

## A new framework for utilizing side information in sparse representation

Seyed Hadi Hashemi Rafsanjani<sup>1</sup>, Saeed Ghazi Maghrebi<sup>2</sup>

<sup>1</sup>Digital Communication Department, ICT Research Center, Tehran, Iran

<sup>2</sup>Yadegar-e-Imam Khomeini (RAH) Shahr-e-Rey Branch, Islamic Azad University, Tehran, Iran

### Article Info

#### Article history:

Received Oct 15, 2020

Revised Aug 27, 2021

Accepted Aug 30, 2021

#### Keywords:

Side information

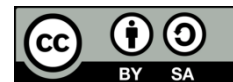
Sparse representation

Weighted sparse

### ABSTRACT

An underdetermined system of linear equation has infinitely number of answers. To find a specific solution, regularization method is used. For this propose, we define a cost function based on desired features of the solution and that answer with the best matches to these function is selected as the desired solution. In case of sparse solution, zero-norm function is selected as the cost function. In many engineering cases, there is side information which are omitted because of the zero-norm function. Finding a way to conquer zero-norm function limitation, will help to improve estimation of the desired parameter. In this regard, we utilize maximum a posterior (MAP) estimation and modify the prior information such that both sparsity and side information are utilized. As a consequence, a framework to utilize side information into sparse representation algorithms is proposed. We also test our proposed framework in orthogonal frequency division multiplexing (OFDM) sparse channel estimation problem which indicates, by utilizing our proposed system, the performance of the system improves and fewer resources are required for estimating the channel.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



### Corresponding Author:

Saeed Ghazi Maghrebi

Yadegar-e-Imam Khomeini (RAH) Shahr-e-Rey Branch

Islamic Azad University

Shahr-e-Rey, Iran

Email: s\_ghazi2002@yahoo.com

## 1. INTRODUCTION

An underdetermined system of linear equation,  $b = Ax$ ,  $A \in \mathbb{R}^{n \times m}$  with  $n < m$  has either infinitely answer or no answer, if  $b$  is not in the span of  $A$ . To guarantee that the underdetermined system of linear equation has infinitely answers,  $A$  the coefficient matrix which is called dictionary matrix in this paper, must be a full rank matrix [1]. To find a specific answer with special features, the regularization method is used. In this method, a cost function is defined based on desired features and looks for those answers that satisfy it [2], [3]. The answer which best satisfies this relation is selected as the desired solution [4]. In case of a sparse solution, the selected cost function for regularization is zero-norm function,  $\|x\|_0 = \sum_i |x_i|^0$  which counts the number of non-zero elements of the  $x$  as [4],

$$\hat{x} = \operatorname{argmin}_x \{\lambda \|x\|_0 + \|Ax - b\|_2\}, \quad (1)$$

where  $\lambda$  is a balancing parameter. The (1) is known as  $P_0$ -problem and it is nondeterministic polynomial time (NP)-hard problem [4]. Solving this problem is important, because in many engineering problems such as [5]-[8], the sparse answer is the desired solution. To find the sparse solution for this problem, two general

approaches are used. The first one, the greedy methods [9], attempts to find the non-zero elements one by one per time. The second one is known as relaxation which replaces the zero-norm function with a smoother one and uses optimization technique to find the solution [10]. Algorithms such as orthogonal matching pursuit (OMP), smoothed  $l_0$  (SL<sub>0</sub>) and  $l_1$ -minimization [11]-[13] are some examples of the proposed algorithms for finding the sparse solution.

Although the zero-norm function is a proper cost function for finding the sparse solution, in case of existing other features or additional information, it suppresses them. Therefore, modification is required to conquer this limitation. To clarify it, let  $C$  be a diagonal cost matrix whose diagonal elements are selected based on side information. To apply side information into the estimation process, the cost matrix  $C$  is multiplied by  $x$  to assign costs to each element of  $x$ , therefore (1) changes as (2).

$$\hat{x} = \operatorname{argmin}_x \{ \lambda \|Cx\|_0 + \|Ax - b\|_2 \} \quad (2)$$

However, since  $\|Cx\|_0 = \sum_i |c_i x_i|^0 = \sum_i |x_i|^0 = \|x\|_0$ , the zero-norm function eliminates the impact of cost matrix. Therefore, to take the impact of side information into account, a new framework is required to manipulate side information and conquers the zero-norm limitation.

Regarding utilizing side information with sparse representation, some researchers proposed to replace zero-norm function with  $l_1$ -norm,  $\|x\|_1 = \sum_i |x_i|^1$ , and try to utilize side information with it [14]-[19]. For example, in [17], an iterative method is proposed that employs a weighted  $l_1$ -minimization method and utilizes the result of previous iterations to calculate the weight of the next iteration. The iterative model in [17] reduces the number of measurements and improves the solution of  $l_1$ -minimization method; however, this method does not utilize prior information to find the sparse solution. Also, in [17], Candes integrated the information via maximizing the correlation between the prior information and the desired solution and found the sparse solution for (3) relation;

$$\operatorname{argmin}_x \{ \lambda (\|x\|_1 + \langle x, \varphi \rangle) + \|Ax - b\|_2 \}, \quad (3)$$

where the  $\|x\|_1$  is the  $l_1$ -norm and the second term calculates the similarity between prior information,  $\varphi$ , and the desired solution,  $x$ . Similarly, in [15] they proposed a framework concentrating on  $l_1$ - $l_1$ -minimization and tried to utilize side information by replacing  $l_0$ -norm with  $l_1$ -norm. These methods and similar frameworks considered  $l_1$ -norm as the cost function to handle side information, and it must be noted they did not address utilization of side information in  $l_0$ -problem.

All the mentioned methods use relaxation to overcome zero-norm limitation. However, in our proposed method, we define a framework to overcome this limitation. In our framework, we define an auxiliary variable,  $x' = Cx$  and consider modified cost function as  $F_C(\cdot)$ . Since  $C$  is a diagonal matrix with non-zero elements, the inverse matrix exists and this matrix is defined as a weight matrix  $W \triangleq C^{-1}$  and consequently,  $x = Wx'$ . Since matrix  $W$  is a diagonal  $N \times N$  matrix and the dictionary  $A$  is an  $L \times N$  matrix, it is possible to combine the weight matrix,  $W$ , with the dictionary,  $A$ , and define a weighted-dictionary,  $B \triangleq AW$ . Note that if the desired solution of  $Ax = b$  is sparse, then the solution of  $Bx' = b$  is also sparse and  $W$  will be a diagonal matrix with non-zero diagonal elements. As a result, by defining the cost matrix  $C$  based on side information and replacing the dictionary matrix  $A$  with the weighted-dictionary  $B = WA$ , the observation relation changes to the weighted-observation. In this case, not only its desired answer is sparse, but also side information is embedded in it. Therefore, the cost function should be selected so that it handles sparse answer. Here since the function must look for a sparse, the best function is a zero-norm function. By replacing the weighted-dictionary and selecting zero-norm as the cost function, the regularization relation changes as (4).

$$\hat{x}' = \operatorname{argmin}_{x'} \{ \|x'\|_0 + \|Bx' - b\| \}. \quad (4)$$

In (4) indicates that in the presence of side information, the weight matrix gives higher weights to those columns of the dictionary matrix which are more probable to be selected. At the next sections we try to provide materials that support above discussion.

The rest of this paper is organized as follows. At the section 2, we use maximum a posterior (MAP) estimation to utilize side information for finding sparse solution and support above argument. Section 3 is devoted to explain the procedure of appending side information into sparse representation problem and in section 4, we test our framework on a channel estimation problem for an orthogonal frequency division multiplexing (OFDM) system and present results.

**2. ESTIMATING SPARSE BY UTILIZING MAP ESTIMATION**

MAP estimation estimates the desired parameter by utilizing prior information. This estimation tries to find the desired answer by maximizing  $F(d|o)$ , where  $d$  is the desired parameter to be estimated,  $o$  is an observation parameter and  $F(.|.)$  is a conditional probability density function (pdf) [20]. To estimate a sparse solution for an underdetermined system of linear equation, by utilizing MAP estimation, the observation vector, desired parameter and observation relation are defined as  $d = x, o = b, b = Ax + z$ , respectively. It is worth mentioning that  $z$  is the observed noise.

By utilizing Bayes rule, the MAP estimation changes as follows,

$$\hat{x} = \operatorname{argmax}_x F(x|b) = \operatorname{argmax}_x \frac{F(b|x)F(x)}{F(b)} \tag{5}$$

where  $F(b)$  can be omitted, since it does not affect maximization process. Considering  $z$  as an identical independent distributed (i.i.d) zero mean white Gaussian noise, the conditional distribution becomes as  $F(b|x) = F(Ax + z|x) = \mathfrak{N}(Ax, \sigma I)$ , where  $I$  is an identical matrix, and  $\sigma$  is the variance of as observation noise and  $\mathfrak{N}(\cdot)$  is normal distribution function. The pdf of  $F(x)$ , which is a prior of the desired parameter, should be selected among those pdf that covers sparse feature. Considering a sparse signal, most of its elements are zero, and its few elements obtain a nonzero amount, or in other word, the probability density function (pdf) of the elements should have a high density at zero and have a heavy tail [21]. High density around zero increases concentration of the mass around zero which results in increasing the probability of selecting zero, and guaranties most elements of the signal gain zero amount. On the other hand, its heavy tail guarantees few elements gain huge amounts. Spike and slab [22], the mixture of narrow and very wide Gaussian distributions, laplace and exponential distribution [21] are some example of sparse pdfs. In this paper, we select generalized Gaussian distribution (GGD) as sparse pdf of the desired parameter, because it covers wide range of sparse and non-sparse pdf's while different amounts are assigned to its parameters. In this paper, it is assumed that the desired parameter is an i.i.d random process containing  $N$  i.i.d parameters,  $x = x_1, x_2, \dots, x_N$ . Consequently,  $F(x) = \prod_{i=1}^N f(x_i)$ . By replacing normal distribution and GGD in (5), the estimation relation changes as (6),

$$\hat{x} = \operatorname{argmax} \mathfrak{N}(Ax, \sigma I) \prod_{i=1}^N \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\frac{\|x_i\|_\alpha^\alpha}{\beta}\right), \alpha > 0, \beta > 0, \tag{6}$$

where  $\alpha$  and  $\beta$  are parameters of the GGD distribution controlling shape and variance of the distribution, respectively. The parameter  $\alpha$  controls sharpness of the GGD distribution and provides a wide range of probability density functions. For example, Laplacian distribution is obtained by selecting  $\alpha = 1$  and sparse distributions are obtained by selecting  $\alpha \leq 1$ . Since normal distribution and GGD are exponential functions and natural logarithmic function ( $\ln$ -function) is strictly increasing, as a result, performing  $\ln$ -function in (6) does not affect the maximization process. After some manipulations, the estimation relation is calculated as follows,

$$\hat{x} = \operatorname{argmax}_x \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \frac{\alpha}{2\beta\Gamma(1/\alpha)}\right) - \frac{\|b-Ax\|_2}{\sigma^2} - \frac{\|x\|_\alpha}{\beta^\alpha}, \tag{7}$$

where  $\|\cdot\|_2, \|\cdot\|_\alpha$  are the norm-2 and norm- $\alpha$ , respectively. Looking closer at (7), it is observed that the first term is a constant amount and can be omitted during maximization problem. By omitting minus sign, maximization process changes to minimization process and the estimation relation changes as follows,

$$\hat{x} = \operatorname{argmin}_x \frac{\|b-Ax\|_2}{\sigma^2} + \frac{\|x\|_\alpha}{\beta^\alpha}. \tag{8}$$

In (8) is also obtained while regularization for an underdetermined system of linear equation is used and  $\alpha$ -norm is selected as a cost function. To consider sparse feature,  $\alpha \leq 1$  should be selected. By selecting  $\alpha = 1$  and performing some calculation, in (8) changes to the following form,

$$\hat{x} = \operatorname{argmin}_x \frac{\|b-Ax\|_2}{\sigma^2} + \frac{\|x\|_1}{\beta^1}. \tag{9}$$

This relation is a  $P_1$ -problem which is convex, but not strictly convex [4], [23]. This relation may have more than one solution, but these solutions are convex and gathered in a bounded convex set [24]. Beside it, as we are looking for the sparse solution (for example  $k$ -sparse), there exists at least one answer with at most  $k$  non-zero elements. The above relation can be formulated as linear programming to find the

sparse solution. Algorithms such as least absolute shrinkage and selection operator (LASSO) [24] are used to solve it [25].

Here, we consider the boundary case where  $\alpha \rightarrow 0$ , and assume that the taps of the channel have GGD distribution with  $\alpha \rightarrow 0$ . In this case, the estimation relation changes as follows,

$$\hat{x} = \underset{x, \alpha \rightarrow 0}{\operatorname{argmin}} \frac{\|b - Ax\|_2}{\sigma^2} + \frac{\|x\|_\alpha}{\beta^\alpha}. \quad (10)$$

As it is clearly indicated in (10), by selecting  $\alpha \rightarrow 0$ , the estimation relation changes to  $P_0$ -problem. Here it is shown that by utilizing MAP estimation and selecting GGD distribution as a pdf for a sparse signal, the obtained relation is similar to the relation obtained by regularization process. In the next section, we explain how to contribute to side information in sparse representation process.

### 3. CONTRIBUTING SIDE INFORMATION IN SPARSE ESTIMATION PROCESS

While sparsity is the only assumption of the desired answer, all elements of  $x = \{x_1, x_2, \dots, x_N\}$  are treated uniformly and equally probable [26]. In case of side information, some atoms are more probable to be selected as non-zero elements of the sparse answer. As  $x$  is considered as an i.i.d signal, those elements that have higher chance of selection (feasible elements) will gain higher variance, and vice versa. The variance of GGD pdf is directly controlled by its scale parameter,  $\beta$ , then to conduct side information into estimation process, those elements that are more probable to be selected, as non-zero elements, have higher amount of  $\beta$ . To indicate it in a mathematical form, a weight element  $w_i$  is selected based on side information and added to the scale parameter as,  $\beta_i \triangleq w_i\beta$ . By replacing the scale factor with the weighted one, the estimated parameter is calculated as follows,

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left( \sum_{i=1}^N \frac{|x_i|^\alpha}{(w_i\beta)^\alpha} + \frac{1}{2\gamma} \|b - Ax\|_2 \right), \quad (11)$$

where  $\|\cdot\|_2$  denotes the  $l_2$ -norm of a vector. Since  $w_i$ 's are positive non-zero values, the term  $\sum_{i=1}^N \frac{|x_i|^\alpha}{(w_i\beta)^\alpha}$  can be re-written in a matrix form as  $\|W^{-1}x\|_\alpha$ , where  $W$  is a diagonal matrix with  $i$ 'th diagonal element defining  $x' \triangleq W^{-1}x$ , in (11) can be re-written as,

$$\hat{x}' = \underset{x'}{\operatorname{argmin}} \left( \|x'\|_\alpha + \frac{1}{2\gamma} \|b - Bx'\|_2 \right). \quad (12)$$

The weight matrix,  $W$ , can be combined with the dictionary matrix,  $A$ , to define the weighted-dictionary,  $B = AW$ . It should be noted that prior side information affects the estimation process by applying weights into the columns of the dictionary matrix and boosting the selection probability of the feasible elements.

The (12) is obtained while regularization method is used for finding a sparse solution of  $b = Bx'$  and selecting  $\|\cdot\|_\alpha$  as the cost function. Then, by selecting  $\alpha \rightarrow 0$ , the (12) changes to the weighted- $P_0$ -problem. In this framework, instead of directly applying side information into the desired parameter, the dictionary matrix is modified and weights are applied to its columns. To indicate one of the benefits of our framework, we compare it with weighted- $l_1$  method [17]. One of the methods to find the sparse solution of an underdetermined system with linear equation is a relaxation method which replaces the  $l_0$ -norm by a smooth function and uses continuous optimization techniques to find the solution [17]. In general, weighted- $l_1$  method is used to contribute prior information into the procedure of finding the sparse solution [26]. However, relaxing the weighted- $P_0$ -problem by a weighted- $l_1$  function is not trivial [27]; if a weighted- $l_1$  solution is adopted as a relaxation of the weighted- $P_0$ -problem. It is not clear whether the solution satisfies the original  $P_0$ -problem or the weighted- $P_0$ -problem due to the fact that  $\sum_{i=1}^N w_i|x_i| = \sum_{i=1}^N |w_i x_i|$ , and  $\sum_{i=1}^N |w_i x_i|$  is a relaxation of  $\sum_{i=1}^N |w_i x_i|_0$  which in turn is equal to  $\sum_{i=1}^N |x_i|_0$ . Therefore, both weighted- $l_1$  and  $l_1$  can be seen as relaxations of the  $P_0$ -problem, with different solutions. On the contrary, our proposed framework contributes prior information into the problem formulation by modifying the dictionary matrix and introduces a weighted observation relation. Hence, the solution satisfies the original  $P_0$ -problem in the presence of prior information.

### 4. SIMULATION RESULT

In this section, we present simulation result of our proposed framework on application of channel estimation for an OFDM system. In this system, we utilized a 4-tap Rayleigh fading channel. The input of the OFDM system is a stream of one thousand OFDM symbols. We simulate each channel realization for

1000 iterations and then, evaluate the mean value of the error rate with respect to different signal to noise ratio (SNR) values and the number of estimation pilots as the final result. To indicate the proficiency of our proposed framework, we use OMP algorithm and  $l_1$ -minimization algorithm to estimate OFDM channel coefficients and compare their results.

In Figure 1, the results of our proposed model are denoted by weighted models, W-OMP, which are compared with OMP sparse solutions across different SNR values. As indicated in this figure, the impact of utilizing side information in low SNR is more than its effects in higher SNR. It has a simple interpretation that, side information compensates the impact of noises of the observed samples. For low SNR's, side information compensates the impact of noises. While for higher SNR's, a good estimation of the channel is obtained from the observed samples. Figure 2 indicates that the performance improvement of the weighted model for both the low and high SNR regimes. Although, for both Figures 1 and 2, by increasing both the SNR value and the number of pilots, the impact of prior information is reduced.

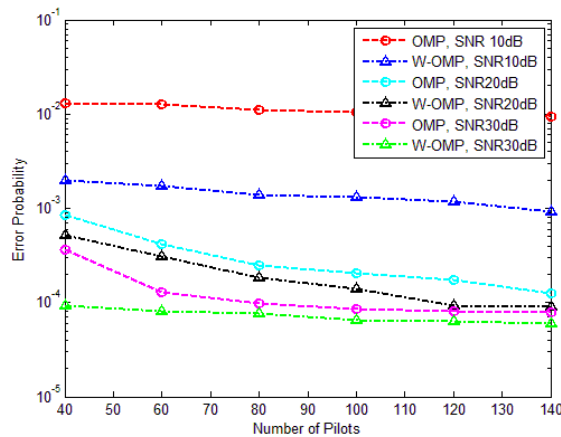


Figure 1. Performance of OFDM system using prior information with fixed SNR, OMP algorithm

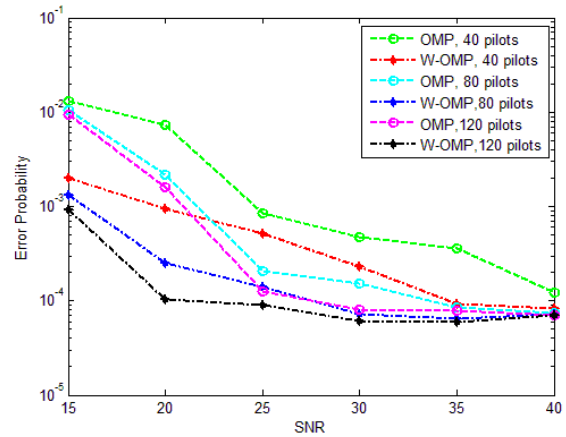


Figure 2. Performance of OFDM system using prior information with fixed number of pilots, OMP algorithm

Figures 1 and 2 clearly show that our proposed weighted model outperforms non-weighted models in terms of error probability for different number of pilots and SNR values. For example at SNR=10 dB, the error probability is reduced around an order of magnitude from  $10^{-1}$  to  $10^{-2}$ . As indicated in these figures, at the same condition by contributing side information in estimation process, the improvement in channel estimation is observed. Then, the system experiences better performance incompare with the ordinary case and the bit error rate is reduced.

To indicate the proficiency of our proposed framework, we test our framework with LASSO algorithm and compare its result while side information is used or not. Impacts of side information across different SNR values are presented in Figure 3. It is clearly observed that utilizing side information improve estimation of the channel and as a result improve performance of the system. Similar to OMP case, by increasing SNR value, the impact of side information will be decreased which has similar interpretation for OMP case. Figure 4 presents the impact of side information on fixed SNR and different amount of pilots. Comparing the performance of the system, it is observed that by utilizing side information, better estimation will be obtained which results in higher performance.

Comparing these figures, it is clearly observed that by considering only sparse feature for the channel, more resources are required to provide same performance in comparison with the case that side information is taken part at the estimation process. It is also indicated that, while the system is in a good condition, such as high SNR value or larger amount of estimation pilots, the need for considering side information is reduced. In other words, considering them will not increased the performance of the system same as before. This means that the mathematical computations will be decreased and also, the processing time will be decreased and as a result this method will be applicable in real-time data communications. Therefore, utilizing the prior channel information gives the opportunity to release some portion of resources of the OFDM system such as the number of pilots without degrading the system performance.

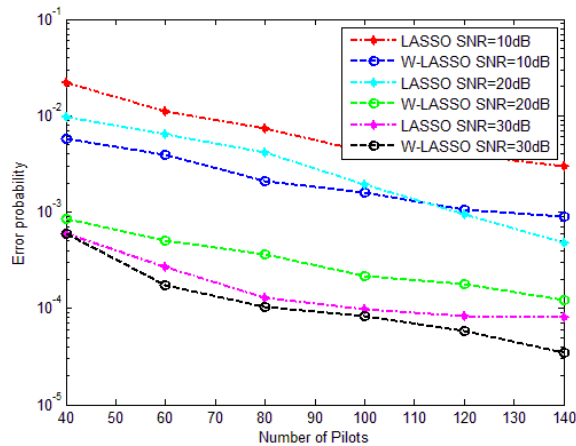


Figure 3. Performance of OFDM system using prior information with fixed SNR, LASSO algorithm

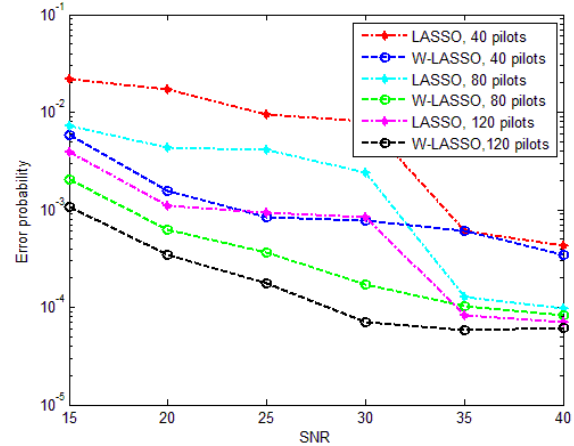


Figure 4. Performance of OFDM system using prior information with fixed number of pilots, LASSO algorithm

## 5. CONCLUSION

In this paper, we proposed a framework using prior information to improve sparse estimation. We showed that because of zero-norm function conventional sparse methods can not manipulate side information. However, to in our framework we applied side information and take advantage of it. In our model, we simultaneously used both side information and sparse feature of the desired parameters and utilized MAP estimation to support our discussion and explained the procedure of appending side information to sparse representation. We have shown that to simultaneously use side information and sparse representation, columns of the dictionary matrix must be modified and side information is added to them. In this regard, we defined weight matrix based on side information and applied it to the dictionary matrix to obtained weighted dictionary and weighted observed relation. To find the sparse solution, regularization was applied on the weighted observed relation and obtained answer was selected as the sparse solution. We have tested our framework for channel estimation problem in an OFDM system and we will have better performance in comparison with conventional sparse representation methods at the same condition. In our experience to reach same performance level, fewer resources (such as pilots and power) are required while side information is used. This shows that by utilizing this framework, for channel estimation, fewer pilots are occupied and more data can be transformed.

## REFERENCES

- [1] M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing" *Springer Science Business Media*, 2010.
- [2] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A Survey of sparse Representation: Algorithms and Applications" *IEEE access*, vol. 3, pp. 490-530, 2015, doi: 10.1109/ACCESS.2015.2430359.
- [3] Z. Qin, J. Fan, Y. Liu, Y. Gao, G. Y. Li, "Sparse Representation for Wireless Communications: A Compressive Sensing Approach," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 40-58, 2018, doi: 10.1109/MSP.2018.2789521.
- [4] M. Elad, "From exact to approximate solutions. In Sparse and Redundant Representations," *Springer*, 2010, doi: 10.1007/978-1-4419-7011-4.
- [5] Z. Chen *et al.*, "A novel sparse representation model for pedestrian abnormal trajectory understanding," *Expert Systems with Applications*, vol. 138, 2019, doi: 10.1016/j.eswa.2019.06.041.
- [6] J. L. Callahan, K. Maeda, and S. L. Brunton, "Robust flow reconstruction from limited measurements via sparse representation," *Physical Review Fluids*, vol. 4, no. 10, pp. 1-31, 2019, doi: 10.1103/PhysRevFluids.4.103907.
- [7] Y. Sun *et al.*, "Discriminative local sparse representation by robust adaptive dictionary pair learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 4303-4317, 2020, doi: 10.1109/TNNLS.2019.2954545.
- [8] M. K. Abboud and B. M. Sabbar, "A novel delay dictionary design for compressive sensing-based time varying channel estimation in OFDM system," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 5, pp. 2284-2291, 2020, doi: 10.12928/TELKOMNIKA.v18i5.14223.
- [9] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, 2006, doi: 10.1002/cpa.20124.
- [10] Y. Saad, "Iterative methods for sparse linear systems," *Society for Industrial and Applied Mathematics*, 2003, doi: 10.1137/1.9780898718003.ch4



- [11] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231-2242, 2004, doi: 10.1109/TIT.2004.834793.
- [12] GH. Mohimani, M. Babaie-Zadeh, and C. Jutten, "Fast sparse representation based on smoothed  $l_0$  norm," In *International Conference on Independent Component Analysis and Signal Separation*, 2007, doi: 10.1007/978-3-540-74494-8\_49.
- [13] P. V. Rossi and R. Vicente, " $l_1$ -minimization algorithm for Bayesian online compressed sensing," *Entropy*, vol. 19, no. 12, 2017, doi: 10.3390/e19120667.
- [14] X. Zhang, W. Cui, and Y. Liu, "Compressed sensing with prior information via maximizing correlation," In *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, doi: 10.1109/ISIT.2017.8006522.
- [15] Y. Nan, L. Zhang, and X. Sun, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for TDD massive MU-MIMO systems," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 345-348, 2015, doi: 10.1109/LWC.2015.2414933.
- [16] J. F. Mota, N. Deligiannis, and M. R. Rodrigues, "Compressed sensing with prior information: Strategies, geometry, and bounds," *IEEE Transactions on Information Theory*, vol. 63, no. 7, pp. 4472-4496, 2017, doi: 10.1109/TIT.2017.2695614.
- [17] E. Candes and M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $l_1$  minimization," *J Fourier Anal Appl*, vol. 14, pp. 877-905, 2008, doi 10.1007/s00041-008-9045-x.
- [18] E. Tsiligianis and N. Deligiannis, "Learning fast sparse representations with the aid of side information," *Tech. Rep.*, 2019.
- [19] G. Gui, W. Peng, and F. Adachi, "Improved adaptive sparse channel estimation based on the least mean square algorithm," In *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, doi: 10.1109/WCNC.2013.6555058.
- [20] J. O. Berger, "Statistical decision theory and Bayesian analysis," *Springer Science and Business Media*, 2013.
- [21] M. Unser and P. Tafti, "An introduction to sparse stochastic processes," *Cambridge University Press*, 2014, doi: 10.1017/CBO9781107415805.
- [22] T. Mitchell and J. Beauchamp, "Bayesian variable selection in linear regression," *Journal of the American Statistical*, vol. 83, no. 404, pp. 1023-1032, 1988, doi: 10.2307/2290129.
- [23] S. Foucart, "A note on guaranteed sparse recovery via  $l_1$ -minimization," *Applied and Computational Harmonic Analysis*, vol. 29, no. 1, pp. 97-103, 2010, doi: 10.1016/j.acha.2009.10.004.
- [24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267-288, 1996.
- [25] Q. Liu and J. Wang, "L1-minimization algorithms for sparse signal reconstruction based on a projection neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 689-707, 2015, doi: 10.1109/TNNLS.2015.2481006.
- [26] J. Huang and S. Ma and C. Zhang, "Adaptive Lasso for sparse high-dimensional regression models," *Statistica Sinica*, vol. 18, no. 4, pp. 1-30, 2008.
- [27] M. Babaie-Zadeh and B. Mehrdad and G. Giannakis, "Weighted sparse signal decomposition," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, doi: 10.1109/ICASSP.2012.6288652.

## BIOGRAPHIES OF AUTHORS



**Seyed Hadi Hashemi Rafsanjani** is a Ph.D. student at ICT research center. He obtained Bachelor Degree in Electronics Engineering from Shahid Bahonar University of Kerman in 2007 and Master degree in digital communication from Sharif University of Technology 2009. His researches are in fields of optical communication, digital video broadcasting and sparse representation.



**Saeed Ghazi-Maghrebi** is a faculty member of Yadegar-e-emam Islamic Azad University of Shahre-Rey. His researches are in fields of digital video broadcasting systems, wireless communication, and digital communication.