2008 International Symposium on Nonlinear Theory and its Applications
NOLTA'08, Budapest, Hungary, September 7-10, 2008

Sparse representation of L-order Markov signals

Zhaoshui HE
Lab. for Advanced Brain Signal Processing
Brain Science Institute, RIKEN
Wako-shi, Saitama, 351-0198, Japan
School of Electronics and Information Engineering
South China University of Technology
Guangzhou, 510641, China
Email: he_shui@brain.riken.jp

Andrzej Cichocki
Lab. for Advanced Brain Signal Processing
Brain Science Institute, RIKEN
Wako-shi, Saitama, 351-0198, Japan
System Research Institute
Polish Academy of Sciences (PAN)
Warsaw, 00-901, Poland
Email: cia@brain.riken.jp

***Abstract*—In engineering applications, many signals are non-white (i.e., they are colored and have temporal structure) and can be modeled as an L-order Markov process. However, the existing sparse representation methods do not consider the L-order Markov property of signals. To fill this gap, we propose a new sparse representation framework: Firstly, we segment (split) the available T samples into several frames, where the length of each frame is $L(1 \leq L \leq T)$; secondly, to make the estimated signals be smooth, we set an appropriate percentage of overlapping between two neighboring frames (typically, 50%-70% overlapping); finally, we perform sparse representation for each frame. Under this framework, a modified-BP algorithm is developed by L-order ℓ_{pq}-norm-like optimization, which can indirectly exploit the L-order Markov property of sources and achieve better results.**

I. Introduction

Sparse representation has already found many applications in electromagnetic and biomagnetic problems (EEG/MEG), time-frequency representation, image processing, fault diagnosis, etc [1]–[3]. Mathematically, sparse representation can be modeled as:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t),\ t = 1, \cdots, T,\ \text{or}\ \boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{1}$$

where $\boldsymbol{X} = [\boldsymbol{x}(1), \cdots, \boldsymbol{x}(T)] \in \mathbb{R}^{m \times T}(T \gg m)$ is a given data (observation) matrix, $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n] \in \mathbb{R}^{m \times n}$ is a given full-row rank basis matrix, $\boldsymbol{S} = [\boldsymbol{s}(1), \cdots, \boldsymbol{s}(T)] \in \mathbb{R}^{n \times T}$ is an unknown matrix representing sparse sources or hidden components, T is the number of available samples, m is the number of observations, and n is the number of sources. The main objective is to find the sparse solutions (sparse sources) $\boldsymbol{S}$ satisfying equations (1). Matrix $\boldsymbol{S}$ is intended to be as sparse as possible. We assume that each column of the sparse solutions $\boldsymbol{S}$ has m or less nonzero elements [4].

Recently, many researchers have paid much attention on sparse representation due to its importance. The ℓ_1-norm was first used to measure the sparsity of sources [5]–[7] and the popular ℓ_1-norm methods include linear programming (LP) [5]–[8], shortest path decomposition (SPD) [7], [9], etc [10], [11]. Later, the ℓ_1-norm was extended to the $\ell^{(p \leq 1)}$-norm-like diversity [2], [4], [10]–[13], and various FOCal Underdetermined System Solvers (FOCUSS) were developed [10], [11]. After this, M-FOCUSS was further developed for multiple measurement vectors (MMV) [14], in which the cost function is $J_p(\boldsymbol{s}(l), l = 1, \cdots, L) = \sum_{i=1}^{n} (\sum_{l=1}^{L} |s_i(l)|^2)^{p/2}$.

It should be mentioned that many signals in practice are not white, but rather have temporal structure and usually can be modeled as an L-order Markov process. For example, speech signals usually can be well described as an L-order Markov model. However, the L-order Markov property of signals is not sufficiently discussed at all in sparse representation. Also, many existing algorithms, such as LP, SPD and FOCUSS etc, can not take advantage of this property.

In this paper, We develop several improved techniques for sparse representation of L-order Markov signals: 1) A new sparse representation framework is proposed; 2) Using power mean, we also propose a new optimization criterion, which can be considered to be a modified version of the objective function of M-FOCUSS [14]. By the proposed objective function, the estimated signals are smoother and usually more physically plausible. Moreover, a modified basis pursuit (BP) algorithm is proposed to optimize the objective function, which outperforms the conventional minimizing ℓ_1-norm methods, FOCUSS and M-FOCUSS in terms of precision.

II. Sparse representation by ℓ_{pq}-norm-like optimization

A. Probability model

Assume n original components $s_1, \cdots, s_n$ are mutually independent[1] and their absolute values $|s_1|, \cdots, |s_n|$ follow one-sided generalized Gaussian distribution (one-sided GGD):

$$\Pr(\rho; p, \beta) = \frac{p}{\beta \Gamma(1/p)} \exp\left(-\left(\frac{\rho}{\beta}\right)^p\right),\ \ \rho \geq 0;\ \beta > 0, \tag{2}$$

where $0 < p \leq 2$ and $\Gamma(\cdot)$ is *Gamma function* given by $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$. In practical applications, many real-world signals follow or approximately follow the above distribution [15], [16]: for example, generalized Gaussian signals (with symmetrical PDF) and nonnegative signals (with

[1]For notational simplicity, the matrix model (1) $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}$ is written in vector format: $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$.

asymmetrical PDF). As mentioned in [15], [16], p is the shape parameter, and β is the scaling parameter. Generally speaking, the smaller the parameter p is, the sparser the corresponding signal is.

From (2), the joint PDF of $|s_1|, \cdots, |s_n|$ can be expressed as

$$\begin{aligned} \Pr(|s_1|, \cdots, |s_n|) &= \prod_{i=1}^{n} \frac{p}{\beta\Gamma(1/p)} \exp\left(-\left|\frac{s_i}{\beta}\right|^p\right) \\ &= \left(\frac{p}{\beta\Gamma(1/p)}\right)^n \exp\left(-\frac{1}{\beta^p}\sum_{i=1}^{n}|s_i|^p\right). \end{aligned} \tag{3}$$

Maximizing a *posteriori probability* (MAP) in (3), we have ℓ_p-norm optimization problem:

$$\begin{cases} \min J_p(\boldsymbol{s}) = \min \|\boldsymbol{s}\|^p = \min \sum\limits_{i=1}^{n} |s_i|^p \\ \text{subject to}: \boldsymbol{A}\boldsymbol{s} = \boldsymbol{x} \end{cases} \tag{4}$$

B. *L-order ℓ_{pq}-norm-like optimization*

The optimization problem (4) has been extensively discussed [4], [10], [11]. It is worth mentioning that many real signals (e.g., speech signals) usually have L-order Markov property. However, the L-order Markov information is not considered at all in problem (4).

For this reason, instead of (4), we consider the following L-order ℓ_{pq}-norm-like optimization problem:

$$\begin{cases} \min J(\cdot; p, q) = \min\limits_{\boldsymbol{s}(l),\, l=1,\cdots,L} \sum\limits_{i=1}^{n} \left(\sum\limits_{l=1}^{L} |s_i(l)|^{pq}\right)^{\frac{1}{q}} \\ \boldsymbol{A}\boldsymbol{s}(l) = \boldsymbol{x}(l),\ l = 1, \cdots, L \end{cases} \tag{5}$$

where $q \geq 1$. The *power mean* (PM) is implicitly incorporated in (5) [17], [18].

Remark 1. *For L nonnegative numbers $y_1, \cdots, y_L$, their q-PM is $M_q(y_1, \cdots, y_L) = \left(\sum_{l=1}^{L} y_l^q / L\right)^{1/q}$, where the 1-PM $M_1(y_1, \cdots, y_L)$ corresponds to the arithmetic mean $A(y_1, \cdots, y_L) = (y_1 + \cdots + y_L)/L$ and $\lim\limits_{q\to+\infty} M_q(y_1, \cdots, y_L) = \max\limits_{l=1,\cdots,L}\{y_l\}$. For any two numbers q_1 and q_2 such that $-\infty \leq q_1 \leq q_2 \leq +\infty$, we have $M_{q_1} \leq M_{q_2}$, where $M_{q_1} = M_{q_2}$ if and only if $y_1 = \cdots = y_L$ [17], [18]. The optimization problem (5) is obtained by substituting the arithmetic mean $A(|s_i(1)|^p, \cdots, |s_i(L)|^p)$ with the q-PM $M_q(|s_i(1)|^p, \cdots, |s_i(L)|^p)$ in (4).*

The minimum ℓ_1-norm sparse representation is the special case of problem (5) with the parameters $q = 1, p = 1$ or $L = 1, p = 1$ [6], [7], [9], [19]. For L-order Markov signals, more accurate and physically meaningful results than the conventional ℓ_1-norm estimations can be achieved by problem (5). For example, the estimated sources are usually smoother and more continuous, whereas the ℓ_1-norm estimations are relatively more random and without any constraints on the temporal structure.

To take advantage of the L-order Markov property of signals, the T samples in model (1) can be segmented into T/L frames with frame length L ($L \ll T$, e.g., $L = 4$). In addition, we propose setting an appropriate percentage of overlapping between two neighboring frames (typically, 50%-70% overlapping). Overlapping is helpful to make the estimated source signals be smoother. Finally, we perform sparse representation by for each frame. Next, we discuss how to set the parameters L, p, q and the overlapping percentage.

TABLE I
SOME FUNCTION VALUES OF $r(p)$

p	0	0.2	0.4	0.6	0.8	1
$r(p)$	0	0.062937	0.23155	0.35624	0.44079	0.5
p	1.2	1.4	1.6	1.8	2	2.2
$r(p)$	0.5431	0.57556	0.60068	0.62057	0.63662	0.64978

C. *Parameter setting*

For PDF model (2), we can obtain its moment estimation of parameter p as follows [15], [16]:

$$\hat{p}(\rho) = r^{-1}\left((E\rho)^2 / E\rho^2\right), \tag{6}$$

where $r^{-1}(\cdot)$ is the inverse function of $r(x) = [\Gamma(2/x)]^2 / [\Gamma(1/x) \cdot \Gamma(3/x)]$. $r(x)$ is monotonously increasing in the internal $(0, +\infty)$ (see Fig.1). So the parameter p can be uniquely estimated by solving nonlinear function (6) using the *bisection method*. Also, we can roughly estimate p by looking up Table I.

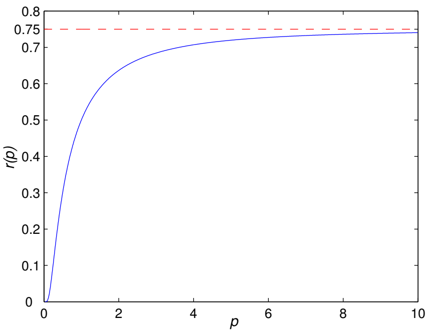

Fig. 1. $r(p)$ is monotonously increasing with p

In addition, by extensive simulations, we found that the parameters L, q and the overlapping percentage can be empirically set by calculating the autocorrelation coefficients of the observed signals. For example, for the observed signal $x_1(t), t = 1, \cdots, T$, we calculate the autocorrelation coefficients $\zeta(x_1(t), x_1(t-1)), \cdots, \zeta(x_1(t), x_1(t-L))$ between the observed signals and their time delays. If the autocorrelation coefficients $\zeta(x_i(t), x_i(t-L_0)),\ i = 1, \cdots, m$ are still significant (typically, larger than 0.02), it is better to set $L \geq L_0$. If all these autocorrelation coefficients are very significant, we can set a relatively big q (e.g., $q = 8$); typically we set $q = 2$. In many situations we can set $L = 4$ with 50% overlapping and much better performance can be achieved, especially when the sources are very sparse.

III. Sparse Representation by Modified-BP

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$ be $m \times n \, (m \le n)$ basis matrix and set $\mathcal{N} = \{\, \boldsymbol{A}_B | \, \boldsymbol{A}_B \text{ is an } m \times m \text{ submatrix of } \boldsymbol{A}.\}$. Obviously $\mathcal{N}$ contains C_n^m submatrices of $\boldsymbol{A}$. We have $\boldsymbol{As} = \boldsymbol{A}_B\boldsymbol{s}_B + \boldsymbol{A}_N\boldsymbol{s}_N = \boldsymbol{x}$, where $\boldsymbol{A}_B$ contains m columns $\boldsymbol{a}_{i1}, \cdots, \boldsymbol{a}_{im}$ of $\boldsymbol{A}$; $m \times (n-m)$ matrix $\boldsymbol{A}_N$ contains the corresponding remaining column vectors of $\boldsymbol{A}$; and $\boldsymbol{s}_B$ and $\boldsymbol{s}_N$ are rearranged according to the corresponding order. In addition, we assume that all $m \times m$ submatrices of $\boldsymbol{A}$ are invertible. Then, problem (5) can be converted to the following optimization problem:

$$\min_{\hat{\boldsymbol{s}}(l),\ l=1,\cdots,L} \sum_{i=1}^{n} \left(\sum_{l=1}^{L} |\hat{s}_i(l)|^{pq} \right)^{1/q}, \tag{7}$$

where $\hat{\boldsymbol{s}}(l) = [(\hat{\boldsymbol{s}}_B(l))^T, (\hat{\boldsymbol{s}}_N(l))^T]^T, \hat{\boldsymbol{s}}_B(l) = (\boldsymbol{A}_B^{(l)})^{-1}\boldsymbol{x}(l)$ and $\hat{\boldsymbol{s}}_N(l) = 0, l = 1, \cdots, L$, and $\boldsymbol{A}_B^{(l)} \in \mathcal{N}, l = 1, \cdots, L$.

Generally speaking, (7) is a *combinatorial* problem and is *NP-hard*. Fortunately, we can reduce the complexity considerably by choosing a small frame window, typically $L \le 4$, so the computation complexity would be acceptable.

The solution of problem (7) has other interesting sparsity properties: there are at most m nonzero entries in the vector $\hat{\boldsymbol{s}}(l)$. We can solve problem (7) by basis pursuit[2]. When parameter p is given, we have *modified-BP* algorithm for problem (5):

1) (Set parameter L, q, and overlapping ratio. Typically, $L \le 5$ to achieve reasonable computational complexity. As mentioned above, we split T samples of the observed signals $\boldsymbol{x}(t)$, $t = 1, \cdots, T$ into a series of frames having frame length L.
2) Compute inverses of all $m \times m$ submatrices of $\boldsymbol{A}$, generate set $\mathcal{N}$, and stack it. Similar to SPD [7], [9], when applied to all samples $t = 1, \cdots, T$, the inverses of all $m \times m$ submatrices of $\boldsymbol{A}$ need to be computed only once.
3) For each frame, search a solution satisfying (7) in $\mathcal{N}$.

The modified-BP algorithm can be viewed as the generalization of SPD [9]. In practice, parameter p is unknown. To solve this problem, we can first set $p = 1$ and run either above algorithm or alternatively FOCUSS to find a rough estimation $\hat{\boldsymbol{s}} = (\hat{s}_1, \cdots \hat{s}_n)^T$ for (4). Usually ℓ_1-norm solution is a first approximate (rough) solution [7] that but usually will be imperfect. To obtain a more accurate solution, we can estimate p from the rough estimation $\hat{\boldsymbol{s}}$ by equation (6) as follows:

$$\hat{p} = \frac{1}{n} \left(\hat{p}\left(\hat{s}_1\right) + \cdots + \hat{p}\left(\hat{s}_n\right) \right). \tag{8}$$

After this, we substitute $\hat{p}$ into (5) to more precisely estimate $\boldsymbol{s}$ using above algorithm. In very sparse and underdetermined cases, ℓ^p-norm solutions are usually more accurate than ℓ_1-norm ones.

[2]Since Chen and Donoho solved minimum ℓ_1-norm problem by basis pursuit [5], here we call above algorithm "modified basis pursuit (modified-BP)".

The algorithm is very fast and very precise, if the combination number C_n^m is not very large. We tested and compared these algorithms for many signals and for many basis matrices. Our experiments confirmed that the proposed algorithm almost always gives more precise solutions. Sometimes, its performance was remarkably better than that of the FOCUSS algorithm, and sometimes only slightly better.

However, if C_n^m is very large (e.g., C_{100}^{40}), the modified-BP needs a lot of storage space to stack C_n^m inverse matrices. In this case, FOCUSS, especially, M-FOCUSS is better.

IV. Numerical Experiments and Result Analysis

In this section, we compare the Modified-BP algorithm with FOCUSS [11], [20] and M-FOCUSS [14]. The experiment is conducted on a PC with an Intel Pentium 4 CPU 2.20GHz. The SIR (Signal-to-Interference Ratio) is calculated to evaluate the estimations:

$$\text{SIR}\,(\hat{s}) = 10 \log_{10}(s^2 \big/ (s-\hat{s})^2) \quad [\text{dB}]. \tag{9}$$

Usually, when SIR $\ge$ 15dB, the estimation is considered to be acceptable.

Example: The four sources (65536 samples in the time domain) are the same as those used in the experiment "FourVoices" in [9]. The mixing matrix was randomly generated as follows:

$$\boldsymbol{A} = \begin{pmatrix} 0.7798 & -0.3703 & 0.1650 & 0.5585 \\ -0.0753 & 0.8316 & 0.6263 & 0.3753 \\ -0.6215 & -0.4319 & 0.7619 & -0.7398 \end{pmatrix}. \tag{10}$$

Then, three mixtures were obtained by model (1). To satisfy the sparseness assumption, we performed sparse representation in the transform domain. As in [9], the short time Fourier transform (STFT) with Hanning window was used and the same parameters as the experiment "FourVoices" in [9] were taken: the window length of STFT was 2048, and hop distance was $d = 614$. In total, we had 425984 samples in the transform domain. After the sources were estimated in the transform domain, we also reconstructed the estimated sources in the same way as the experiment "FourVoices" in reference [9].

Here, we found that the signals were not white. By calculation, the autocorrelation coefficients were $\zeta(x_2(t), x_2(t-1)) = -0.6646$, $\zeta(x_2(t), x_2(t-2)) = 0.1836$, $\zeta(x_2(t), x_2(t-3)) = -0.0386$, $\zeta(x_2(t), x_2(t-4)) = 0.0327$ and $\zeta(x_2(t), x_2(t-5)) = -0.0082$. We set $L = 4$. Also we could determine $L = 4$ by computing the autocorrelation coefficients of two other observed signals, x_1, x_3, and their corresponding delays. Since these autocorrelation coefficients are not very significant, we set $q = 2$. Overlapping was set to 66.667%. By rough estimation, we found $\hat{p} = 0.1679$.

From Table II, we can see that the proposed algorithm considerably outperformed the ℓ_1-norm solutions and M-FOCUSS [14]. We performed many trials, in which the parameter p took values ranging from 0.1 to 1, and found that the modified-BP algorithm worked best with 75% overlapping and $p = 0.4$, which was a little larger than its true value 0.1679. When $p = 0.1679$, the SIRs were 15.9499dB, 17.5057dB, 16.8212dB, and 20.6749dB. M-FOCUSS obtained the best SIRs (14.6311dB, 16.1179dB, 15.5644dB and 19.6240dB) when the parameters were set to $L = 7, p = 0.7$.

TABLE II
COMPARISON OF PERFORMANCE FOR $\hat{p} = 0.1679$; NOISE IS FREE.

	SIRs (dB)			
ℓ_1-norm solutions	11.3125	12.7976	12.3106	16.3585
M-FOCUSS	13.3211	14.7750	13.3805	18.4096
The proposed algorithm	**15.4856**	**17.0709**	**16.3693**	**20.5713**

In this example, M-FOCUSS was faster than the modified-BP algorithm. For the modified-BP algorithm, It cost 2707.6 seconds. After about 50 iterations, M-FOCUSS found the solution. It took 705.1719 seconds.

Moreover, we also set $L = 1$ to re-test the modified-BP algorithm and compared it with FOCUSS, where the parameter p was taken value from 0.1 to 1. We found that the results were not as good as when $L = 4$. The SIRs of the estimations are shown in Table III.

TABLE III
COMPARISON OF PERFORMANCE OF FOCUSS AND THE MODIFIED-BP FOR $\hat{p} = 0.1679$ AND NOISE IS FREE.

p		1.0	0.8	0.6	0.4	0.2	0.1679
FOCUSS	SIR($\hat{s}_1$)	**11.31**	11.11	10.63	10.31	10.17	10.18
	SIR($\hat{s}_2$)	**12.80**	12.59	12.11	11.79	11.66	11.66
	SIR($\hat{s}_3$)	**12.31**	12.11	11.70	11.41	11.27	11.28
	SIR($\hat{s}_4$)	**16.36**	16.16	15.72	15.44	15.34	15.34
Modified-BP	SIR($\hat{s}_1$)	11.31	12.64	13.70	**13.93**	13.49	13.50
	SIR($\hat{s}_2$)	12.80	14.10	15.23	**15.49**	15.10	15.11
	SIR($\hat{s}_3$)	12.31	13.58	14.65	**14.85**	14.40	14.40
	SIR($\hat{s}_4$)	16.36	17.63	18.71	**18.98**	18.59	18.59

Next, we test the robustness of the modified-BP algorithm. Some white Gaussian noise was added to the mixtures with SNR=20dB. The parameter setting was the same as above: $L = 4, \hat{p} = 0.1679, q = 2$ and overlapping was 66.667%. Table IV shows the SIRs of the estimated sources, where we can see that the performance of the modified-BP algorithm is better than both M-FOCUSS and minimum ℓ_1-norm solution.

TABLE IV
COMPARISON OF PERFORMANCE FOR DIFFERENT ALGORITHMS $\hat{p} = 0.1679$, SNR=20DB.

	SIRs (dB)			
ℓ_1-norm solutions	10.9281	12.0371	11.6393	14.3458
M-FOCUSS	12.7266	13.5789	13.1990	15.7021
Modified-BP	**14.4028**	**15.0861**	**14.5127**	**16.5016**

V. CONCLUSIONS

Sparse representation of multiple vector measurement was discussed in this paper. A new framework was proposed for this problem: Firstly, the T samples are split into a serial of frames with frame length $L(1 \leq L \leq T)$; secondly, to make the estimated signals be smooth, we set an appropriate percentage of overlapping between two neighboring frames (typically, 50%-70% overlapping); finally, we perform sparse representation for each frame. The proposed framework is particularly suitable for L-order Markov signals. Under this framework, a modified-BP algorithm was developed. In addition, since the modified-BP algorithm is based on the BP, it cannot avoid combinatorial explosion when L is very large. So further research is required to develop more computationally efficient and faster algorithms for large scale problems in the future. Fortunately, we can achieve better results by modified-BP algorithm than minimum ℓ_1-norm solutions, even for a relatively small frame length L (e.g., $L = 3$ or $L = 4$).

REFERENCES

[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, June 1996.
[2] B. D. Rao, "Signal processing with the sparseness constraint," in *Proceedings of the ICASSP*, vol. III, Seattle, WA, 1998, pp. 1861–1864.
[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: John Wiley & Sons, 2003.
[4] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A recursive weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.
[5] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
[6] Y. Q. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Computation*, vol. 16, pp. 1193–1234, 2004.
[7] I. Takigawa, M. Kudo, and J. Toyama, "Performance analysis of minimum ℓ_1-norm solutions for underdetermined source separation," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 582–591, March 2004.
[8] Y. Q. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. L. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 423–437, February 2006.
[9] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001.
[10] B. D. Rao and I. F. Gorodnitsky, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Processing*, vol. 47, no. 1, pp. 187–200, 1999.
[11] K. Kreutz-Delgado, J. F. Murry, B. D. Rao *et al.*, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, pp. 349–396, 2003.
[12] I. F. Gorodnitsky, J. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4, pp. 231–251, October 1995.
[13] B. D. Rao and K. Kreutz-Delgado, "Deriving algorithms for computing sparse solutions to linear inverse problems," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, 1997, pp. 955–959.
[14] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.
[15] Z. S. He, S. L. Xie, and Y. L. Fu, "Sparsity analysis of signals," *Progress in Natural Science*, vol. 16, no. 8, pp. 879–884, August 2006.
[16] J. A. Molina and R. M. Dagnino, "A practical procedure to estimate the shape parameter in the generalized gaussian distribution," Technique report I-01-18_eng.pdf, 2002.
[17] P. S.Bullen, *The Power Means. Ch. 3 in Handbook of Means and Their Inequalities*. Dordrecht, Netherlands: Kluwer, 2003.
[18] Http://en.wikipedia.org/wiki/Generalized_mean.
[19] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, pp. 863–882, 2001.
[20] J. Murry, "Matlab codes of FOCUSS," http://dsp.ucsd.edu/˜jfmurray/software.htm.