Efficient Minimization for Dictionary Based Sparse Representation and Signal Recovery *

Xiaojing Ye
School of Mathematics,
Georiga Institute of Technology
Atlanta, GA 30332, USA
xye33@math.gatech.edu

Kefei Liu
Department of Electronic Engineering, City University of Hong Kong
Kowloon, Hong Kong, China
kefeilau@gmail.com

Meizhu Liu
Department of Computer and Information Science and Engineering, University of Florida
Gainesville, FL 32603, USA
mliu@cise.ufl.edu

ABSTRACT
This paper provides an efficient minimization algorithm for dictionary based sparse representation and its application in some signal recovery problems. Dictionary has shown great potential in effectively representing various kinds of signals sparsely. However the computational cost associated with dictionary based sparse representation can be tremendous, especially when the representation problem is coupled with the complex encoding processes of the signals. The proposed algorithm tackles this problem by alternating direction minimizations with the use of Barzilai-Borwein's optimal step size selection technique to significantly improve the convergence speed. Numerical experiments demonstrate the high efficiency of the proposed algorithm over traditional optimization methods.

Categories and Subject Descriptors
G.1.6 [**Optimizations**]: Nonlinear Programming; I.4 [**Image Processing and Computer Vision**]: Miscellaneous

General Terms
Algorithms

1. INTRODUCTION
1.1 Sparse Representation and Dictionary
Sparse representation is an important research topic in modern signal/image processing community. Given a set of n discrete signals $u_i \in \mathbb{R}^M$, it is usually desirable to find a sparse representation using $D \in \mathbb{R}^{M\times N}$ such that

$$u_i = Dc_i, \quad i = 1, \cdots, n, \tag{1}$$

where the representation coefficients c_i are sparse in the sense that the support of each c_i has much less cardinality than N, namely, there are only few nonzero entries in each c_i. Hence, this system of equations in (1) implies that any u_i in the sample set is a linear combination of *only few* columns in D.

Sparse representation has extensive applications in signal processing. For instance, it can be used to compress signals as locations and values of those *few* coefficients are suffice to recover the original signal provided the representation matrix D [1]. It can also be utilized to classify signals based on their corresponding representation coefficients [19, 20]. In addition, sparse representation can also be used to remove background noises as the main features should be sparsely represented already [13, 23, 6]. There have been a number of successful examples showing the efficiency of sparse representations, using tools such as discrete cosine transforms, wavelets, curvelets and etc. See, for example, [1, 13, 8, 14, 10] and references therein.

Dictionary is an advanced tool for sparse representation, and has proved its great potential in various applications. It attracts much more attentions these years due to the recent boost in compressed sensing which also requires effective tools to explore the intrinsic sparsity of underlying signals. Unlike traditional orthogonal and universal sparsifying transforms (e.g. wavelets), dictionary is usually overcomplete ($M < N$) and can be trained to absorb features from the learning samples, and hence it can more adaptively represent the objective data.

Due to the great importance of dictionaries, there have been a number of methods proposed for training dictionaries from given sample set in recent years. See, e.g. [18, 17, 1]. The resulting dictionary can then sparsely represent every sample, as well as new samples that have similar features to those in the original sample set. The general formulation of these methods is as follows,

$$\min_{D,c_i} \sum_{i=1}^{n} \|u_i - Dc_i\|^2, \quad \text{s.t. } \|c_i\|_0 \le T_0, \quad i = 1, \cdots, n, \tag{2}$$

ISABEL '11, October 26-29, Barcelona, Spain.
Copyright©2011 ACM ISBN 978-1-4503-0913-4/11/10... $10.00

where $\|\cdot\| \equiv \|\cdot\|_2$ is the regular Euclidean norm, and T_0 is a prescribed sparsity level. With properly chosen parameters, (2) is supposed to return a dictionary D that can sparsely represent all vectors u_i in the give sample set. The solution to (2) provides an effective tool to find sparse representation of any given type of signals, even if they are not that sparse in conventional transform domains such as wavelets. Note that the practical way of processing this minimization (2) varies in different methods. In this paper, we chose K-SVD algorithm [1] which proves to be a very robust and efficient method for training dictionaries. K-SVD algorithm alternately updates D and c when solving (2). In particular, it applies the singular value decomposition (SVD) of the residual matrices to update the columns of D in order. For more details about K-SVD algorithm, we refer readers to [1].

1.2 Application in Signal Recovery

In real applications, a signal is usually encoded into f, which is the observed data contaminated by some noises. The encoding process can usually be described as a linear transform via a sensing matrix H, and the noises γ are usually assumed to be additive Gaussian. Hence, the data acquisition process is formulated as follows,

$$f = Hu + \gamma. \tag{3}$$

If u itself is a sparse signal, it can be recovered by solving the minimization problem

$$\min_u \left\{ \alpha\|u\|_0 + \frac{1}{2}\|Hu - f\|^2 \right\}, \tag{4}$$

where $\|\cdot\|_0$ is 0-norm that counts the number of nonzeros in its argument. This is the routine of sparse signal recovery in compressed sensing [11, 7].

However, in most applications, the signal u is not sparse in its space domain, but rather in the frequency domain (e.g. Fourier, cosine transform), or under certain transform (e.g. wavelet). In case it is not sparse in either case, one can still train a dictionary D from a set of samples that are similar to u using (2), and then use this D to sparsely represent the signal: provided this dictionary D, we can enforce the sparsity of representation coefficient c by minimizing

$$\alpha\|c\|_0 + \frac{1}{2}\|Dc - u\|^2, \tag{5}$$

where $\alpha > 0$ is a balance term between the sparsity level and representation error. However, the 0-norm minimization in (5) is proved to be an NP-hard problem and hence not tractable. In practice, the 0-norm can be substituted by the convex 1-norm and this relaxed formulation is shown to be equivalent under certain conditions such as high sparsity level of c as well as some desirable properties of D. There are many references and fast algorithms proposed in this context. See, e.g. [7, 16, 5, 21, 12, 3, 15, 4, 25].

Utilizing (5) as the sparsity constraint for the underlying signal u, one can recover u by solving the following minimization problem

$$\min_{u,c} \left\{ \alpha \left(\beta\|c\|_0 + \frac{1}{2}\|Dc - u\|^2 \right) + \frac{1}{2}\|Hu - f\|^2 \right\}. \tag{6}$$

Here α weights the regularization term against the data fitting term, and β weights the sparsity of the representation coefficients against the representation error.

In most imaging applications, such as magnetic resonance imaging (MRI) reconstruction, it is usually not the images that have sparse representation, but the small patches in the images can be sparsely represented by a trained dictionary. In this case, one can solve for the image u by

$$\min_{u,c_j} \left\{ \alpha \sum_{j=1}^{J} \left(\beta\|c_j\|_0 + \frac{1}{2}\|Dc_j - R_j u\|^2 \right) + \frac{1}{2}\|Hu - f\|^2 \right\}, \tag{7}$$

where $R_j \in \mathbb{R}^{M\times P}$ is a binary matrix that extracts the j-th patch of size M from the image u consisting of P pixels in total, and c_j is the representation coefficient for this patch under the dictionary D [13, 26, 9, 22]. Nevertheless, methods proposed for (6) can be readily modified to use for (7).

2. PROPOSED METHOD

2.1 Algorithm Derivation

In this section, we propose an alternating direction minimization algorithm for solving the dictionary based signal recovery problem

$$\min_{u,c} \left\{ \alpha \left(\beta\|c\|_1 + \frac{1}{2}\|Dc - u\|^2 \right) + \frac{1}{2}\|Hu - f\|^2 \right\}, \tag{8}$$

with the relaxed convex 1-norm to enforce sparsity in coefficient c. As the objective function is convex with respect to both of u and c, it is guaranteed that an alternating minimization scheme leads to a global minimum from any starting point. However, conventional alternating scheme requires the minimizations of

$$\begin{cases} c^{k+1} = \arg\min_c \left\{ \beta\|c\|_1 + \frac{1}{2}\|Dc - u^k\|^2 \right\}, \\ u^{k+1} = \arg\min_u \left\{ \frac{\alpha}{2}\|Dc^{k+1} - u\|^2 + \frac{1}{2}\|Hu - f\|^2 \right\}, \end{cases} \tag{9}$$

which requires tremendous iterations in each subproblem (note that H may *not* have a specific form that makes the u-step directly solvable). In this paper, we propose to relax minimizations in each subproblem by the first order Taylor expansion of the the quadratic term with the penalization on the distance between the next iterate and the previous one. Namely, we substitute the representation term $(1/2)\cdot\|Dc - u^k\|^2$ by

$$\frac{1}{2}\|Dc^k - u^k\|^2 + \langle D^T(Dc^k - u^k), c - c^k\rangle + \frac{1}{2\delta_k}\|c - c^k\|^2, \tag{10}$$

where δ_k acts as a stepsize and controls the penalty of the distance between two concatenated iterates c^{k+1} and c^k. Note that this can be readily simplified to

$$\frac{1}{2\delta_k}\|c - (c^k - \delta_k D^T(Dc^k - u^k))\|^2, \tag{11}$$

by completing squares since the previous iterate c^k is merely a constant while minimizing with respect to c. Similarly, we can also substitute the data fitting term $(1/2)\cdot\|Hu - f\|^2$ by

$$\frac{1}{2\mu_k}\|u - (u^k - \mu_k H^T(Hu^k - f))\|^2, \tag{12}$$

where μ_k is the step size of the u-problem in the k-th iteration. Then the minimization scheme (9) becomes

$$\begin{cases} \bar{c}^k = c^k - \delta_k D^T(Dc^k - u^k), \\ c^{k+1} = \arg\min_c \left\{ \beta\|c\|_1 + \frac{1}{2\delta_k}\|c - \bar{c}^k\|^2 \right\}, \\ \bar{u}^k = u^k - \mu_k H^T(Hu^k - f), \\ u^{k+1} = \arg\min_u \left\{ \frac{\alpha}{2}\|Dc^{k+1} - u\|^2 + \frac{1}{2\mu_k}\|u - \bar{u}^k\|^2 \right\}. \end{cases} \tag{13}$$

Here the intermediate steps for $\bar{c}^k$ and $\bar{u}^k$ can be computed directly. The correction steps c^{k+1} and u^{k+1} also have closed form solutions for the minimizations as follows.

First, c^{k+1} can be obtained by componentwise soft shrinkages

$$c_i^{k+1} = \max\{|\bar{c}_i^k| - \beta\delta_k, 0\} \cdot \mathrm{sign}(\bar{c}_i^k). \tag{14}$$

Second, by taking derivative with respect to u in the u^{k+1} step, one gets

$$u^{k+1} = (\alpha\mu_k + 1)^{-1}(\alpha\mu_k Dc^{k+1} + \bar{u}^k), \tag{15}$$

where α and μ_k are positive scalars so the inverse on the right hand side is well defined and trivial to compute.

It is clear that the algorithm (13) requires two operations on each of $D \in \mathbb{R}^{M\times N}$ and $H \in \mathbb{R}^{m\times M}$, as well as their transposes D^T and H^T. Other operations include soft shrinkages of complexity N, which are usually negligible compared to that of D and H.

2.2 Algorithm Acceleration

Given that the complexity in each iteration is low, it remains to improve the convergence rate of the proposed algorithm (13) to achieve high efficiency.

Conventional gradient descent algorithms require that the step size δ_k and μ_k have strict bounds to ensure the iterates converge to a fixed point. These bounds also apply to (13). It can be readily shown that

$$\delta_k \le \frac{1}{\|D^TD\|_2}, \quad \mu_k \le \frac{1}{\|H^TH\|_2}, \tag{16}$$

where $\|\cdot\|_2$ is the 2-matrix-norm, or the largest singular value of the matrix. However, these bounds are in some sense too conservative such that the convergence under the optimal cases, i.e. set δ_k and μ_k to these bounds in (16), is still very slow. Here, we propose to use the Barzilai-Borwein (BB) step size selection method [2] to compute δ_k and μ_k to improve the convergence rate.

BB step size mimics the inverse of the Hessian of the objective over the most recent step via a multiple of identity matrix in the following way:

$$\delta_k = \arg\min_\delta \|(c^k - c^{k-1}) - \delta D^T(Dc^k - Dc^{k-1})\|^2. \tag{17}$$

Hence δ_k can be explicitly computed by

$$\delta_k = \|D(c^k - c^{k-1})\|^2 / \|D^TD(c^k - c^{k-1})\|^2, \tag{18}$$

which is no smaller than the conservative bounds $1/\|D^TD\|_2$. However, this step size selection significantly improves the efficiency of the algorithm (13) and the iterates appear to converge empirically in all circumstances. We also apply this BB method to the update of μ_k by

$$\mu_k = \|H(u^k - u^{k-1})\|^2 / \|H^TH(u^k - u^{k-1})\|^2. \tag{19}$$

To sum up, we obtain the Algorithm 1 for solving the minimization problem (8).

Algorithm 1 Alternating Direction Minimization for (8)

Input D, H, f, α, β.
Initialize $c^{-1} = -1$, $u^{-1} = -1$, $c^0 = 0$, $u^0 = 0$, $k = 0$.
Main loop:

1. $\bar{c}^k = c^k - \delta_k D^T(Dc^k - u^k)$.
2. $\delta_k = \|D(c^k - c^{k-1})\|^2 / \|D^TD(c^k - c^{k-1})\|^2$.
3. $c_i^{k+1} = \max\{|\bar{c}_i^k| - \beta\delta_k, 0\} \cdot \mathrm{sign}(\bar{c}_i^k)$.
4. $\bar{u}^k = u^k - \mu_k H^T(Hu^k - f)$.
5. $\mu_k = \|H(u^k - u^{k-1})\|^2 / \|H^TH(u^k - u^{k-1})\|^2$.
6. $u^{k+1} = (\alpha\mu_k + 1)^{-1}(\alpha\mu_k Dc^{k+1} + \bar{u}^k)$.
7. Stop if converged, else set $k \leftarrow k + 1$ and go to 1.

3. NUMERICAL EXPERIMENTS

3.1 Comparison Algorithms

In this section, we test the performance of the proposed algorithm 1 on simulated data with comparison to the commonly used orthogonal matching pursuit (OMP) algorithm [24]. The OMP algorithm solves the minimization problem in the c-step in (9) with 1-norm replaced by 0-norm. Although an exact solution is not guaranteed, OMP can usually generate a very close approximation to a true solution efficiently. Here we use OMP to denote the algorithm (9) with the c-step solved by OMP algorithm and u-minimization replaced by those updates in (13) (the latter is a must as H can be of any form that make the minimization in u-step have no direct solution). Therefore the per-iteration computational cost for OMP is much higher than the proposed method, as the latter only involves componentwise shrinkage without any iterations as OMP does. In addition, to show the efficiency brought by BB method, we also compare the performance of the algorithm with δ_k and μ_k updated by BB method or kept constants as $1/\|D^TD\|_2$ and $1/\|H^TH\|_2$, respectively.

All the algorithms are implemented in MATLAB (R2010b) and tested under GNU/Linux operating system (kernel version 2.6.35) on a Lenovo ThinkPad laptop with Intel Dual Core 2.53GHz CPU and 3GB memory.

3.2 Data Simulation

We first generate a set of $n = 1,500$ random samples of length $M = 100$ generated by the build-in MATLAB function `randn`. Then we train a dictionary D of size 100×400 using the K-SVD algorithm [1]. Next, we generate a random sensing matrix H of size $m \times M$ with $m = 60$ and normalize its columns, and then randomly pick a sample as the ground

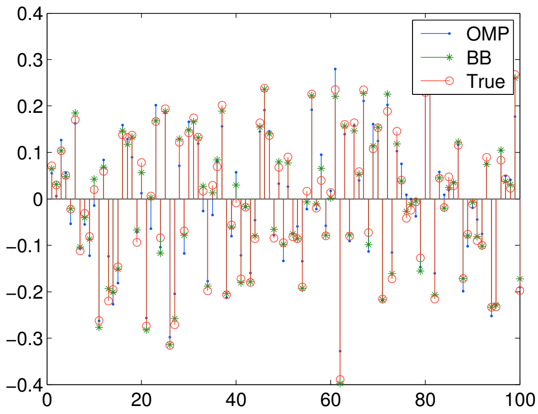

Figure 1: Recovered signals by OMP and BB in test 1.

truth signal u_0 from the $n = 1,500$ samples. Finally we obtain $f = Hu_0 + \gamma$ where γ is a white Gaussian noise with standard deviation 0.01.

We set the parameters $\alpha = 10^{-5}$ and $\beta = 10^{-3}$ which appear to give optimal results of all the tested algorithms. We also found that the proposed algorithm is not quite sensitive to these two parameters as the results do not change much when their values vary by a multiple of numbers between 0.1 and 10. The default termination criterion for all algorithms is that $\|u^k - u^{k-1}\|/\|u^k\|$, the relative change of the most recent two iterates, is less than the prescribed tolerance 10^{-5}.

3.3 Experimental Results

The recovered signals by OMP and the proposed algorithm are shown in Figure 1. The red circles show the original data u_0, and blue dots and green stars represent the recovered values by OMP and the proposed algorithm, respectively. It is clear that the signal recovered by the proposed algorithm is closer to the ground truth signal. The relative reconstruction error $\|u - u_0\|/\|u_0\|$ is 9.0% by the proposed algorithm 1, which is much smaller than 23.9% obtained by OMP. To show the efficiency of BB step size optimization technique, we also test the proposed algorithm with δ_k and μ_k set to constant values $1/\|D^T D\|_2$ and $1/\|H^T H\|_2$, respectively. Although these two bounds are the optimal choices for *constant* δ_k and μ_k with guaranteed convergence, BB method can easily outperform them by exhibiting much faster convergence rate as shown in Figure 2. As the per-iteration computational cost are almost identical with or without BB updates, Figure 2 demonstrates a significantly improved efficiency by BB updates as it requires much less iterations to reach the same level of relative error.

The second test replicates the previous one with a larger data dimension. In this test, we set $m = 150$, $M = 250$ and $N = 1,000$. The sample set consists of 15,000 vectors of length M. The simulated observation error γ is still white Gaussian with standard deviation 0.01. The parameters α, β, and termination criterion are all set to the same as in previous test.

The first 100 entries of the recovered signals (of length 250)

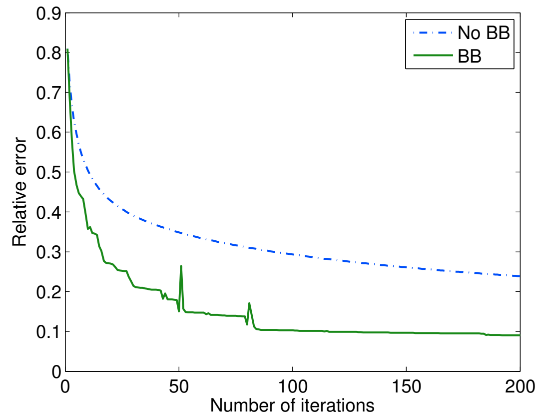

Figure 2: Relative reconstruction error with and without BB step size optimization in test 1.

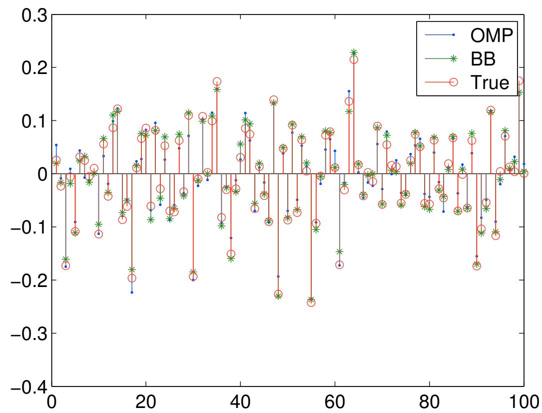

Figure 3: Recovered signals by OMP and BB in test 2. Only the first 100 of the total 250 entries are shown for each signal.

by OMP and the proposed algorithm are shown in Figure 3. Again the recovery by the proposed method (green star) appears to be closer to the ground truth (red circle) than that of OMP (blue dot). The relative error is 11.4% for the proposed algorithm and 18.4% for the OMP. Figure 4 shows the performance of algorithm (13) with and without BB updates. It can be seen again that BB method can significantly improve the effectiveness of the algorithm by using this quasi-Newton like gradient descent scheme.

4. CONCLUSIONS

In this paper, we developed a fast numerical algorithm for solving dictionary based signal recovery problem arising from sparse representation and compressed sensing. To tackle the computational difficulty associated to dictionary representation and arbitrary data acquisition process, we proposed to use an alternating direction minimization algorithm to split the problematic minimizations and make the scheme involve only few simple updates in each iteration. Moreover, the proposed algorithm benefits from the quasi-Newton like property by utilizing the Barzilai-Borwein's step size selection method. The numerical tests showed that the proposed algorithm has significantly improved efficiency and accuracy.

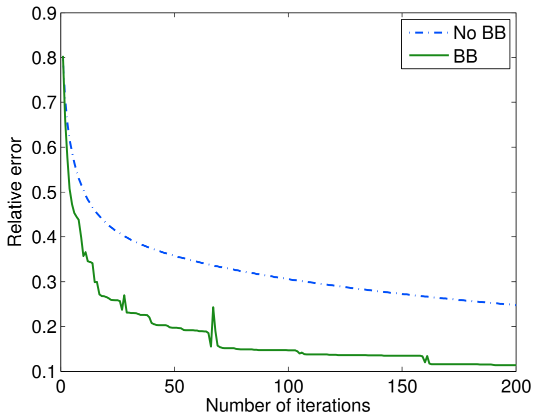

Figure 4: Relative reconstruction error with and without BB step size optimization in test 2.

5. REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[2] J. Barzilai and J. M. Borwein. Two point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[4] S. Becker, J. Bobin, and E. J. Candès. NESTA: a fast and accurate first-order method for sparse recovery. Technical report, Applied and Computational Mathematics, Caltech, Pasadena, CA, April 2009.

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

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[7] E. J. Candes, J. K. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[8] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7:319–335, 1998.

[9] Y. Chen, X. Ye, and F. Huang. A novel method and fast algorithm for mr image reconstruction with significantly under-sampled data. *Inverse Probl. Imag.*, 4(2):223–240, 2010.

[10] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete represenation in the presence of noise. *IEEE Trans. Inf. Theory*, 52:6–18, 2006.

[11] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 59(7):907–934, 2006.

[12] M. Elad. Why simple shrinkage is still relevant for redundant representations? *IEEE Trans. Inf. Theory*, 52(12):5559–5569, 2006.

[13] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15(12):3736–3745, 2006.

[14] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8):906–916, 2003.

[15] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Sign. Proces.*, 1(4):586–598, 2007.

[16] S. Kim, K. Koh, M. Lustig, and S. Boyd. An efficient method for compressed sensing. *IEEE Proc. Intl. Conf. on Image Process.*, 3:117–120, 2007.

[17] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neural Comp.*, 15:349–396, 2003.

[18] K. Kreutz-Delgado and B. D. Rao. FOCUSS-based dictionary learning algorithms. *Wavelet Applications in Signal and Image Process. VIII*, 4119:459–473, 2000.

[19] M. Liu, L. Lu, X. Ye, and S. Yu. Coarse-to-fine classification using parametric and nonparametric models for computer-aided diagnosis. In *Proceedings of ACM Conference on Information and Knowledge Management (CIKM)*, 2011.

[20] M. Liu, L. Lu, X. Ye, S. Yu, and M. Salganicoff. Sparse classification for computer aided diagnosis using learned dictionaries. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2011.

[21] Y. Nesterov. Gradient methods for minimizing composite objective function. CORE Discussion Papers 2007/76, UniversitÃľ catholique de Louvain, Center for Operations Research and Econometrics (CORE), Sept. 2007.

[22] S. Ravishankar and Y. Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging*, 30(5):1028–1041, 2011.

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

[24] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, 2007.

[25] S. J. Wright, R. D. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57(7):2479–2493, 2009.

[26] X. Ye, Y. Chen, and F. Huang. MR image reconstruction via sparse representation: Modeling and algorithm. *International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, pages 10–16, 2009.