A Sparsity Basis Selection Method for Compressed Sensing

Dongjie Bi, Yongle Xie, Xifeng Li, and Yahong Rosa Zheng, *Fellow, IEEE*

***Abstract*—This letter presents a new sparsity basis selection compressed sensing method (SBSCS) for improving signal reconstruction from compressed sensing (CS) measurements. Based on the observation that different classes of transform cause different sparsity expressions and better sparsity expression leads to better signal recovery, the proposed SBSCS method searches the best class of transform and basis in a set of redundant tree-structured dictionaries by nesting sparsity maximization within the CS minimization. The SBSCS method adaptively selects the class of transform and basis with the best sparsity measure at each ℓ^1 iteration and converges quickly to the final class of transform and basis. Numerical experiments show that the proposed SBSCS method improves the quality of signal recovery over the existing best basis compressed sensing method (BBCS) proposed by Peyré in 2010.**

***Index Terms*—Basis selection, compressed sensing (CS), sparsity maximization, sparsity.**

I. Introduction

COMPRESSED sensing introduces a new signal acquisition framework which uses a fixed set of linear measurements together with a nonlinear recovery process, and goes beyond the traditional Nyquist sampling paradigm [1], [2]. Considering an unknown signal $f \in \mathbb{R}^N$, compressed sensing uses a fixed set of $M \ll N$ linear measurements $y = \Phi f = \{\langle f, \phi_i \rangle\}_{i=1}^M \in \mathbb{R}^M$ to reconstruct the signal f, where Φ is the measurement matrix with $\phi_i \in \mathbb{R}^N$. The sparsity of f in an orthogonal basis $\mathcal{B} = \{\psi_j\}_{j=1}^N$ and $\psi_j \in \mathbb{R}^N$ can be measured by using the ℓ^0 norm $\|\Psi f\|_0$, where $\Psi f = \{\langle f, \psi_j \rangle\}_j$ with Ψ being the sparsifying operator. To guarantee sparse recovery, the compressed sensing theory requires the sensed signal f to be sparse in the given orthogonal basis $\mathcal{B}$ and the sensing matrix Φ to be incoherent with this basis.

Ideally, the optimally recovered signal $f^\star \in \mathbb{R}^N$ is the solution of the minimization $\|\Psi f\|_0$, subject to the equality constraint $y = \Phi f$. However the ℓ^0 norm minimization problem is combinational and intractable [3]. The common approach to the sparse signal recovery is to replace the ℓ^0 norm by the ℓ^1 norm and relax the equality constraint to deal with noisy measurements. The recovered optimal signal becomes

$$f^\star = \arg\min_{f \in \mathbb{R}^N} E(y, \mathcal{B}, t)$$
$$E(y, \mathcal{B}, t) = \frac{1}{2}\|\Phi f - y\|^2 + t\|\Psi f\|_1, \quad (1)$$

where $t \in \mathbb{R}_+$ is the *regularization parameter*.

If the orthogonal basis $\mathcal{B}$ is fixed, then the optimization problem (1) can be solved through an iterative procedure [3]. However, fixed bases are often inflexible in capturing the regularity of sounds or natural images. For instance, orthogonal wavelet bases define optimal approximations for classes of piecewise regular functions, but are inefficient in compressing regular edges; local cosine bases divide the time axis in segments that adapt to the local frequency content of a sound, but fail to capture transient parts of a sound [3], [4]. To improve the sparsity of complicated sounds or images, several orthogonal bases that compose a large dictionary of atoms [5]–[7] are often used.

In [1], the author proposes a best basis compressed sensing method (BBCS) to search for the best basis through a tree-structured dictionary with objective energy minimization. In contrast, our approach enhances the BBCS method in two aspects: one is that we consider different classes of transform Ψ which lead to different sparsity expression Ψf, and the best sparsity expression Ψf during the iteration of solving (1) leads to the best recovery of the signal $f^\star$; another aspect is that our method selects the basis based on sparsity maximization instead of energy minimization. Hence we name our algorithm Sparsity Basis Selection Compressed Sensing (SBSCS). The numerical experiments show that the SBSCS method converges quickly to the best class of transform and basis within a few tens of iterations and the quality of signal recovery outperforms the BBCS method.

II. Best Basis Compressed Sensing Method (BBCS)

Complex signals and natural images often include structures that require large redundant dictionaries [8], such as wavelet packet, local cosine, and bandlet orthonormal bases [9], to incorporate many signal patterns and improve sparsity approximation. The efficiency of sparsifying signal in different classes of wavelet transform, such as Haar, Coiflet and Daubechies, depends mostly on the regularity of f, the number of vanishing moments of ψ_j and its size of support [3]. When selecting the best basis, the effect of using different classes of transform are

Manuscript received April 01, 2015; accepted April 28, 2015. Date of publication May 05, 2015; date of current version May 11, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Negar Kiyavash.

D. Bi, Y. Xie, and X. Li are with the Department of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: dongjiebiee@gmial.com; xieyongle@uestc.edu.cn; xifenglee@126.com).

Y. R. Zheng is with the Department of Electrical & Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: zhengyr@mst.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2429748

often investigated. Let θ be the index of a specific class of transform to be used, and denote Θ as the set of θ. Tree dictionaries are constructed with a recursive split of orthogonal vector spaces and by defining specific orthonormal bases in each subspace. For 1D signal, a tree dictionary is a set $\mathcal{D}_\Lambda^\theta = \{\mathcal{B}_\lambda^\theta\}_{\lambda\in\Lambda}$ of orthogonal bases $\mathcal{B}_\lambda^\theta$ of $\mathbb{R}^N$. The atoms of $\mathcal{D}_\Lambda^\theta$ span subspaces $W_{j,i}^\theta$ of $\mathbb{R}^N$ for scale $0 \leq j \leq J = \log_2(N)$ and position $0 \leq i < 2^j$, that obey a refinement relationship $W_{j,i}^\theta = \oplus_{\epsilon=0}^1 W_{j+1,2i+\epsilon}^\theta$ with $j < J$. Each subspace $W_{j,i}^\theta$ has dimension $N/2^j$ and is equipped with one or several orthogonal bases $\mathcal{B}_{j,i}^\theta = \{\psi_{j,i,s}^\theta \backslash \forall 0 \leq s < \frac{N}{2^j}\}$.

The parameter λ that indexes a basis $\mathcal{B}_\lambda^\theta \in \mathcal{D}_\Lambda^\theta$ is a binary tree for a 1D signal or a quad-tree for a 2D signal. The set of nodes of λ is denoted as $\mathcal{N}(\lambda)$ and each node $(j,i) \in \mathcal{N}(\lambda)$ is located in the jth row and the ith column of the tree. For 1D signal, each interior node $(j,i) \in \mathcal{I}(\lambda) \subset \mathcal{N}(\lambda)$ has 2 children $\{(j+1,2i),(j+1,2i+1)\}$. The leaf nodes $(j,i) \in \mathcal{L}(\lambda)$ have no child.

Instead of using a fixed basis with fixed θ and λ, the BBCS method finds a parameter $\lambda^\star \in \Lambda$ adapted to the structures of the signal and uses the basis $\mathcal{B}_{\lambda^\star}^\theta$ with a fix class of transform θ. The BBCS method tries to solve the following optimization problem

$$(f^\star, \lambda^\star) = \underset{(f,\lambda)\in\mathbb{R}^N\times\Lambda}{\arg\min}\ \varepsilon\left(y, \mathcal{B}_\lambda^\theta, t\right)$$
$$\varepsilon\left(y, \mathcal{B}_\lambda^\theta, t\right) = \min_{f\in\mathbb{R}^N} \frac{1}{2}\|\Phi f - y\|^2 + t\left\|\Psi_\lambda^\theta f\right\|_1. \tag{2}$$

The BBCS method solves (2) through a 3-stage iterative process: updating the estimate of the signal, updating the best basis $\mathcal{B}_{\lambda^\star}^\theta$, and denoising the estimate of the signal. The updating best basis stage is a particular instance of the Classification and Regression Tree (CART) algorithm [10], [11], which first computes the cost function $\varepsilon(y, \mathcal{B}_{j,i}^\theta, t)$ for each subspaces $W_{j,i}^\theta$, then explores all tree nodes from bottom to top, and finds the leaf node $\mathcal{L}(\lambda)$ with the minimum cost. A best basis $\mathcal{B}_\lambda^\theta$ is then obtained by aggregating bases $\mathcal{B}_{j,i}^\theta$ for (j,i) that are leaves of λ [3].

III. Sparsity Basis Selection Compressed Sensing Method (SBSCS)

A. Sparsity Basis Selection Method For Compressed Sensing

Considering both sets of θ and λ, we try to solve the sparsity based optimization problem (3) instead of (2)

$$(f^\star, \lambda^\star, \theta^\star) = \underset{(f,\lambda,\theta)\in\mathbb{R}^N\times\Lambda\times\Theta}{\arg\max}\ S_P\left(\Psi_\lambda^\theta f_{\mathcal{B}_\lambda^\theta}\right)$$
$$f_{\mathcal{B}_\lambda^\theta} = \underset{f\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\Phi f - y\|^2 + t\left\|\Psi_\lambda^\theta f\right\|_1 \tag{3}$$

with $\mathcal{S}_P()$ representing the sparisty measure of the signal, which we choose to use the Gini index as detailed in Section III-B. For a fixed setting of $\lambda \in \Lambda$ and $\theta \in \Theta$, the optimization problem (3) degrades to (1) which can be solved through an iterative soft-thresholding algorithm by minimizing its surrogate function $E_{\mathcal{B},t}^{sur}(f; f_k) = \frac{1}{2}\|\Phi f - y\|^2 + t\|\Psi f\|_1 + \frac{1}{2}\|f - f_k\|^2 - \frac{1}{2}\|\Phi f - \Phi f_k\|^2$ during the iterations [12]–[15]. With $\|\Phi\|_2 \leq 1$, this surrogate objective function is a majorization of the original problem and minimization of this surrogate function leads to a majorization minimization (MM) algorithm [15]. This surrogate function $E_{\mathcal{B},t}^{sur}(f; f_k)$ has a unique minimizer $f_{k+1} = S_t(f_k + \Phi^T(y - \Phi f_k), \mathcal{B})$, with $S_t(f,\mathcal{B}) = \sum_m s_t(\langle f, \psi_m\rangle)\psi_m$ and $s_t(x) = \max(0, 1 - t/|x|)x$.

To choose the best class of transform and basis during the iterations of solving (3), the proposed SBSCS method also uses a 3-stage optimization process, as shown in Algorithm 1. The main differences between the SBSCS and BBCS methods are that the proposed SBSCS method uses an optimal basis selection process, as shown in Step 4 of Algorithm 1, to search for both the best class of transform $\theta^\star$ and the best basis $\lambda^\star$ during the iterations of ℓ^1 optimization.

Algorithm 1 The Sparsity Basis Selection Compressed Sensing (SBSCS) Method

---

1: **Initialization:** Choose $\lambda_1 \in \Lambda, \theta_1 \in \Theta, f_1 = 0$, iteration index $k = 1$.

2: **repeat**

3: **Update the estimate:** $\hat{f}_k = f_k + \Phi^T(y - \Phi f_k)$.

4: **Update the best class of transform and basis:**

$$(\lambda_{k+1}, \theta_{k+1}) = \underset{(\lambda,\theta)\in\Lambda\times\Theta}{\arg\max}\ \mathcal{S}_P(\Psi_\lambda^\theta \hat{f}_k).$$

This step is detailed in Section III-B.

5: **Denoise the estimate:** $f_{k+1} = S_t(\hat{f}_k, \mathcal{B}_{\lambda_{k+1}}^{\theta_{k+1}})$.

6: **Until** $\|f_{k+1} - f_k\| \leq \eta$.

The steps of Algorithm 1 are repeated until a user defined tolerance η is reached. For noisy measurements $y = \Phi f_0 + \omega$, the regularization parameter t is adjusted so that the final residual error satisfies $\|y - \Phi f^\star\| = \sqrt{M}\sigma_{noise}$, where ω is a Gaussian white noise of variance σ_{noise}^2. For noiseless measurements, Algorithm 1 is also applicable by reducing the value $t = t_k$ linearly to zero during the iterations in a fashion similar to the Morphological Components Algorithm (MCA) [16]. Similar to the BBCS method, the convergence of the SBSCS method is difficult to prove mathematically, but our extensive numerical experiments indicate that the SBSCS method always converges to a stable class of transform θ after a small number of iterations and the quality of the recovered signal gradually converges to a stable value after a larger number of iterations. Therefore, the computational complexity can be further reduced by stopping the optimal basis selection process after θ is converged to its final value which often takes 20-50 iterations. In comparison, the BBCS method uses a best basis search process via energy minimization during the iteration, which causes over-segmentation of the recovered signal [1]. Algorithm 1 uses the sparsity maximization instead of energy minimization, thus reducing the over-segmentation drawback of the BBCS method.

B. Optimal Basis Selection via Sparsity Maximization

Intuitively, a sparse representation is one in which a small number of coefficients contain a large proportion of signal energy [17]. The authors of [18] examine and compare several commonly-used sparsity measures quantitatively, based on intuitive and desirable attributes which they call *Robin Hood*, *Scaling*, *Rising Tide*, *Cloning*, *Bill Gate*, and *Babies*. Only two of these measures satisfy all these six attributes: the $pq-$ mean with $p \leq 1, q > 1$ and the Gini Index.

In this letter, the Gini Index is used to measure the sparsity of the signal $\sigma = [\sigma_1, \cdots, \sigma_N]$, with its elements re-ordered and represented by $\sigma_{[n]}$ for $n = 1, 2, \cdots, N$, where $|\sigma_{[1]}| \leq |\sigma_{[2]}|, \cdots, \leq |\sigma_{[N]}|$, then the Gini index is defined as [18]

$$\mathrm{GI}(\sigma) = 1 - 2\sum_{n=1}^{N} \frac{|\sigma_{[n]}|}{\|\sigma\|_1}\left(1 - \frac{n - 1/2}{N}\right). \tag{4}$$

Denote $\sigma(f, W_{j,i}^{\theta}) = \{\langle f, \psi_{j,i,s}^{\theta}\rangle\}$ as the projections of signal f on subspace $W_{j,i}^{\theta}$, and we concatenate the child vectors to form $\sigma_u(f, W_{j,i}^{\theta}) = [\sigma(f, W_{j+1,2i}^{\theta}), \sigma(f, W_{j+1,2i+1}^{\theta})]$. Since the vectors $\sigma(f, W_{j,i}^{\theta})$ and $\sigma_u(f, W_{j,i}^{\theta})$ have the same length and energy, the sparsity of these two vectors can be directly compared. Taking advantage of the tree-structured dictionaries, a fast optimal basis selection process is proposed to overcome this undesirable characteristic and quickly search the best class of transform $\theta^\star$ and basis $\lambda^\star$. If the GI Index of the root-node $\mathrm{GI}(\sigma(f, W_{j,i}^{\theta}))$ is greater than the child-node $\mathrm{GI}(\sigma_u(f, W_{j,i}^{\theta}))$, then node $(j, i) \in \mathcal{L}(\lambda^\star)$. The basis $\mathcal{B}_{\lambda^\star}^{\theta}$ is then obtained by aggregating bases $\mathcal{B}_{j,i}^{\theta}$ for (j, i) that are leaves of $\lambda^\star$. The optimal basis selection based on sparsity maximization is shown in Algorithm 2, where U is the size of Θ or the number of classes of transform, $\mathcal{S}_P()$ is selected as the Gini index.

Algorithm 2 Optimal Basis Selection via Sparsity Maximization: $(\bar{\lambda}, \bar{\theta}) = \underset{(\lambda,\theta)\in\Lambda\times\Theta}{\arg\max}\ \mathcal{S}_P(\Psi_\lambda^\theta f)$

```
Let S_P(Ψ_λ^θ f) be the Gini index
for 1 ≤ u ≤ U do
    for j = 0, ..., J − 1 do
        for i = 0, ..., 2^j do
            If GI(σ(f, W_{j,i}^{θ_u})) > GI(σ_u(f, W_{j,i}^{θ_u})) then
                Declare (j, i) as leaf (j, i) ∈ L(λ_u).
            else
                Declare (j, i) as interior (j, i) ∈ I(λ_u).
    B_{λ_u}^{θ_u} = [B_{j,i}^{θ_u}] : (j, i) ∈ L(λ_u).
  [GĪ, ū] = max(GI(Ψ_{λ_1}^{θ_1} f), ..., GI(Ψ_{λ_U}^{θ_U} f)).
  λ̄ = λ_ū, θ̄ = ū.
```

This optimal basis selection process searches the dictionaries from top to bottom, requiring decomposition of f onto each atom $\psi_{j,i,s}$ and signal sparsity computation GI on different classes of transform. The computational complexity of the algorithm is $O(UP)$, where P is the total number of atoms in $\mathcal{D}_\Lambda^\theta$. If implemented with the local cosine or wavelet packets, the algorithm has an overall computational complexity of $O(UN\log_2(N))$.

IV. Experimental Results

To evaluate the performance of the SBSCS method, we selected $U = 29$ different classes of local cosine and wavelet packet transform for CS. We used Sine and Trivial as the bell filters which determine the local cosine tree [3]. We used wavelet packet transform with different quadrature mirror filters, such as Haar, Beylkin, Coiflet, Daubechies, Symmlet, Vaidyanathan, Battle, and with different support and vanishing moments. We denote Coiflet_8 as the wavelet packet transform with Coiflet filter containing 8 vanishing moments. To generate compressed sensing measurements, we used a fast operator $\Phi f = (P_1 H P_2 f)\downarrow_M$ to sample the original fully-sampled data, where P_1 and P_2 are realizations of a random permutation of the N entries of a vector in $\mathbb{R}^N$, H is a 1D orthogonal Hadamard transform, and $\downarrow_M$ selects the M first entries of a vector.

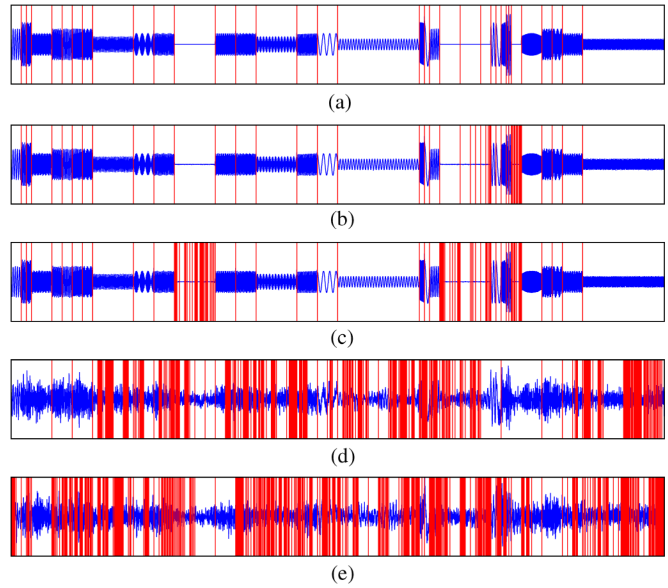

Fig. 1. (a) Synthetic signal with 30 random cosine atoms $N = 4096$, with best spatial segmentation λ_0. (b) Recovery using SBSCS method with Haar wavelet filter as the initial guess θ_1, $M = N/3$. (c) Recovery using BBCS method with local cosine basis with Sine filter, $M = N/3$. (d) Recovery using BBCS method with wavelet packet with Coiflet_8 filter, $M = N/3$. (e) Recovery using BBCS method with wavelet packet with Daubechies_8 filter, $M = N/3$.

As the first example, a synthetic sparse signal was generated as $f = (\Psi_{\lambda_0}^{\theta_0})^{-1} h$ using a random local cosine basis with Sine filter $\mathcal{B}_{\lambda_0}^{\theta_0}$ and a random signal of spikes h with $\|h\|_0 = 30$, as shown in Fig. 1(a). In the noiseless measurements $y = \Phi f$, both SBSCS and BBCS methods used a uniform basis λ_1 as the first guess, and the SBSCS used the Haar wavelet filter as the first guess of θ. Using the SBSCS method, the recovered signal $f^\star$, shown in Fig. 1(b), is nearly identical to f and $\theta^\star$ converged to the best class of transform (local cosine tree with Sine filter), and the final basis segmentation $\lambda^\star$ differed slightly from the best basis segmentation λ_0.

If the BBCS method chose the best class of transform as the first guess of θ, then the BBCS method performed well in signal recovery, but suffered more over-segmentation problem than the SBSCS method in the small-amplitude signal interval, as shown in Fig. 1(c). This was because the best basis updating of the BBCS method is sensitive to small changes in the signal, which leads to poor performance of signal recovery, while the proposed optimal basis selection method via sparsity maximization eases this problem. Furthermore, if the BBCS method chose the wrong class of transform, such as wavelet packet with Coiflet_8,

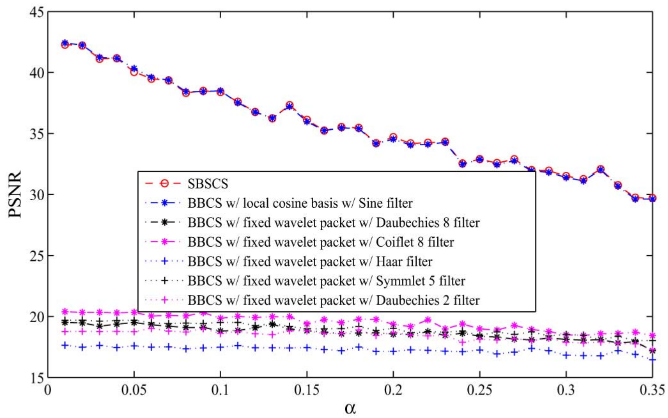

Fig. 2. PSNR results of the SBSCS and BBCS method used in a synthetic signal contaminated by different level of noise, with $\|\omega\| = \alpha\|\Phi f\|$.

Daubechies_8 filter as the θ, then the recovered signals suffered far more over-segmentation problem and resulted in worse signal quality, as shown in Fig. 1(d)–(e).

In addition, the proposed SBSCS method is particularly useful in dealing with noisy measurements. We added different level of noise ω to the synthetic signal with Signal to Noise Ratio (SNR) $1/\alpha$ and processed the CS signal recovery again. The Peak SNR (PSNR) results are shown in Fig. 2. It is easily seen that the SBSCS method performed equally well as the BBCS method when the best class of transform was chosen as the first guess of θ for the BBCS, but outperformed the BBCS when the wrong class of transform was initially chosen.

The second example used a sound of a tiger howling as shown in Fig. 3(a), with $y = \Phi f + \omega$ and $\|\omega\| = 0.03\|\Phi f\|$. In this example, both SBSCS and BBCS methods used a uniform basis as the first guess of λ_1, and the SBSCS method used local cosine basis with Sine filter as the first guess of θ_1. To demonstrate the performance of the proposed optimal basis selection process, we fix θ to a specific class of transform, as shown in Fig. 3(c), (e). Thus the SBSCS method reduces to a simple basis selection process just like the BBCS method which does not search through the different classes of transform. From the PSNR results in Fig. 3(c)–(f), we can see that the BBCS method suffered more over segmentation problem in $\lambda^\star$ than the proposed SBSCS method. Instead of using an energy function as the cost function to search for the best basis, which is sensitive to the noisy interference, the proposed optimal basis selection process was more robust against noisy interference and resulted in better $\lambda^\star$ basis segmentation.

When the optimal basis selection was used to search through the $U = 29$ different classes of transform, the SBSCS method searched the best class of transform $\theta^\star$ and basis $\lambda^\star$ through a much larger set of dictionaries. The SBSCS method achieved a higher PSNR than other methods, as shown in Fig. 4. Within the fist few iterations of the SBSCS method, the class of transform parameter θ changed from local cosine basis with Sine filter to the wavelet packet with Coiflet_10 filter after one iteration, then changed to the wavelet packet with Vaidyanathan filter after 27 iterations, and finally stayed at that transform afterwards. When the signal f is sparse in at least one class of transform $\theta \in \Theta$, the adaptability of the λ, θ during the iterations allowed the SBSCS method to find the best class of transform and basis.

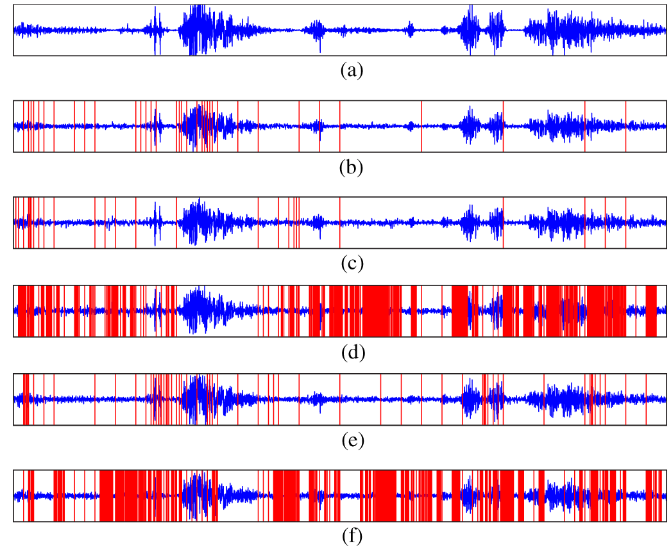

Fig. 3. (a) Sound of a tiger howling with $N = 32768$. (b) Recovery from $M = N/3$ using the SBSCS method with local cosine basis with Sine filter as the initial guess θ_1, achieved PSNR $= 22.36$ dB. (c) Recovery from $M = N/3$ using the SBSCS method with fixed wavelet packet with Daubechies_2 filter, achieved PSNR $= 21.01$ dB. (d) Recovery from $M = N/3$ using the BBCS algorithm with wavelet packet with Daubechies_2 filter, achieved PSNR $= 19.61$ dB. (e) Recovery from $M = N/3$ using the SBSCS method with fixed local cosine basis with Sine filter, achieved PSNR $= 21.48$ dB. (f) Recovery from $M = N/3$ using the BBCS method with local cosine basis with Sine filter, achieved PSNR $= 21.46$ dB.

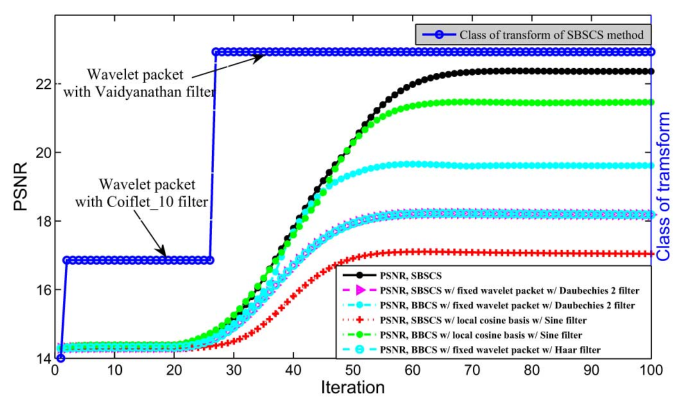

Fig. 4. PSNR convergence of the SBSCS and BBCS method used in sound of a tiger howling.

V. Conclusion

In this letter, a sparsity basis selection compressed sensing method, called SBSCS, is proposed for improving the reconstruction from compressed sensing measurements. The proposed SBSCS method takes advantage of the fact that different classes of transform lead to different sparsity expressions, and embeds a fast optimal basis selection via sparsity maximization in the CS iterations to search for the best class of transform and basis. Numerical results show that the proposed SBSCS method converges fast to the optimal transform and basis within $20 \sim 50$ iterations, and outperforms the best basis compressed sensing method (BBCS)[1] in terms of the quality of recovered signals.

REFERENCES

[1] G. Peyré, "Best basis compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2613–2622, May 2010.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory.*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd Ed ed. New York, NY, USA: Elsevier, 2009.

[4] E. Candès and D. Donoho, "New tight frames of curvelets and optimal representations of objects with C^2 singularities," *Commun. Pure Appl. Math.*, vol. 57, no. 2, pp. 219–266, Feb. 2004.

[5] M. Aharon, "K - SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[6] J. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[7] R. Carrillo, J. McEwen, D. V. D. Ville, J. Thiran, and Y. Wiaux, "Sparsity averaging for compressive imaging," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 591–594, Jun. 2013.

[8] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Ann. Statist.*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[9] E. L. Pennec and S. Mallat, "Bandelet image approximation and compression," *SIAM Multiscale Model. Simul.*, vol. 4, no. 3, pp. 992–1039, Jul. 2005.

[10] R. R. Coifman and M. V. Wickerhause, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. IT-38, no. 2, pp. 713–718, Mar. 1992.

[11] D. L. Donoho, "Cart and best-ortho-basis: A connection," *Ann. Statist.*, vol. 25, no. 5, pp. 1870–1911, 1997.

[12] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[13] K. K. Herrity, A. C. Gilbert, and J. A. Tropp, "Sparse approximation via iterative thresholding," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, May 2006.

[14] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, "A wide-angle view at iterated shrinkage algorithms," in *Proc. SPIE 6701, Wavelets XII*, Sep. 2007.

[15] M. A. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 1057–7149, Dec. 2007.

[16] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Adv. Imag. Electron Phys.*, vol. 132, 2004.

[17] D. Zonoobi, A. A. Kassim, and Y. V. Venkatesh, "Gini index as sparsity measure for signal reconstruction from compressive samples," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 927–932, Sep. 2011.

[18] N. Hurley, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, Oct. 2009.