

# Image Denoising Using Sparse Representations

SeyyedMajid Valiollahzadeh<sup>1,\*</sup>, Hamed Firouzi<sup>1</sup>, Massoud Babaie-Zadeh<sup>1</sup>,  
and Christian Jutten<sup>2</sup>

<sup>1</sup> Department of Electrical Engineering, Sharif University of Technology,  
Tehran, Iran

<sup>2</sup> GIPSA-lab, Grenoble, France

**Abstract.** The problem of removing white zero-mean Gaussian noise from an image is an interesting inverse problem to be investigated in this paper through sparse and redundant representations. However, finding the sparsest possible solution in the noise scenario was of great debate among the researchers. In this paper we make use of new approach to solve this problem and show that it is comparable with the state-of-art denoising approaches.

## 1 Introduction

Being a simple inverse problem, the denoising is a challenging task and basically addresses the problem of estimating a signal from the noisy measured version available from that. A very common assumption is that the present noise is additive zero-mean white Gaussian with standard deviation  $\sigma$ . Many solutions have been proposed for this problem based on different ideas, such as spatial adaptive filters, diffusion enhancement [1], statistical modeling [2], transfer domain methods [3], [4], order statistics [5] and yet many more. Among these methods, methods based on with sparse and redundant representations has recently attracted lots of attentions [8]. Many researchers have reported that such representations are highly effective and promising toward this stated problem [8]. Pioneered by Donoho [5], sparse representations firstly examined with unitary wavelet dictionaries leading to the well-known shrinkage algorithm [5]. A major motivation of using overcomplete representations is mainly to obtain translation-invariant property [6]. In this respect, several multiresolutional and directional redundant transforms are introduced and applied to denoising such as curvelets, contourlets, wedgelets, bandlets and the steerable wavelet [5] [8].

The aim of all such transforms is to provide a redundant sparse decomposition of the signal. In parallel, beside providing a suitable redundant transform, representation of a signal with these transforms is also of high value, since such a representation is not necessarily unique. Several methods are then proposed to

---

\* This work has been partially supported by Iran National Science Foundation (INSF) under contract number 86/994 and also by ISMO and French embassy in Iran in the framework of a Gundi-Shapour collaboration program.

find the best possible representation of a signal from a redundant, overcomplete dictionary obtained by these transforms, namely Matching Pursuit(MP), Basis Pursuit(BP), FOCUSS, and Smoothed  $\ell^0$ - Norm (SL0) [7]. All these approaches basically try to find the sparsest possible solution among all the possible representations a signal can obtain. As an alternative point of view to obtain the sparse representation, example-based dictionary learning of K-SVD which is introduced by Aharon, *et. al.* [8] attempts to find the sparse dictionary over the image blocks. When using the Bayesian approach to address this inverse problem with the prior of sparsity and redundancy on the image, it is the dictionary to be used that we target as the learned set of parameters. Instead of the deployment of a pre-chosen set of basis functions like the curvelet or contourlet, this process of dictionary learning can be done through examples, a corpus of blocks taken from a high-quality set of images and even blocks from the corrupted image itself. This idea of learning a dictionary that yields sparse representations for a set of training image blocks has been studied in a sequence of works [8] and specifically the one using K-SVD has shown to outperform in both providing the sparse representation and capability of denoising. While the work reported here is based on the same idea of sparsity and redundancy concepts, a different method is used to solve the sparsest possible solution in presence of noise. An example-based dictionary learning such as K-SVD along with here used technique can provide better solutions in estimation of the original clean signal.

The paper is organized as follows. In section 2, we briefly present modeling of the scenario in decomposing a signal on an overcomplete dictionary in the presence of noise. In section 3 we discuss this algorithm in the real image denoising task. At the end we conclude and give a general overview to future's work.

## 2 Finding the Sparse Representation in Presence of Noise

Consider the problem of estimation of  $\mathbf{x}$  from the observed signal

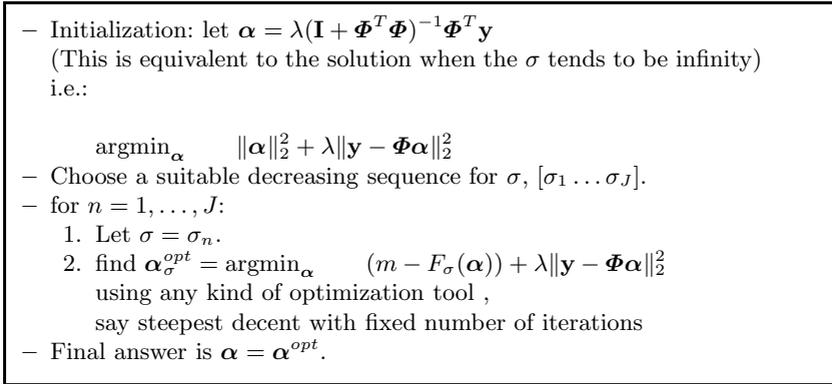
$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

where  $\mathbf{n}$  denotes the observation noise. Assume that  $\mathbf{x}$  has a sparse representation over the dictionary  $\Phi$ , i.e.  $\mathbf{x} = \Phi\alpha$  with a small  $\|\alpha\|_0^0$  (the number of nonzero elements of a vector) and also assume that a good estimation on the energy of the present noise,  $\|\mathbf{n}\|_2^2 \leq \epsilon^2$  is provided.

The sparsest representation we are looking for, is simply

$$P_0 : \quad \min \|\alpha\|_0^0 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\alpha\|_2^2 \leq \epsilon^2 \quad (1)$$

Note that the above-stated problem rarely has a unique solution [11], since once the sparsest solution is found, many feasible variants of it sharing the same support can be built. Since the above-stated problem is highly nonconvex and hard to deal with, many researchers pursue a strategy of convexification with



**Fig. 1.** Algorithm for finding the sparse coefficients in presence of noise

replacing  $\ell^0$  norm with  $\ell^1$ - norm. so simply try to solve the following problem instead:

$$P_1 : \quad \min \|\alpha\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi \alpha\|_2^2 \leq \epsilon^2 \quad (2)$$

where  $\|\alpha\|_1 = \sum \alpha_i$  is the  $\ell^1$ -norm of  $\alpha$ . Note that the replacing  $\ell^0$ -norm by other convex cost functions such as  $\ell^1$ -norm is only asymptotic and the equivalence does not always hold [9]. Hereafter, motivated by the recently stated work of Mohimani, *et al.* [7] we seek to find the sparsest possible answer without such a replacement and instead, attempt to relax the replacing  $\ell^0$ - norm by a continuous, differentiable cost function, say  $F_{\sigma}(\alpha) = \sum_i \exp(-\alpha_i^2/2\sigma^2)$ .

This function tends to count the number of zero elements of a vector. So, as stated in [7] the above problem can be converted to:

$$P_0 : \quad \min_{\alpha} (m - F_{\sigma}(\alpha)) \quad \text{subject to} \quad \|\mathbf{y} - \Phi \alpha\|_2^2 \leq \epsilon^2 \quad (3)$$

The above optimization task can be converted to optimizing the Lagrangian:

$$P_0 : \quad \min_{\alpha} (m - F_{\sigma}(\alpha)) + \lambda \|\mathbf{y} - \Phi \alpha\|_2^2 \quad (4)$$

So that the constraint becomes a penalty and the parameter  $\lambda$  is dependent on  $\epsilon$ . Solution toward this problem was recently proposed in [12] and it is shown that for a proper choice of  $\lambda$ , these two problems are equivalent. The  $\sigma$  parameter determines the smoothness of the approximated cost function. By gradual decrease in this parameter it is highly probable to skip trapping in local minimum. The overall algorithm which is used through this paper is shown in Fig. 1 is a slight modification of the same idea presented in [12].

Once the sparsest solution of (3) has been found with the stated algorithm summarized in Fig. 1, we can retrieve the approximate image by  $\hat{\mathbf{x}} = \Phi \hat{\alpha}$ .

### 3 Image Denoising

The problem of estimation of  $\mathbf{X}$  from an observed noisy version of it under the sparsity prior has two essential issues: first, to find a dictionary  $\Phi$  which permits a sparse representation regarding the fact that the sample are noisy and second to find the coefficients of this sparse representation. The second phase was what explained so far. As it was shown by Aharon [8], *et. al.*, the K-SVD learning is a very efficient strategy which leads to satisfactory results. This method along with all other types of dictionary learning fails to act properly [8] when the size of dictionary grows. Beside that, the computational complexity and thus time needed for training will grow awesome.

When we are dealing with larger size images we are still eager to apply this method but as stated it is computationally costly and both dictionary learning and optimization to find the coefficients of sparse representation are sometimes intractable. To overcome this difficulty, an image with size  $\sqrt{N} \times \sqrt{N}$  is divided to blocks of size of  $\sqrt{n} \times \sqrt{n}$ . These blocks are chosen highly overlapped for two reasons: first, to avoid blockiness and second to have better estimate in noise removal process. Then a dictionary is tried to be found over these blocks and all these blocks are cleaned with algorithm of Fig. 1. Let  $\mathbf{L}_{ij}$  be a matrix representing each block to be located in  $(ij)$ -th position of the image.  $\mathbf{L}_{ij}$  is a matrix of size  $n \times N$  which provides the location information of all possible blocks of the images. So in this respect, the noise removal process changes to:

$$\{\hat{\mathbf{X}}, \hat{\boldsymbol{\alpha}}\} = \operatorname{argmin}_{\mathbf{X}, \boldsymbol{\alpha}} \lambda \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{ij} \gamma \|\boldsymbol{\alpha}_{ij}\|_0 + \sum_{ij} \|\Phi \boldsymbol{\alpha} - \mathbf{L}_{ij} \mathbf{X}\|_2^2 \quad (5)$$

in which  $\mathbf{X}$  is the original image to be estimated and the  $\mathbf{Y}$  is the observed available noisy version of it. This equation is similar to (1) with this slight difference that local analysis was taken into account and a linear combination of  $\ell^0$ -norm and  $\ell^2$ -norm of all sparse representation and error between the original signal and the reconstructed one tried to be minimized. In this process, visible artifacts may occur due to blocking phenomena. To avoid this, we choose the blocks with overlap and at the end average the results in order to prevent blockiness artifact. After determining all the approximated coefficients, we estimate the original image by solving the following equation:

$$\hat{\mathbf{X}} = \operatorname{argmin}_{\mathbf{X}} \lambda \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{ij} \|\Phi \boldsymbol{\alpha} - \mathbf{L}_{ij} \mathbf{X}\|_2^2 \quad (6)$$

This quadratic equation has the solution:

$$\hat{\mathbf{X}} = (\lambda \mathbf{I} + \sum_{ij} \mathbf{L}_{ij}^T \mathbf{L}_{ij})^{-1} (\lambda \mathbf{Y} + \sum_{ij} \mathbf{L}_{ij}^T \Phi \hat{\boldsymbol{\alpha}})^{-1} \quad (7)$$

This estimated modified image can be interpreted as a relaxed averaging between the noisy observed image with the cleaned estimated one. The summarized overall algorithm is shown in Fig. 2.

- Goal: denoise a given image  $\mathbf{Y}$  from additive white Gaussian noise with variance of  $\|\mathbf{n}\|_2^2$
- parameters:  $n$  block-size,  $k$  dictionary,  $\lambda$  Lagrangian multiplier. the task is to optimize
 
$$\{\hat{\mathbf{X}}, \hat{\alpha}\} = \operatorname{argmin}_{\mathbf{X}, \alpha} \lambda \|\mathbf{Y} - \mathbf{X}\|_2^2 + \sum_{ij} \gamma \|\alpha_{ij}\|_0^0 + \sum_{ij} \|\Phi \alpha - \mathbf{L}_{ij} \mathbf{X}\|_2^2$$
- train a dictionary  $\Phi$  of size  $n \times k$  using K-SVD.
- find the sparse noisy coefficients of  $\alpha$  using algorithm stated in Fig. 1.
- Final estimation is  $\hat{\mathbf{X}} = (\lambda \mathbf{I} + \sum_{ij} \mathbf{L}_{ij}^T \mathbf{L}_{ij})^{-1} (\lambda \mathbf{Y} + \sum_{ij} \mathbf{L}_{ij}^T \Phi \hat{\alpha})^{-1}$ .

Fig. 2. The final denoising algorithm

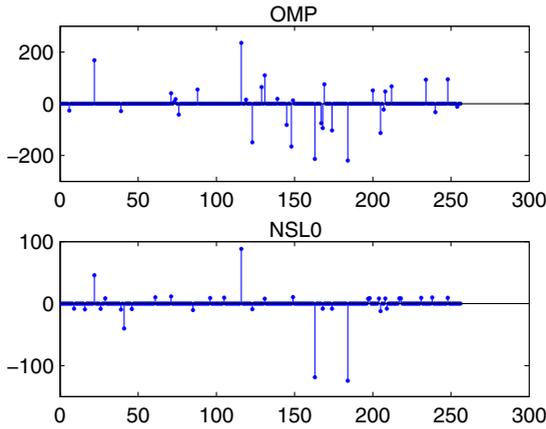
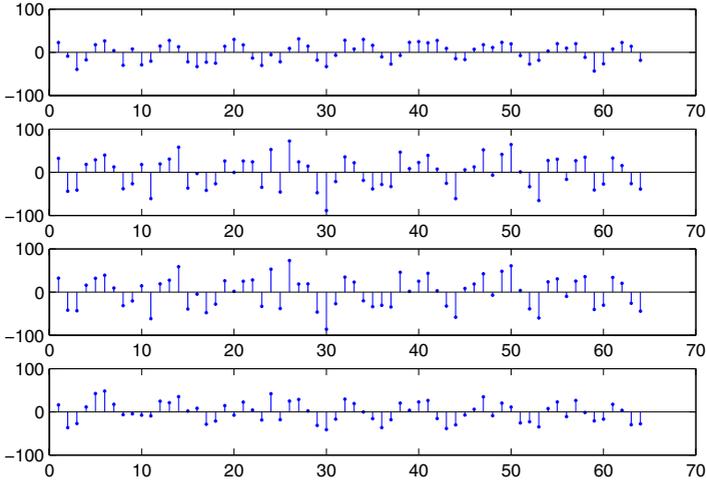


Fig. 3. Coefficients of a sample block represented with OMP above and in bottom. The latter, leads to the same result or sparsely superior one.

### 4 Experimental Results

In this work, the underlying dictionary was trained with the K-SVD method and once the learning is done, the image blocks was represented sparsely via Fig. 1. The algorithm of section 2 was used for such a representation. The overall denoising method explained above was examined with numerous test images mainly of size  $256 \times 256$  and  $512 \times 512$  with different noise levels. Blocks of size  $8 \times 8$  was driven by the synthesis noisy image and a dictionary of size  $64 \times 256$  was learned through this blocks using K-SVD method. Then we applied the algorithm of Fig. 1 to represent each block on the provided dictionary, while the similar approach done by Aharon [8] make use of Orthogonal Matching Pursuit (OMP) [10] for this stage. The tested images are all the same ones as those used in the denoising experiments reported in [8], in order to enable a fair comparison. Table 1 summarizes the denoising results in the same database of



**Fig. 4.** Coefficients of a sample block. From top to bottom: the original clean signal, the signal corrupted with additive white Gaussian noise of  $\|\mathbf{n}\|_2 = 20$ , recovered block via OMP and the recovered block with the algorithm of Fig. 1.



**Fig. 5.** From left to right: original image, noisy image with zero-mean white gaussian noise of  $\|\mathbf{n}\|_2 = 20$ , the cleaned image via sparse representation described

images. In a quite large experiments we found sparser solution and better quality of representations. Every result reported is an average over 5 experiments, having different realizations of the noise. To show a comparison in sparsity yielded with different methods coefficients in representations of a sample block with OMP and the stated algorithm was depicted in Fig. 3. The quite same results is valid for other blocks as well.

The denoised blocks were averaged, as described in Fig.2 .In Fig. 5 the results of the overall algorithm for the image "Barbara" for  $\|\mathbf{n}\|_2 = 20$  is shown. By refereing to Table 1, as it is seen, when the level of noise grows, our approach outperforms K-SVD with OMP and we can conclude the mentioned algorithm is suitably designed for noisy cased with known energy.

Also a comparison was done with other types of sparse coding phase such as FOCUSS and SL0 [8] and yet the proposed algorithm outperforms them. A

**Table 1.** Summary of denoising PSNR results. In each column the bottom is corresponding to our approach and the above is corresponding to the K-SVD with OMP. the bold one corresponds with better response.

$\frac{\sigma}{PSNR}$	LENA	BARBARA	BOAT	Fgrpt	House	Peppers	Average	$\sigma_{PSNR}$
2/42.11	<b>43.58</b> 42.11	<b>43.67</b> 42.38	<b>43.14</b> 42.17	<b>42.99</b> 41.85	<b>44.47</b> 42.92	<b>43.33</b> 42.51	<b>44.47</b> 42.92	<b>43.33</b> 42.51
5/34.15	<b>38.60</b> 38.18	<b>38.08</b> 37.41	<b>37.22</b> 36.68	<b>36.65</b> 36.17	<b>39.37</b> 38.25	<b>37.78</b> 37.08	<b>39.37</b> 38.25	<b>37.78</b> 37.08
10/28.13	<b>35.47</b> 35.42	34.42 <b>34.51</b>	<b>33.64</b> 33.62	<b>32.39</b> 32.31	<b>35.98</b> 35.60	34.28 <b>34.53</b>	35.98 35.60	<b>34.28</b> <b>34.53</b>
15/24.61	33.70 <b>33.91</b>	32.36 <b>32.79</b>	31.73 <b>32.13</b>	30.06 <b>30.258</b>	34.32 <b>34.40</b>	32.22 <b>32.79</b>	34.32 <b>34.40</b>	32.22 <b>32.79</b>
20/22.11	32.38 <b>33.46</b>	30.83 <b>32.01</b>	30.36 <b>31.29</b>	28.47 <b>29.16</b>	33.20 <b>34.19</b>	30.82 <b>31.58</b>	33.20 <b>34.19</b>	30.82 <b>31.58</b>
25/20.17	31.32 <b>32.72</b>	29.60 <b>31.01</b>	29.28 <b>30.46</b>	27.26 <b>28.90</b>	32.15 <b>33.61</b>	29.73 <b>30.83</b>	32.15 <b>33.61</b>	29.73 <b>30.83</b>
50/14.15	27.79 <b>28.98</b>	25.47 <b>26.93</b>	25.95 <b>27.30</b>	23.24 <b>24.43</b>	27.95 <b>28.69</b>	26.13 <b>27.70</b>	27.95 <b>28.69</b>	26.13 <b>27.70</b>
75/10.63	25.80 <b>26.93</b>	23.01 <b>24.71</b>	23.98 <b>25.33</b>	19.97 <b>21.69</b>	25.22 <b>26.83</b>	23.69 <b>24.28</b>	25.22 <b>26.83</b>	23.69 <b>24.28</b>
100/8.13	24.46 <b>26.32</b>	21.89 <b>23.55</b>	22.81 <b>24.36</b>	18.30 <b>22.19</b>	23.71 <b>25.08</b>	21.75 <b>23.14</b>	23.71 <b>25.08</b>	21.75 <b>23.14</b>

sample comparison has been done in Fig. 4. In this experiment after providing the dictionary, the sparse representation coefficients are found with different approaches. The coefficients of the original clean signal, the signal corrupted with additive white gaussian noise of  $v\|\mathbf{n}\|_2 = 20$ , recovered block via OMP and the recovered signal via Fig. 1 is depicted in Fig. 4 and as it can be seen the our recovered signal resembles more to the original signal.

## 5 Discussions and Conclusions

In this paper a simple algorithm for denoising application of an image was presented leading to state-of-the-art performance, equivalent to and sometimes surpassing recently published leading alternatives. It is basically on the basis of sparse representation of an image in the presence of noise. The stated algorithm considers local approach, splits the noisy observed image to several blocks and learns a dictionary over these blocks and attempts to find the best possible sparse representation of each block with this dictionary. In order to find the cleaned image some averaging is needed to avoid the blocking effect in boundaries. Experimental results show satisfactory recovering of the image. Future theoretical work on the general behavior of this algorithm is on our further research agenda.

## References

1. Gilboa, G., Sochen, N., Zeevi, Y.: Forward-and-Backward Diffusion Processes for Adaptive Image Enhancement and Denoising. *IEEE Trans. on Image Processing* 11(7), 689–703 (2002)
2. Mihcak, M.K., Kozintsev, I., Ramchandran, K., Moulin, P.: Low complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters* 6(12), 300–303 (1999)
3. Grace Chang, S., Yu, B., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. on Image Processing* 9(9), 1532–1546 (2000)
4. Starck, J., Candes, E., Donoho, D.L.: The curvelet transform for image denoising. *IEEE Trans on Image Processing* 11(6), 670–684 (2002)
5. Donoho, D.L.: De-noising by soft thresholding. *IEEE Trans. Inf. Theory* 41(3), 613–627 (1995)
6. Coifman, R., Donoho, D.L.: Translation invariant de-noising. In: *Wavelets and Statistics. Lecture Notes in Statistics*, pp. 125–150. Springer, New York (1995)
7. Mohimani, H., Babaie-Zadeh, M., Jutten, C.: A Fast approach for overcomplete sparse decomposition based on smoothed L0 norm. *IEEE Trans on Signal Processing* (to appear)
8. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. on Signal Processing* 54(11) (November 2006)
9. Candes, E.J., Wakin, M.B., Boyd, S.: Enhancing Sparsity by Reweighted l1 Minimization. *Journal of Fourier Analysis and Applications*, special issue on sparsity (October 2008)
10. Davis, G., Mallat, S., Zhang, Z.: Adaptive time-frequency approximations with matching pursuits, Tech. Rep. TR1994-657 (1994)
11. Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory* 52(1), 6–18 (2006)
12. Firouzi, H., Farivar, M., Babaie-Zadeh, M., Jutten, C.: Approximate Sparse decomposition Based on Smoothed L-0 norm. In: *ICASSP 2009* (2009), <http://www.arxiv.org>