

# Improved Iterative Curvelet Thresholding for Compressed Sensing and Measurement

JIANWEI MA <sup>1,2</sup>

<sup>1</sup> School of Aerospace, Tsinghua University, Beijing 100084, China

<sup>2</sup> Centre of Geoscience, Ecole des Mines de Paris, 35 rue Saint-Honore 77305, Fontainebleau, France

## Abstract

A new theory named compressed sensing (CS) for simultaneous sampling and compression of signals indicates novel mechanism and design for measurement instrumentation. In this paper, we concern the recovery methods for the CS measurements. Firstly we investigate how the iterative curvelet thresholding (ICT) can be improved for sparse reconstruction of CS undetermined linear inverse problem, by considering several accelerated strategies including Bioucas-Dias and Figueiredo's two-step iteration, Beck and Teboulle's fast method, and Osher et al's linearized Bregman iteration. Secondly, we propose a two-stage active-set anisotropic-total-variation (ATV) minimization based ICT. In the first stage, a curvelet thresholding is applied to obtain a rough approximation of objects, and the index of remained significant coefficients is labeled as an active set. A Barzilai-Borwein-Dai-Yuan step size is used to accelerate the gradient line search. In the second stage, an active-set constrained ATV minimization is applied, in which only insignificant coefficients beyond the active set are changed into small values subjecting to ATV minimization of reconstructed objects. Numerical experiments show good performance of the improved ICT methods for single-pixel imaging and Fourier-domain CS imaging in remote sensing and medical engineering.

**Keywords:** Compressed sensing, iterative shrinkage/thresholding, curvelets, anisotropic total variation, Bregman iteration, Barzilai-Borwein step sizes

## 1 Introduction

In many engineering fields, e.g., geophysical exploration, medical magnetic resonance imaging (MRI) and remote sensing, we have to deal with problems on incomplete and inaccurate measurements limited by physical constraints or extreme expenses. CS, see [10, 11, 12, 22], is a new concept to solve these problems. The CS theory says that compressible unknown signals can be recovered from a small number of random or pseudo-random measurements by sparsity-promoting nonlinear recovery algorithms. Instead of taking usual samples, the measurements are inner products of signals with measurement vectors. The data acquisition now depends on its sparsity rather than its bandwidth limited by the Nyquist-Shannon sampling theorem. CS might have an important impact for measurement mechanisms and design of devices in various engineering fields, e.g., compressive optical imaging [23], medical MRI [34], machine learning [53], surface metrology [35], and remote sensing [7, 36, 37], etc.

The CS includes two steps: compressed measurement (i.e., so-called encoding) and recovery (i.e., so-called decoding). Because we use random-like matrices in the encoding step, the measured/observed data are also random-like signals where there are not visual geometric features (it should be noted that in this paper we only consider the most popular CS measurement matrix, random-like measurements, though determinate measurement [21], adaptive CS [1, 16, 28] and learning measurement [20] are other possibilities for CS encoding). Therefore, one crucial step for the CS is its nonlinear recovery. It involves sparsity-promoting reconstruction of ill-posed undetermined linear inverse problem. Different recovery algorithms have been proposed, e.g., linear

programming [10], orthogonal matching pursuit (OMP) [49], Bregman iteration [55], fixed-point continuation [27], Bayesian method [46], gradient methods [29], iterative shrinkage/thresholding (IST) algorithms [4, 18, 5, 6, 37], etc. The CS essentially shifts online measurement cost to offline recovery cost.

Among the existing methods, the iterative shrinkage methods are quite universal, robust and simple to be implemented by engineers. Another advantage is that rich existing sparse transforms can be incorporated easily into the IST framework. Therefore, it became one of the most popular tools for solving of linear inverse problems. However, a drawback of the IST methods is its slow convergence. For applications involving real-time problem and large-scale problem, the speed of decoding is very important.

Many accelerated techniques for IST methods have been proposed. Here we just mention a few literatures including Bioucas-Dias and Figueiredo's two-step IST algorithm [5], Beck and Teboulle's first-order fast IST method [4], Elad, Matalon, and Zibulevsky's sequential subspace optimization techniques [26], Nesterov's multistep gradient-like method [40], Daubechies, Fornasier, and Loris's accelerated projected gradient method [19], Fornasier's domain-decomposition method [30], and Vonesch and Unser's fast multilevel thresholded Landweber algorithm [51]. Recently, a fast algorithm for solving the  $\ell_1$ -regularized sparse reconstruction was proposed by combining the shrinkage, subspace optimization and continuation by Wen et al [52]. This method mainly includes two stages that are performed repeatedly. The first stage is based on an iterative shrinkage scheme to produce an active set of indices corresponding to the zero and nearly zero components of reconstructed objects. In the second stage, a subspace optimization problem is applied by fixing the components in this active set to zeros.

On the other hand, most existing literatures on IST methods pay attention to the mathematical analysis of convergence and use the Haar wavelets for sparse representation. Other wavelets have been rarely used in the IST methods. We show how the curvelet frames can gain the performance in the CS recovery.

In this paper, specially considering curvelet transform, we firstly investigate several accelerated iterative curvelet thresholding (ICT) methods for CS, by combining with the two-step IST [5], first-order fast IST [4], and the linearized Bregman iteration[55], respectively. The classic IST operators the (soft) shrinkage on image pixels or wavelet coefficients, while ICT operators the shrinkage on curvelet coefficients. Secondly, we extend the ICT to a more effective method by considering Barzilai-Borwein-Dai-Yuan (BBDY) step sizes [17] and an additional active-set anisotropic total variation (ATV) minimization [31] as an inner iteration. The algorithm mainly includes two iterative stages. In the first stage, a curvelet thresholding is applied to obtain a rough approximation of objects, and the index of remained significant coefficients is labeled as an active set. In the second stage, an active-set constrained ATV minimization is applied, in which only the insignificant coefficients (that have been set zeros in the thresholding step) beyond the active set are changed into small values subjecting to ATV minimization of reconstructed objects. The motivation of this stage is to reduce the noises and artifacts resulted from the curvelet shrinkage while preserving the features and edges. This stage can be also seen an ATV-minimization-promoting special shrinkage strategy. We decrease the thresholding values and the step sizes of the ATV-minimization algorithm motivated by the continuation (homotopy) strategy, and repeat two stages iteratively until a stop rule is satisfied. The ATV is not new, but it is the first time to be applied for curvelet shrinkage and further for CS. Below, we abbreviate the proposed active-set ATV-minimization based iterative curvelet thresholding as ATVICT.

It should be noted that the ATVICT method has some commons with recent two-stage method proposed by Wen et al [52]. Both methods use shrinkage, active set, BB (Barzilai-

Borwein)-like step sizes. But they have essential different idea and motivation. The motivation of [52] is to combine the respective good features of greedy algorithms and convex optimization approaches. In the first stage, a spatial-domain iterative shrinkage (using multistep iterations to achieve an active set) is applied without consideration of sparse transforms. The classic BB step sizes are used in the embedded gradient method. The motivation of our method is to combine the curvelet thresholding and anisotropic TV-minimization approaches. We use single-step curvelet thresholding to obtain an active set for the index of significant coefficients. Usually, the curvelet thresholding is much better than the spatial-domain shrinkage for general images. An improved BB step size, BBDY step size, is used for the embedded gradient components in our method. Furthermore, unlike a subspace optimization based on the active set is applied in the second stage in [52], we apply a particular ATV minimization with the active-set coefficient constraints. We can call it as a subspace ATV minimization. In [52], the shrinkage step is for providing a good initial point for subspace optimization, while in our method the ATV-minimization step is for providing an additional post-processing for the curvelet shrinkage step.

The contribution of our paper includes: 1) We combine the recent presented accelerated techniques (Bioucas-Dias and Figueiredo’s TwIST, Beck and Teboulle’s first-order FIST, linearized Bregman iteration) with the framework of ICT, and investigate the gain of the improved ICT for CS recovery. 2) We proposed a new two-stage iteration method where the curvelet thresholding in the first stage and an active-set ATV minimization in the second stage. 3) A new step size, BBDY step size, is applied in our proposed method.

The rest of this paper is organized as follows. In section 2, we introduce an abstract framework for compressed sensing. In section 3, we review basic ICT and some accelerated techniques. The curvelet-based linearized Bregman iteration is presented in section 4. The new ATVICT algorithm is proposed in section 5. Numerical results and comparisons for respective methods are given in section 6.

## 2 The basic theory for compressed sensing

Let us consider the compressed sensing problem with incomplete measurements [10, 11, 12, 22]

$$f = \Phi u + \epsilon, \quad (2.1)$$

where  $\Phi \in \mathbb{R}^{K \times N}$  ( $K \ll N$ ) is a given so-called CS measurement matrix. 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  $\Phi$ . 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  $\Psi u$  contains only a small set of significant entries. Further, let the measurement matrix  $\Phi$  be not correlated with  $\Psi$ . Usually, one assumes that  $\Phi$  satisfies the so-called Restricted Isometry Property (RIP), see [10, 12]. Then  $u$  can be reconstructed with high accuracy from the incomplete measurements  $f$ .

Frequent-used measurement matrices  $\Phi$  are random matrices because they are incoherent to most transforms and also physically realizable. Generally, one can take an orthogonal transform followed by randomly downsampling, to build the measurement matrix. In this paper we consider the most popular Gaussian random matrices and randomly downsampling Fourier matrices. Our method can be also applied for other measurement matrices that have been recently presented, including Toeplitz matrices/circulant matrices [2], jittered sampling/Poisson disk sampling/farthest point sampling [48], noiselet measurement [15], random convolution [42], adaptive CS sampling [28, 1], optimal matrices [25] and learning measurement matrices [20], etc.

Applying a transform  $\Psi$  to  $u$ , the CS-problem reads in coefficient domain

$$f = \Phi\Psi^{-1}\Psi u + \epsilon = \tilde{\Phi}\vartheta + \epsilon \quad (2.2)$$

with  $\vartheta = \Psi u$  and  $\tilde{\Phi} = \Phi\Psi^{-1}$ , where the  $\Psi$  is a forward curvelet transform in this paper and  $\Psi^{-1}$  is its inversion.

One of classical method to find a solution of the CS-problem (2.1) (resp.(2.2)) is the sparsity-promoting  $\ell_1$  minimization

$$\min_{\vartheta} \left\{ \frac{1}{2} \|f - \Phi\Psi^{-1}\vartheta\|_2^2 + \lambda \|\vartheta\|_{\ell_1} \right\}, \quad (2.3)$$

or

$$\min_u \left\{ \frac{1}{2} \|f - \Phi u\|_2^2 + \lambda \|\Psi u\|_{\ell_1} \right\}, \quad (2.4)$$

or more general formulation

$$\min_u \left\{ \frac{1}{2} \|f - \Phi u\|_2^2 + \lambda J(u) \right\}. \quad (2.5)$$

The  $J(u)$  could be various choices, e.g.,  $\|u\|_{\ell_1}$ ,  $\|\Psi u\|_{\ell_1}$ , total variation  $TV(u)$ , and Bregman distance. In the  $\ell_1$ -minimization problems, the first term is a penalty that represents closeness of the solution to observed scene. The second term is a regularization term that represents a priori sparse information of original scene.  $\lambda$  is a regularization parameter that can be tuned. Eq. (2.3) is a synthesis formulation where one seeks the sparse coefficients, while the Eq. (2.4) is an analysis formulation where one directly seeks the images  $u$  whose coefficients have  $\ell_1$  minimization. They are equivalent when the transform  $\Psi$  is orthogonal. These minimizing problems can be solved easily by IST methods or projected gradient methods [18]. In this paper, we specially consider the curvelet transform for  $\Psi$ , and show how one can improve the ICT with the help of some additional techniques.

### 3 Iterative curvelet thresholding

#### 3.1 Basic ICT

The curvelets [13, 14] constitute a directional frame, 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  $\theta_{j,l} = 2\pi l \cdot 2^{-\lfloor j/2 \rfloor}$ ,  $0 \leq l \leq 2^{\lfloor j/2 \rfloor} - 1$ ; 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

$$\Psi_{j,l,k}(x) := \Psi_j(R_{\theta_{j,l}}(x - x_k^{(j,l)})), \quad x = (x_1, x_2) \in \mathbb{R}^2, \quad (3.6)$$

where  $\Psi_j$  are smooth functions with compact support on wedges in Fourier domain. Unlike the isotropic elements of wavelet bases, the needle-shaped elements of this transform possess very high directional sensitivity and anisotropy. Such elements are very efficient in representing line-like edges and texture-like features.

Let  $\Lambda = (j, l, k)$  be the collection of the triple index. The family of curvelet functions forms a tight frame, and thus each function  $f$  has a representation

$$f = \sum_{\Lambda} \langle f, \Psi_{\Lambda} \rangle \Psi_{\Lambda}$$

where  $\langle f, \Psi_\Lambda \rangle$  denotes the scalar product of  $f$  and  $\Psi_\Lambda$ . The coefficients  $\vartheta_\Lambda = \langle f, \Psi_\Lambda \rangle$  are called curvelet coefficients of the function  $f$ . In this paper, we use the second-generation discrete curvelet transform [14, 39] for our implementation.

The basic ICT (see e.g., [18, 37]) can be written

$$u^{k+1} = \mathcal{S}_{\sigma, \mathcal{T}}(u^k + \alpha^k \Phi^T(f - \Phi u^k)). \quad (3.7)$$

Here  $\alpha^k$  is a step size of line search, and the shrinkage operator  $\mathcal{S}_{\sigma, \mathcal{T}}$  is given by

$$\mathcal{S}_{\sigma, \mathcal{T}}(f) = \sum_{\Lambda} \mathcal{T}_{\sigma}(\langle f, \Psi_{\Lambda} \rangle) \Psi_{\Lambda}, \quad (3.8)$$

where  $\mathcal{T}_{\sigma}$  can 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

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

Actually, the ICT can be also considered in coefficient domain

$$\vartheta^{k+1} = \mathcal{T}_{s, \sigma}(\vartheta^k + \alpha^k \Psi \Phi^T(f - \Phi \Psi^{-1} \vartheta^k)) \quad (3.9)$$

where  $\vartheta^k = \Psi u$ , and  $u^k = \Psi^{-1} \vartheta^k$ .

### 3.2 linear-search accelerated ICT

The two-step accelerated IST [5] has a more general formulation

$$u^{k+1} = (1 - \alpha)u^{k-1} + (\alpha - \beta)u^k + \beta \mathcal{S}_{\Psi, \mathcal{T}}(u^k + \gamma \Phi^T(f - \Phi u^k)). \quad (3.10)$$

Here the relaxed parameters  $\alpha$  and  $\beta$  can be taken as

$$\alpha = \rho^2 + 1, \quad \beta = 2\alpha/(1 + \xi) \quad (3.11)$$

where  $\xi$  is a small value, e.g.,  $10^{-1}$  or  $10^{-3}$ , and

$$\rho = \frac{1 - \sqrt{\kappa}}{1 + \sqrt{\kappa}} < 1, \quad \kappa = \zeta_1 / \max(1, \zeta_m). \quad (3.12)$$

The  $\zeta_1$  and  $\zeta_m$  denote lower and upper bound of eigenvalue of  $\Phi^T \Phi$ , respectively. Let  $\alpha = 1$  and  $\beta = 1$ , the two-step method degenerates to basic ICT.

As the same to use the previous two-step information, Beck and Teboulle [4] presented the following fast IST (FIST) version

$$\begin{aligned} u^{k+1} &= \mathcal{S}_{\Psi, \mathcal{T}}(\tilde{u}^k + \gamma \Phi^T(f - \Phi \tilde{u}^k)), \\ \tilde{u}^k &= u^k + \left(\frac{t^k - 1}{t^{k+1}}\right)(u^k - u^{k-1}). \end{aligned} \quad (3.13)$$

Here  $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$  starting with  $t^0 = 1$ .

These techniques can be seen as special cases of line search. In our work, these accelerated algorithms have been extended to use the ICT for remote sensing imaging by applying the  $\Psi$  as the curvelet transform. We will give detailed comparisons in numerical section (section 6.1) to show how they can gain the ICT for CS recovery. Below, we rename the two methods as two-step ICT and FICT, respectively.

## 4 Curvelet-based linearized Bregman iteration

In the general  $\ell_1$ -norm minimization in Eq. (2.5), we consider a specific regularized functional, so-called Bregman distance based on convex functional  $J(\cdot)$  between points  $u$  and  $v$  [44, 55]

$$D_J^q(u, v) = J(u) - J(v) - \langle q, u - v \rangle$$

where  $q \in \partial J(v)$  is some subgradient in the subdifferential of  $J$  at the point  $v$ . The Bregman iterative regularization needs to solve a sequence of convex problems

$$u^{k+1} = \min_u \{D_J^q(u, u^k) + \frac{1}{2}\|f - \Phi u\|_2^2\} \quad (4.14)$$

where  $D_J^q(u, u^k) = J(u) - J(u^k) - \langle u - u^k, q^k \rangle$ . The Bregman distance is not a distance in the usual sense. But it does measure the closeness between  $u$  and  $v$  in the sense that  $D_J^q(u, v) \geq 0$  and  $D_J^q(u, v) \geq D_J^q(w, v)$  for all points  $w$  on the line segment connecting  $u$  and  $v$  [55]. Solving of Eq. (4.14) is to minimize the Bregman distance of  $u$  to a previous solution  $u^k$ . This is equivalent to solve [55]

$$\begin{aligned} u^{k+1} &= \min_u \{J(u) + \frac{1}{2}\|\Phi u - f^{k+1}\|_2^2\}, \\ f^{k+1} &= f + f^k - \Phi u^k \end{aligned} \quad (4.15)$$

with initial value  $f^0 = 0$  and  $u^0 = 0$ .

Recently, the linearized Bergman method, as a modification to the Bregman iteration that is simpler and sometimes faster, was proposed in [55]. The convergence of the linearized Bregman method then was studied in [8, 9]. A so-called *kicking* technique was added in [45] and, more recently, the algorithm was thoroughly reviewed and generalized in [56]. The linearized Bregman iteration that is generated by linearizing the penalty term  $\frac{1}{2}\|\Phi u - f^{k+1}\|_2^2$  is given as follows

$$\begin{aligned} u^{k+1} &= \min_u \{J(u) + \frac{1}{2\delta}\|u - \delta v^k\|_2^2\} \\ v^{k+1} &= v^k + \Phi^T(f - \Phi u^k). \end{aligned} \quad (4.16)$$

In comparison with classic  $TV - L_2$  model [43], the Bregman iteration in (4.15) is adding back the residual, while the linearized Bregman iteration in (4.16) is adding back a “linearized noise” [45].

Considered  $J(u) = \|u\|_1$  for sparsity-promoting CS problem, the first equation in (4.16) can be solved explicitly by soft thresholding. One can get concise algorithm

$$\begin{aligned} u^{k+1} &= \delta \cdot \mathcal{T}_s(v^{k+1}) \\ v^{k+1} &= v^k + \Phi^T(f - \Phi u^k). \end{aligned} \quad (4.17)$$

To make the linearized Bregman iteration works for general images that are not necessarily sparse in the spatial domain but in a transform domain, we apply the linearized Bregman algorithm for ICT.

$$\begin{aligned} u^{k+1} &= \delta \cdot \mathcal{S}_{\sigma, \mathcal{T}}(v^{k+1}) \\ v^{k+1} &= v^k + \alpha^k \Phi^T(f - \Phi u^k). \end{aligned} \quad (4.18)$$

Also one can consider the Bregman-ICT in the curvelet coefficient domain

$$\begin{aligned}\vartheta^{k+1} &= \delta \cdot \mathcal{T}_s(v^{k+1}) \\ v^{k+1} &= v^k + \alpha^k \Psi \Phi^T (f - \Phi \Psi^{-1} \vartheta^k).\end{aligned}\tag{4.19}$$

The step sizes  $\alpha^k$  relate to acceleration of ICT.

Let us look at the connections between the Bregman-curvelet method and the basic ICT. By incorporating the above two equations to one equation, we have

$$v^{k+1} = \delta \cdot \mathcal{T}_s(v^k + \alpha^k \Psi \Phi^T (f - \Phi \Psi^{-1} \vartheta^k)).\tag{4.20}$$

Comparing with the coefficient-domain ICT in (3.9), the first term in right side is the  $v^k$  instead of  $\vartheta^k$ . The  $v^k$  denotes coefficient components of  $\vartheta^k$  before the thresholding operation. That is to say, in comparison with “clean”  $\vartheta^k$ , the  $v^k$  adds back the noise/residual to some extent.

On the other hand, we can incorporate the two equations in (4.19) to another equation by killing the variable  $\vartheta$

$$v^{k+1} = v^k + \alpha^k \Psi \Phi^T (f - \Phi \Psi^{-1} \delta^k \mathcal{T}_s(v^k)).\tag{4.21}$$

Comparing with the coefficient-domain ICT again, we find that the only difference is the location of the thresholding function. In (4.21), we first apply thresholding for the results obtained from latest iteration, and then add the residuals.

We outline the Bregman-curvelet algorithm in Table 1. The main difference between the Bregman-curvelet and the linearized Bregman iteration is that curvelet shrinkage (instead of simple soft shrinkage in image spatial domain) is applied for thresholding step in the Bregman-curvelet method. The analogous analysis of convergence of the Bregman-curvelet method can be also referred to [41] where the split Bregman iteration is further considered.

---



---

<b>Input:</b> $(u^0, \vartheta^0, v^0) = (0, 0, 0)$ , $\sigma = \sigma_0$ , $\delta = \delta_0$ .
<b>While</b> stopping conditions not satisfied <b>do</b>
1) $v^{k+1} = v^k + \alpha^k \Psi \Phi^T (f - \Phi \Psi^{-1} \vartheta^k)$
2) $\vartheta^{k+1} = \delta \cdot \mathcal{T}(v^{k+1})$
3) Decrease $\sigma$
4) $k \leftarrow k + 1$
<b>end while</b>
$u = \Psi^{-1} \vartheta^{k+1}$
<b>Output:</b> $u$

---

Table 1: Bregman-curvelet algorithm.

## 5 Active-set ATVICT

For two dimensional data  $u := (u_p)_{p \in I}$  we define the *discrete isotropic TV* for a symmetric neighborhood  $N(p)$  of the pixel  $p$  by

$$TV(u) = \sum_{p \in I} \sum_{p' \in N(p)} |u_{p'} - u_p|.\tag{5.22}$$

Usual neighborhoods are the 4-star-neighborhood  $N_4(p) = \{p' \in I : \|p - p'\|_2^2 = 1\}$  or the 8-neighborhood  $N_8(p) = \{p' \in I : \|p - p'\|_2^2 \leq 2\}$ . The discrete ATV functional is the weighted

sum [47]

$$ATV(u) = \sum_{p \in I} \sum_{p' \in N(p)} \sqrt{w_{p,p'}} |u_{p'} - u_p|, \quad (5.23)$$

where the weights  $w_{p,p'} \geq 0$  cause the anisotropic smoothing taking the local image geometry into account. That means,  $w_{p,p'}$  depends on the intensity difference  $|u_p - u_{p'}|$ . More precisely, the smoothing process across discontinuities (i.e. for large  $|u_p - u_{p'}|$ ) should be stopped while smoothing in flat regions (i.e. for small  $|u_p - u_{p'}|$ ) should be encouraged. Therefore we claim that  $w_{p,p'}$  is monotone decreasing for  $u_p - u_{p'} \geq 0$ , and  $w_{p,p'} \rightarrow 0$  for  $|u_p - u_{p'}| \rightarrow \infty$ . Secondly,  $w_{p,p'}$  is symmetric, i.e.  $w_{p,p'} = w_{p',p}$  for all  $p, p' \in I$ . Thirdly, we normalize the weights setting  $w_{p,p} = 1$ .

We use bilateral-filter-like weights which contain of an intensity and a spatial component [31]

$$w_{p,p'} = w_{p,p'}(u_p - u_{p'}, p - p') = \exp\left(-\frac{1}{\sigma_i^2}(u_p - u_{p'})^2\right) \cdot \exp\left(-\frac{1}{\sigma_s^2}\|p - p'\|_2^2\right).$$

The first term is the *intensity weight* which punishes pixel values  $u_{p'}$  from another image region (and thus ensures edge preserving smoothing), the second term is the *spatial weight* which punishes pixels  $p'$  that are far away from  $p$  in the spatial domain. The parameters  $\sigma_i$  and  $\sigma_s$  are control parameters.

Let  $M_{u,\sigma} := \{\Lambda : |\vartheta_\Lambda(u)| \geq \sigma\}$  be an index set of significant curvelet coefficients.  $u^0$  denotes the reconstructed image after curvelet thresholding, i.e.,  $u^0 = \sum_{\Lambda \in M_{u,\sigma}} \langle u, \Psi_\Lambda \rangle \Psi_\Lambda$ . Let  $V$  be a subspace to index the insignificant coefficients that have been set to zeros by hard/soft thresholding. Then we define a subspace  $U = \{u^0\} + V$ . The signals  $u \in U$  differ from the given  $u^0$  at most in the small coefficients which do not contain important information of original images.

The active-set ATV-minimization is to find

$$u^* \in U \text{ such that } ATV(u^*) = \min_{u \in U} ATV(u).$$

The active-set ATV-minimization means during the ATV minimization algorithm we always keep the main components of objects unchanged, but change the detail components that belong to the subspace  $V$  (see e.g., [24, 38]).

This ATV minimization can be solved by projected gradient descent scheme [24, 31]

$$u^{k+1} = u^k - t^k P_V(g_{ATV}(u^k)) \quad (5.24)$$

where  $g_{ATV}$  denotes gradient of the ATV functional. The sequence of step size ( $t^k$ ) can be taken e.g.,  $t^k = \frac{1}{k+1}$  or  $(1 - k/N_{numb})$  where  $N_{numb}$  denotes the number of total iterations. The projection  $P_V$  denotes the subspace active-act constraint. That means,  $P_V(g)$  is computed by the inverse curvelet transform only from the index of insignificant coefficients resulted from the previous curvelet thresholding

The subgradient at the point  $p \in I$  can be computed by

$$\begin{aligned} g_{ATV}(u)|_p &= \frac{\partial ATV(u)}{\partial u_p} \\ &= 2 \sum_{p' \in N(p)} \sqrt{w_{p,p'}} \text{sgn}(u_p - u_{p'}) \end{aligned} \quad (5.25)$$

with symmetric weights  $w_{p',p} = w_{p,p'}$ . For more details on the ATV, we refer to [31] where we applied the ATV minimization for tetrolet image approximation.

We outline the proposed active-set ATVICT algorithm in Table 2. In this algorithm, we use curvelet thresholding in the first stage and then apply the active-set ATV-minimization in the second stage. Generally, We just need to apply two or three iterations for active-set ATV-minimization algorithm. In addition, unlike the Barzilai-Borwein step sizes used in [29, 52, 54, 33], we apply the Barzilai-Borwein-Dai-Yuan (BBDY) step sizes in our algorithm. We give details on the BBDY step sizes in the following subsection.

---



---

<b>Input:</b> $u^0 = 0, \sigma = \sigma^0, t^0 = 1$
<b>While</b> stopping conditions not satisfied <b>do</b>
1) Compute the BBDY step size $\alpha_k^{BBDY}$ .
2) $u^{k+1} = u^k + \alpha_k^{BBDY} \Phi^T(f - \Phi u^k)$ .
3) Apply thresholding $u^{k+1} = \mathcal{S}_{\sigma, \mathcal{T}}(u^{k+1})$ and record the active set $M_{\mu, \sigma}$ .
4) Apply active-set ATV minimization $u^* = \min_{u \in U} ATV(u)$ using $u^{k+1}$ as an initial value.
5) Decrease threshold values $\sigma$ and step sizes $t^k$ in the ATV algorithm.
6) $u^{k+1} \leftarrow u^*, k \leftarrow k + 1$
<b>end while</b>
<b>Output:</b> $u^*$

---



---

Table 2: ATVICT algorithm.

One can also incorporate the Bregman-curvelet algorithm proposed in section 4 into the ATVICT, by simply replacing the above 2nd and 3th step with 2)  $v^{k+1} = v^k + \alpha_k^{BBDY} \Phi^T(f - \Phi v^k)$ , and 3)  $u^{k+1} = \mathcal{S}_{\sigma, \mathcal{T}}(v^{k+1})$  and record the active set.

### 5.1 Barzilai-Borwein-Dai-Yuan step sizes

The steepest descent (SD) method defines the next iteration by

$$u_{k+1} = u_k - \alpha_k^* g_k$$

where  $g_k = g(u_k)$  denotes the gradient vector of object function  $\mathcal{F}$  at  $u_k$ , and the step size  $\alpha_k^* > 0$  satisfies

$$\mathcal{F}(u_k - \alpha_k^* g_k) = \min_{\alpha} \mathcal{F}(u_k - \alpha g_k).$$

For  $\mathcal{F} = \frac{1}{2} \|\Phi u - f\|_2^2$  in our case, we have classical formulation

$$g_k = \Phi^T(\Phi u_k - f), \quad \alpha_k^* = \frac{\|g_k\|_2^2}{g_k^T \Phi g_k}.$$

In 1988, the well-known Barzilai-Borwein step size was proposed [3]

$$\alpha_k^{BB1} = \frac{s_{k-1}^T g'_{k-1}}{\|g'_{k-1}\|_2^2},$$

and

$$\alpha_k^{BB2} = \frac{\|s_{k-1}\|_2^2}{s_{k-1}^T g'_{k-1}}$$

where  $s_{k-1} = u_k - u_{k-1}$  and  $g'_k = g_k - g_{k-1}$ . These step sizes are derived by minimizing function

$$\min \|s_{k-1} - D_k g'_{k-1}\|_2,$$

and

$$\min \|D_k^{-1}s_{k-1} - g'_{k-1}\|_2, \quad D_k = \alpha_k I.$$

It has been verified that the gradient method with BB step sizes converges faster than that with classic SD step sizes (see e.g., [57]).

Based on the seminal work, Dai and Yuan [17, 57] proposed an improved step size (we call Barzilai-Borwein-Dai-Yuan (BBDY) step size below) by combining the BB step sizes and a new one

$$\alpha_k^{BBDY} = \begin{cases} \alpha_k^*, & \text{mod}(k, 4) = 1 \text{ or } 2 \\ \alpha_k^{DY}, & \text{otherwise} \end{cases} \quad (5.26)$$

where

$$\alpha_k^{DY} = \frac{2}{\sqrt{(1/\alpha_{k-1}^* - 1/\alpha_k^*)^2 + 4\|g_k\|_2^2/(\alpha_{k-1}^*\|g_{k-1}\|_2)^2 + 1/\alpha_{k-1}^* + 1/\alpha_k^*}}. \quad (5.27)$$

The BBDY scheme produced better numerical results than BB method [17, 57]. In order to avoid the parameter  $\alpha^{BBDY}$  being either too small or too large, we can limit their bound by

$$\alpha_k^{BBDY} = \min\{\max\{\alpha_{min}, \alpha_k^{BBDY}\}, \alpha_{max}\}$$

where  $\alpha_{min}$  and  $\alpha_{max}$  are the lower and upper limitation. Usually we take  $\alpha_{min} = 10^{-2}$  and  $\alpha_{max} = 10^2$ . The  $\alpha_k^{BBDY}$  step sizes have been used in ATVICT algorithm described in Table 2.

## 6 Numerical experiments

### 6.1 Two-step ICT and FICT

In the first example, we tested the curvelet-based methods in Eq. (3.10) and (3.13) with different thresholding values for computer simulation of single-pixel remote sensing imaging of Chinese Chang'e-1 lunar probe for moon surface. The single-pixel remote sensing imaging can be interpreted as follows [23, 36]. We firstly reshape the 2D  $n \times n$  image into a 1D signal  $x = \{x_1, \dots, x_i, \dots, x_N\}$ ,  $N = n^2$ , and assemble the 2D  $\Phi$  matrix using random vector  $\Phi_m = \{\Phi_{m,1}, \dots, \Phi_{m,i}, \dots, \Phi_{m,N}\}$ , ( $m = 1, \dots, K$ ) in each row. The random  $\Phi_m$  is steered by orienting the bacterium-sized mirrors in digital micro-mirror device. The reflecting light field is focused on the single photodiode by lens, in order to get one measurement  $y_m = \sum_{i=1}^N \Phi_{m,i}x_i$ . Repeat  $K$  times using different  $\Phi_m$  to obtain all measurements  $y = \Phi x$ . The camera is single-pixel sequential imaging. But the number of total data is much smaller than the data acquired by traditional CCD camera. In this case, the measurement matrix  $\Phi$  is a  $K \times N$  random matrix. We consider 30% measurements, i.e.,  $K = 30\% \cdot N$ . Figure 1 (a) is an original scene (with  $80 \times 80$  sizes) of moon surface taken by Chinese Chang'e-1 lunar probe. Fig. 1 (b) is obtained directly by  $u = \Phi^T f$ . Fig. 1 (c) is obtained by iterative Daubechies Db4 wavelet thresholding (IWT) with decay threshold values  $\sigma = \sigma_0(1 - k/k_{numb})$ . The initial thresholding value  $\sigma_0 = 0.1$ ,  $k$  is an index of iterations, and the number of total iterations  $k_{numb} = 40$ . Fig. 1 (d) shows a comparing result obtained by a state-of-the-art algorithm/code named TVAL3 where the TV minimization is implemented by augmented lagrangian and alternating direction algorithms [32]. The TVAL3 uses 90 iterations with elapsed time 20.82 seconds. Fig. 1 (e) is obtained by ICT with the same decay threshold values. Fig. 1 (f) is the result obtained by FICT method in Eq. (3.13) with the same computational parameters. We also apply the two-step ICT in Eq. (3.10) with parameters  $\alpha = 1.2$  and  $\beta = 2.18$ , and using different strategies of threshold decay, i.e.,

(A) :  $\sigma = \sigma_0(1 - k/k_{numb})$ , (B) :  $\sigma = \sigma_0(1 - \sqrt{k/k_{numb}})$ , and (C) :  $\sigma = \sigma_0(1 - \sqrt[3]{k/k_{numb}})$ , respectively. Fig. 1 (g),(h) and (i) show the results by the two-step ICT using the decay threshold strategy (A), (B), and (C), respectively. The elapsed time of our two-step ICT is 63.27 seconds. The most time is consumed by curvelet transform in each iteration. Indeed, the state-of-the-art TVAL3 is three times faster than our methods, but our methods achieves much higher SNR values and visual quality. The curvelets gain the performance obviously. The two-step ICT converges faster than simple ICT to achieve a given SNR values.

Figure 2 (a) and (b) show comparisons of the change of SNR and recovery errors as the iteration increases, by using the ICT and two-step ICT with different thresholding values, respectively. The dash-dotted line denotes the ICT. The solid line, dotted line, and dashed line denote the two-step ICT with above three different strategies of thresholding decay (A)-(C), respectively. It can be seen that the two-step ICT achieves higher SNR values than original ICT. In terms of the two-step ICT, the faster decay threshold values result in the higher SNR values in this case. Fig. 2 (c) displays their pseudo-Pareto curves (vertical coordinate  $\|f - \Phi u^k\|_2^2$  versus horizontal coordinate  $\|u^k\|_1$ ). The Pareto curve [50] is a curve that traces optimal trade-off between one-norm of the solution  $\|u^k\|_1$  and two-norm of the residual  $\|f - \Phi u^k\|_2^2$  in solving of least-square problems. The ideal Pareto curve can be seen as an optimal reference for one-norm solvers. In this paper, we use the name of pseudo-Pareto to refer to such a similar curve but not optimal. The dash-dotted circle line denotes the ICT method where the circle denotes the iteration point. The solid square line, dotted point line, dashed cross line denote the two-step ICT methods with three thresholding strategies (A)-(C), respectively. It is clear to see that the two-step ICT converges faster than the ICT method. Faster decay thresholds result in a little faster convergence for the residual  $\|f - \Phi u^k\|_2^2$  but slower convergence for the  $\|u^k\|_1$ . Fig. 2 (d) is a close-up of (c) in the dense trail part.

Figure 2 (e) shows comparisons of the SNR curves of the ICT, FICT, and two-step ICT with thresholding strategy (A). Fig. 2 (f) shows their pseudo-Pareto curves. The dash-dotted line, dashed line, and solid line denote the ICT, FICT, and two-step ICT, respectively. The FICT has the similar convergence speed with the two-step ICT. Both of them are faster than the original ICT. All three methods use the linear decay threshold values  $\sigma = \sigma_0(1 - k/k_{numb})$ .

## 6.2 Bregman-ICT

We consider the Bregman-ICT in Eq. (4.19) for Fourier-domain CS imaging [34]. Fig. 3 (a) is an original cloud scene with sizes  $512 \times 512$ . Fig. 3 (b) shows the Fourier-domain pseudo-random sampling where the white points are the sampling points. We consider 30% measurements. Fig. 3 (c) is a result directly recovered by  $u = \Phi^T f$ . Fig. 3 (d) shows the results by iterative wavelet-TV regularization that was used in [34]. Due to the limitation of wavelets for line singularities, the method can not recover the edges well. Fig. 3 (e) is obtained by the coefficient-domain ICT in Eq. (3.9) with 40 iterations, and Fig. 3 (f) is obtained by Bregman-ICT with parameter  $\delta = 1$  in Eq. (4.19) only using 7 iterations. Continually, in Figure 4, we show the pseudo-Pareto curves of the two ICT methods, in comparison with spatial-domain ICT method. It is obvious that the Bregman-ICT method converges very fast. The computation complexity of the Bregman-ICT is very close to original ICT. In the case of  $512 \times 512$  images, each iteration of Bregman-ICT takes 10.72 seconds.

## 6.3 ATVICT

Because the ATV curvelet thresholding itself is new, we first show its performance by denoising tests for a piecewise image and MRI image, in comparison with some existing related methods.

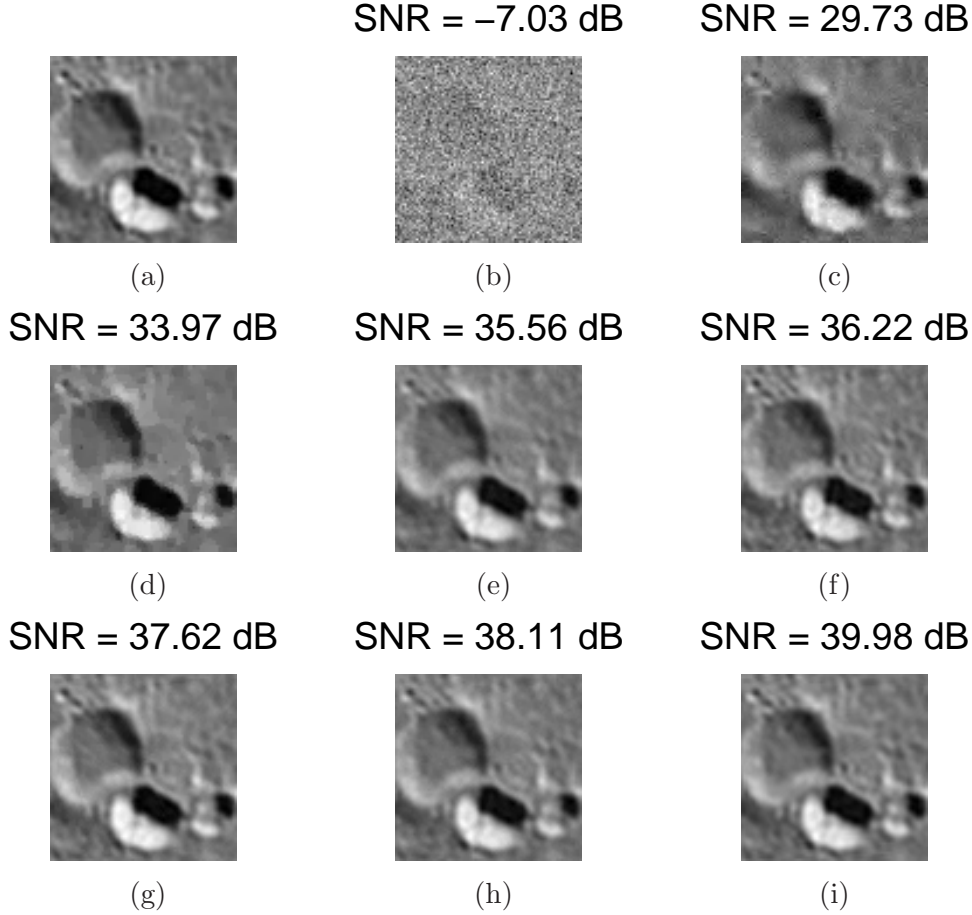


Figure 1: Single-pixel CS imaging of moon surface. (a) Original scene. (b) Recovery directly by  $u = \Phi^T f$ . (c) IWT. (d) TVAL3. (e) ICT. (f) FICT. (g)-(i) two-step ICT with thresholding strategy (A)-(C), respectively. The coordinates denote  $80 \times 80$  spatial sampling.

Fig. 5 (a) shows the original image. Fig. 5 (b) shows the noisy image where we consider Gaussian white noise. Fig. 5 (c) and (d) are obtained by Db4 wavelet hard thresholding and curvelet thresholding, respectively. The curvelet transform preserves the edges much better than wavelet transform. But there is still some obvious pseudo-Gibbs and element-like artifacts. Fig. 5(e) and (f) are obtained by TV-based curvelet thresholding and ATV-based curvelet thresholding, respectively. The proposed method obtains more clean results with well preserved edges. Both iterative methods use step size  $t_k = 0.002$  and 20 iterations. The additional parameters for ATV-based curvelets are  $\sigma_i = 40$ ,  $\sigma_s = 1$ , and the neighborhood size  $5 \times 5$ . Since the anisotropic functional enforces edge preserving we can increase the step size  $t^k$  without fear that edges were smoothed. This leads to the fact that in practice one possibly needs fewer iterations to get a convergence result in contrast to the isotropic case.

Now we test the proposed ATVICT method for CS recovery. Figure 6 (a) shows an original MRI brain image with  $366 \times 366$  sizes. Fig. 6 (b) shows a result obtained by ICT. Fig. 6 (c) shows the recovered result by the ATVICT. The computational parameters are listed as follows. Threshold values  $\sigma = \sigma_0(1 - k/N_{numb})$  with an initial value  $\sigma_0 = 0.06$  and total iterations  $N_{numb} = 20$ .  $t^k = t^0(1 - k/N_{numb})$  with an initial value  $t^0 = 0.04$ , in each iteration, the ATV minimization implements two internal iterations. Fig. 6 (d) shows comparisons of SNR changes

as the number of iterations increases. The dotted line denotes ICT. The solid line denotes the proposed ATVICT with BBDY step sizes. As comparison, we also give a result by ATVICT with BB step sizes, as shown by dot-dashed line. It can be seen that the ATVICT's results have better visual quality and higher SNR values.

Figure 7 shows performance of the ATVICT for a textural Barbara image with sizes  $256 \times 256$ . Fig. 7 (a) is the original image, and Fig. 7 (b) is a direct recovery by  $u = \Phi^T f$ . Fig. 7 (c) is a result by using a  $TV - \ell_1$  model [34]. Fig. 7 (d) is obtained by ICT. Fig. 7 (e) is obtained by the proposed ATVICT method. Fig. 7 (f) shows its recovery error. It can be seen that only small detailed features can not be recovered. We did not consider the measurement noises in above tests.

From many other experiments for the ATVICT, we observed that for piecewise smooth images (e.g., MRI brain image) the ATV leads to main gains for the improved results, while for detailed texture images (e.g., Barbara image) the gains mainly benefit from the BBDY step sizes.

The computational time of ATVICT is mainly taken by curvelet shrinkage and ATV minimization. For a  $64 \times 64$  image, the curvelet shrinkage consumes 0.72 seconds and ATV-minimization algorithm consumes 0.79 seconds. For  $128^2$ ,  $256^2$ ,  $512^2$  images, the curvelet shrinkage consumes 2.42, 9.21, 31.93 seconds, and the ATV algorithm consumes 2.76, 10.40, 40.55 seconds, respectively. Basically, the curvelet shrinkage and ATV algorithm have the same-order computational complexity. In the experiments shown in Fig. 7, we use 20 iterations and each iteration uses two inner-iteration ATV algorithm. The elapsed time is 594.80 seconds. Instead of using the gradient descent method, one can use advance methods (e.g., split Bregman iteration [58]) for solving the ATV-minimization model to speedup our algorithm in future. On the other hand, how to build fast curvelet transform with fewer redundance is open.

Besides these illustrative examples, we also give a numerical evaluation of the accuracy and robustness of the ATVICT as the number of measurements increases. Figure 8 shows the change of SNR and recovery error as the number of measurements increases from 15% to 60% measurements. The solid line denotes ATVICT, and the dash-dotted line denotes ICT. It can be seen clearly that the accuracy of ATVICT is indeed higher than that of ICT method. The change curves of both methods are robust monotonic functions as the number of measurements increase.

From the above experiments and comparisons, we observed two aspects: 1) The choice of sparse transforms is one of the most important factors for CS recovery. The curvelet transform used in our work greatly gains the performance of CS recovery in terms of SNR values and visual quality, especially for textural images. 2) By combining some accelerated techniques, the ICT indeed improves the efficiency to some extent. 3) There is still big room to reduce the computational time by speeding the curvelet transform and ATV algorithm themselves.

Finally, we should note that all experiments are implemented by using laptop with 1.86GHz Intel Pentium processing and 512 MB memory.

## 7 Conclusion

The motivation of this paper is to investigate how we can improve the ICT for CS recovery. Some additional techniques including linearized Bregman algorithm and active-set ATV-minimization are incorporated into the ICT. Iterative shrinkage methods essentially extract useful features and kick out noisy components in each iteration. The methods in (3.10) and (3.13) add the scaling differences between the successive iterations into next iteration, while the Bregman iteration method incorporates the residual information between the recovery and observation into the

next iteration. The ATVICT can be simply interpreted as an ICT with a special curvelet shrinkage.

Each method has advantages and drawbacks. For instance, the Bregman-ICT is much faster than the ATVICT, but the latter achieves higher SNR values, especially for textural images. The proposed algorithm could be further combined each other not only for image reconstruction problems, but also for other CS problems in real applications in remote sensing, medical engineering, optical engineering, and metrology, by modifying the CS measurement matrices.

## 8 Acknowledgments

The author would like to thank Wotao Yin from the Rice University (USA) for his many constructive comments, which substantially improved the presentation of the first version of this paper, and Jens Krommweh from University of Duisburg-Essen (Germany) for worthy work on ATV. J.M. also thanks financial support from NSFC (40704019), CACS Innovation Fund (20094030085), Tsinghua Autonomous Research Plan (2009THZ02-1), and Invited Professorship at Ecole des Mines de Paris.

## References

- [1] A. Aldroubi, H. Wanf, K. Zarringhalam, Sequential adaptive compressed sampling via Huffman codes, preprint, 2009.
- [2] W. Bajwa, J. Haupt, G. Raz, S. Wright, R. Nowak, Toeplitz structured compressed sensing matrices, *IEEE Workshop SSP*, 2007.
- [3] J. Barzilai, J. Borwein, Two point step size gradient methods, *IMA J. Numer. Anal.*, **8**, 141-148 (1988).
- [4] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, **2**, 183-202 (2009).
- [5] 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).
- [6] T. Blumensath, M. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.*, **27** (3), 265-274 (2009).
- [7] J. Bobin, J. Starck, R. Ottensamer, Compressed Sensing in Astronomy, *IEEE J. Selected Topics in Signal Process.*, **2** (5), 718-726 (2008).
- [8] J. Cai, S. Osher, Z. Shen, Linearized Bregman iterations for compressed sensing, *Math. Comput.*, **78**, 1515-1536 (2009).
- [9] J. Cai, S. Osher, Z. Shen, Convergence of the linearized Bregman iteration for  $\ell_1$ -norm minimization, *Math. Comput.*, **78** (268), 2127-2136 (2009).
- [10] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Infor. Theory*, **51**, 4203-4215 (2005).
- [11] E. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate information, *Comm. Pure Appl. Math.*, **59**, 1207-1233 (2005).
- [12] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Infor. Theory*, **52** (2), 489-590 (2006).
- [13] E. Candès, D. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise singularities, *Comm. Pure Appl. Math.*, **57**, 219-266 (2004).
- [14] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, *Multiscale Model. Simul.*, **5** (3), 861-889 (2006).

- [15] R. Coifman, F. Geshwind, Y. Meyer, Noiselets, *Appl. Comput. Harmon. Anal.*, **10** (1), 27-44 (2001).
- [16] W. Coulter, C. Hillar, F. Sommer, Adaptive compressed sensing - a new class of self-organizing coding models for neuroscience, preprint, 2009.
- [17] Y. Dai, Y. Yuan, Analysis of monotone gradient method, *J. Industrial and Management Optimization*, **1**, 181-192 (2005).
- [18] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, **57** (11), 1413-1457 (2004).
- [19] I. Daubechies, M. Fornasier, I. Loris, Accelerated projected gradient method for linear inverse problems with sparsity constraints, *J. Fourier Anal. Appl.*, **14** (5-6), 764-792 (2008).
- [20] J. Duarte-Carvajalino, G. Sapiro, Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization, *IEEE Trans Image Process.*, **18** (7), 1395-408 (2009).
- [21] R. De Vore, Deterministic constructions of compressed sensing matrices, *J. of Complexity*, **23**, 918-925 (2007).
- [22] D. Donoho, Compressed sensing, *IEEE Trans. Infor. Theory*, **52** (4), 1289-1306 (2006).
- [23] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, Single-pixel imaging via compressive sampling, *IEEE Signal Process. Magazine*, **25** (2), 83-91 (2008).
- [24] S. Durand, J. Froment, Reconstruction of wavelet coefficients using total variation minimization, *SIAM J. Sci. Comput.*, **24**, 1754-1767 (2003).
- [25] M. Elad, Optimized projections for compressed sensing, *IEEE Trans. Signal Process.*, **55** (12), 5695-5702 (2007).
- [26] M. Elad, B. Matalon, M. Zibulevsky, Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization, *Appl. Comput. Harmon. Anal.*, **23**, 346-367, (2007).
- [27] E. Hale, W. Yin, Y. Zhang, Fixed-point continuation for  $\ell_1$ -minimization: methodology and convergence, *SIAM J. Optimization*, **19** (3), 1107-1130 (2008).
- [28] J. Haupt, R. Castro, R. Nowak, Distilled sensing: selective sampling for sparse signal recovery, *Proc. 12th conf. Artificial Intelligence and Statistics*, FL, April 2009.
- [29] 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).
- [30] M. Fornasier, Domain decomposition methods for linear inverse problems with sparsity constraints, *Inverse Problems*, **23**, 2505-2526 (2009).
- [31] J. Krommweh, J. Ma, Tetrolet shrinkage with anisotropic total variation minimization for image approximation, *Signal Processing*, 2010, to appear.
- [32] C. Li, W. Yin, Y. Zhang, TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms, <http://www.caam.rice.edu/optimization/L1/TVAl3/>.
- [33] I. Loris, M. Bertero, C. De Mol, R. Zanella, L. Zanni, Accelerating gradient projection methods for  $\ell_1$ -constrained signal recovery by steplength selection rules, *Appl. Comput. Harmon. Anal.*, **27** (2), 247-254 (2009).
- [34] 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).
- [35] J. Ma, Compressed sensing for surface characterization and metrology, *IEEE Trans. Instrum. Measure.*, 2010, to appear.
- [36] J. Ma, Single-pixel remote sensing, *IEEE Geosci. Remote Sensing Lett.*, **6** (2), 199-203 (2009).

- [37] J. Ma, F.-X Le Dimet, Deblurring from highly incomplete measurements for remote sensing, *IEEE Trans. Geosci. Remote Sensing*, **47** (3), 792-802 (2009).
- [38] J. Ma, M. Fenn, Combined complex ridgelet shrinkage and total variation minimization, *SIAM J. Sci. Comput.*, **28** (3), 5984-1000 (2006).
- [39] J. Ma, G. Plonka, A review of curvelets and recent applications, *IEEE Signal Processing Magazine*, **27** (2), 2010.
- [40] Y. Nesterov, Gradient methods for minimizing composite objective function, preprint, 2007.
- [41] G. Plonka, J. Ma, Curvelet-wavelet regularized split Bregman method for compressed sensing, *Int. J. Wavelets, Multi. Inform. Process.*, 2010, to appear.
- [42] J. Romberg, Compressive sensing by random convolution, *SIAM J. Image Sci.*, **2** (4), 1098-1128 (2009).
- [43] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithm, *Physica D*, **60**, 259-268 (1992).
- [44] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, *Multi. Model. Simul.*, **4**, 460-489 (2005).
- [45] S. Osher, Y. Mao, B. Dong, W. Yin, Fast linearized Bregman iteration for compressive sensing and sparse denoising, *Commun. Math. Sci.*, **8** (1), 93-111 (2010).
- [46] S. Ji, Y. Xue, L. Carin, Bayesian compressed sensing, *IEEE Trans. Signal Process.*, **56** (6), 2346-2356 (2008).
- [47] V. Ta, S. Bougleux, A. Elmoataz, and O. Lezoray, Nonlocal anisotropic discrete regularization for image, data filtering and clustering, preprint, 2007.
- [48] G. Tang, R. Shahidi, J. Ma, F. Herrmann, Design of two-dimensional randomized sampling schemes for curvelet-based sparsity-promoting seismic data recovery, *Geophysical Prospecting*, revised, 2010.
- [49] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, **53** (12), 4655-4666 (2008).
- [50] E. van den Berg, M. Friedlander, Probing the pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.*, **31** (2), 890-912 (2008).
- [51] C. Vonesch, M. Unser, A fast multilevel algorithm for wavelet-regularized image restoration, *IEEE Trans. Image Process.*, **18** (3), 509-523 (2009).
- [52] Z. Wen, W. Yin, D. Goldfarb, Y. Zhang, A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation, submitted, 2009.
- [53] 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).
- [54] S. Wright, R. Nowak, M. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process.*, **57** (7), 2479-2493 (2009).
- [55] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for  $\ell_1$  minimization with applications to compressed sensing, *SIAM J. Imaging Sci.*, **1** (1), 143-168 (2008).
- [56] W. Yin, The analysis and generalizations of the linearized Bregman method, submitted, 2009.
- [57] Y. Yuan, A new stepsize for the steepest descent method, *J. Comput. Math.*, **24**, 149-156 (2006).
- [58] X. Zhang, M. Burger, X. Bresson, S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, preprint, 2009.

**Jianwei Ma** received the Ph.D. degree in solid mechanics from Tsinghua University in 2002. He has been a Visiting Scholar, Research Fellow, and Guest Professor with the University of Cambridge, University of Oxford, University of Huddersfield, University of Grenoble (INRIA), University of Duisburg-Essen, and EPFL at Switzerland. He has been an Assistant Professor (since 2006) and Associate Professor (since 2009) with the School of Aerospace, Tsinghua University. From June to September 2007, he was a Visiting Professor in Florida State University and University of Colorado at Boulder. In May 2008, and from March to September 2009, he was an Invited Professor at INRIA-Grenoble and Ecole des Mines de Paris. He is an Editor of Int. J. Artificial intelligence, and a chairman/organizer of the 2th Conference on Applied Geophysics at Beijing, 2009. His research interests include curvelets, image processing, compressed sensing, remote sensing and seismic exploration. Contact him at [jma@tsinghua.edu.cn](mailto:jma@tsinghua.edu.cn).

Caption of figures:

Fig 1: Single-pixel CS imaging of moon surface. (a) Original scene. (b) Recovery directly by  $u = \Phi^T f$ . (c) IWT. (d) TVAL3. (e) ICT. (f) FICT. (g)-(i) two-step ICT with thresholding strategy (A)-(C), respectively. The coordinates denote  $80 \times 80$  spatial sampling.

Fig 2: Comparisons of the ICT, FICT and two-step ICT for single-pixel imaging. (a) Recovery SNR (vertical coordinate) versus the number of iterations (horizontal coordinate) for ICT and two-step ICT with different thresholding strategies. (b) Recovery error versus iterations for the ICT and two-step ICT methods. (c) Comparisons of their pseudo-Pareto curves (vertical:  $\|f - \Phi u^k\|_2^2$ , horizontal:  $\|u^k\|_1$ ). (d) A close-up of (c) in the dense trail part. (e) SNR versus iterations by the ICT, FICT, and two-step ICT methods. (f) Pseudo-Pareto curves.

Fig 3: Fourier-domain CS imaging. (a) Original scene. (b) Sampling model. (c) Direct reconstruction. (d) Recovery by iterative wavelet-TV regularization. (e) and (f) ICT with 40 iterations and Bregman-ICT with 7 iterations. The coordinates denote  $512 \times 512$  spatial sampling.

Fig 4: Pseudo-Pareto curves for spatial-domain IST (solid cross line), coefficient-domain IST (dotted circle line), and Bregman-IST (dot-dashed square line). All methods display 40 iterations. The vertical coordinate denotes  $\|f - \Phi u^k\|_2^2$  and the horizontal coordinate denotes  $\|u^k\|_1$ .

Fig 5: Denoising of a piecewise smooth object. (a) Original image. (b) Noisy image. (c) Wavelet thresholding (Db4 wavelets). (d) Curvelet thresholding. (e) TV-curvelet thresholding. (f) ATV-curvelet thresholding.

Fig 6: Decoding of CS for MRI brain image. (a) Original image. (b) ICT recovery. (c) ATVICT recovery. (d) Recovery SNR (vertical coordinate) verse the number of iterations (horizontal coordinate). The dotted line denotes ICT. The solid line denotes the proposed ATVICT with BBDY step sizes. The dot-dashed line denotes ATVICT with BB step sizes.

Fig 7: Decoding of CS for Barbara image by using different methods. (a) Original image. (b) Direct recovery. (c) TV- $\ell_1$  recovery. (d) ICT recovery. (e) ATVICT recovery. (f) The ATVICT recovery errors.

Fig 8: (a) Recovery SNR (vertical coordinate) versus the number of measurements. (b) Recovery error (vertical coordinate) versus the number of measurements. The horizontal coordinate denotes the ratio of number of measurements to total number of pixels. Solid line and dash-dotted line denote ATVICT and ICT, respectively.

The Broad Topics is:

1. Algorithms, Artificial Intelligence, SOFT Computing and Informatics

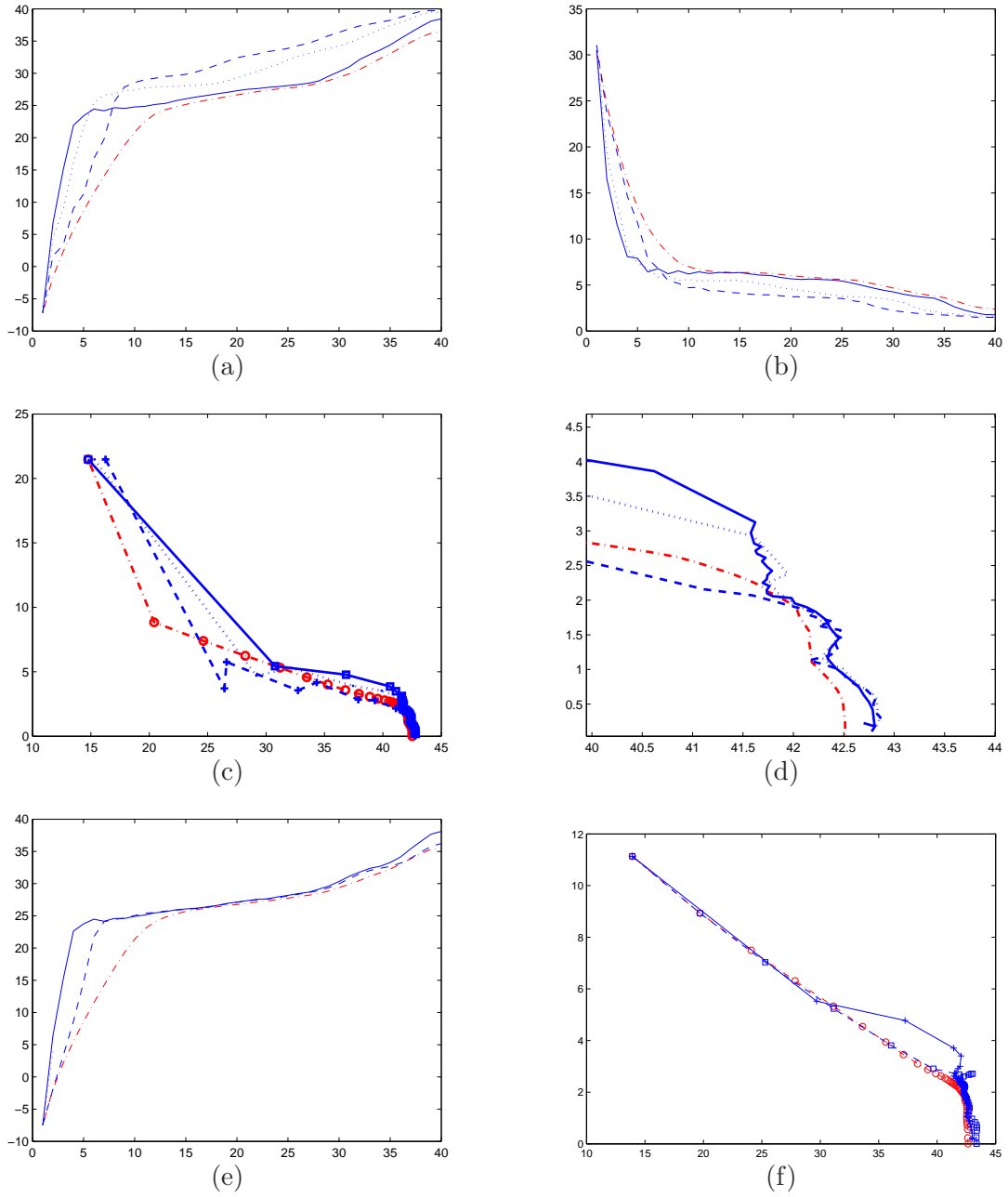
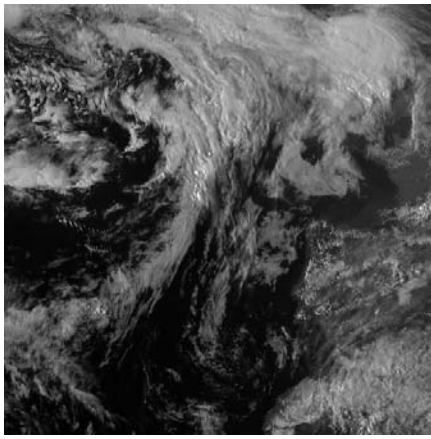
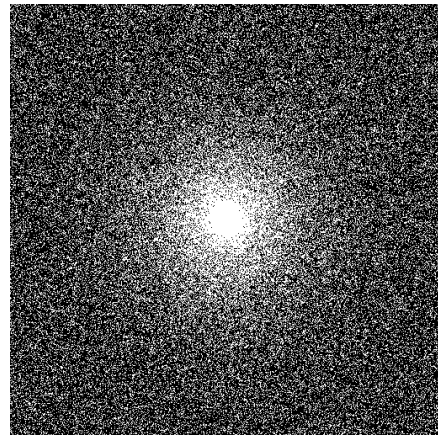


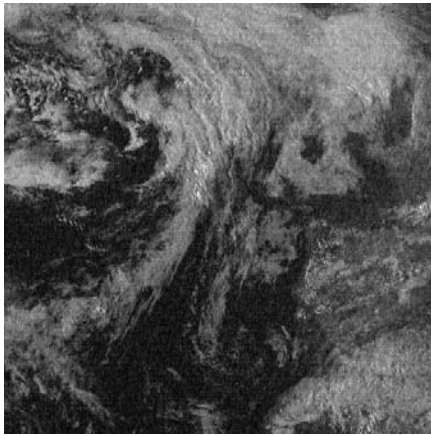
Figure 2: Comparisons of the ICT, FICT and two-step ICT for single-pixel imaging. (a) Recovery SNR (vertical coordinate) versus the number of iterations (horizontal coordinate) for ICT and two-step ICT with different thresholding strategies. (b) Recovery error versus iterations for the ICT and two-step ICT methods. (c) Comparisons of their pseudo-Pareto curves (vertical:  $\|f - \Phi u^k\|_2^2$ , horizontal:  $\|u^k\|_1$ ). (d) A close-up of (c) in the dense trail part. (e) SNR versus iterations by the ICT, FICT, and two-step ICT methods. (f) Pseudo-Pareto curves.



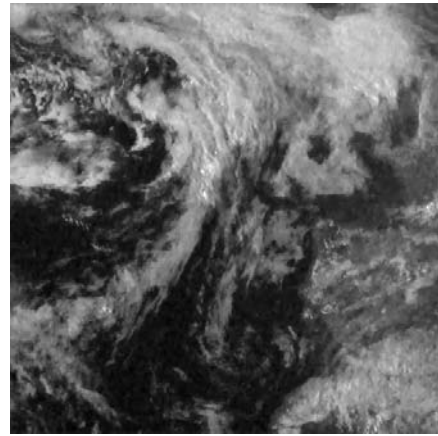
(a)  
SNR = 27.98 dB



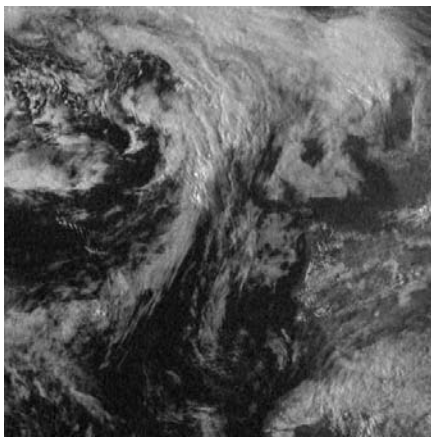
(b)  
SNR = 29.88 dB



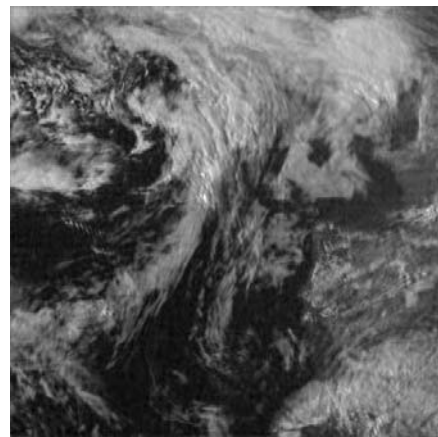
(c)  
SNR = 39.89 dB



(d)  
SNR = 36.47 dB



(e)



(f)

Figure 3: Fourier-domain CS imaging. (a) Original scene. (b) Sampling model. (c) Direct reconstruction. (d) Recovery by iterative wavelet-TV regularization. (e) and (f) ICT with 40 iterations and Bregman-ICT with 7 iterations. The coordinates denote  $512 \times 512$  spatial sampling.

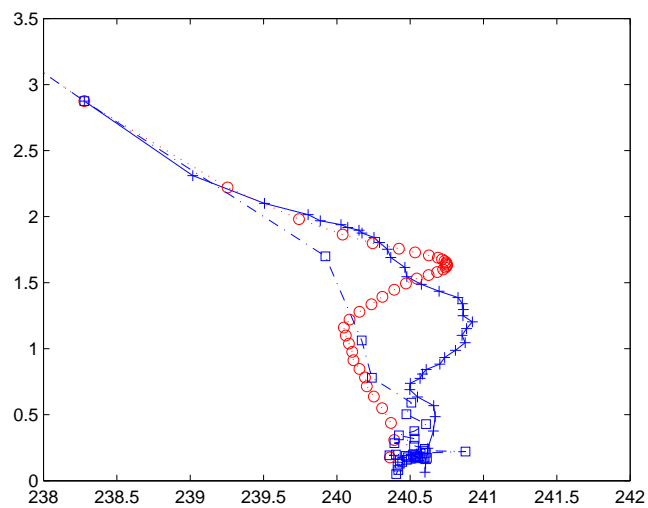


Figure 4: Pseudo-Pareto curves for spatial-domain IST (solid cross line), coefficient-domain IST (dotted circle line), and Bregman-IST (dot-dashed square line). All methods display 40 iterations. The vertical coordinate denotes  $\|f - \Phi u^k\|_2^2$  and the horizontal coordinate denotes  $\|u^k\|_1$ .

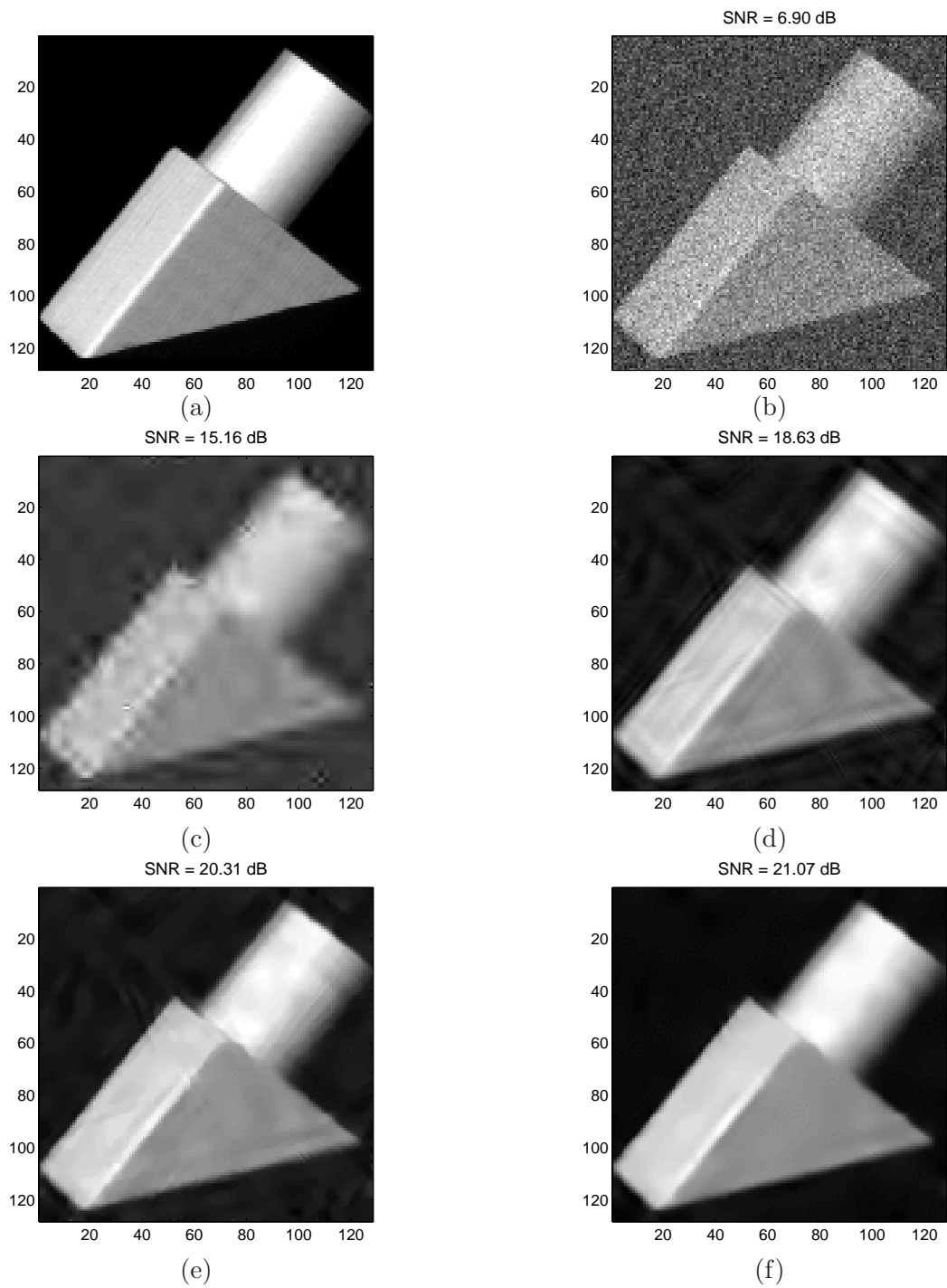


Figure 5: Denoising of a piecewise smooth object. (a) Original image. (b) Noisy image. (c) Wavelet thresholding (Db4 wavelets). (d) Curvelet thresholding. (e) TV-curvelet thresholding. (f) ATV-curvelet thresholding.

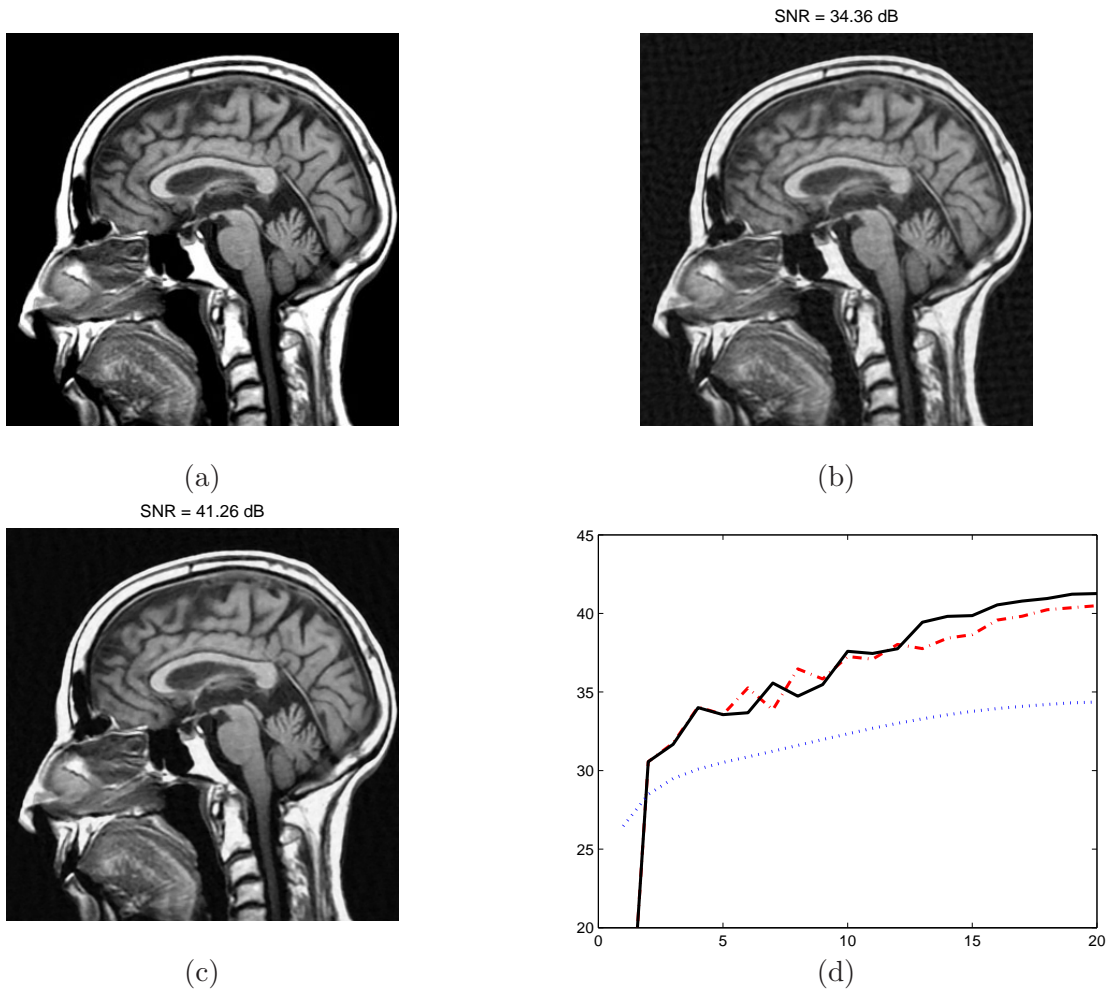


Figure 6: Decoding of CS for MRI brain image. (a) Original image. (b) ICT recovery. (c) ATVICT recovery. (d) Recovery SNR (vertical coordinate) versus the number of iterations (horizontal coordinate). The dotted line denotes ICT. The solid line denotes the proposed ATVICT with BBDY step sizes. The dot-dashed line denotes ATVICT with BB step sizes.

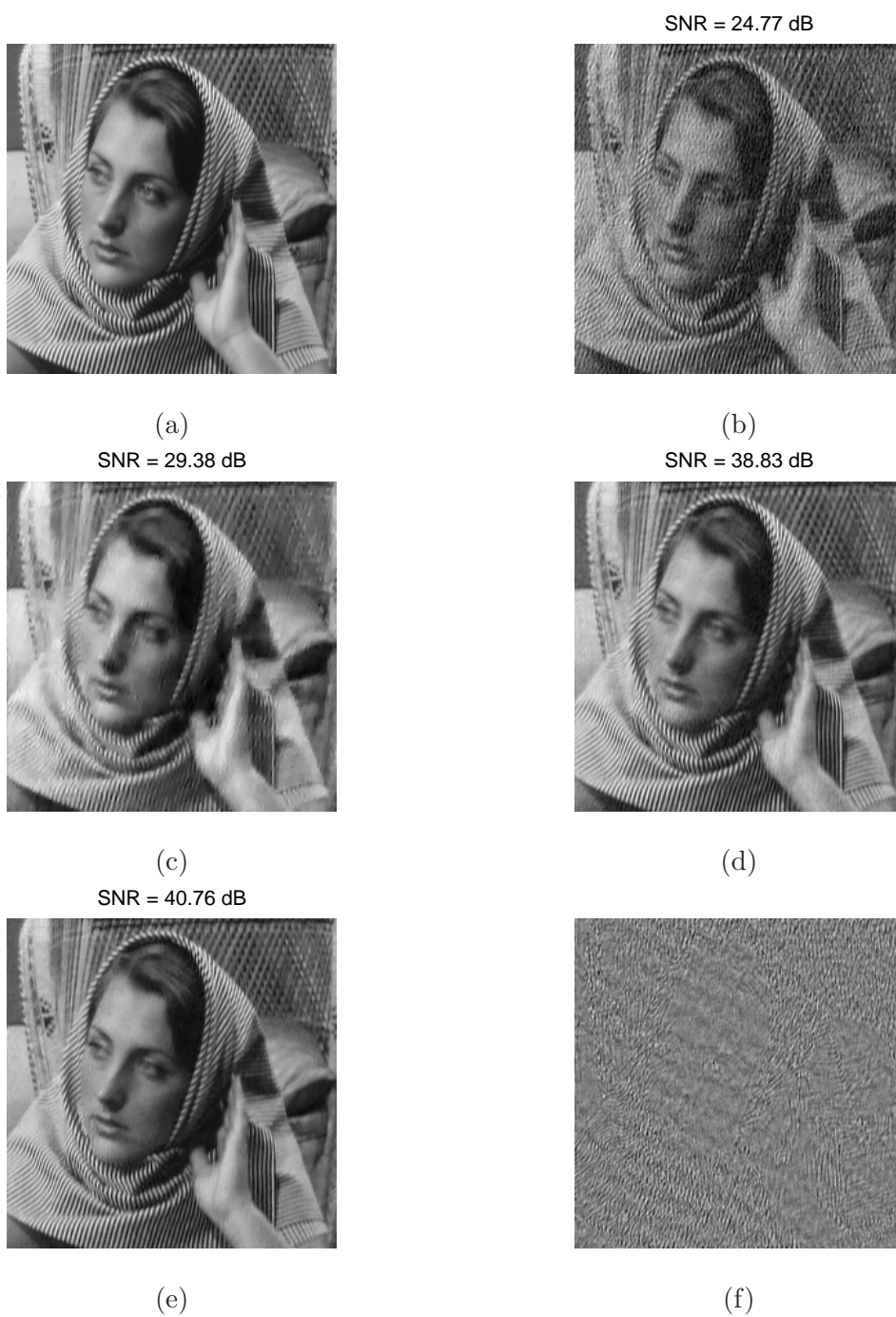


Figure 7: Decoding of CS for Barbara image by using different methods. (a) Original image. (b) Direct recovery. (c) TV- $\ell_1$  recovery. (d) ICT recovery. (e) ATVICT recovery. (f) The ATVICT recovery errors.

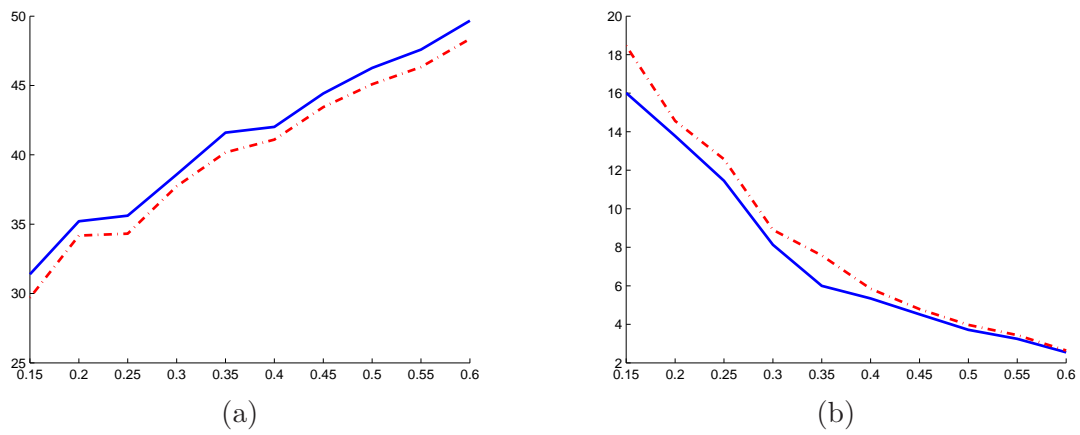


Figure 8: (a) Recovery SNR (vertical coordinate) versus the number of measurements. (b) Recovery error (vertical coordinate) versus the number of measurements. The horizontal coordinate denotes the ratio of number of measurements to total number of pixels. Solid line and dash-dotted line denote ATVICT and ICT, respectively.