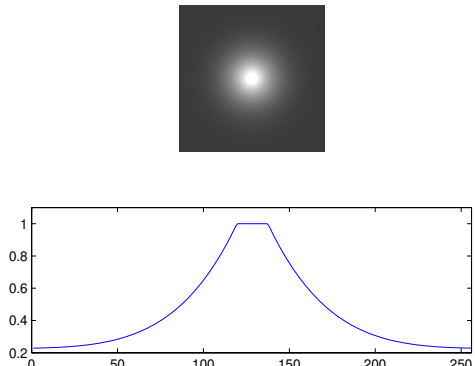
[54] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intel. **31** (2), 210-227 (2009).

[55] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1 minimization with applications to compressed sensing, SIAM J. Imaging Sci. **1** (1), 143-168 (2008).

Figure 1: Element functions of curvelets at different spatial locations, directions, and scales.



(a)



(b)

Figure 2: Pseudo-random Fourier-domain downsampling is used as a CS measurement matrix. (a) Sampling pattern in Fourier domain. (b) 2D probability density function and its 1D center crossed profile.

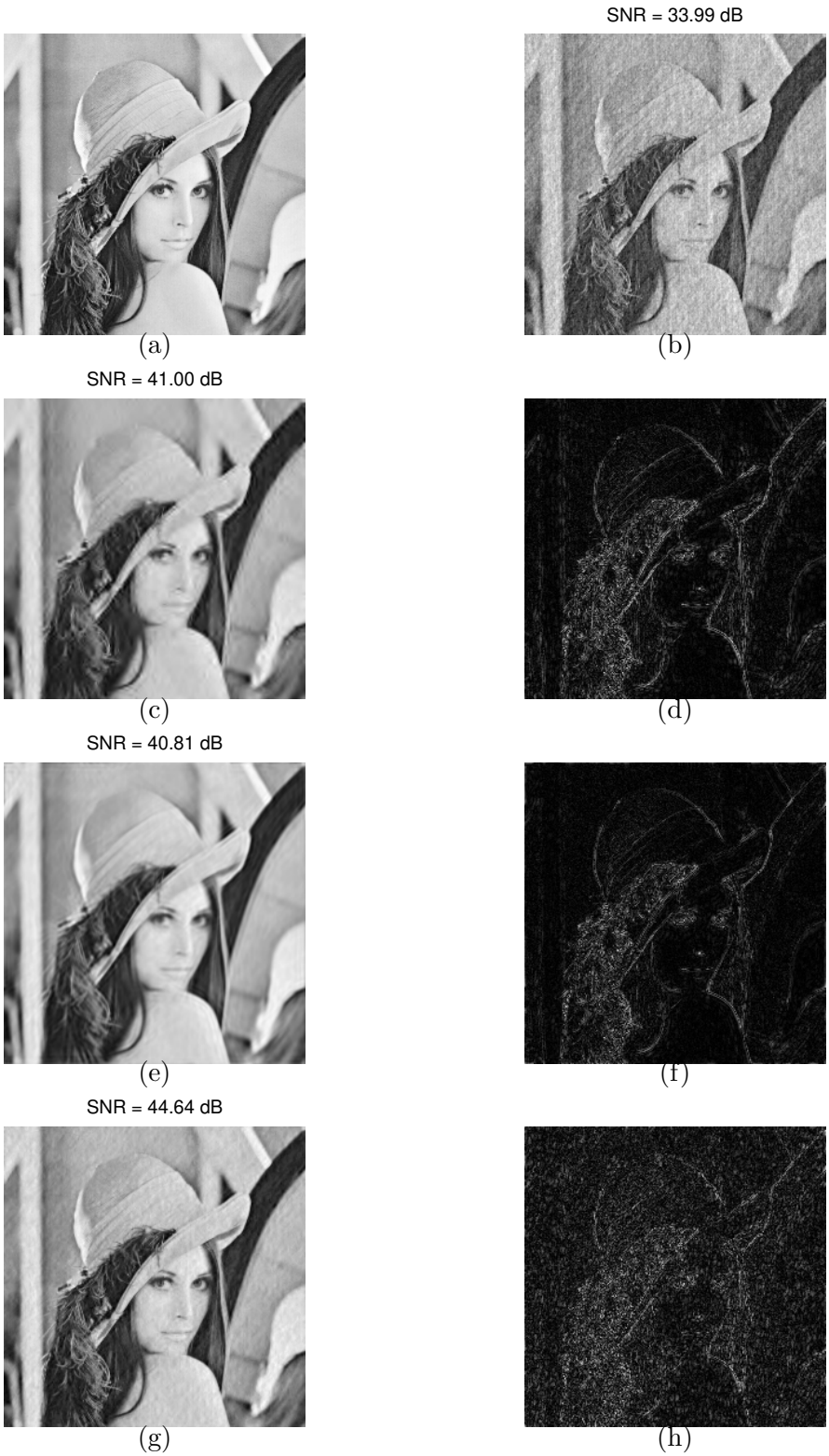


Figure 3: Fourier-domain imaging for the standard Lena image. (a) Original Lena image. (b) Directly zero-filling recovery $u = \Phi^T f$. (c) and (d) IWT recovery and its recovery error. (e) and (f) ICT recovery and error. (g) and (h) CWSpB recovery and error.

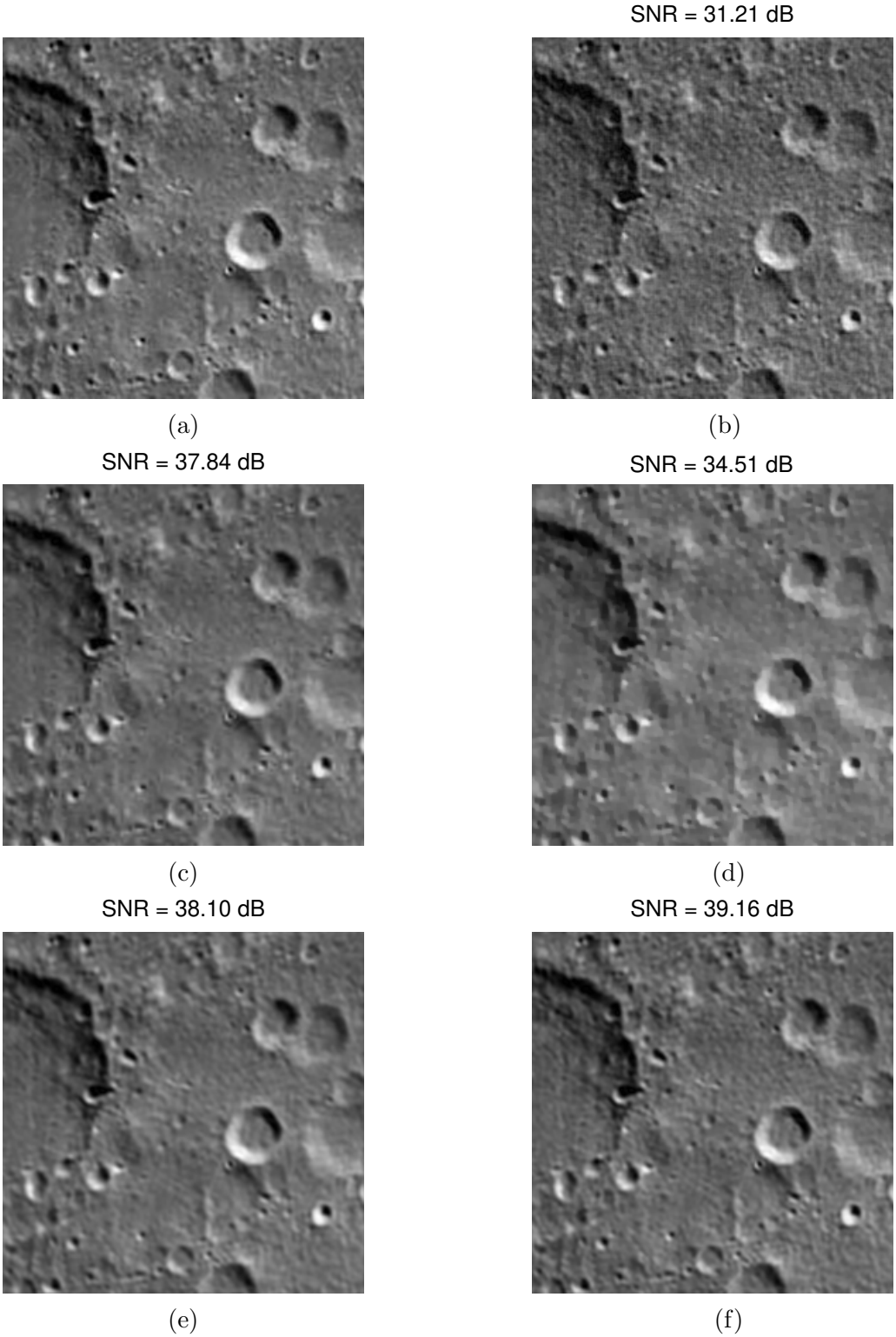
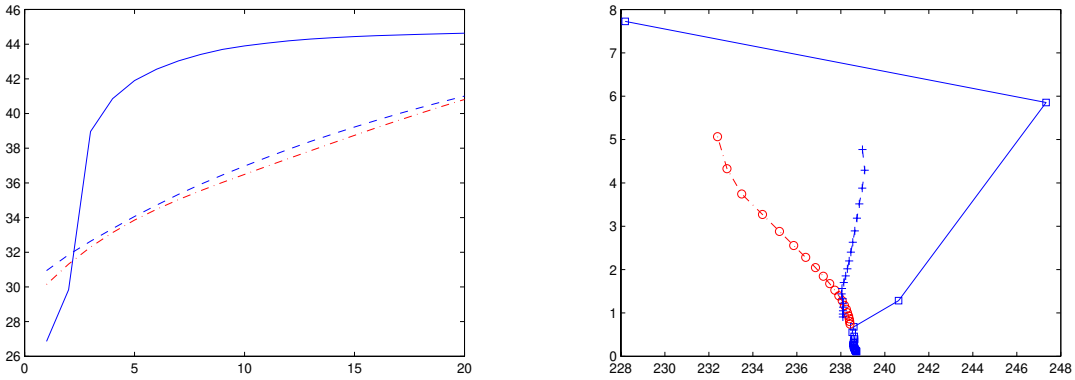
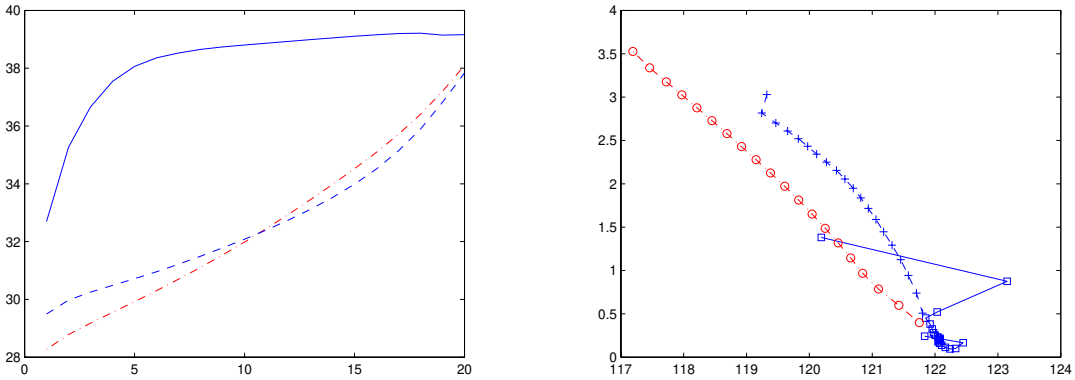


Figure 4: CS imaging for a moon surface. (a) Original unknown scene. (b) Recovery by $u = \Phi^T f$. (c) IWT. (d) Wavelet-TV. (e) ICT. (f) CWSpB.



(a) (b)

Figure 5: (a) SNR-to-iteration curves for the Lena image in Figure 3. (b) Pseudo-Pareto curves for the Lena image. Solid lines (with squares), dot-dashed lines (with circles), and dashed lines (with crosses) denote CWSpB, ICT, and IWT, respectively.



(a) (b)

Figure 6: (a) SNR-to-iteration curves for the moon image in Figure 4. (b) Pseudo-Pareto curves for the moon image. Solid line with squares, dot-dashed line with circles, and dashed line with crosses denote CWSpB, ICT, and IWT, respectively.

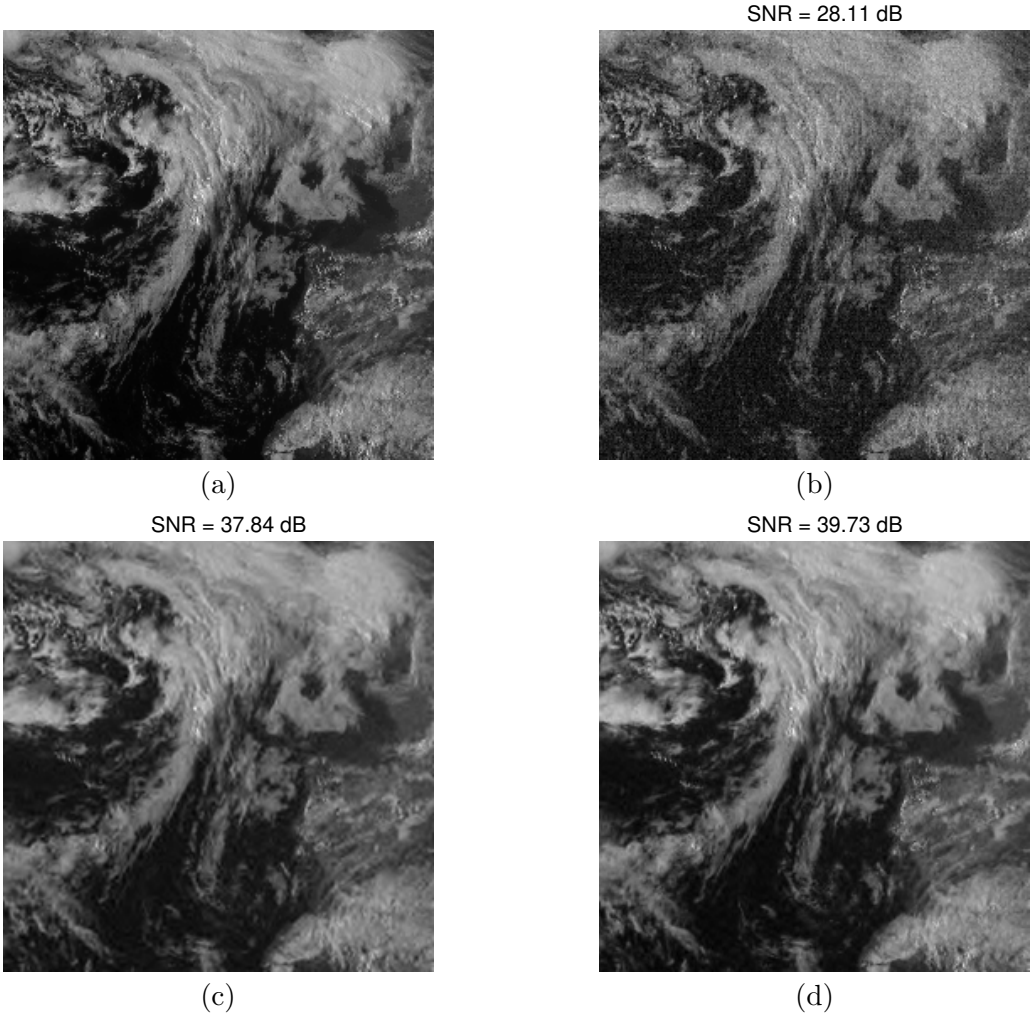


Figure 7: Performance of the proposed CWSpB method for a cloud scene. (a) Original cloud scene. (b) Recovery by $u = \Phi^T f$ (SNR=28.11 dB). (c) IWT (SNR=37.84 dB). (d) CWSpB (SNR=39.73 dB).