Sparse Representation Based Anomalies Detection in Electrocardiography Signals

Tomasz Andrysiak[(✉)]

Faculty of Telecommunications and Electrical Engineering, Institute of Telecommunications, University of Technology and Life Sciences (UTP), ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland
tomasz.andrysiak@utp.edu.pl

Abstract. In this article, we present the use of sparse representation of signal and dictionary learning method for solving the problem of anomaly detection. The analyzed signal was presented as a set of correct ECG structures and outliers (characterizing different types of disorders). In the course of learning we used the modified Method of Optimal Directions (MOD) to find a dictionary that would reflect correct structures of an ECG signal. The dictionary found this way became a basis for sparse representation of the analyzed ECG signal. In the process of anomaly detection based on decomposition of the analyzed signal onto correct values and outliers, there was used a modified Alternating Minimization Algorithm (AMA). Performance of the proposed method was tested using a widely available database of ECG signals - MIT–BIH Arrhythmia Database. The obtained experimental results confirmed the effectiveness of the method of anomaly detection in the analysed ECG signals.

Keywords: Sparse representation · Outlier data · Analysis of ECG signal · Overcomplete dictionaries · Anomaly detection · MIT–BIH Arrhythmia Database

1 Introduction

ECG signal analysis, for which the subject is a human, consists in replication of the electrical activity of cardiac muscle cells. The ability of these cells to produce electrical impulses determines functioning of heart as a mechanical pump stamping blood in the circulatory system, due to which the remaining biological systems in a human organism work. Electrocardiographic signals are usually registered as changes in electrical potentials in time, and are obtained from various characteristic points of a human body (these points are directly connected to 3, 6 or 12-channel acquisition of ECG signals). Signals acquired in this manner contain vast essential diagnostic information [1].

A standard record of ECG, which is a set of multi lead signal, presents in result a series of repeated wave structures. Each of these structures consists of characteristic waves (P, Q, R, S, T, U), segments (P-R, S-T) and intervals (P-R, R-R, S-T, Q-T), which correspond to different stages of cardiac muscle electrical activity.

Morphological changes in shape, or disturbances in the structure of the ECG signal may have various reasons of morbid or physiological character. In an electrocardiographic record, these changes most often appear as distortions of shape or location of waves P and T, and segments ST or PQ. Anomalies may also emerge in localization and structure of the QRS complexes, which are basis for measuring the R-R interval. Distortion in the heart beat rhythm, directly connected to its irregular contractions, may be symptomatic of a disease entity called arrhythmia. One of the most often appearing factors of heart rhythm disturbance (especially in old people) are premature ventricular complexes (PCVs). They arise as a result of ectopic hyperactivity of myocardium and they cause disturbances in its synchronization [2].

In this article, we present the use of outlier dictionary learning method and sparse representation of signal for specific time series describing the analysed ECG signals. Anomaly detection is realized as a solution to the ECG signal's sparse representation issue on the base of the modified AMA method. As a result, we obtain the ECG signal decomposition onto the correct values and outliers (anomalies).

This paper is organized as follows. After the introduction, in Sect. 2, the related work is presented. Next, in Sect. 3 sparse representation of the ECG signal is described in detail. Then, in Sect. 4 the outlier dictionary learning method based on modified MOD algorithm estimation is shown. Next, Sect. 5 presents the details of the proposed solution consisting of formulation the sparse representation problem with outliers, and the modified AMA method for solving the problem of sparse representation. Implementation details and experimental results are described in Sect. 6. Conclusion are given thereafter.

2   Related Work

Until now, there have been published numerous experimental works directly describing different stages of processing, analysis and detection of given structures in an ECG signal [3]. It was performed by means of: (*i*) classical methods of signal processing, (*ii*) artificial neural networks, (*iii*) high-end classifiers (Support Vector Machine), (*iv*) genetic algorithms [4–8]. However, the current promising approach towards analysis and detection of anomalies in an ECG signal are the machine learning methods, in particular, the sparse representations of the signals realized with the use of dictionaries with redundancy (which describe essential structures of the analyzed signal). These methods, however, are most often used at the early stages of the ECG signal analysis, i.e. quality enhancement, or detection of the QRS complex [9]. The problem of anomaly detection for electrocardiography record should not be understood in the same manner as the traditional anomaly detection solutions. This issue is discussed in detail in the existing literature [10, 11]. For symbolic sequences, several anomaly detection methods were proposed by Chandola et al. [12] survey. On the other hand, in the process of broadly understood detection of anomalies (especially in a sparse representation) included in the ECG signal, they are occasionally used [13]. This fact has become inspiration for the present work.

3 Sparse Representation of ECG Signal

Representations of an ECG signal realized as linear extensions against a defined set of base functions well localized in time and/ or frequency, in many cases, are not precise and optimal enough. Therefore, a natural solution is introduction of more diverse, numerous, and adjusted to the signal's character sets of functions, called dictionaries with redundancy [14]. In result, we obtain representations which are more universal and flexible.

Sparse representation is searching for the sparse solution of decomposition coefficients c representing the signal s_i over the redundant dictionary when the remainder is smaller than the given constant δ, which can be stated as [15]:

$$\min \|C_0\| \quad s.t. \quad \left\| s_i - \sum_{p=0}^{P-1} c_p d_p \right\| < \delta, \tag{1}$$

where $\|\cdot\|_0$ is the ℓ_0 normcounting the nonzero entries of a vector, $c_p \in C$ representing a set of projection coefficients, and $d_p \in D$ are the elements of the redundant dictionary D.

Optimal representation can be defined as such a subset of dictionary D elements, whose linear combination explains the biggest percentage of the signal s_i energy among all the subsets of the same count. The choice of such a representation is computationally NP-hard [16].

3.1 The Proposed Data Model

We assume that $S \in R^{Q \times L}$ is a set of signals (including L signals $\left\{ s_i \in R^n \right\}_{i=1}^{L}$) which are well approximated by a sparse representation model, omitting the outliers (signals which vary from the model ones). Such a set S is presented as

$$S = DC + O + N, \tag{2}$$

where D is an assumed known, O is the matrix of outliers that have few non-zero columns that equal to the deviation of each outlier from the sparse representations model, and N is a low-energy noise component.

3.2 Sparse Representation Models with Outliers

Our aims are: detection of outliers in the set/collection S, and recovery of the sparse representations. For the case of the Single Measurement Vector (SMV) [17], this aim can be achieved by finding a solution to the problem below

$$\{C, O\} = \mathrm{argmin}_{C,O} \|S - DC - O\|_F^2 \quad s.t. \quad \|C\|_0 \leq LT_0 \quad \text{and} \quad \|O\|_{2,0} \leq P_0, \tag{3}$$

where $\|O\|_{2,0}$ is a count of the number of non-zero columns in O, the constraints ensure at most T_0 is the maximum number of non-zero entries, and at most P_0 non-zero columns in the outlier matrix. The solution encouraged by problem (3) involves sparse C.

However, it allows nonzero columns in O in case some outliers cannot be represented solely by D.

For the case of Multiple Measurement Vector (MMV) [18], our aim can be achieved by finding a solution to the problem below

$$\{C, O\} = \text{argmin}_{C,O} \|S - DC - O\|_F^2 \quad \text{s.t.} \quad \|C\|_{0,2} \le T_0 \quad \text{and} \quad \|O\|_{2,0} \le P_0, \qquad (4)$$

where $\|C\|_{0,2}$ is a count of the number of non-zero rows in C.

The basic difference between the models (3) and (4) consists in the fact that each signal in SVM is assumed to be a single measurement associated with a unique non-zero pattern of its sparse representation (i.e. a unique combination of atoms). In MMV, however, we are dealing with joint-sparsity for the collection S.

4   Dictionary Learning Method of Resistant-to-Outlier

The question of dictionary learning in the scope of sparse representation has been a common field of analysis over the last years [19]. What differentiates dictionary learning algorithms is their attitude towards the dictionary update process, i.e. some update all atoms at once (e.g. Method of Optimal Directions [20]), other update atoms one after another (e.g. K-Singular Value Decomposition [21]).

The main function of these solutions is finding the best dictionary D (resistant-to-outlier) which would represent the signal S as sparse composition, by solving

$$\min_{D,C} \left[\|S - DC - O\|_F^2 + \lambda \|o_i\|_{2,1} \right], \qquad (5)$$

where $\|\cdot\|_F^2$ is the Frobenius norm and $O = [o_1, \dots, o_L]$ is the matrix of outliers, $\|o\|_{j2,1}$ is the mixed $\ell_{2,1}$ norm of vector $o = [\|o\|_{12}, \dots, \|o\|_{L2}]^T$, in which is defined as the ℓ_1 norm of matrix of outliers, λ is a threshold parameter.

For solving the resultant optimization problem (5), the MOD (Method of Optimal Directions) is used. This process is conducted by means of minimizing the objective function iteratively over one variable, whilst the remaining two are fixed. Firstly, D and O are initialized, then minimization over C is performed – the iterative optimization starts. The regular course of initializing D implies the use of a predefined dictionary, for instance Gabor's [22], or the dictionary is composed of atoms randomly chosen from the training signals. The second option does not apply to our solution-finding process because some outliers might be taken as atoms, and as a result, they could possibly affect the whole process (in the sequel iterations). O is initialized by using the zero matrix. By way of explanation, all the training signals are accepted as not "outliers" at first iteration.

The modified MOD algorithm is divided into three stages:

- **Sparse Coding Stage:** In this phase, the coefficients c_i are set on the basis of the redundant dictionary D and signal s_i. In every step, we seek the minimum number of coefficients allowing to satisfy the Eq. (6). The given D is fixed. The Orthogonal Matching Pursuit algorithm [23] is used to calculate M sparse coefficients c_i for every sample of signal s_i, by approximating the solution of

$$c_i \leftarrow \arg\min_{c_i} s\|_i - Dc_i - o_i\|_2^2 \quad s.t \quad \|c_i\|_0 \leq T \quad i = 1, 2, \ldots, M. \tag{6}$$

- **Outlier Update Stages:** In this phase, we update the o outlier vector (in every step we seek the minimum number of outliers) to satisfy the Eq. (7).

$$o_i \leftarrow \operatorname{argmin}_{o_i} \left[\|s_i - Dc_i - o_i\|_2^2 + \lambda\|o_i\|_2 \right] \quad i = 1, 2, \ldots, M. \tag{7}$$

For making an update of an outlier vector, it is necessary to solve

$$\min_o \left[\|r - o\|_2^2 + \lambda\|o\|_2 \right], \tag{8}$$

where $r = s - Dc$ is the residual vector. When the derivative of the objective function is set to be equal to zero, at the optimal point $\hat{o}$, the result is

$$\hat{o} = \left\{ \begin{array}{c} \left(1 - \dfrac{\lambda}{2\|r\|_2}\right) r, \ if \|r_2\| > \dfrac{\lambda}{2} \\ 0, \qquad\qquad otherwise \end{array} \right\}. \tag{9}$$

The conclusion from the above is that when the sparse representation error (the residual) norm of the training data exceeds the threshold λ, this point is identified as an outlier. In other case, it is perceived as a relevant data vector. Another inference that can be drawn about the importance of the trade-off parameter λ is that the smaller values it has, the more points are recognized as outliers.

- **Dictionary Update Stages:** In this phase, we update the D dictionary atoms. The alternative values of an atom and coefficients are computed in order to lower the possibility of an error between the signal S and the sparse representation $D * C$ with outliers.

$$D \leftarrow \arg\min_D \sum_{i=1}^{L} \|s_i - Dc_i - o_i\|_2^2 \tag{10}$$

In the dictionary update stage, the following problem has to be solved

$$\min_D \|F - DC\|_F^2, \tag{11}$$

where $F = S - DC$. We chose the MOD algorithm to solve the above problem (for its simplicity) [20], however, any dictionary update algorithm may be used here.

All dictionary atoms are updated in this manner. Iterating according to the three above steps will create a dictionary (without outliers) that approximates the provided signal S in a sparse and accurate way. The outcome of the MOD algorithm will be a dictionary D (resistant-to-outlier used in the modified AMA method) containing often correlated atoms that best represent the analyzed signal S regarding its sparse representation $D * C$.

5   The Proposed Solution

Representation of the signal S, described by dependency (2) requires solving the problems of sparse representations of SVM and MMV types (see Subsect. 3.2), which are presented by the examples (3) and (4), and also finding a resistant-to-outlier dictionary D (see Sect. 4). To realize so stated task, the alternating minimization algorithm was modified so as to achieve a set of outliers O describing anomalies in the signal S.

5.1   Problem Formulation of Sparse Representation with Outlier

The solutions to problems described by Eqs. (3) and (4) can be approximated by solving the following unconstrained convex problem:

$$\min_{C,O} \left[\frac{1}{2} \|S - DC - O\|_F^2 + \alpha \|C\|_{1,d} + \beta \|O\|_{2,1} \right], \tag{12}$$

where d characterized the norm $\|*\|_{2,d}$ (i.e. $d = 1$ for SMV, and $d = 2$ for MMV), α, β are small positive parameters.

We propose to solve problem (12) by the modified Alternating Minimization Algorithm [24] because this algorithm is simpler than Alternating Direction Method of Multipliers.

5.2   Modified AMA Algorithm for Anomaly Detection

The classical alternating minimization algorithm was designed to solve the following structured optimization problem [24]:

$$\min_{x,y} f(X, Y) \quad s.t. \quad AX + BY = Z, \tag{13}$$

where X, Y, Z, A, B are matrices and the objective function is either separable $f(X, Y) = \theta_1(X) + \theta_2(Y)$, and $\theta_1(\cdot), \theta_2(\cdot)$ are two real-valued functions.

To use AMA for solving the case of Eq. (12), an auxiliary variable V and equality constraint are added

$$\min_{C,O,V} \frac{1}{2} \|S - DC - O\|_F^2 + \alpha \|V\|_{1,d} + \beta \|O\|_{2,1} \quad s.t. \quad V = C. \tag{14}$$

It is necessary to notice that changing (12) into a constrained issue, the first and second terms of (12) have been decoupled. Such action allowed for evasion of necessity of an iterated-shrinkage solution for C [25]. By adding the auxiliary variable C, we obtained a close-form solution for C, and a one-shot shrinkage solution for V. The Lagrangian function and the augmented-Lagrangian function for (14) are respectively:

$$\mathcal{L}(C, V, O, \lambda) = \frac{1}{2}\|S - DC - O\|_F^2 + \alpha\|V\|_{1,d} + \beta\|O\|_{2,1} - \langle\lambda, V - C\rangle \tag{15}$$

and

$$\mathcal{L}_\mu(C, V, O, \lambda) = \frac{1}{2}\|S - DC - O\|_F^2 + \alpha\|V\|_{1,d} + \beta\|O\|_{2,1} - \langle\lambda, V - C\rangle + \frac{\mu}{2}\|V - C\|_F^2, \tag{16}$$

where λ is a Lagrange multiplier and μ is a penalty parameter that controls the penalty level of deviation from the equality constraint.

Then, the framework of the AMA (with anomaly detection) for solving (14) can be expressed as follows:

$$\begin{cases} C^{k+1} = argmin_C \, \mathcal{L}\left(C, V^k, O^k, \lambda^k\right) \\ V^{k+1} = argmin_V \, \mathcal{L}_\mu\left(C^{k+1}, V, O^k, \lambda^k\right) \\ O^{k+1} = argmin_O \, \mathcal{L}_\mu\left(C^{k+1}, V^{k+1}, O, \lambda^k\right) \\ \lambda^{k+1} = \lambda^k + \delta\mu^k\left(V^{k+1} - C^{k+1}\right) \end{cases}, \tag{17}$$

where $\delta \in (0, 2)$, there is a guarantee of convergence of the above on condition that the sub-problem is solved ensuring high precision at every iteration [24]. Algorithm stop criterion is given by

$$\frac{\|V - C\|_F^2}{\|C\|_F^2} < \varepsilon, \tag{18}$$

where ϵ is the threshold parameter.

As a result of the AMA algorithm performance, we achieve: (*i*) a set of coefficients C, which together with the dictionary D (obtained by means of the MOD algorithm) represent the correct structure of the signal S, (*ii*) outliers (anomalies included in the signal S) reflected by the matrix O. The whole represents the proposed model of data described by dependency (2).

6   Experimental Results

In the conducted research, there was used the Massachusetts Institute of Technology – Beth Izrael Hospital Arrhythmia Database (MIT–BIH) consisting of eight sets of data burdened with various disorders, which can be detected with the use of the Holter's method. Forty-eight records from the MIT–BIH were tested, in which there are: a file with the subject signal, a header file (containing data about the patient), and a file with diagnostic data (including the signal's description in the binary form). The sampling

frequency of the signal equaled 360 Hz, and the resolution was set to 12 bits. The detailed outcomes, however, were presented exclusively for three selected signals. The mentioned signals have consecutive numbers, and are characterized by (respectively): #108 – high P wave amplitude (often mistakenly detected as R), #203 – high T wave amplitude (often mistakenly detected as R), #109 – contains a lot of PVC events. Exemplary fragments of those signals are presented in Fig. 1, as follows: (a) signal #105, (b) signal #203, (c) signal #109.
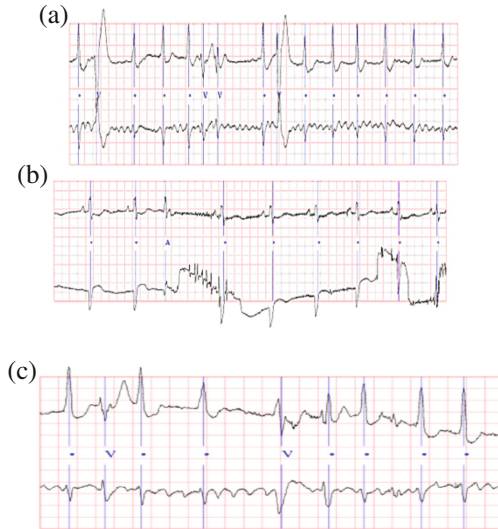

Fig. 1. The fragments of electrocardiographic signals' records from the MIT–BIH base (a) signal #108, (b) signal #203, (c) signal #109.
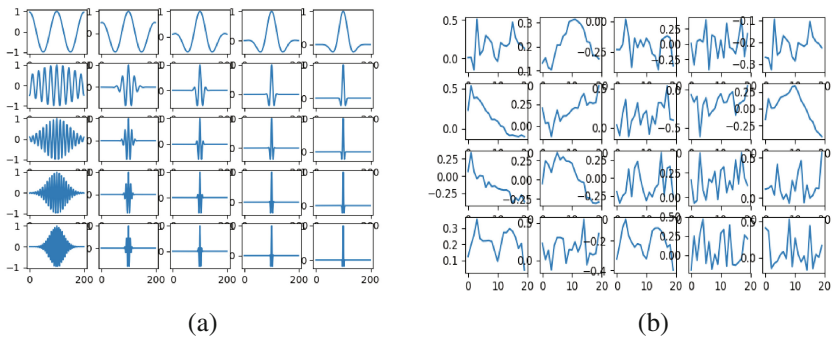

Fig. 2. Exemplary atomic dictionaries: (a) Gabor dictionary for the MOD initialization, (b) MOD dictionary for 20 atoms of the #203 signal.

The simulations were performed on a dual-processor Dell work station composed of eight core i7 processors and 32 GB of RAM.

For the proposed data model (2), and for solving the problem of sparse representation with separated outliers (12), a proper dictionary composed of 200 atoms was sought. Figure 2b presents chosen atoms of a dictionary obtained for the signal #203 by means of methodology described in Sect. 4. For initialization of the initial dictionary of the MOD algorithm, the Gabor's dictionary proposed by Mallat and Zhang in [22] was used. The parameters of the atoms were chosen from dyadic sequences of integers. Chosen Gabor's dictionary atoms are presented in Fig. 2a.

To solve the problem (14) we used the modified AMA method for the following parameters: $\mu^0 = 1.0$, $\delta = 1.5$, $\alpha = 1.0$, $\beta = 2.5$, $\varepsilon = 0.005$, $d = 1$. Occurrences of arrhythmia (anomalies in the analyzed signal) were detected as columns in the matrix O, with the use of norm ℓ_2, and the threshold value 0.1. The obtained results for the analyzed signals #105, #203 and #109 are presented in Table 1. (DR stands for a detection rate, and FP – a false positive rate).

Table 1. Results of anomaly detection in ECG signals.

Signal	DR [%]	FP [%]
#105	95,57	2,89
#203	94,46	3,75
#109	96,71	2,90

7   Conclusions

Contemporary medicine offers a series of myocardium diagnostic methods, ranging from physical examinations, through biochemical, and ending with diagnostic imaging. It is worth noticing that in the process of diagnosing it is often necessary to use a quick and noninvasive diagnostic method. In such a case, an electrocardiograph is the most often used equipment (in a hospital or a surgery). It is easy to imagine that if such diagnostics were common and possible to utilize remotely (where the patient can perform the survey themselves and obtain an automatically generated and understandable result, or connect with the cardiological monitoring center to evaluate the ECG record more precisely), it would certainly create new quality to the widely understood medical care. In terms of automatic analysis of ECG, the question of detection and identification of selected ECG parameters determines the need to use complex methods in the field of processing, analysis and recognition of signals.

In this article, we present the use of sparse representation of signal and outlier dictionary learning method for the analysed electrocardiography records. To find appropriate structure of the dictionary, there was used the modified MOD algorithm. The dictionary obtained this way constituted basis for the sparse representation of the analyzed ECG signal. The process of anomaly detection was realized as a solution of the task of sparse representation of the ECG signal described by means of correct values and outliers. To realize this task, there was used the modified ADA algorithm. Efficiency

of our method was examined using a widely available database of ECG signals - MIT–BIH Arrhythmia Database. The received experimental results confirm the efficiency of the presented method.

References

1. Einthoven, W.: The string galvanometer and the measurement of the action currents of the heart. Nobel Lecture (1925)
2. Daskalov, I.K., Christov, H.: Electrocardiogram signal preprocessing for automatic detection of QRS boundaries. Med. Eng. Phys. **21**, 37–44 (1999)
3. de Chazal, P., O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans. Biomed. Eng. **51**(7), 1196–1206 (2004)
4. Tang, X., Lan, S.: A frame work of automatic analysis system of electrocardiogram signals. Int. J. Signal Process. Image Process. Pattern Recognit. **7**(2), 211–222 (2014)
5. Shyu, L.Y., Wu, Y.H., Hu, W.: Using wavelet transform and fuzzy neural networks for VPC detection from the Holter ECG. IEEE Trans. Biomed. Eng. **51**(7), 1269–1273 (2004)
6. Poli, R., Cagnoli, S.: Genetic design of optimum linear and non-linear QRS detection. IEEE Transl. Biomed. Eng. **42**, 1137–1141 (1995)
7. Deng, Z., Zhang, S., Yang, L., Zong, M., Cheng, D.: Sparse sample self-representation for subspace clustering. Neural Comput. Appl., 1–7 (2016)
8. Lehnert, W., Soderland, S., Aronow, D., Feng, F., Shmueli, A.: Inductive text classification for medical applications. J. Exp. Theor. Artif. Intell. **7**(1), 49–80 (1995)
9. Ning, X., Selesnick, I.W.: ECG enhancement and QRS detection based on sparse derivatives. Biomed. Signal Process. Control **8**, 713–723 (2013)
10. Agyemang, M., Barker, K., Alhajj, R.: A comprehensive survey of numeric and symbolic outlier mining techniques. Intell. Data Anal. **10**(6), 521–538 (2006)
11. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artif. Intell. Rev. **22**(2), 85–126 (2004)
12. Banerjee, A., Chandola, V., Kumar, V.: Anomaly detection: a survey. To Appear in ACM Computing Surveys (2009)
13. Andrysiak, T.: Machine learning techniques applied to data analysis and anomaly detection in ECG signals. Appl. Artif. Intell. **30**, 610–634 (2016)
14. Elad, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer, New York (2010)
15. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. J. Constr. Approx. **13**, 57–98 (1997)
16. Tropp, J.A.: Greed is good: algorithmic results for sparse approximation. ICES Report 03-04, The University of Texas at Austin (2003)
17. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. **51**(1), 34–81 (2009)
18. Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans. Signal Process. **53**(7), 2477–2488 (2005)
19. Rubinstein, R., Bruckstein, M., Elad, M.: Dictionaris for sparse representation modeling. Proc. IEEE **98**, 1045–1057 (2010)
20. Engan, K., Aase, S.O., Husoy, J. H.: Method of optimal directions for frame design. In: Proceedings of IEEE ICASSP, vol. 5, pp. 2443–2446 (1999)

21. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
22. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. **41**(12), 3397–3415 (1993)
23. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: Asilomar Conference on Signals, Systems and Computers, vol. 1, pp. 40–44 (1993)
24. Tseng, P.: Applications of splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**, 119–138 (1991)
25. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)