Matrix of Polynomials Model based Polynomial Dictionary Learning Method for Acoustic Impulse Response Modeling

Jian Guan[1], Xuan Wang[1], Pengming Feng[2], Jing Dong[3] and Wenwu Wang[4]

[1]Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, 518055, China
[2]Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[3]Nanjing Tech University, Nanjing, 211800, China
[4]University of Surrey, Guildford, GU2 7XH, UK

`{j.guan, wangxuan}@cs.hitsz.edu.cn`, `p.feng2@newcastle.ac.uk`, `jingdong@njtech.edu.cn`,
`w.wang@surrey.ac.uk`

Abstract

We study the problem of dictionary learning for signals that can be represented as polynomials or polynomial matrices, such as convolutive signals with time delays or acoustic impulse responses. Recently, we developed a method for polynomial dictionary learning based on the fact that a polynomial matrix can be expressed as a polynomial with matrix coefficients, where the coefficient of the polynomial at each time lag is a scalar matrix. However, a polynomial matrix can be also equally represented as a matrix with polynomial elements. In this paper, we develop an alternative method for learning a polynomial dictionary and a sparse representation method for polynomial signal reconstruction based on this model. The proposed methods can be used directly to operate on the polynomial matrix without having to access its coefficients matrices. We demonstrate the performance of the proposed method for acoustic impulse response modeling.

Index Terms: polynomial dictionary learning, sparse representation, acoustic modeling, denoising

1. Introduction

Sparse representation aims to represent a signal by the linear combination of a few atoms from an overcomplete dictionary [1, 2]. The dictionary can be either pre-defined using Fourier basis or wavelet basis, or adapted from training data, using dictionary learning algorithms [2]. Dictionary learning methods usually employ a two-step alternating optimization strategy to learn a dictionary: the first step is sparse coding, which finds the sparse representation coefficients of a signal with a given dictionary [3–6]; and the second step is dictionary update, where the dictionary is updated to better fit the signal with the sparse representation coefficients found in the previous step [2, 7–10].

Conventional dictionary learning has been studied extensively, and used in a variety of applications, including speech denoising [2], and source separation [11–13]. However, it cannot be used directly to deal with the signals having time delays, such as convolutive (reverberant) signals or acoustic room impulse responses. In order to deal with such signals, we developed a polynomial dictionary learning technique in our recent work [14], where the polynomial matrix is employed to model the signals with time lags (e.g. room impulse responses). A polynomial matrix can be represented as a polynomial with matrix coefficients (so-called polynomial of matrices model) or alternatively a matrix with polynomial elements (i.e., the matrix of polynomials model) [15]. In [14], we converted the proposed polynomial dictionary learning model to the convention-al dictionary learning model based on the polynomial of matrices model, so that any conventional dictionary learning methods can be applied for obtaining sparse representation of the polynomial "signal". In [14], the K-SVD algorithm [7] and OMP algorithm [3] are employed for learning the dictionary and for calculating the sparse representation coefficients, respectively.

In this paper, we propose a polynomial MOD algorithm (P-MOD) and a polynomial OMP algorithm (POMP) based on the matrix of polynomials model, which are the extension of the MOD algorithm [8] and OMP algorithm, respectively. Different from the method in [14], the new methods can operate on the polynomial matrix directly without converting the polynomial model to a conventional model. We evaluate the proposed methods for acoustic impulse responses denoising, where the PMOD is used to learn the polynomial dictionary from a polynomial matrix modeled by acoustic impulse responses. Both the proposed POMP algorithm and OMP algorithm are applied for the "signal" reconstruction by using the learned dictionary. The proposed PMOD can obtain a better performance for acoustic signal denoising, when compared with the method in [14].

The remainder of the paper is organized as follows: Section 2 reviews the background and previous work about polynomial dictionary learning; Section 3 presents the proposed methods in details; Section 4 shows the simulation results; and Section 5 gives the conclusion and potential future work.

2. Background and previous work

2.1. Polynomial matrix

A polynomial matrix can be expressed as either a matrix with polynomial elements, or alternatively a polynomial with matrix coefficients [15]. The polynomial of matrices model of a $p \times q$ polynomial matrix $\mathbf{A}(z)$ can be represented as

$$\mathbf{A}(z) = \sum_{\ell=0}^{L-1} \mathbf{A}(\ell) z^{-\ell}, \qquad (1)$$

where $\mathbf{A}(\ell) \in \mathbb{C}^{p \times q}$ is the coefficients matrix at lag ℓ, and L is the length of a polynomial element. The Frobenius norm (F-norm) of $\mathbf{A}(z)$ is defined as

$$\|\mathbf{A}(z)\|_F = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{\ell=0}^{L-1} |a_{ij}(\ell)|^2}, \qquad (2)$$

where $a_{ij}(\ell)$ is the coefficient of $a_{ij}(z)$, and $a_{ij}(z)$ is the (i, j)th polynomial element of $\mathbf{A}(z)$.

2.2. Previous work

In our previous work [14], we have introduced a polynomial dictionary learning technique to deal with the signals having time delays. The aim is to learn a polynomial dictionary $\mathbf{D}(z) \in \mathbb{R}^{n \times K}$ from the training data $\mathbf{Y}(z) \in \mathbb{R}^{n \times N}$, which represents signals with time delays (such as acoustic impulse responses), so that the given "signals" can be represented by the learned dictionary $\mathbf{D}(z)$. The proposed model is given as follows

$$\mathbf{Y}(z) = \mathbf{D}(z)\mathbf{X}, \tag{3}$$

where $\mathbf{X} \in \mathbb{R}^{K \times N}$ is the sparse representation coefficients.

According to (1), (3) can be rewritten as

$$\sum_{\ell=0}^{L-1} \mathbf{Y}(\ell)z^{-\ell} = \sum_{\ell=0}^{L-1} \mathbf{D}(\ell)z^{-\ell}\mathbf{X}, \tag{4}$$

where $\mathbf{Y}(\ell) \in \mathbb{R}^{n \times N}$ and $\mathbf{D}(\ell) \in \mathbb{R}^{n \times K}$ are the coefficients matrices of $\mathbf{Y}(z)$ and $\mathbf{D}(z)$ at lag ℓ, respectively. From (4), it can be seen that $\mathbf{Y}(\ell)$ can be sparsely represented by $\mathbf{D}(\ell)$ with the same $\mathbf{X}$ for each lag ℓ, so that (3) can be converted to the conventional dictionary learning model as

$$\underline{\mathbf{Y}} = \underline{\mathbf{D}}\mathbf{X}, \tag{5}$$

where $\underline{\mathbf{Y}} \in \mathbb{R}^{nL \times N}$ and the new dictionary $\underline{\mathbf{D}} \in \mathbb{R}^{nL \times K}$ are obtained by concatenating the coefficient matrices of $\mathbf{Y}(z)$ and $\mathbf{D}(z)$ at all time lags, respectively, which are

$$\underline{\mathbf{Y}} = [\mathbf{Y}(0); \ldots ; \mathbf{Y}(\ell); \ldots ; \mathbf{Y}(L-1)], \tag{6}$$

$$\underline{\mathbf{D}} = [\mathbf{D}(0); \ldots ; \mathbf{D}(\ell); \ldots ; \mathbf{D}(L-1)]. \tag{7}$$

In [14], the K-SVD algorithm is employed to learn $\underline{\mathbf{D}}$, and the OMP algorithm is used to calculate $\mathbf{X}$ for the reconstruction of $\underline{\mathbf{Y}}$.

3. Proposed methods

In this section, we present methods for polynomial dictionary learning and sparse representation based on the matrix of polynomials model. More specifically, the MOD algorithm and the OMP algorithm are extended to the polynomial cases, for dictionary update and sparse approximation, respectively.

3.1. Polynomial MOD

The proposed polynomial dictionary learning method updates the dictionary by optimizing the following cost

$$\mathbf{D}(z)^{(n+1)} = \underset{\mathbf{D}(z)}{\operatorname{argmin}} \|\mathbf{Y}(z) - \mathbf{D}(z)\mathbf{X}^{(n)}\|_F^2, \tag{8}$$

where $\mathbf{X}^{(n)}$ is the sparse representation matrix at the nth iteration. Here, (8) can be seen as a polynomial least-squares problem, so that the same strategy as in MOD can be employed to solve (8), where the dictionary can be updated by calculating the least-squares solution of (8), which is

$$\mathbf{D}(z)^{(n+1)} = \mathbf{Y}(z)\mathbf{X}^{(n)^T}\left(\mathbf{X}^{(n)}\mathbf{X}^{(n)^T}\right)^{-1}. \tag{9}$$

By using the polynomial least-squares, the proposed polynomial MOD method can update the dictionary directly without operating on the coefficient matrices as in our previous work [14]. The proposed PMOD algorithm is given in Algorithm 1, and a polynomial sparse representation method is discussed next.

Algorithm 1 Polynomial MOD (PMOD)

Input: $\mathbf{Y}(z), I_n$
Initialization: $\mathbf{D}(z)^{(0)} = \mathbf{Y}(z)(:, 1 : K), I_n = 80$.
Iterations:
for $i = 1, \cdots, I_n$
    Sparse coding:
    Calculating $\mathbf{X}$ by using OMP as in [14].
    Polynomial dictionary update:
    Updating $\mathbf{D}(z)$ by using (9).
end for
Output: $\mathbf{D}(z)$ and $\mathbf{X}$

3.2. Polynomial OMP

As the aim of our work is to deal with the signals with time delays, once the polynomial dictionary $\mathbf{D}(z)$ is learned by our proposed methods, the sparse representation coefficients matrix $\mathbf{X}$ also needs to be calculated for the reconstruction of the polynomial matrix $\mathbf{Y}(z)$. Here, we present a polynomial sparse representation method by extending the OMP algorithm [3] to the polynomial case.

The polynomial sparse representation coefficients can be calculated by optimizing the following cost

$$\min_{\mathbf{x}} \|\mathbf{y}(z) - \mathbf{D}(z)\mathbf{x}\|_F^2$$
$$\text{subject to } \|\mathbf{x}\|_0 \le K_{max}, \tag{10}$$

where the "signal" $\mathbf{y}(z)$ is a column of $\mathbf{Y}(z)$, and K_{max} is the max number of non-zero elements in $\mathbf{x}$.

The proposed POMP algorithm employs the same strategy as the OMP algorithm to calculate the sparse representation coefficients $\mathbf{x}$ to approximate the "signal" $\mathbf{y}(z)$. The OMP algorithm calculates the sparse representation coefficients by iteratively selecting the best-matched atoms from the dictionary to approximate the signal, where the best-matched atom selected is the one that is most correlated with the residual at each iteration. However, the similarity measure between two polynomial vectors (e.g. polynomial residual and polynomial dictionary atom) cannot be directly achieved with the inner product between the residual and the atoms as in the original OMP.

In our proposed method, we calculate the distance between the polynomial residual and polynomial dictionary atom by using the F-norm as the similarity measure, so that the best-matched atom is the one that has the smallest F-norm error with the polynomial residual at each iteration, which can be formulated as follows

$$k_0 = \underset{k}{\operatorname{argmin}} \|\mathbf{d}_k(z) - \mathbf{r}(z)^{(j-1)}\|_F^2, \ k = 1, \cdots, K, \tag{11}$$

where $\mathbf{d}_k(z)$ is an atom in $\mathbf{D}(z)$, $\mathbf{r}(z)^{(j-1)}$ is the polynomial residual vector obtained at the $(j-1)$th iteration, and k_0 is the index of the selected atom at the jth iteration. Then, k_0 is added to the support set $S^{(j)}$, and the sub-dictionary $\mathbf{D}_{S^{(j)}}(z)$ is also updated by adding the selected atom $\mathbf{d}_{k_0}(z)$. Then, the representation coefficients can be updated by minimizing the following cost function

$$\min_{\mathbf{x}} \|\mathbf{y}(z) - \mathbf{D}_{S^{(j)}}(z)\mathbf{x}\|_F^2$$
$$\text{subject to } \|\mathbf{x}\|_0 \le K_{max}, \tag{12}$$

where (12) is a polynomial least-squares problem. Note that, according to (4) and (5), the coefficients $\mathbf{x}$ should satisfy the

linear combination between $\mathbf{y}(z)$ and $\mathbf{D}_{S(j)}(z)$ at all lags. So that, the coefficient can be updated by the least-squares solution of (12), which is

$$\mathbf{x}^{(j)} = (\underline{\mathbf{D}}_{S(j)}^T \underline{\mathbf{D}}_{S(j)})^{-1} \underline{\mathbf{D}}_{S(j)}^T \underline{\mathbf{y}}, \qquad (13)$$

where $\underline{\mathbf{y}}$ and $\mathbf{D}_{S(j)}(z)$ are constructed according to (6) and (7). Then, the polynomial residual can be updated as

$$\mathbf{r}^{(j)}(z) = \mathbf{y}(z) - \mathbf{D}_{S(j)}(z)\mathbf{x}^{(j)}. \qquad (14)$$

The proposed POMP algorithm is summarized in Algorithm 2.

Algorithm 2 Polynomial OMP (POMP)

Input: $\mathbf{y}(z)$, $\mathbf{D}(z)$, K_{max}
Initialization: residual $\mathbf{r}(z)^{(0)} = \mathbf{y}(z)$, $\mathbf{x} = 0$, $S^0 = \emptyset$, $\epsilon = 10^{-6}$.
Iteration:
for $j = 1, \ldots, K_{max}$
    Atom selection:
    $k_0 = \underset{k}{\mathrm{argmin}} \|\mathbf{d}_k(z) - \mathbf{r}(z)^{(j-1)}\|_F^2$, $k = 1, \cdots, K$
    Support set update: $S^{(j)} = S^{(j-1)} \bigcup \{k_0\}$
    Coefficient update: $\mathbf{x}^{(j)} = (\underline{\mathbf{D}}_{S(j)}^T \underline{\mathbf{D}}_{S(j)})^{-1} \underline{\mathbf{D}}_{S(j)}^T \underline{\mathbf{y}}$
    Residual update: $\mathbf{r}(z)^{(j)} = \mathbf{y}(z) - \mathbf{D}_{S(j)}(z)\mathbf{x}^{(j)}$
    Stopping rule: If $\|\mathbf{r}(z)^{(j)}\|_F^2 \leq \epsilon$, then $\mathbf{x}_{opt} = \mathbf{x}^{(j)}$, and break, else continue.
end for
Output: $\mathbf{x}_{opt}$

---

4. Simulations and results

In this section, we apply our proposed polynomial dictionary learning method to represent the signals with time delays, e.g. acoustic impulse responeses. Polynomial dictionaries are learned from an acoustic impulse responses modeled polynomial matrix. The learned dictionaries are then used to recover the noise corrupted acoustic impulse responses, where the method in [14] is employed as a baseline for learning the polynomial dictionaries for performance comparison. The performance is measured by the polynomial "signal" reconstruction error, which is defined as

$$R_{err} = \frac{\|\mathbf{Y}(z) - \hat{\mathbf{Y}}(z)\|_F^2}{\|\mathbf{Y}(z)\|_F^2}, \qquad (15)$$

where $\mathbf{Y}(z)$ is the original signal and $\hat{\mathbf{Y}}(z)$ is the reconstructed signal.

4.1. Acoustic impulse responses generation and modeling

The proposed method is evaluated on acoustic impulse response signals generated by the room image model [16] as in [14]. A total number of 1000 impulse responses are used as training signals, the length of each impulse response is 14400. By applying our proposed method, a 10×72000 polynomial matrix with 20 lags is designed to model the training acoustic signals, where each element of the polynomial matrix has 20 lags, which can be seen as a finite impulse response (FIR) with a length of 20 samples. Each test acoustic signal is split into 720 segments, where the length of each segment is set to be 20, so that the test signal can be modeled by a 10×72 polynomial matrix with 20 lags.

4.2. Results and analysis

First, we carried out an experiment to compare the performance between the proposed method and method in [14] for acoustic signal denoising, where the polynomial dictionaries were trained by PMOD and the method in [14], respectively. The number of iterations for training the dictionaries in both PMOD and the method in [14] were set as 80. The size of the dictionaries were designed to be the same, which was 10×400 with 20 lags. The learned dictionaries were used to recover the noise corrupted acoustic signal, where white Gaussian noise with zero-mean and unit variance, set at different levels of signal-to-noise (SNR) were added to the test signal. By applying the PMOD trained dictionary, both OMP and POMP were used to calculate the sparse representation coefficients for acoustic signal reconstruction, which were denoted by PMOD+OMP and PMOD+POMP, respectively. The sparsity for training the dictionaries and calculating the sparse representation coefficients was all set to be 3. The experiments were conducted for 20 realizations at each noise level, and the average reconstruction errors were given in Table 1.

Table 1: *Performance comparison in terms of Reconstruction Error ($\times 10^{-2}$) between the proposed method and the baseline method in [14] for acoustic signal denoising at different noise levels.*

Noise levels (dB)	-10	0	10	20	30
Method in [14]	249.71	37.19	15.62	15.45	15.43
PMOD+OMP	248.92	37.05	15.23	15.07	15.05
PMOD+POMP	228.63	37.26	20.27	20.16	20.15

From Table 1, we can see that the PMOD+OMP can obtain better performance than the other two methods for acoustic signal recovery when the noise is at the level from 0 dB to 30 dB, whereas the PMOD+POMP gives better recovery accuracy for SNR at -10 dB. We found that a larger dictionary tends to give better reconstruction accuracy for most noise levels tested by using all these methods when using the training polynomial matrices with same lags (e.g. the polynomial matrices modeled with 20 lags). Therefore, dictionaries with the size of 10×400 were trained. The proposed methods can achieve better reconstruction accuracy in this situation.

Then, another experiment was conducted to give an illustration of the acoustic signal denoising. The size of the learned dictionaries was the same as that in the previous experiment, and 5 dB white Gaussian noise was added to the test signal. Figure 1 shows the result of the polynomial dictionary learning methods for noisy acoustic signal reconstruction, using the method in [14], PMOD+OMP, and PMOD+POMP. From Figure 1, we can see that the denoising performance of the polynomial dictionary learning methods used are quite similar, all these methods can recover the source signal to a certain degree.

In order to show how the proposed polynomial dictionary learning method used for the sparse representation of the signals with time delays, we randomly selected two polynomial elements from the polynomial matrix used to model the acoustic signal in the last experiment. The selected polynomial elements can be seen as a polynomial sub-matrix modeled by two FIRs, which are two segments of the clean test acoustic impulses. Figure 2 shows the clean FIRs in (a), their corresponding noisy FIRs in (b), and the denoised FIRs by the method in [14] in (c), PMOD+OMP in (d), and PMOD+POMP in (e), respec-
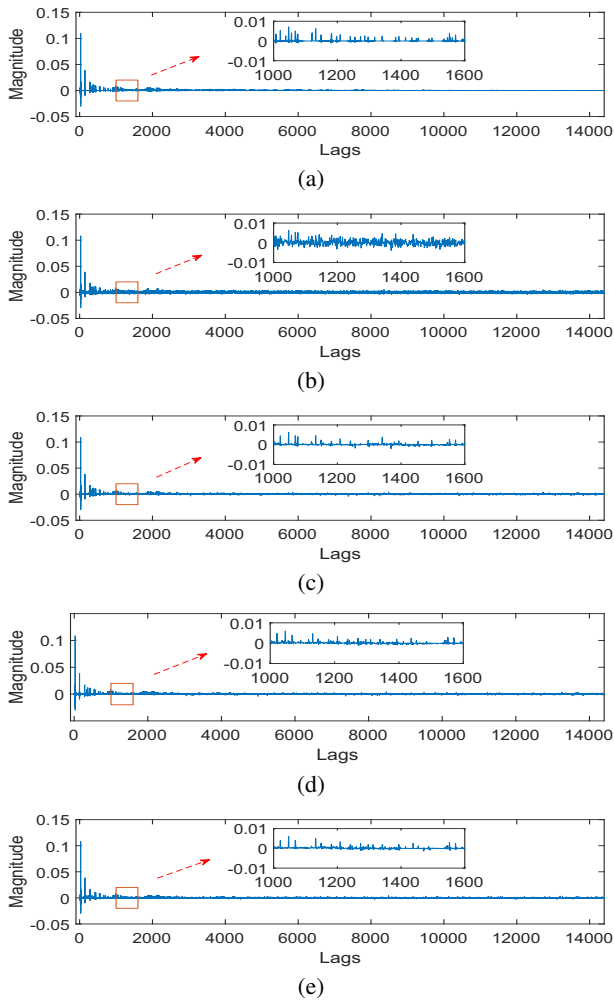
Figure 1: *Illustration of the polynomial dictionary learning methods for noise corrupted acoustic impulse reconstruction: Clean acoustic impulse response (a); Noisy acoustic impulse response (b); Reconstructed acoustic impulse response by the method in [14] (c); by PMOD+OMP (d); and by PMOD+POMP (e).*

Figure 2: *An example of polynomial dictionary learning methods for the polynomial matrix modeled acoustic impulse denoising, where two polynomial elements are randomly selected to illustrate the denoising performance: Clean FIRs (a); Noisy FIRs (b); Reconstructed FIRs by the method in [14] (c); Reconstructed FIRs by PMOD+OMP (d); Reconstructed FIRs by PMOD+POMP (e).*

tively. From Figure 2, we can see that both the proposed method and the method in [14] are capable of denoising the acoustic impulses.

5. Conclusions and future work

In this paper, we presented a new polynomial dictionary learning method based on the matrix of polynomials model to deal with the signals with time lags, such as acoustic impulse responses. As a byproduct, a polynomial OMP algorithm was also proposed to represent the signals with the learned polynomial dictionary. The experiments showed the proposed methods can achieve better performance for acoustic impulse denoising when compared with the baseline methods.

In our future work, we will further evaluate the two types of polynomial dictionary learning methods, e.g., for the polynomial matrix with different lags. Another potential future work is to extend the current polynomial dictionary learning model, by using a polynomial matrix as the sparse representation coefficient matrix.
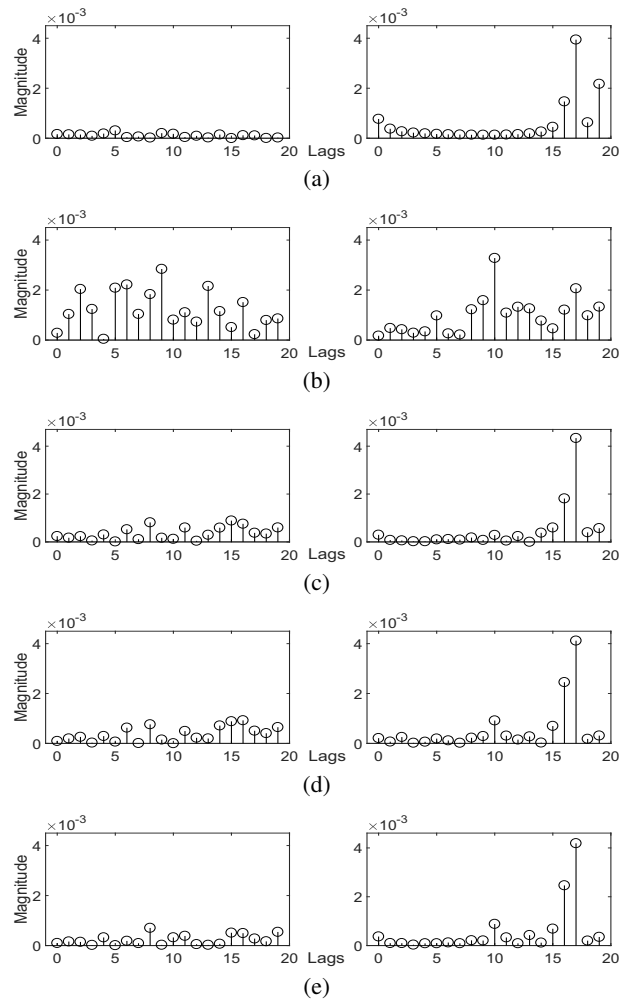
6. Acknowledgements

7. References

[1] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, 2003.

[2] M. G. Jafari and M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1025–1031, 2011.

[3] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of the 27th Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 40–44.

[4] 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*, vol. 51, no. 1, pp. 34–81, 2009.

[5] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[6] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[8] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5. IEEE, 1999, pp. 2443–2446.

[9] W. Dai, T. Xu, and W. Wang, "Simultaneous codeword optimization (SimCo) for dictionary update and learning," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6340–6353, 2012.

[10] Z. Sadeghipoor, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse decomposition: a new criterion and algorithm," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2013, pp. 5855–5859.

[11] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, no. 4, pp. 863–882, 2001.

[12] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, 2004.

[13] T. Xu and W. Wang, "Methods for learning adaptive dictionary in underdetermined speech separation," in *Proc. of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2011, pp. 1–6.

[14] J. Guan, J. Dong, X. Wang, and W. Wang, "A polynomial dictionary learning method for acoustic impulse response modeling," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science. Springer, 2015, vol. 9237, pp. 211–218.

[15] J. A. Foster, J. G. McWhirter, M. R. Davies, and J. A. Chambers, "An algorithm for calculating the QR and singular value decompositions of polynomial matrices," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1263–1274, 2010.

[16] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.