

Improving Mutual Coherence with Non-Uniform Discretization of Orthogonal Function for Image Denoising Application

Hani Nozari¹, Alireza Siamy²

¹Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran; ²Department of Electrical Engineering, Mazandaran University, Babol, Iran.
Email: hani_nozari@elec.iust.ac.ir

Received April 27th, 2011; revised May 20th, 2011; accepted May 27th, 2011.

ABSTRACT

This paper presented a novel method on designing redundant dictionary from known orthogonal functions. Usual way of discretization of continuous functions is uniform sampling. Our experiments show that dividing the function definition interval with non-uniform measure makes the redundant dictionary sparser and it is suitable for image denoising via sparse and redundant dictionary. In this case the problem is to find an appropriate measure in order to make each atom of dictionary. It has shown that in sparse approximation context, incoherent dictionary is suitable for sparse approximation method. According to this fact we define some optimization problems to find the best parameter of distribution measure (in our study normal distribution). For better convergence to optimum point we used Genetic Algorithm (GA) with enough diversity on initial population. We show the effect of this type of dictionary design on exact sparse recovery support. Our results also show the advantage of this design method on image denoising task.

Keywords: Grassmannian Frames, Normal Distribution, Mutual Coherence, Genetic Algorithm, Image Denoising

1. Introduction

Since the sparse representation of signals has led to improvements in many applications such as coding, denoising, feature extraction and etc. in signal and image processing, there has been a growing interest in the study of this method in recent years. Sparse and redundant representation modeling of data assumes an ability to describe signal as linear combinations of a few atoms from a pre-specified dictionary [1], where the linear coefficients are sparse. In this case we use a matrix

$D_{d \times N} \in \mathbb{R}^{d \times N} : d < N$, as a dictionary, either to compactly express or efficiently approximate a signal. Let $y \in \mathbb{R}^d$ and $x \in \mathbb{R}^N$ be the given signal and the coefficient vector respectively. We can express sparse equation as

$$\begin{aligned} (P_{0,\varepsilon}) : \min_x \|x\|_0 \\ \text{subject to } \|y - Dx\|_2 \leq \varepsilon \end{aligned} \quad (1)$$

where $\|x\|_0$ is the l^0 norm, counting the nonzero entries of a vector. With $\varepsilon = 0$, we have sparse representation which actually is a NP-hard problem and cannot be solved in a reasonable time [2]. Many algorithms have

been proposed in order to reach an approximate solution and all of them can be categorized as greedy algorithms like matching pursuit (MP) [1] and its derivations [3], and relaxation method, like basis pursuit (BP) [4]. Image denoising via sparse and redundant representation uses this fact. Suppose Z is zero-mean white Gaussian noise with known standard deviation σ , and X be a cleaned image. $Y = X + Z$ would be noisy image. Image denoising via redundant dictionary is a kind of energy minimization [5] and it can be expressed by following functional expression

$$\begin{aligned} f(\{\alpha_{ij}\}_{i,j}, X) = \lambda \|X - Y\|_2^2 + \sum \|D\alpha_{ij} - R_{ij}X\|_2^2 \\ + \sum \mu_{ij} \|\alpha_{ij}\|_0 \end{aligned} \quad (2)$$

The first term of above expression says the difference between measured image Y and its denoised version X . second and third part of this expression represent sparse approximation. Choosing a dictionary D which have to be an overcomplete dictionary is one of the most important fundamental choices that must be considered [6]. An overcomplete dictionary D that leads to sparse represen-

tations can either be chosen as a pre-specified set of functions (dictionary of this type include wavelets [7], Curvelets [8], Contourlets [9] and bandelets [10], among others.) or designed by adapt its content to fit a given set of sample signals such as method of optimal direction (MOD) [11], the K-SVD [12] and others. Using a pre-specified dictionary leads to creating simple and fast algorithms for the evaluations of the sparse representation. These types of dictionaries are called non-adaptive dictionaries. Adaptive dictionary is time consuming for sparse approximation so working with pre-specified functions still have its own interest.

In this paper we used different ways to signals atom from pre-specified functions. Each column of the dictionary is selected from a set of orthogonal functions in a specific dimension. In order to make a dictionary of size $d \times N$ we should sample d times from each N atoms. The main question is how one can sample the function definition interval in order to achieve a sparser dictionary. To divide the interval of function definition in order to select each sample of function formed with normal distribution $\mathcal{N}(\mu, \sigma)$, we used some optimization techniques to find the best parameter of normal distribution. For confidence in convergence we implement our optimization task with Genetic Algorithm (GA). As describe later our objective function has several extremum so enough diversity on initial population decrease the possibility of trapping in local minimum. With this discretization, the result dictionary has smaller mutual coherence than uniform one as a result of this fact better Peak-Signal-to-Noise-Ratio (PSNR) achieve in image processing. This type of dictionary design is also improved the required time for image denoising.

2. Tight Frames

The concept of a frame is a generalized concept of an orthonormal basis. Each vector in the space can be represented as the sum of the elements in the frame, but it is not necessarily unique. The frame theory gives energy equivalence conditions to solve both synthesis and analysis problems with stable operators. A family $\{\phi_p\}$ is a frame of the Hilbert space $V \in \mathbb{C}^d$. If $B \geq A > 0$ then we can have [7]:

$$\forall h \in V, A \|h\|^2 \leq \sum_{m \in \Lambda} |\langle h, \phi_p \rangle|^2 \leq B \|h\|^2 \quad (3)$$

where the usual Hermitian inner product is denoted with $\langle \cdot, \cdot \rangle$, and $\|\cdot\|$ is written for the associated norm. when $A = B$ we will have tight frames. We can achieve the unit norm tight frames when $\|\phi_p\| = 1$. With concatenating N frames of \mathbb{C}^d we have overcomplete dictionary. In following part of this section we explain the process of de-

signing overcomplete dictionary from a specific family of frames.

2.1. Overcomplete Dictionary

Suppose $\{\phi_1, \phi_2 \dots \phi_N\}$ are a set of N orthogonal continuous functions in the $[a, b]$ interval. In order to make N vectors of size " d " ($d < N$), we should sample these functions d times on its definition interval. The outcome of this process is a $D_{d \times N} \in \mathbb{R}^{d \times N} : d < N$ matrix called overcomplete dictionary that is used in sparse representations. Which distribution we should use to sample these frames? The answer of this question completely depends on the type of orthonormal frames, but experiments show that normal distribution is suitable for majority of orthonormal functions. Normal distribution $\mathcal{N}(\mu, \sigma)$ has two parameters; mean μ and standard deviation σ . Different parameters of this distribution define different type of discretization so we have distinct dictionary for each pair of parameters. In this paper we want to discuss different methods in order to find the best parameters of normal distribution, after that we define some optimization problem to find these parameters. At first we mention some definitions for dictionary selection in sparse representations. A substantial parameter in dictionary selection for sparse representation is mutual coherence. The mutual coherence of a dictionary D , $\mu(D)$, express the coherency between atoms of dictionary. This parameter defined by inner product between each distinct normalized atoms of dictionary D [13]

$$\mu(D) = \max_{i \neq j} \frac{|a_i^T a_j|}{a_i^T a_i a_j^T a_j} \quad (4)$$

Minimum $\mu(D)$ is suitable for sparse representation. For an orthogonal matrix D , we have $\mu(D) = 0$. But for a redundant dictionary mutual coherence is greater than zero because of its redundancy in known dimension. For an overcomplete dictionary ($N > d$) it has shown that there exist a lower bound. It has been shown in [14] that for all full rank matrix of size $d \times N$

$$\mu \geq \mu_G \sqrt{\frac{k-n}{n(k-1)}} \quad (5)$$

μ_G is the mutual coherence of Grassmannian frames which a dictionary with minimum coherence [14]. In order to have sparse representation, finding a dictionary with minimum of $\mu(D)$ is the competing area [15], [16]. As mention in previous section different sampling of orthogonal set make different redundant dictionary so with suitable tuning in sampling process we can hope to find a dictionary that has minimum distance to Grassmannian frames. So we need an objective function to

minimize this distance. Following part represent construction of our objective function.

2.2. Optimization Problem

Let $D = [\phi_1, \phi_2, \dots, \phi_N]$ be a dictionary in $\mathbb{R}^{d \times N}$, we define $\mathcal{G}_D = D^T \cdot D$ as gram matrix of D . Each element of this matrix represents the coherency of each two atoms that obtain with an inner product between corresponding atoms. Suppose \mathcal{G}_G be a square matrix in which the absolute value each of off diagonal elements are μ and the diagonal element is 1 (Gram matrix of Grassmannian frames). In order to make a dictionary, which have minimum distance in frobenious norm to Grassmannian frames, we can define our objective function as a distance minimization of \mathcal{G}_D and \mathcal{G}_G

$$OBJ : \|\mathcal{G}_D - \mathcal{G}_G\|_F \quad (6)$$

$\|\cdot\|_F$ stands for frobenious norm; \mathcal{G}_D and \mathcal{G}_G is Gram matrix of dictionary D and Grassmannian frames respectively $\|\cdot\|_F$ stands for frobenious norm. This is a big objective function to minimization, so we need to break down this function in order to achieve better optimization problem. In other words, we want to link this problem to sampling process of orthogonal set. The main ongoing question is how to sample orthogonal functions. Suppose $\{\varphi_k\}$, $k = 1 \dots N$ be a set of N orthogonal functions in the $[a, b]$ interval. To divide this interval to d sample we use methods that are explained as

$$t_0 = \delta_0, t_k = t_{k-1} + \delta_k, k = 1, 2, \dots, d-1 \quad (7)$$

where $\sum_{k=0}^{d-1} \delta_k \leq b-a$, and the values of $\{\delta_k\}$ are determined by normal distribution $\mathcal{N}(t, \mu, \sigma)$. As it has been explained before, each element of \mathcal{G}_D is an inner product of corresponding atoms $(\mathcal{G}_D(i, j) = \langle \varphi_i(\mathcal{N}(t, \mu, \sigma)), \varphi_j(\mathcal{N}(t, \mu, \sigma)) \rangle)$. We can rewrite the objective function as below:

$$\|\mathcal{G}_D - \mathcal{G}_G\|_F = \left[\sum_i \sum_j (\mathcal{G}_D(i, j) - \mathcal{G}_G(i, j))^2 \right]^{1/2} = F(u, s) \quad (8)$$

So with this assumption the objective function only depends on μ, σ . So we can use any optimization algorithm to minimize this objective function [17]. The value of $F(u, s)$ or distance between Gram matrix of designed dictionary and Grassmannian frames implicitly depend on orthogonal function, so different set of orthogonal function has different value for $F(u, s)$.

2.3. Solving the Problem: GA

Objective function that defined in (7) is a non-convex problem. We might be in trouble if we want to solve this problem with gradient methods; because these methods

need an initial intelligent guess on initial point for confidence in real convergence and not trapped in local minima. GA is a method that we should use to conquest to this problem. The main part of GA is its initial population that should have enough diversity. Our experiments show that uniform distribution on μ and σ is suitable for finding absolute minimum. This algorithm repeatedly modifies a population of individual solutions. At each step, it randomly selects individuals from the current population to be parents and uses them to produce children for the next generation [18]. Number of variables in optimization process is two. In the next section we implement our algorithm on trigonometric function as an orthogonal set.

3. Implementation and Results

This section illustrated our simulation results. As mention above we can apply this algorithm on any set of orthogonal function. We implement our algorithm on Sinusoid functions $\{\sin(nx)\}$ as a specific set of orthogonal function. At first we perform our optimization process and find the best parameters of normal distribution and compare the distance to Grassmannian frames with uniform version of this function. Second part of this section demonstrated the ability of this kind of dictionary design on sparse coding. The last part shows the effect of normal sampling type dictionary design on image denoising.

3.1. Design Dictionary

$\{\sin(nx)\}$ is an orthogonal set in $\mu[0, \pi]$. We want to compare uniform and normal distributions in order to make overcomplete dictionary in sparse representations. To optimize the (8) function we used “ga” command of Matlab software. We implement our algorithm on several lengths of dictionary’s atom and different redundancy factor and plotted the distance of Gram matrix of each dictionary from Grassmannian frames in **Figure 1**. As it is obvious, the dictionary that has been designed with normal distribution is closer to Grassmannian frames than the uniform one. Optimize parameters for normal distribution, are listed in **Table 1** for different atom length and dictionary redundancy.

3.2. Sparse Coding

This part demonstrated the ability of sampling with normal distribution in terms of exact sparse recovery. In this experiment we use the method that presented in [19] and plotted the probability of success in sparse coding versus the cardinality of solution and our result be depicted in **Figure 1**. As it clear from the **Figure 2**, normal type distribution has better performance.

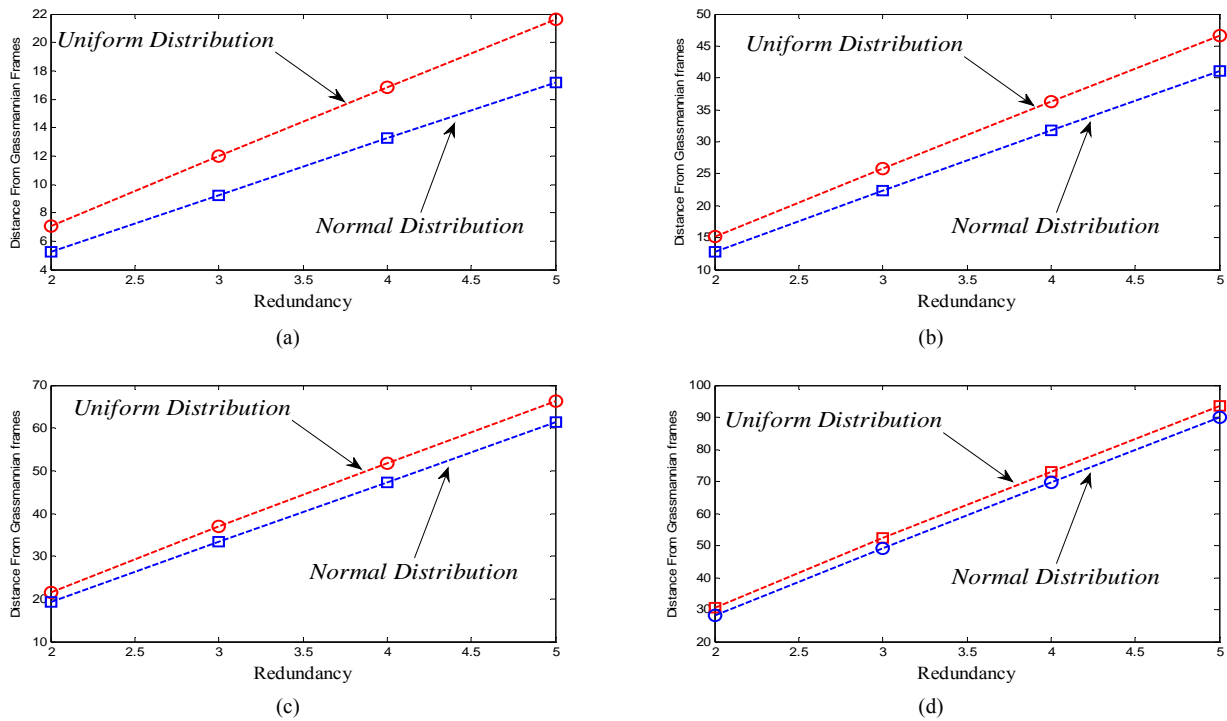


Figure 1. Distance from Grassmannian frames for different length of atom and redundancy factor (a) $d = 16$, (b) $d = 64$, (c) $d = 128$, (d) $d = 256$.

3.3. Image Denoising

As a result of this type of dictionary design, we performed our algorithm on image denoising. We used image denoising method which was proposed in [17]. In this type of image denoising a controversial point is selecting overcomplete dictionary. Signal to noise ratio (PSNR) of resulted image with these two types of overcomplete dictionaries are illustrated in Table 2 (the value of PSNR are average for 10 times running the program)The length of each atom of dictionary in this experiment selected to 64 and redundancy factor is 4. A considerable point in this method of image denoising is decreasing of denoising process time in normal type dictionary design. Figure 3 shows the time of denoising

process for Boat image in different noise level. Paying attention to the figures the advantages of using this type of design is obvious. In Figure 4 these two types of overcomplete dictionaries are shown as an element in gray scale.

4. Conclusions

In this paper we defined new method on overcomplete dictionary design from orthogonal functions. Our experiment showed that non-uniform sampling of these function makes a dictionary that is closer to Grassman nian frames

Table 2. PSNR of denoised image for uniform and normal type sampling dictionary (bold numbers are related to normal type dictionary design).

Table 1. Optimized $\mathcal{N}(\mu, \sigma)$ for each dictionary.

d	Redundancy			
	2	3	4	5
16	$\mathcal{N}(-0.25, 0.64)$	$\mathcal{N}(-0.3, 0.88)$	$\mathcal{N}(1.76, 0.1)$	$\mathcal{N}(-0.6, 0.7)$
64	$\mathcal{N}(-0.25, 0.8)$	$\mathcal{N}(-0.45, 0.58)$	$\mathcal{N}(1.02, -0.35)$	$\mathcal{N}(-0.65, 1.95)$
128	$\mathcal{N}(-0.45, 1)$	$\mathcal{N}(-0.5, 0.8)$	$\mathcal{N}(1.4, -0.35)$	$\mathcal{N}(-0.65, 1.22)$
256	$\mathcal{N}(-0.5, 1.16)$	$\mathcal{N}(-0.5, 1.26)$	$\mathcal{N}(1, -0.5)$	$\mathcal{N}(-0.70, 1.22)$

Image	Noise Level				
	10	20	30	40	50
barbara	33.40/	29.80/	27.71/	26.16/	25.01/
	33.59	30.00	27.83	26.34	25.13
lena	34.61/	31.51/	29.65/	28.34/	27.29/
	34.73	31.64	29.76	28.41	27.41
boat	33.01/	29.61/	27.72/	26.48/	25.50/
	33.10	29.69	27.83	26.49	25.54
pepper	34.07/	31.43/	29.73/	28.40/	27.46/
	34.18	31.51	29.87	28.51	27.54

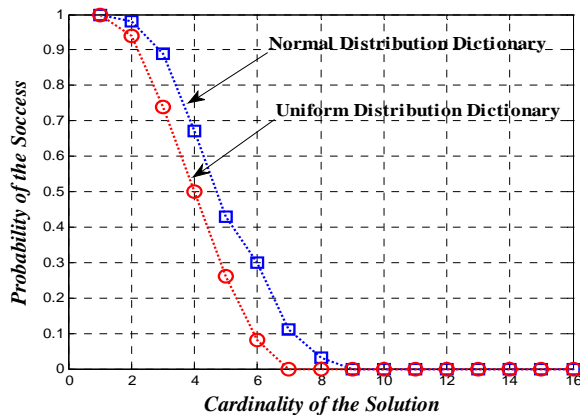


Figure 2. Comparison of two type dictionary design in exact sparse recovery.

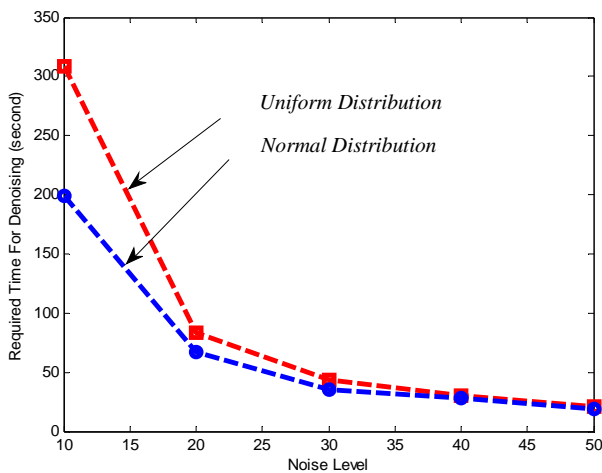


Figure 3. Time required for denoising process for two type dictionary design.

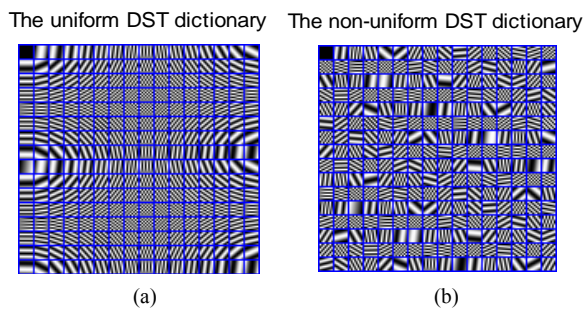


Figure 4. (a) uniform DST dictionary, (b) non-uniform (normal) DST dictionary.

compared to the uniform one. Normal distribution is used for sampling and its parameters (mean and standard deviation) are optimized through some optimization processes. Proposed method design has some advantage in sparse coding and we show this result with plotted the

probability of success in exact sparse recovery. In this framework we showed that this type of dictionary design has better performance than uniform sample dictionary design in image denoising application with improving the PSNR of result image and decrease the required time for image denoising. As mention in paper, we can apply this type redundant dictionary design on any orthogonal set and also we can use other distribution for sampling the interval of function definition.

REFERENCES

- [1] S. Mallat and Z. Zhang, "Matching Pursuits with Time Frequency Dictionaries," *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, 1993, pp. 3397-3415. [doi:10.1109/78.258082](https://doi.org/10.1109/78.258082)
- [2] G. Davis, "Adaptive Nonlinear Approximations," Ph.D. Dissertation, New York University, New York, 1994.
- [3] T. Blumensath and M. Davies, "Gradient Pursuits," *IEEE Transactions on Signal Processing*, Vol. 56, No. 6, 2008, pp. 2370-2382. [doi:10.1109/TSP.2007.916124](https://doi.org/10.1109/TSP.2007.916124)
- [4] S. Chen, D. Donoho and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, 1998, pp. 33-61. [doi:10.1137/S1064827596304010](https://doi.org/10.1137/S1064827596304010)
- [5] M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing," Springer, New York, 2010.
- [6] R. Rubinstein, A. M. Bruckstein and M. Elad, "Dictionaries for Sparse Representation Modeling," *IEEE Proceedings—Special Issue on Applications of Sparse Representation & Compressive Sensing*, Vol. 98, No. 6, April 2010, pp. 1045-1057.
- [7] S. Mallat, "A Wavelet Tour of Signal Processing," 3rd Edition, Academic, New York, 2009.
- [8] E. J. Candès and D. L. Donoho, "Curvelets—A Surprisingly Effective Nonadaptive Representation for Objects with Edges," Vanderbilt University Press, Nashville, 1999.
- [9] M. N. Do and M. Vetterli, "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation," *IEEE Transaction on Image Processing*, Vol. 14, No. 12, 2005, pp. 2091-2106. [doi:10.1109/TIP.2005.859376](https://doi.org/10.1109/TIP.2005.859376)
- [10] E. LePennec and S. Mallat, "Sparse Geometric Image Representations with Bandelets," *IEEE Transaction on Image Processing*, Vol. 14, No. 4, 2005, pp. 423-438. [doi:10.1109/TIP.2005.843753](https://doi.org/10.1109/TIP.2005.843753)
- [11] K. Engan, S. O. Aase and J. H. Husoy, "Method of Optimal Directions for Frame Design," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, 15-19 March 1999, pp. 2443-2446.
- [12] M. Aharon, M. Elad and A. M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, Vol. 54, No. 11, 2006, pp. 4311-4322. [doi:10.1109/TSP.2006.881199](https://doi.org/10.1109/TSP.2006.881199)

- [13] M. Aharon, M. Elad and A. M. Bruckstein, "On the Uniqueness of Overcomplete Dictionaries, and a Practical Way to Retrieve Them," *Journal of Linear Algebra and Applications*, Vol. 416, No. 1, 2006, pp. 48-67.
- [14] T. Strohmer and R. Heath, "Grassmannian Frames with Applications to Coding and Communication," *Applied and Computational Harmonic Analysis*, Vol. 14, No. 3, 2003, pp. 257-275. [doi:10.1016/S1063-5203\(03\)00023-X](https://doi.org/10.1016/S1063-5203(03)00023-X)
- [15] J. Tropp, I. Dhillon, R. Heath Jr. and T. Strohmer, "Designing Structural Tight Frames via an Alternating Projection Method," *IEEE Transactions on Information Theory*, Vol. 51, No. 1, 2005, pp. 188-209. [doi:10.1109/TIT.2004.839492](https://doi.org/10.1109/TIT.2004.839492)
- [16] M. Yaghoobi, T. Blumensath and M. Davies, "Dictionary Learning for Sparse Approximations with the Majorization Method," *IEEE Transactions on Signal Processing*, Vol. 57, No. 6, 2009, pp. 2178-2191. [doi:10.1109/TSP.2009.2016257](https://doi.org/10.1109/TSP.2009.2016257)
- [17] J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, Vol. 9 No. 1, 1998, pp. 112-147. [doi:10.1137/S1052623496303470](https://doi.org/10.1137/S1052623496303470)
- [18] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Addison Wesley, Reading, 1989.
- [19] M. Elad and M. Aharon, "Image Denoising via Sparse and Redundant Representations over Learned Dictionaries," *IEEE Transactions on Image Processing*, Vol. 15, No. 12, December 2006, pp. 3736-3745.